HAL
open science

ROLE DE DEUX ARN DANS LE CONTROLE DE L'EXPRESSION DES GENES: REGULATIONS DE LA REPLICATION DU PLASMIDE R1 PAR UN ARN ANTISENS ET DES GENES DE VIRULENCE DE STAPHYLOCOCCUS AUREUS PAR L'ARN-III

Fabrice Kolb

To cite this version:

Fabrice Kolb. ROLE DE DEUX ARN DANS LE CONTROLE DE L'EXPRESSION DES GENES: REGULATIONS DE LA REPLICATION DU PLASMIDE R1 PAR UN ARN ANTISENS ET DES GENES DE VIRULENCE DE STAPHYLOCOCCUS AUREUS PAR L'ARN-III. Biochimie [q-bio.BM]. Université Louis Pasteur - Strasbourg I, 2001. Français. NNT : . tel-00002806

HAL Id: tel-00002806

https://theses.hal.science/tel-00002806v1

Submitted on 6 May 2003

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE LOUIS PASTEUR
STRASBOURG I

ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

THESE

Présentée
Pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE LOUIS PASTEUR
Domaine : BIOLOGIE MOLECULAIRE ET CELLULAIRE

par

Fabrice KOLB

ROLE DE DEUX ARN DANS LE CONTROLE DE L'EXPRESSION DES GENES : REGULATIONS DE LA REPLICATION DU PLASMIDE R1 PAR UN ARN ANTISENS ET DES GENES DE VIRULENCE DE *STAPHYLOCOCCUS AUREUS* PAR L'ARN-III

Soutenue le 27 septembre 2001 devant la Commission d'examen

Docteur Christiane Branlant	Rapporteur externe
Docteur Chantal Ehresmann	Directeur de thèse
Professeur Bernard Ehresmann	Examinateur
Professeur Witold Filipowicz	Rapporteur externe
Professeur Gérard Jonard	Rapporteur interne
Docteur Pascale Romby	Directeur de thèse

UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, Strasbourg

SOMMAIRE

SOMMAIRE **1**

ABREVIATIONS **6**

INTRODUCTION **8**

I. INTRODUCTION GENERALE : LA NOTION DU « MONDE TOUT ARN » 8

II. LES RIBOREGULATEURS : MECANISMES GENERAUX 12

A. Ribor égulateurs modulant l'activité de protéines régulatrices *12*

B. Les ARN antisens *13*

C. Les ribor égulateurs multifonctionnels *14*

D. Pourquoi utiliser des ribor égulateurs ? *15*

1. *Economie d'énergie et temps de synthèse du ribor égulateur* 15

2. *Dégradation du régulateur* 15

3. *Versatilité des structures des ribor égulateurs* 15

4. *Rapidité d'action* 16

E. Les ARN régulateurs comme outils génétiques *16*

1. *Développement de ribor égulateurs artificiels* 16

a. Les ARN antisens artificiels 16

b. Les aptamères 18

2. *Vers d'autres types de régulations ARN-dépendantes : l'exemple de l'ARN interférence* 18

a. Historique de l'ARN interférence (ARNi) 18

b. Modèle d'action pour l'ARNi 19

c. Rôle(s) de l'ARNi 21

III. LES RIBOREGULATEURS MODULANT L'ACTIVITE DE FACTEURS PROTEIQUES 23

A. Contrôle de la biosynthèse du glycogène chez E. coli *23*

B. Mécanismes viraux de résistance aux interférons. *27*

IV. LES RIBOREGULATEURS MULTIFONCTIONNELS 32

A. Implication d'ARN régulateurs dans le contrôle de l'expression du facteur de transcription alternatif σ^S *32*

1. *Le facteur de transcription alternatif* σ^S 32

2. *La protéine Hfq* 33

3. *La protéine H-NS* 34

4. *Contrôle de l'expression de* rpoS *par l'ARN DsrA* 34

a. DsrA répresseur de *hns* 34

b. DsrA active directement l'expression de *rpoS* ... 37
5. *L'ARN RprA* ... 37
6. *L'ARN OxyS répresseur de* rpoS ... 38
B. Rôle central de DsrA dans le contrôle de l'expression génétique ... *40*
C. Plusieurs cibles pour OxyS ... *40*
D. Caractéristiques générales des ARN multirégulateurs ... *41*
V. LES ARN ANTISENS ... 43
A. Les ARN antisens eucaryotiques ... *43*
1. *Régulation traductionnelle : l'exemple du contrôle du développement larvaire chez* C. elegans ... 45
a. L'ARN Lin-4 ... 45
b. L'ARN let-7 ... 47
2. *Régulation de la stabilité d'un ARNm* ... 49
a. Contrôle de la formation des préspores chez *Dictyostelium discoideum* ... 49
b. Modification post-transcriptionnelle d'un ARNm : l'exemple de *bFGF* ... 49
3. *Conclusions* ... 50
B. Les ARN antisens procaryotiques ... *51*
1. *Les ARN antisens procaryotiques chromosomiques* ... 52
a. Le contrôle du transport du Fer chez *Vibrio anguillarum* ... 53
b. Contrôle de l'expression de la porine OmpF chez *Escherichia coli* ... 54
c. Contrôle de la division cellulaire chez *Escherichia coli* ... 55
2. *Les ARN antisens codés par les éléments génétiques accessoires* ... 56
a. La régulation de la transposition de l'élément *IS10* ... 57
b. Les ARN antisens chez les bactériophages ... 59
VI. CAS PARTICULIERS DES CONTROLES DE FONCTIONS PLASMIDIQUES ... 61
A. Biologie des plasmides bactériens ... *62*
B. Contrôle du transfert conjugatif ... *62*
C. Contrôle de la ségrégation plasmidique ... *64*
D. Mécanismes de contrôle de la réplication plasmidique ... *67*
1. *Contrôle de la formation de l'amorce de réplication* ... 68
2. *Contrôle transcriptionnel du gène de la protéine initiatrice de la réplication* ... 71
3. *Mécanismes de contrôles post-transcriptionnels de la réplication plasmidique* ... 74
a. Mécanisme de traduction de la protéine de réplication : l'exemple de Col1b-P9 ... 74
b. Contrôle de la traduction de *repZ* par un ARN antisens ... 77
VII. LE CONTROLE DE LA REPLICATION DU PLASMIDE R1 : ETAT DES LIEUX ... 79
A. La biologie du plasmide R1 ... *79*
B. Contrôle transcriptionnel de l'expression de repA ... *80*

C. Contrôles post-transcriptionnels de l'expression de repA *81*

1. *Inhibition de la traduction du peptide leader* tap 81

2. *Dégradation de l'ARN messager* repA 84

VIII. MOTIVATION DU TRAVAIL DE THESE 85

PRESENTATION DU TRAVAIL DE THESE 86

I. CHAPITRE 1 : RELATION STRUCTURE-FONCTION DU COMPLEXE ARN ANTISENS – ARNM INHIBITEUR RESPONSABLE DU CONTROLE DE LA REPLICATION DU PLASMIDE R1 86

II. CHAPITRE 2 : ROLE DE L'ARN-III DANS LA REGULATION DE L'EXPRESSION DES GENES DE VIRULENCE CHEZ *STAPHYLOCOCCUS AUREUS* 87

RESULTATS & DISCUSSION

1ER CHAPITRE : STRUCTURE ET FONCTION DU COMPLEXE INHIBITEUR COPA-COPT RESPONSABLE DU CONTRÔLE DE LA RÉPLICATION DU PLASMIDE R1 88

I. ARTICLE I : ETUDE STRUCTURALE DU COMPLEXE INHIBITEUR COPA-COPT 89

A. Objectifs *89*

B. Stratégie expérimentale *89*

C. Résultats *90*

D. Conclusions *92*

II. ARTICLE II : ROLE DES NUCLEOTIDES NON APPARIES DES HELICES DE COPA ET DE COPT DANS LA FORMATION DU COMPLEXE IRREVERSIBLE 108

A. Objectifs *108*

B. Stratégie expérimentale *108*

C. Résultats *109*

D. Conclusions *110*

III. ARTICLE III : ETUDE MECANISTIQUE DE LA FORMATION DU COMPLEXE IRREVERSIBLE COPA-COPT .. 121

A. Objectifs *121*

B. Stratégie expérimentale *121*

C. Résultats *122*

D. Conclusions *123*

IV. ARTICLE IV : CONSERVATION DE LA JONCTION A QUATRE HELICES DANS LES COMPLEXES IMPLIQUES DANS LE CONTROLE DE LA REPLICATION PLASMIDIQUE 135

A. Objectifs *135*

B. Résultats *135*

C. Conclusions *136*

V. CRISTALLOGENESE DU COMPLEXE COPA-COPT 148
A. Stratégie expérimentale *148*
1. *Conception des ARN* 148
2. *Préparation des ARN $CopA_{61}$ et $CopT_{61}$* 149
B. Analyse biochimique du complexe $CopA_{61}$-$CopT_{61}$ *150*
1. *Le complexe $CopA_{61}$-$CopT_{61}$ se forme avec une constante de vitesse similaire à celle du complexe sauvage* 150
2. *Le complexe $CopA_{61}$-$CopT_{61}$ adopte une topologie identique au complexe sauvage* 150
C. Essais de cristallisation du complexe $CopA_{61}$-$CopT_{61}$ *153*
1. *Renaturation des ARN et formation du complexe* 153
2. *Essais de cristallisation à l'aide de la matrice éparse Natrix™ d'Hampton Research* 153
3. *Optimisation des conditions de cristallisation* 154
4. *Analyse du contenu des cristaux* 157
VI. CONCLUSIONS ET PERSPECTIVES 158
A. Conclusions *158*
1. *Caractéristiques de l'étape de reconnaissance ARN-ARN* 158
2. *De l'interaction initiale au complexe stable* 161
B. Perspectives *165*
1. *Etude de la structure du complexe CopA-CopT formé* in vivo 165
2. *Recherche d'autres intermédiaires réactionnels dans le processus d'appariement de CopA à CopT* 165
3. *Cristallogenèse du complexe CopA-CopT* 166

2ÈME CHAPITRE : ETUDE STRUCTURALE ET FONCTIONNELLE DE L'ARN-III DE STAPHYLOCOCCUS AUREUS 167

I. INTRODUCTION 168
A. La virulence des Staphylocoques *168*
B. Le locus agr *171*
1. *L'opéron P2* 171
2. *L'opéron P3* 173
C. L'ARN-III : un riborégulateur à effet pléiotropique *173*
1. *L'ARN-III, régulateur transcriptionnel* 173
2. *L'ARN-III, régulateur traductionnel* 174
3. *L'ARN-III, riborégulateur dynamique* 175
II. ARTICLE V : ETUDE STRUCTURALE ET FONCTIONNELLE DE L'ARN-III 176
III. RECHERCHE DE FACTEURS SE FIXANT A L'ARN-III PAR PONTAGES AUX UV 190
IV. CONCLUSIONS ET PERSPECTIVES 193

A. Les différents domaines fonctionnels de l'ARN-III *193*

B. Facteurs interagissant avec l'ARN-III *194*

C. Vers de nouvelles cibles de l'ARN-III *195*

CONCLUSION GENERALE 197

REFERENCES 199

ABREVIATIONS

5'-UTR	: région 5' non traduite
A	: adénine
Å	: angström
Ac	: acétate
ADN	: acide désoxyribonucléique
ARN	: acide ribonucléique
ARNi	: ARN interference
ATP	: adénosine triphosphate
C	: cytosine
°C	: degré Celsius
CMCT	:1-cyclohexyl-3-(2N-morpholinoéthyl)-carbodiimide-metho-p-toluène sulfonate
db	: double-brin
DEPC	: diéthylpyrocarbonate
DMS	: diméthyl sulfate
DO	: densité optique
dsRBM	: domaine de fixation à l'ARN double-brin
EDTA	: éthylène diamine tétraacétate de sodium
ENU	: N-éthyl-N-nitrosourée
fmet	: formylméthionine
G	: guanine
h	: heure
Hepes	: acide N,2-hydroxylethylpipérazine-N'-éthane-solfonique
His_6	: étiquette peptidique de six histidines
k*app*	: contante de vitesse apparente
kb	: kilobase
KCl	: chlorure de potassium
kDa	: kilodalton
M	: molaire
MES	: acide ß-morpholino-éthanesulfonique
Mg	: magnésium
µg	: microgramme
min	: minute
mJ	: millijoule
µl	: microlitre
µm	: micromètre
MPD	: 2-méthyl 2,4-pentanediol
Na	: sodium
NH_4	: ammonium

NiCR	: perchlorate de (2,12-diméthyl-3,7,11,17-tétraazabicyclo[11.3.1]heptadéca-1(17),2,11,13,15-pentaénato)-nickel-(II)
nm	: nanomètre
nt	: nucléotide
ORF	: phase ouverte de lecture
pb	: paire de bases
PEG	: polyéthylène glycol
RBS	: site de fixation du ribosome
RNase	: ribonucléase
s	: seconde
S :	: svedberg
SDS	: dodécylsulfate de sodium
Stains-all™	: bromure de 1-éthyl-2-[3-1(-éthylnaphtho[1,2-d]thiazolin-2-ylidène)-2-méthylpropenyl]naphtho[1,2-d]thiazolium
T	: thymidine
U	: uracile
UV	: ultraviolet

Code des acides aminés :

A	alanine
C	cystéine
D	acide aspartique
E	acide glutamique
F	phénylalanine
G	glycine
H	histidine
I	isoleucine
K	lysine
L	leucine
M	méthionine
N	asparagine
P	proline
Q	glutamine
R	arginine
S	sérine
T	thréonine
V	valine
W	tryptophane
Y	tyrosine

INTRODUCTION

I. Introduction Générale : La Notion du « Monde tout ARN »

La question « qui de la poule ou de l'œuf apparut en premier ? » peut être appliquée au couple acide nucléique – protéine. En effet, les protéines sont codées par le génome mais celui-ci nécessite l'intervention d'enzymes pour sa réplication. Une hypothèse alléchante est donc qu'à l'origine de la vie, une seule et même molécule pouvait assurer à la fois des fonctions codantes et des fonctions catalytiques. Dès 1968, Francis Crick proposa que l'origine de la vie reposait sur la molécule d'ARN [1]. Il manquait néanmoins à l'époque des évidences que l'ARN pouvait remplir des fonctions catalytiques. Il avait été établi jusqu'alors que l'ARN messager était utilisé comme intermédiaire entre l'ADN et les protéines et qu'il servait de matrice à la synthèse de celles-ci. Parallèlement, les ribosomes avaient été caractérisés ainsi que leur contenu en ARN. Enfin, en 1956, les ARN de transfert ont été définis comme étant les adaptateurs entre les ARN messagers et les chaînes polypeptidiques.

Il a fallu attendre les années 1980, et les découvertes presque simultanées de Sydney Altman et Thomas Cech, pour mettre en évidence que des ARN naturels pouvaient remplir des fonctions catalytiques. Il s'agissait alors de l'identification de l'ARN M1 de la RNase P bactérienne [2], impliqué dans la maturation des ARN de transfert, et de l'intron autocatalytique de *Tetrahymena thermophila* [3]. Depuis, un grand nombre d'ARN catalytiques ou ribozymes ont été caractérisés dans des organismes aussi variés que les organelles des plantes, les eucaryotes inférieurs, les amphibiens, les procaryotes, les bactériophages, les viroides ou les virus de plantes (pour une revue, [4]). Un ribozyme a également été identifié chez le virus de l'hépatite delta qui est un pathogène humain sévère. L'hypothèse émise par Francis Crick paraissait alors tout à fait plausible et en 1986, Walter Gilbert proposa le terme « RNA World » pour décrire une période pendant laquelle l'ARN était à la fois le support primaire de l'information génétique et une molécule catalytique [5]. En faveur de cette dernière idée, de nombreuses autres fonctions catalytiques que celle de phosphodiestérase ont depuis été assignées aux molécules d'ARN, qu'elles soient naturelles, telles que dans le cas de l'activité peptidyl-transférase du ribosome [6], ou artificielles [7-11].

Parallèlement aux molécules d'ARN catalytiques, et contrairement à l'idée pré-établie des années 1960-1970 qu'il n'existait que trois classes d'ARN (à savoir messagers, ribosomiques et de transfert), un ARN à fonction non catalytique et ne faisant pas partie des trois classes majeures a été identifié. Il s'agissait de l'ARN 4.5S d'*Escherichia coli* [12] (Tableau 1). Depuis, un grand nombre d'autres ARN non codants ont été caractérisés [13,14]. Le tableau I répertorie les exemples les plus étudiés d'ARN non codants à fonctions non régulatrices. Certains d'entre eux remplissent cependant indirectement des fonctions régulatrices de l'expression génétique. Ceci est le cas des ARN 6S, 10Sa (ou tmARN), hsrΩ et Xist qui interviennent dans des mécanismes de surveillance, de défense cellulaire ou d'empreinte parentale. Ils ne modulent pas spécifiquement l'expression d'un gène cible et c'est pour cette raison qu'ils sont distingués des ARN régulateurs qui seront décrits dans l'introduction de ce manuscrit.

La plupart des ARN non codants sont des petits ARN nucléaires très conservés (snRNA), impliqués dans l'épissage chez les eucaryotes, ou nucléolaires (snoRNA) impliqués dans la maturation et la modification post-transcriptionnelle des ARNr [15]. Tel que le montre le tableau ci-contre, ces ARN agissent grâce à des modes d'action divers. Ils font souvent partie intégrante de complexes ribonucléoprotéiques et certains d'entre eux ont un ARN pour cible. De même, les fonctions cellulaires dans lesquelles ils interviennent sont multiples. La plupart des ARN non codants caractérisés ne l'ont été que très récemment. Ceci est principalement lié au fait que les critères de recherche de gènes dans les génomes sont principalement axés sur l'identification de phases ouvertes de lecture. Les différentes méthodes informatiques ou expérimentales qui permettraient d'identifier ces ARN non codants seront discutées plus en détail dans la conclusion de ce manuscrit. Cependant, nous pouvons d'ores et déjà affirmer que le nombre d'ARN non codants caractérisés va augmenter de façon significative ces prochaines années.

ARN	Fonction cellulaire	Facteurs Associés	Mode d'action	Référence
Procaryotiques				
4.5 S	Sécrétion protéique	Ffh	Facilite l'interaction SRP/récepteur	[16]
10Sa (tmRNA)	Cible les ARNm tronqués	SmpB / S1/ Ribosome	Code pour une séquence peptidique signal de dégradation	[17]
pRNA	Empaquetage de l'ADN viral de ϕ29	GTPase gp16	Hexamérisation	[18]
6S	Module l'activité du complexe σ^{70} - ARN Polymérase	σ^{70} et sous-unités β et β' de l'ARN Polymérase	*ND*	[19]
Spot 42	*ND*	Ribosome	*ND*	[20]
ARN a-d	Protection contre les UV chez *Deinococcus radiodurans*	Rsr	*ND*	[21]
Eucaryotiques				
7SL	Sécrétion protéique	Srp54p/Srp19 Srp72,68/Srp9,14	Core de la SRP	[22]
U1, U2, U4, U5, U6, U11, U15	Epissage des pré-ARNm	Protéines Prp et autres	Interactions ARN-ARN et ARN-protéine	[23]
ARN Box C/D	Méthylation des pré-ARNr	Nop56p/Nop58p Nop1p	Appariement avec séquence cible	[24]
ARN Box H/ACA	Pseudouridylation des pré-ARNr	Cbf5p/Gar1p/ Nhp2p/Nop10p	Appariement avec séquence cible	[24]
gRNA	Editing des ARNm - Modification de l'anticodon ARN_t^{Trp}	gBP21	Appariement avec séquence cible	[25]
ARN Y	Contrôle de la biogenèse des ribosomes	Ro hnRNP I & K	*ND*	[26,27]
TER	Synthèse des télomères	TERT	Amorce pour la synthèse d'ADN des télomères	[28]
meiRNA	Initiation de la méiose	Mei2	Transport de Mei2 dans le noyau	[29]
hsr Ω	Choc Thermique	Hsp90	*ND*	[30]
roX1/roX2	Stimulation de l'expression à partir du chromosome Y	MSL1/MSL2/ MSL3/MLE/MOF	Hyperacétylation de la chromatine ?	[31]
Xist	Inactivation du chromosome X	*ND*	Hypoacétylation de la chromatine	[32]
H19	Inactivation de l'allèle maternel d'Igf2	*ND*	Déméthylation de l'ADN	[33]

Tableau 1 : ***Exemples d'ARN non codants***. *ND* : Mode d'action non déterminé.

Récemment une nouvelle classe d'ARN non codants a été mise en évidence. Ce sont les ARN régulateurs, ou riborégulateurs, dont le rôle est de contrôler spécifiquement une fonction biologique. Suivant leur mode d'action, ceux-ci peuvent être regroupés en trois catégories : certains riborégulateurs séquestrent des facteurs protéiques et modulent leur fonction, d'autres agissent comme ARN antisens sur des ARN cibles et les derniers contiennent plusieurs domaines fonctionnels et sont donc des multirégulateurs. Cette thèse porte sur l'étude de deux riborégulateurs bactériens. L'un est un ARN antisens (CopA) qui contrôle le taux de réplication du plasmide R1. Le second est un ARN multifonctionnel (ARN-III) qui contrôle l'expression des gènes de virulence chez *Staphylococcus aureus*. Notre but a été de définir les mécanismes d'actions de ces ARN ainsi que de déterminer les relations existant entre leur structure et leur fonction. L'ARN antisens CopA interagit avec son ARN cible par un mécanisme plus élaboré qu'une simple hybridation entre séquences complémentaires. Le processus de fixation implique plusieurs étapes réactionnelles dont le produit final et fonctionnel adopte une topologie originale qui n'est pas un duplexe étendu. Concernant l'ARN-III, nous avons déterminé sa structure secondaire à la fois *in vitro* et *in vivo*. Celle-ci nous a permis de caractériser un domaine fonctionnel indépendant impliqué dans le contrôle de l'expression d'une toxine de *S. aureus*. Nous présenterons également des résultats préliminaires qui ont pour but d'identifier les facteurs (protéines ou ARN) reconnus par l'ARN-III.

Ces deux systèmes sont des exemples caractéristiques parmi la grande variété des mécanismes de contrôle impliquant des ARN régulateurs. Avant de décrire nos résultats, il nous semblait important de présenter les propriétés générales des riborégulateurs dans l'introduction de ce manuscrit et d'illustrer la grande diversité de leurs fonctions par quelques exemples.

II. Les riborégulateurs : mécanismes généraux

Les principaux mécanismes d'action des riborégulateurs sont schématisés dans la figure 1. Trois grandes classes peuvent être distinguées, tant chez les procaryotes que chez les eucaryotes. Cependant, les séquences génomiques maintenant disponibles pour de nombreux organismes vont probablement conduire à l'identification de nouveaux ARN régulateurs. Leur étude fonctionnelle devrait aussi mener à la caractérisation de nouvelles voies de régulation de l'expression des gènes.

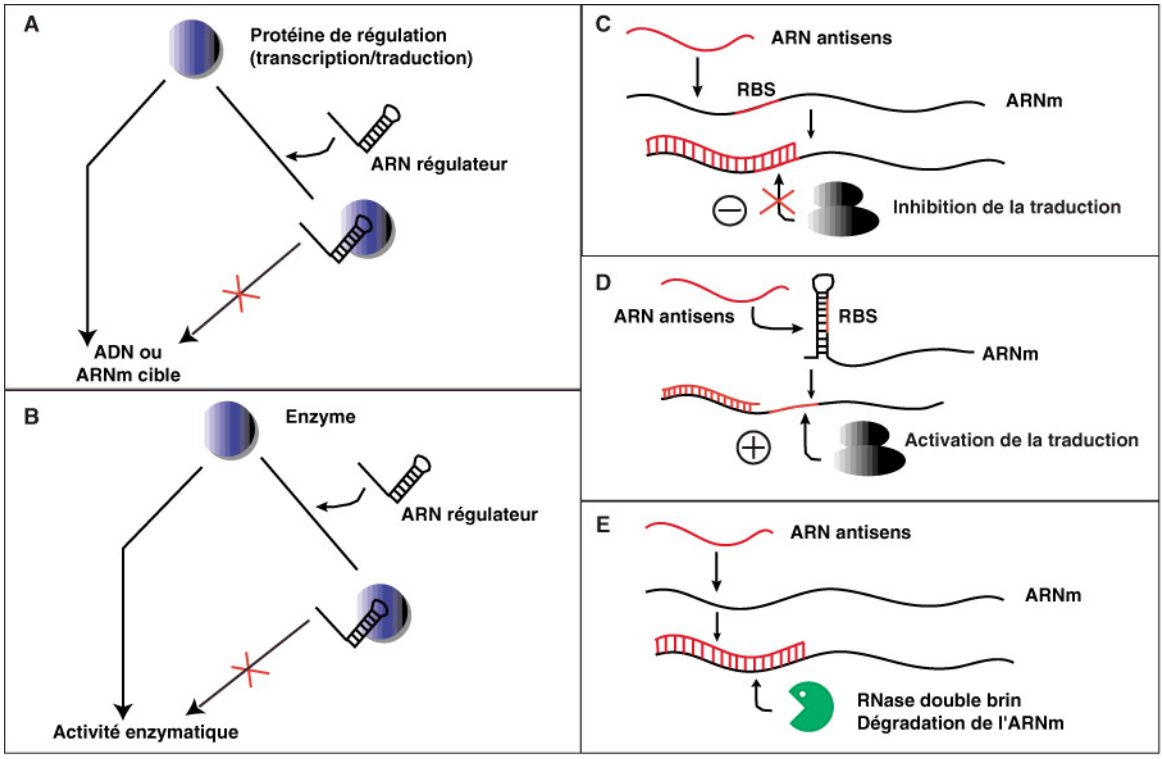

Figure 1 : ***Quelques modes d'action des riborégulateurs***. (A) et (B) Contrôle de l'activité d'une protéine par un riborégulateur. (A) Répression d'une protéine de liaison à l'ADN ou à l'ARN par un riborégulateur. La transcription ou la traduction d'un gène cible, respectivement, est inhibée par la fixation de l'ARN régulateur à la protéine activatrice. (B) Répression d'une enzyme par un riborégulateur. (C-E) Contrôles de l'expression génétique au niveau post-transcriptionnel par des ARN antisens. (C) Inhibition de la traduction d'un ARNm. Le site de fixation du ribosome est indiqué par « RBS ». Le duplexe ARN antisens-ARNm empêche la fixation du ribosome et donc la traduction. (D) Activation traductionnelle. Le RBS, séquestré dans une structure en tige-boucle, est rendu accessible par la fixation d'un ARN antisens ce qui permet la fixation du ribosome. (E) Dégradation favorisée par un ARN antisens. La fixation de l'antisens à l'ARNm crée un site de clivage par une RNase spécifique de régions bicaténaires. Les ARN peuvent aussi contrôler l'expression des gènes au niveau transcriptionnel (voir texte). Dans le cas des ARN à domaines fonctionnels multiples constituant la troisième classe de riborégulateurs, ceux-ci peuvent agir par différentes combinaisons des mécanismes représentés dans cette figure (voir texte).

A. Riborégulateurs modulant l'activité de protéines régulatrices

Un mode d'action possible d'un riborégulateur est de fixer spécifiquement un facteur protéique (Figure 1A, B). Cette séquestration de la protéine (enzyme ou facteur de régulation)

module alors son activité. Souvent, la protéine cible a un rôle central dans le contrôle de l'expression de plusieurs gènes. L'ARN joue par conséquent un rôle indirect sur la régulation de ces derniers, mais ceci de façon pléiotropique. A titre d'exemples, nous décrirons le mécanisme de contrôle de la biosynthèse du glycogène chez les bactéries et le mécanisme viral de résistance aux interférons chez les mammifères (§ III). De manière intéressante, cette classe de riborégulateurs tire profit des propriétés d'interaction avec l'ARN de leur protéine cible. Dans ces deux systèmes, l'ARN régulateur joue le rôle d'inhibiteur compétitif. Il va en effet moduler l'activité de la protéine grâce à de subtils mécanismes de mimétisme moléculaire avec le substrat naturel, tant au niveau de la séquence que de la structure.

B. Les ARN antisens

Parmi les ARN régulateurs, les ARN antisens représentent la classe majoritaire. C'est en 1972 que l'intervention d'un ARN antisens dans le contrôle d'une fonction biologique a pour la première fois été postulée grâce à la caractérisation d'une transcription bidirectionnelle d'un gène du bactériophage λ [34]. Il a fallu attendre près d'une dizaine d'années pour que ce nouveau mode de contrôle soit démontré dans le cas de la réplication du plasmide bactérien ColE1, où un ARN antisens régule la maturation de l'amorce de réplication du plasmide [35,36]. La plupart d'entre eux ont été caractérisés dans les organismes procaryotiques et plus précisément dans leurs éléments accessoires (plasmides, phages et transposons : pour une revue : [37]). Quelques exemples ont été décrits chez les archaebactéries [38] et chez les eucaryotes, notamment dans le contrôle du développement chez *Dictyostelium discoideum* [39] et *Caernorhabditis elegans* [40,41]. Les ARN antisens naturels identifiés jusqu'alors chez les procaryotes sont généralement courts (environ 100 nucléotides), non traduits, diffusibles et hautement structurés. Dans les cas décrits chez les eucaryotes, les antisens sont généralement plus longs (> 1 kb, à deux exceptions près, les ARN Lin-4 et Let-7 caractérisés chez *C. elegans*) et peuvent contenir une phase ouverte de lecture. Tous interagissent spécifiquement avec des séquences complémentaires d'ARN cibles pour contrôler une fonction biologique. Ils sont le plus souvent synthétisés à partir du même locus génétique que l'ARN cible, à partir du brin d'ADN opposé. Dans ce cas, la complémentarité avec leur cible est parfaite. Ceci est généralement le cas pour les ARN antisens codés à partir d'éléments génétiques accessoires chez les procaryotes. Ils peuvent également être synthétisés à partir d'un locus distinct. Dans ce cas la complémentarité entre les deux ARN n'est que partielle. Cette situation est généralement observée pour les antisens codés à partir d'un locus chromosomique.

Les mécanismes d'action des ARN antisens sont multiples et sont présentés dans la figure 1C-E. Ils sont bien documentés chez les procaryotes et sont moins bien connus chez les eucaryotes. Les ARN antisens peuvent contrôler chaque étape de l'expression des gènes.

Le plus souvent, l'interaction ARN antisens–ARNm cible réprime l'initiation de la traduction de l'ARNm (Figure 1C) soit en inhibant la fixation du ribosome [42-46] soit en inhibant la formation d'une structure activatrice de l'initiation de la traduction [47,48]. Il faut également noter que deux ARN antisens activateurs de la traduction ont été décrits à l'heure actuelle (Figure 1D) : l'ARN DsrA d'*E. coli* [49] et l'ARN-III de *S. aureus* [50]. Les ARN antisens peuvent également affecter la stabilité d'un ARN en augmentant sa susceptibilité vis-à-vis d'une endoribonucléase (Figure 1E) [38,39,46,51-55]. Enfin, ils peuvent moduler la transcription d'un gène par des mécanismes d'atténuation [56,57] ou inhiber la formation d'un complexe amorce/matrice productif lors de l'initiation de la réplication du plasmide bactérien ColEI [58].

Ces différents modes de contrôle de l'expression des gènes par des ARN antisens seront illustrés au § V. Parmi les antisens naturels identifiés, un grand nombre régule des fonctions plasmidiques chez les procaryotes. Un des objets de cette thèse étant l'étude d'un ARN antisens contrôlant le taux de réplication d'un plasmide, une attention particulière sera donné à ces systèmes à la fin de cette introduction (§ VI).

C. Les riborégulateurs multifonctionnels

La plupart des ARN régulateurs identifiés à ce jour interviennent par l'un des mécanismes décrits ci-dessus et n'agissent que sur une cible unique. Quelques exemples ont néanmoins été décrits où un seul et même ARN peut agir de façon pléiotropique en contrôlant l'expression de divers gènes. Dans certains cas bien précis, il peut y avoir nécessité pour une cellule de réguler un ensemble de gènes de façon coordonnée, en particulier lorsque les cellules sont soumises à un stress ou plus généralement lors de conditions de croissance particulières. Ceci peut être effectué en synthétisant une molécule d'ARN régulatrice unique pouvant interagir avec des cibles variées. Ces ARN sont généralement au cœur de cascades d'intégration de signaux extracellulaires. A l'heure actuelle, seuls trois régulateurs de ce type ont été caractérisés. Les mode d'action des ARN OxyS et DsrA identifiés chez *E. coli* seront décrits dans cette introduction. Par ailleurs, l'étude de l'ARN-III de *S. aureus* étant l'un des objets de cette thèse, ce système ainsi que nos résultats seront présentés dans le deuxième chapitre de la partie *Résultats et Discussion*.

D. Pourquoi utiliser des riborégulateurs ?

1. *Economie d'énergie et temps de synthèse du riborégulateur*

Le premier argument en faveur d'une régulation par une molécule d'ARN plutôt que par une protéine est l'économie d'énergie. Cette notion est applicable à tous les systèmes de riborégulation décrits ci-dessus. Il est effectivement beaucoup moins coûteux pour la cellule de synthétiser un ARN intrinsèquement fonctionnel, plutôt que de synthétiser un ARN messager devant être traduit en protéine qui pourrait remplir une fonction équivalente. Par ailleurs, au moins chez les procaryotes, les riborégulateurs caractérisés sont généralement de petite taille (< 500 nts). Les phases ouvertes de lecture des ARN messagers sont généralement plus grandes, auxquelles il faut ajouter les régions 5' et 3' non traduites. De plus, s'il est moins coûteux en énergie de synthétiser une molécule d'ARN plutôt qu'une protéine, le temps de synthèse est lui aussi beaucoup plus court.

2. *Dégradation du régulateur*

Si la vitesse de synthèse des riborégulateurs est primordiale pour leur efficacité, la rapidité à laquelle ils sont dégradés l'est également. En jouant à la fois sur la synthèse et la dégradation du régulateur, la fonction biologique peut être modulée de façon précise en réponse à des signaux divers. Ainsi, de nombreuses ribonucléases jouent un rôle important dans les mécanismes de riborégulation [53,59,60]. Dans le cas de réponses des cellules à des stress, l'ARN régulateur impliqué est généralement stable de manière à protéger les cellules contre le stress et à adapter la croissance cellulaire. En revanche, les ARN contrôlant le taux de réplication de plasmides sont hautement instables et synthétisés constitutivement de manière à suivre en permanence et à ajuster la concentration plasmidique si nécessaire. Ce sont généralement des caractéristiques structurales des ARN régulateurs qui vont moduler leur stabilité.

3. *Versatilité des structures des riborégulateurs*

Dans tous les cas décrits, la structure des ARN antisens est cruciale et optimisée pour leur activité en leur garantissant une interaction rapide avec leur ARN cible [61-64], (pour une revue, [65]). Dans tous ces systèmes, il y a une corrélation stricte entre la vitesse d'interaction avec la cible *in vitro* et l'efficacité du contrôle *in vivo*. De manière intéressante, dans la plupart des systèmes, l'ARN antisens et/ou sa cible contient un motif en tige-boucle par

lequel l'interaction est initiée. Le faible nombre de nucléotides non appariés présents dans les boucles permet probablement d'augmenter la spécificité de reconnaissance antisens-cible en diminuant le nombre de points de nucléation. De plus, les boucles peuvent adopter des topologies facilitant les interactions ARN-ARN. Ainsi, il a été suggéré pour de nombreux systèmes procaryotiques que la boucle par laquelle la reconnaissance est effectuée adopte une structure particulière [66]. L'ARN antisens ou l'ARN cible contiendrait un motif de type U-turn (5'-YUNR-3' où Y=pyrimidine, N=nucléotide quelconque et R=purine) trouvé dans la boucle de l'anticodon des ARNt et connu pour promouvoir l'appariement codon-anticodon lors de la traduction [67].

4. *Rapidité d'action*

Les riborégulateurs étant synthétisés rapidement, leur utilisation dans une cascade régulatrice permet à la cellule de contrôler une fonction biologique dans un intervalle de temps beaucoup plus fin. L'efficacité des riborégulateurs est par ailleurs accrue par l'observation que la plupart d'entre eux agissent au niveau post-transcriptionnel. Ces propriétés permettent ainsi un contrôle rapide et optimal. Dans les exemples qui seront décrits plus loin, cet aspect apparaît primordial. Par exemple, l'expression des gènes contrôlant le métabolisme du glycogène chez *E. coli* doit être contrôlée de façon précise en fin de phase exponentielle. De même, l'expression des gènes de virulence chez *E. carotovora ssp. carotovora* et *S. aureus* doit être régulée de manière fine en fonction de la phase de croissance. Les ARN OxyS et DsrA quant à eux interviennent dans des cascades de réponse à différents stress. Leur action rapide sur l'expression des gènes cibles est donc essentielle pour la survie des cellules. Finalement, comme cela sera détaillé plus loin, les ARN antisens régulant le taux de réplication des plasmides bactériens doivent également agir dans une fenêtre de temps très étroite.

E. Les ARN régulateurs comme outils génétiques

1. *Développement de riborégulateurs artificiels*

a. Les ARN antisens artificiels

Depuis l'identification des premiers ARN antisens il y a près d'une trentaine d'années, de nombreux oligonucléotides (ADN ou ARN) antisens artificiels à but thérapeutique ont été élaborés (pour une revue, [68]). Ces systèmes artificiels sont généralement conçus pour

inhiber la traduction du gène d'intérêt et promouvoir sa dégradation. Cependant, d'autres fonctions cellulaires comme l'épissage des pré-ARNm ou la localisation cellulaire des ARNm ont également été ciblées (pour une revue, [69,70]). Peu d'oligonucléotides ADN antisens ont été retenus pour une application clinique et la plupart d'entre eux présentent des problèmes qui n'étaient pas soupçonnés. Le premier obstacle est leur mode de délivrance dans les cellules. En effet, leur nature polyanionique ne favorise pas leur internalisation à travers les membranes. Une fois au sein des cellules, un autre problème auquel il a fallu pallier a été leur stabilité vis-à-vis des endonucléases cellulaires. Des modifications chimiques au niveau de la chaîne ribose-phosphate permettent de partiellement contourner cet obstacle, néanmoins non sans effet sur leur affinité envers leur ARN cible [71]. Finalement, l'efficacité de reconnaissance entre l'oligonucléotide et son ARN cible est rarement prévisible du fait des problèmes de compartimentation et d'accessibilité de la cible. Cette étape est néanmoins cruciale pour l'activité de la molécule. La plupart des antisens artificiels ADN ou ARN ont été conçus pour que le complexe aie une énergie libre minimale, en d'autres termes qu'il soit particulièrement stable [72], ce qui n'est pas nécessairement une réalité biologique. En effet, il s'avère que les ARN antisens artificiels les plus efficaces *in vivo* sont ceux possédant des cinétiques d'appariement rapides *in vitro* et non pas ceux formant les complexes les plus stables avec leur cible [73,74]. Cette caractéristique est conservée parmi les ARN antisens naturels, d'où l'intérêt de les étudier et de s'en inspirer pour la conception de molécules artificielles. Par ailleurs, la formation de duplexes étendus dans les systèmes naturels n'est pas requise pour l'efficacité des antisens (pour une revue, [65]). Durant de nombreuses années, les ARN antisens artificiels utilisés étaient long de plusieurs kilobases et l'importance de leur structure ou de celle de la région ciblée de l'ARNm n'a jamais été prise en compte. Une sélection d'ARN antisens artificiels en fonction de leur taille a récemment montré que de courts ARN ($\approx$ 70 nts) sont 100 fois plus efficaces que des longs fragments et que leur structure est déterminante pour leur activité [75]. Par ailleurs, une étude comparative a permis de révéler que près de la moitié des ARNm ciblés de façon efficace contiennent un motif conservé 5'-GGGA-3', localisé dans une structure particulière [76]. Finalement, une étude récente a permis d'identifier un ARN antisens artificiel efficace dirigé contre l'ARNm *BCL-ABL* dont le produit est impliqué dans des leucémies [77]. Cet ARN antisens est court (103nts), très structuré et l'interaction initiale avec l'ARNm fait intervenir huit appariements impliquant la région 5' simple brine de l'ARN antisens. Après l'interaction initiale, les appariements se propagent dans l'hélice de l'ARN antisens. Ce type de mécanisme ressemble à ceux décrits dans les systèmes CopA-CopT du plasmide R1 (§ VII) et ARN-IN-ARN-OUT de l'élément transposable IS10 (§V.B.2.a) et de la même manière, nécessite des nucléotides non appariés dans la tige de l'ARN antisens. La constante d'association ARN-ARN atteint

celle observée pour les systèmes naturels (*i.e.* ≈ 10^6 $M^{-1}.s^{-1}$) et est corrélée avec une bonne efficacité de l'antisens dans des cellules HeLa [77]. Il est donc intéressant de noter que les systèmes artificiels les plus efficaces présentent des caractéristiques structurales et cinétiques très proches des ARN antisens naturels.

b. Les aptamères

La technique de sélection *in vitro* est largement utilisée pour rechercher de nouveaux oligonucléotides se fixant spécifiquement à des protéines ou à des ARN cibles de diverses fonctions (pour une revue, [78-80]). Le but est de modifier les propriétés de la cible, ceci pour des applications thérapeutiques. A l'heure actuelle, aucun aptamère n'a cependant été retenu pour un traitement clinique. En effet, les aptamères présentent les mêmes inconvénients que ceux décrits pour les ARN antisens artificiels (§ E.1.a).

2. *Vers d'autres types de régulations ARN-dépendantes : l'exemple de l'ARN interférence*

a. Historique de l'ARN interférence (ARNi)

A l'origine, la découverte de ce phénomène biologique a été fortuite. En effet, deux groupes ont constaté que les injections d'ARN monocaténaires sens *ou* antisens dans les oocytes de *Caenorhabtidis elegans* étaient aussi efficaces pour inhiber spécifiquement l'expression d'un gène cible [81,82]. Si l'effet de l'ARN antisens était attendu, celle de l'ARN sens était plus surprenante. Fire et collaborateurs ont plus tard montré que les effets des ARN antisens et sens étaient tout deux dus à une contamination des milieux d'injection par de l'ARN double-brin (ARNdb) issu de la transcription par les ARN polymérases phagiques utilisées [83]. Ils ont appelé ce phénomène ARN interférence, ou ARNi. Il consiste en l'inhibition séquence-spécifique de l'expression de gènes par de l'ARNdb. Ce mécanisme est fascinant dans la mesure où chaque gène d'un organisme peut potentiellement être contrôlé de cette manière. Depuis il a été montré que l'ARNi fonctionne dans *Drosophila melanogaster* [84], dans *Trypanosoma brucei* [85], dans *Planaria gonocephala* [86], dans l'Hydre [87] et dans *Brachydanio rerio* [88].

Parallèlement, plusieurs groupes ont observé que l'injection de transgènes dans différentes espèces de plantes [89-91], dans le champignon *Neurospora crassa* [92], dans *Drosophila melanogaster* [93] ou dans des cellules de souris [94], conduisait à une chute d'expression du transgène et du locus chromosomique homologue correspondant. Ce

phénomène est ARN-dépendant [95] et est connu sous le nom de co-suppression ou PTGS (Post-Transcriptional Gene Silencing). Il apparaît maintenant que ces deux mécanismes d'extinction génique, à savoir ARNi et PTGS, reposent sur un mécanisme similaire impliquant des gènes homologues.

b. Modèle d'action pour l'ARNi

L'ARNi est un processus réversible puisqu'à part dans quelques rares cas [96], il n'est pas transmis à la descendance des animaux injectés au-delà de la première génération. Il a récemment été montré que lors de l'ARNi chez les plantes, l'ADN était méthylé au niveau des séquences homologues à l'ARN double-brin injecté, ceci par un mécanisme inconnu [97,98]. Cependant, le mode d'action principal de l'ARNi est post-transcriptionnel, que ce soit chez *C. elegans* [99] ou chez *T. brucei* [85]. En effet, il a été montré que seuls des ARN double-brins dirigés contre les régions codantes des gènes homologues permettaient l'ARNi, tandis que ceux dirigés contre des séquences introniques étaient inefficaces [83,85].

Une caractéristique intéressante et conservée de l'ARNi est que l'ARN double-brin introduit dans les organismes hôtes agit dans des conditions sub-stoechiométriques par rapport à son ARNm cible [83]. Ceci exclu donc un mécanisme de type antisens classique. En revanche, cette observation suggère l'existence d'une étape d'amplification de l'ARN double-brin ou que l'ARNi interviendrait par un mécanisme de type catalytique. De plus, le phénomène d'ARNi est mobile, puisque quel que soit le lieu d'injection de l'ARN double-brin, tous les tissus de l'animal ou de la plante sont touchés. Il est maintenant clairement établi que l'inhibition spécifique de l'expression d'un gène par l'ARNi est due à la dégradation de l'ARNm dépendante de l'ARN double-brin correspondant. Ceci a été montré chez *C. elegans* [83,99], *T. brucei* [85], *in vitro* dans des extraits cellulaires d'embryon de *D. melanogaster* [100], dans les cellules HeLa [101,102] et chez les plantes (pour une revue : [103]).

Plusieurs modèles d'action ont été proposés pour l'ARNi et la co-suppression mais celui intégrant le mieux toutes les données expérimentales est présenté dans la figure 2.

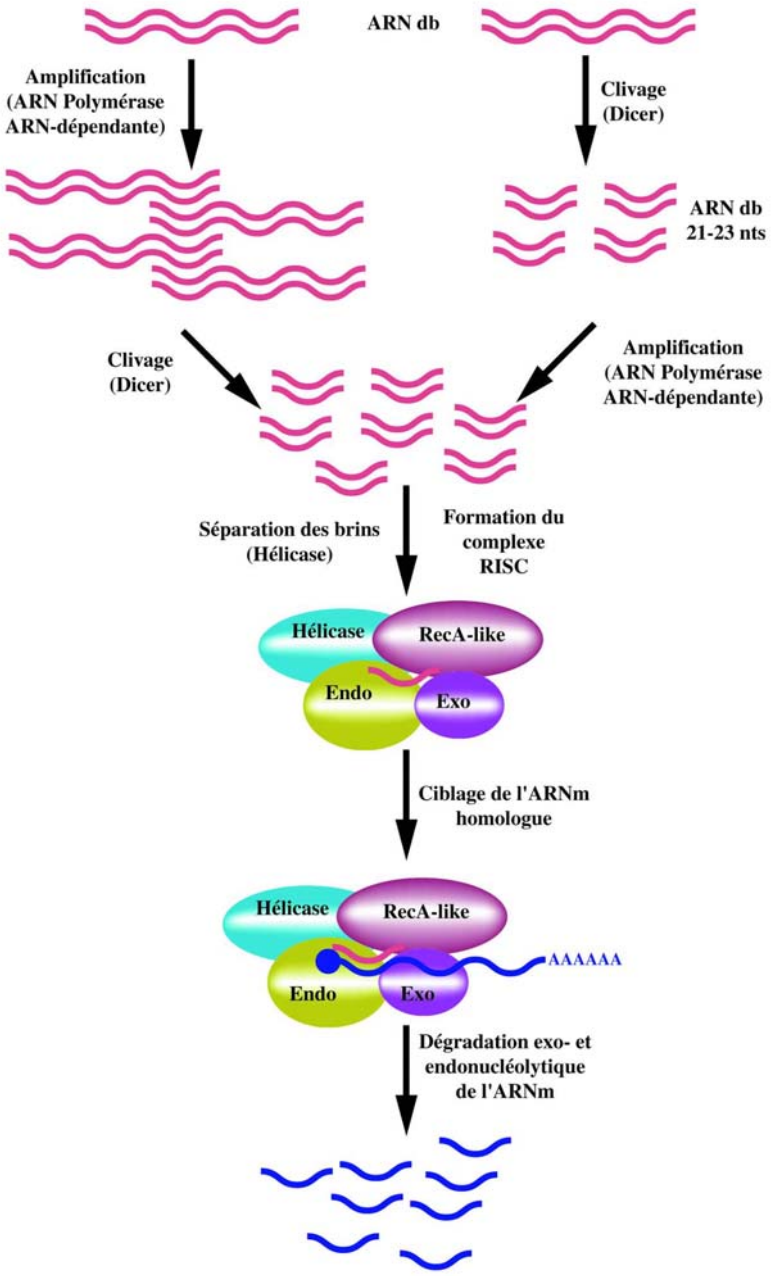

Figure 2 : ***Modèle d'action de l'ARNi d'après [104]***. Dans le cas de la première étape, les deux voies alternatives sont représentées. L'ordre dans lequel la génération de petits fragments et l'amplification de l'ARN double-brin (ARNdb) a lieu n'est pas encore établi. Les activités enzymatiques requises pour chaque étape sont indiquées entre parenthèses. Endo : Endoribonucléase ; Exo : Exoribonucléase. Voir le texte pour le détail.

Il a été montré dans des systèmes acellulaires que l'ARN double-brin introduit est clivé en fragments de 21-23 nts et amplifié [105,106]. Il ne peut cependant pas être établi laquelle des deux étapes de clivage ou d'amplification est effectuée la première. L'enzyme Dicer qui présente deux domaines RNase III est responsable de la génération des fragments de 21-23 nts chez *D. melanogaster* [107]. Ce modèle implique ensuite une séparation des deux brins des

fragments de 21-23 nts. De manière intéressante, Dicer contient un domaine à fonction ARN-hélicase qui pourrait être impliqué dans cette étape. Les ARN guides simple-brins sont alors pris en charge par le complexe RISC (RNA-induced Silencing Complex) dont la composition en facteurs protéiques reste hypothétique [107]. Il contiendrait des activités endoribonucléasique, exoribonucléasique et hélicase. De plus on pourrait imaginer la présence d'un facteur de type RecA (impliqué dans la reconnaissance de séquences d'ADN homologues dans le phénomène de recombinaison) facilitant la reconnaissance de l'ARNm homologue aux fragments de 21-23 nts et dirigeant ainsi ce dernier vers la dégradation.

Si ce modèle est correct, il repose sur une activité permettant d'amplifier l'ARNdb activateur du processus. Plusieurs gènes impliqués dans les processus d'ARNi chez les animaux, et de co-suppression chez les plantes et chez les champignons ont été caractérisés par mutagenèse aléatoire. Parmi ces gènes on retrouve des homologues d'une ARN-polymérase ARN-dépendante que ce soit chez *C. elegans* (*ego1*, [108]), chez *A. thaliana* (*sgs2* et *sde1*, [109,110]) et chez *N. crassa* (*qde1*, [111]) potentiellement impliqués dans cette étape d'amplification. Très récemment, Hammond et collaborateurs ont purifié la protéine Argonaute2 à partir d'extraits cellulaires de *D. melanogaster* et ont montré qu'elle correspondait à la nucléase du complexe RISC [112]. Parallèlement, un crible génétique chez *C. elegans* a permis d'identifier le gène *mut7* qui est essentiel à l'ARNi et est homologue à la RNase D d'*E. coli* [113,114]. Son produit est donc également un bon candidat pour l'étape de dégradation de l'ARNm par le complexe RISC. Enfin, que ce soit chez *Chlamydomonas reinhardtii* (*mut6*, [115]), *N. crassa* (*qde3*, [116]) ou *A. thaliana* (*sde3*, [117]) des gènes codant pour des enzymes de type hélicases sont essentiels pour les phénomènes de co-suppression et pourraient donc faire partie du complexe RISC. Un certain nombre d'autres gènes de fonction inconnue, impliqués dans les mécanismes d'ARNi ou PTGS, ont été caractérisés. Parmi eux, *rde1* de *C. elegans*, *ago1* de *A. thaliana* et *qde2* de *N. crassa* appartiennent à la famille Argonaute [118]. Ces protéines contiennent toutes un domaine d'interaction protéine-protéine de type PAZ, également trouvé dans Dicer. Ceci laisse supposer une interaction physique entre Dicer et un ou plusieurs composants du complexe RISC.

c. Rôle(s) de l'ARNi

L'ARNi est maintenant largement utilisée comme outil pour réprimer spécifiquement l'expression d'un gène, même si son mode d'action n'est pas encore clairement élucidé [119]. Par rapport aux ARN antisens artificiels décrits au § II.E.1.a, le contrôle par l'ARNi est probablement moins dépendant de la structure de l'ARN cible. De plus, dans la mesure où des

outils génétiques ne sont pas toujours disponibles, cette approche s'avère efficace pour étudier la fonction des gènes dans une grande variété d'organismes eucaryotiques. L'ARNi ne semble cependant pas fonctionner chez les procaryotes (E. G. H. Wagner, Uppsala, Suède, communication personnelle). Elle permet un criblage systématique de l'expression de gènes et peut être utilisée dans des projets de génomique fonctionnelle. Par ailleurs, il a récemment été montré que l'ARNi peut être contrôlée chez *D. melanogaster* lorsque les transgènes exprimant les ARN double-brin sont placés sous la dépendance d'un promoteur inductible [120]. Curieusement, l'ARNi n'a pas encore été décrite chez les archaebactéries et il serait intéressant de vérifier si ce phénomène est conservé dans ces organismes.

Il reste à comprendre le rôle de l'ARNi pour les cellules. Plusieurs études de mutagenèse ont montré que la plupart des gènes impliqués dans l'ARNi ou la co-suppression étaient également impliqués dans l'inhibition de la transposition chez *C. elegans* et dans la résistance des plantes aux virus [98,109,113]. Dans la mesure où les transposons sont des éléments mobiles, ils peuvent s'insérer de façon aléatoire dans le génome. La transposition doit par conséquent être contrôlée pour ne pas être délétère pour l'hôte. Un fort degré de transposition augmente la probabilité que deux transposons soient insérés à proximité l'un de l'autre, en orientations opposées. Dans ce cas, un faible degré de trans-lecture lors de la transcription des deux gènes convergents conduirait à la formation d'ARN double-brin activant l'ARNi. Ceci inactiverait le gène de la transposase et conduirait à l'inhibition de la transposition. De façon alternative, un même transposon pourrait être transcrit dans les deux orientations s'il existe un promoteur cryptique à sa proximité. En ce qui concerne les virus de plantes, un grand nombre d'entre eux ont une étape de leur cycle réplicatif lors de laquelle le génome viral se trouve sous forme ARN double-brin. Si ces molécules sont capables d'activer l'ARNi, ceci explique pourquoi la mutation de gènes impliqués dans ce processus accroît la sensibilité des plantes à l'agent viral. Dans ces deux cas l'ARNi serait un mécanisme de défense des cellules eucaryotes. Enfin, un grand nombre de gènes impliqués dans l'ARNi pourraient jouer un rôle clé dans le contrôle du développement embryonnaire chez de nombreux organismes (§ V.A.1). En effet, de courts ARN antisens sont impliqués dans le contrôle du développement embryonnaire chez *C. elegans*. Il a récemment été montré que l'homologue à Dicer de cet organisme et deux homologues à *rde1* étaient nécessaire à la maturation de courts ARN antisens à partir de structures en tiges-boucles des ARN précurseurs, ainsi qu'à leur activité régulatrice [121,122].

III. Les riborégulateurs modulant l'activité de facteurs protéiques

Les ARN régulateurs de cette sous-famille remplissent leur fonction en interagissant spécifiquement avec une protéine cible. Les exemples de tels riborégulateurs sont encore peu documentés mais les quelques systèmes connus permettent de faire ressortir des caractéristiques générales pour ces ARN (Tableau 2). Leur structure est un élément essentiel pour leur rôle de régulateur. Certains de ces ARN font partie intégrante de cascades régulatrices intégrant des signaux extracellulaires variés. Ces ARN ont alors un rôle pléiotropique en modulant l'expression de plusieurs gènes de manière coordonnée mais de façon indirecte en interagissant avec une protéine régulatrice.

Riborégulateur	Protéine Cible	Fonction contrôlée	Substrat naturel de la protéine	Organisme	Référence
Procaryotiques					
CsrB	CsrA	Biosynthèse du Glycogène	séquence SD d'ARN cibles	*E. coli*	[123]
RsmB	RsmA	Expression des gènes de virulence	séquence SD d'ARN cibles	*E. carotovora*	[124]
Prrb	ND	Expression des métabolites secondaires	séquence SD d'ARN cibles ?	*P. fluorescens*	[125]
Eucaryotiques					
VAIs	PKR	Défense antivirale	ARNdb	Adénovirus	[126]
EBERs	PKR	Défense antivirale	ARNdb	Virus d'Epstein-Barr	[127]

Tableau 2 : ***Riborégulateurs modulant l'activité de protéines.***

A. Contrôle de la biosynthèse du glycogène chez *E. coli*

Lors de la transition entre la phase exponentielle de croissance et la phase stationnaire, les bactéries non sporulantes réajustent leur statut physiologique de manière à compenser la déplétion en substrats du milieu. Une des cascade activée est celle régulant la biosynthèse du glycogène, servant de source de carbone et donc d'énergie lors de la phase stationnaire. La production de glycogène est influencée par différents gènes. Parmi eux, le gène *csrA* joue un rôle inhibiteur puisque son inactivation entraîne une augmentation de plus de vingt fois du niveau de glycogène intracellulaire. Cette augmentation repose sur la dérepression d'au moins six gènes impliqués dans la biosynthèse du glycogène [128-130]. Parallèlement à son rôle de

répresseur, CsrA contrôle positivement certains gènes impliqués dans la glycolyse tels que ceux codant pour la triose phosphate isomérase et l'énolase [128,129]. La manière dont est effectuée cette régulation positive n'est à l'heure actuelle pas comprise. Finalement CsrA a un rôle pléiotropique dans les cellules puisqu'elle influence la mobilité cellulaire [131] ainsi que la taille et les propriétés d'adhérence des cellules [132].

Le gène *csrA* code pour une protéine de 61 acides aminés contenant un motif de liaison à l'ARN de type KH [133]. Dans une souche *csrA*$^+$, les transcrits des gènes cibles impliqués dans la biosynthèse du glycogène ont une demi-vie de l'ordre d'une minute. En revanche, ces mêmes transcrits sont significativement stabilisés dans une souche *csrA*-. Ceci indique que CsrA régule l'expression des gènes cibles au niveau post-transcriptionnel et que la protéine est essentielle pour promouvoir une dégradation rapide des ARN messagers correspondants. Liu et collaborateurs ont purifié un complexe constitué d'une molécule d'ARN de 360 nts et de 18 molécules CsrA [123]. L'ARN a été purifié et l'ADNc correspondant a été cloné et séquencé. Le gène identifié a été nommé *csrB*. Le transcrit CsrB ne contient pas de phase ouverte de lecture et comporte un signal de terminaison de transcription ρ-indépendant à son extrémité 3'. Un modèle de repliement bidimensionnel de l'ARN CsrB est présenté dans la figure 3.

Une des caractéristiques les plus frappantes de cet ARN est la présence de 18 séquences imparfaitement répétées de type 5'-CAGGA(U,C,A)G-3'. Parmi elles, 17 sont présentes dans des régions non appariées et 10 sont localisées dans des boucles apicales de 5 nts fermées par une paire de bases C≡G entre le premier et le dernier nucléotide de la séquence conservée. La présence de ces 18 séquences répétées est à corréler avec l'interaction de 18 polypeptides CsrA avec CsrB et suggère que ces séquences constituent le site de fixation de CsrA sur l'ARN. Elles miment partiellement la séquence Shine et Dalgarno, postulée comme étant le site de fixation de CsrA sur ses ARNm cibles [123]. Ceux-ci doivent néanmoins comporter des déterminants additionnels afin de les différencier des autres ARNm non régulés par CsrA mais dont les séquences Shine et Dalgarno sont identiques.

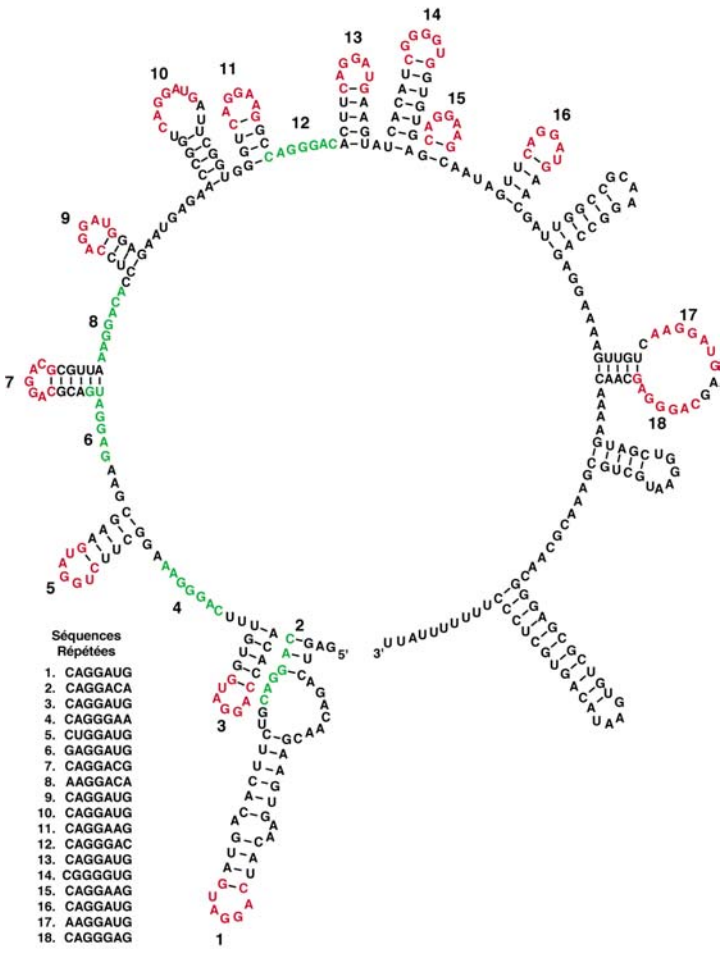

Figure 3 : ***Modèle de repliement secondaire de l'ARN CsrB d'*E. coli.** Les 18 séquences répétées sont numérotées sur la structure. Les séquences présentes dans des boucles sont indiquées en rouge tandis que les séquences présentes dans des régions non structurées ou appariées sont indiquées en vert. Un alignement des 18 séquences est représenté.

Liu et collègues ont montré que la surproduction de l'ARN CsrB dans *E. coli* conduit à un phénotype semblable à celui observé par inactivation de *csrA,* à savoir une augmentation du taux de glycogène intracellulaire [123]. Par ailleurs, les autres fonctions régulées par CsrA sont également affectées par la surexpression de CsrB. La protéine CsrA seule est plus inhibitrice *in vitro* que ne l'est le complexe CsrA-CsrB. Ensemble, ces observations suggèrent que l'ARN CsrB réprime les fonctions régulatrices de *csrA* par séquestration du facteur protéique. CsrB agit probablement comme inhibiteur compétitif de la fixation de CsrA sur ses substrats naturels. Le fait que ces séquences répétées soient préférentiellement localisées dans des boucles apicales et que cette localisation soit conservée (voir plus bas), suggère que de telles structures ont une importance pour la fonction de CsrB. Il manque cependant des évidences expérimentales confirmant cette hypothèse. Finalement, le fait que les séquences

conservées soient présentes en 18 exemplaires amplifie le contrôle négatif de la fonction de CsrA par une molécule d'ARN unique.

```
E. coli          LILTRRVGETLMIGD-EVTVTVLGVKGNQVRIGVNAPKEVSVHREEIY
A. carotovora    LILTRRVGETLIIGD-EVTVTVLGVKGNQVRIGVNAPKEVSVHREEIY
Y. enterocolitica LILTRRVGETLMIGD-EVTVTVLGVKGNQVRIGVNAPKEVSVHREEIY
P. fluorescens   LILTRRCAESLIIGDGEITVTVLGVKGNQVRIGVNAPKEVAVHREEIY
H. influenzae    LILTRKVGESVLIGD-DISITVLSVRGNQVKLGVEAPKEVSVHREEIY
B. subtilis      LVLSRKINEAIQIGA-DIEVKVIAVEGDQVKLGIDAPKHIDIHRKEIY
H. pylori        LILSRKVNEGIVIDD-NIHIKVISIDRGSVRLGFEAPESTLILRAELK
B. burgdorferi   LVLSRKANESIKINS-DIEVLILEIKKDAVKIAIKAPENIKIFRSEIY

E. coli          QRIQAEKSQQSSY--------------------
E. carotovora    QRIQAEKSQPTSY--------------------
Y. enterocolitica QRIQAEKSQPTTY--------------------
P. fluorescens   LRIKKEKDEEPSH--------------------
H. influenzae    QRIKGTKDEPYLGSS------------------
B. subtilis      LTIQEENNRAAALSSDVISALSSQKK-------
H. pylori        EAIVSENQKASVCVDESLLENIKKVIKP-----
B. burgdorferi   EFIIEENKKSLLKDKHNISKIKSLFNHYFKNEN
```

Figure 4: ***Comparaison des séquences en acides aminés des homologues à CsrA identifiés chez différents organismes procaryotiques***. Les acides aminés conservés sont encadrés (d'après [134]).

Des homologues à la protéine CsrA d'*E. coli* ont été caractérisées dans de nombreuses autres souches eubactériennes (Figure 4). Leur fonction n'est pas conservée et parmi elles, la protéine RsmA d'*Erwinia carotovora* ssp. *carotovora*, codée par le gène *rsmA,* est responsable de la répression de l'expression des gènes de virulence chez ce pathogène de plante [135] et chez de nombreuses souches voisines [136]. Tout comme CsrA chez *E. coli*, ce contrôle est post-transcriptionnel et affecte la stabilité des ARN messagers [137,138]. Les modes de régulation dépendants de *csrA/rsmA* dans ces deux organismes sont donc similaires, même si les gènes cibles ont des fonctions distinctes. En faveur de cette hypothèse, le locus *rsmB* a été caractérisé [139]. Il réprime les fonctions inhibitrices de RsmA et il code pour un ARN non traduit de 479 nts [124]. Cet ARN (RsmB) contient 9 séquences imparfaitement répétées en partie identiques aux séquences répétées de CsrB. Une protéine recombinante His_6-RsmA reconnaît spécifiquement RsmB au niveau de plusieurs sites et des mutations dans les séquences répétées abolissent la fonction de l'ARN [139]. Finalement, la surexpression du riborégulateur RsmB dans *E. coli* conduit à une augmentation de la concentration intracellulaire en glycogène. Cette dernière observation suggère que RsmB et CsrB sont des orthologues fonctionnels, même s'ils ne sont pas reliés au niveau évolutif.

Si ce mécanisme de contrôle de l'expression des gènes à deux composants est si conservé dans différentes souches eubactériennes, et s'il est impliqué dans la régulation de fonctions très diverses, c'est qu'il doit être particulièrement efficace. Apparemment, lors de l'évolution, les cellules bactériennes ont sélectionné de façon indépendante de petits ARN capables de se fixer et d'inhiber la fonction de protéines homologues entre elles. D'après les exemples décrits ci-dessus, la fonction de ces ARN repose sur la présence d'une séquence précise présentée dans une structure bien définie. Ces systèmes mettent donc en avant l'importance de la structure des riborégulateurs pour leur fonction. Même si ce n'est pas encore le cas, il est probable que d'autres ARN modulant l'activité des protéines orthologues à CsrA présentées dans la figure 4 seront identifiés à l'avenir. Récemment, un ARN partiellement homologue à CsrB a été identifié dans *Salmonella typhimurium* [140] et *Pseudomonas fluorescens* [125]. L'ARN Prrb de *P. fluorescens* [125] joue un rôle dans le contrôle de la synthèse des métabolites secondaires. Tout comme CsrB et RsmB, il contient 5 séquences imparfaitement répétées localisées dans des boucles apicales ou dans des régions simple-brin. Il a également été montré que Prrb intervient dans la même cascade régulatrice que l'orthologue à RsmA de *P. fluorescens* [141]. L'interaction ARN-protéine reste néanmoins à établir dans ce dernier exemple. Un dernier point intrigant est la présence d'un nombre variable de séquences répétées dans chacun de ces ARN homologues (18, 9 et 5 respectivement pour CsrB, RsmB et PrrB). En modulant ce nombre, les propriétés séquestrantes des ARN changent ce qui pourrait permettre d'adapter le contrôle de manière fine selon le système. Il serait également intéressant de vérifier si le nombre de séquences répétées est proportionnel à la concentration de la protéine cible dans les cellules.

B. Mécanismes viraux de résistance aux interférons.

Chez les mammifères, la réponse antivirale induite par les interférons (IFN) est la première barrière contre l'attaque par un virus. Les virus eucaryotiques ont développé de nombreux mécanismes permettant la reproduction virale sans détruire la cellule infectée. De manière a développer une infection productive, les virus ont donc à contourner l'inhibition induite par l'expression des IFN. Les IFN de type I se fixent à leurs récepteurs et activent une cascade stimulant la transcription de plus d'une trentaine de gènes [142]. Parmi eux, deux gènes codent pour des enzymes normalement exprimées à un faible degré dans les cellules eucaryotiques et activées par de l'ARN double-brin produit comme intermédiaire réplicatif du cycle viral. Il s'agit d'une part de la 2'-5'-oligoadénylate synthétase nécessaire à l'activation de la RNase L et d'autre part de la protéine kinase PKR.

PKR joue un rôle central dans les mécanismes antiviraux des cellules en réponse aux IFN de type I. Elle agit par différents mécanismes dont l'induction de l'apoptose [143,144] mais le principal étant l'inhibition de la traduction de la cellule infectée [145]. Il s'agit d'une kinase à serine-thréonine de 68 kDa appartenant à la famille des kinases de la sous-unité α du facteur eucaryotique d'initiation de la traduction, eIF-2. Elle est composée d'un domaine N-terminal régulateur et d'un domaine C-terminal catalytique et sa forme active est un dimère. Présente de façon ubiquitaire dans tous les tissus, PKR est rapidement activée par fixation d'un ARN double-brin d'origine virale ou par des ARN possédant de longs segments de structure secondaire (pour une revue : [146]). PKR fixe les ARN double-brin par ses deux domaines dsRBM présents dans sa région régulatrice N-terminale ce qui modifie la conformation du domaine catalytique et résulte en l'autophosphorylation de la protéine. PKR ainsi activée peut phosphoryler plusieurs protéines, mais sa cible majeure est la sous-unité α du facteur de traduction eIF-2. La phosphorylation de eIF-2α augmente son affinité pour le GDP de plus de 100 fois ce qui a pour conséquence d'empêcher l'échange GDP-GTP et de séquestrer toute l'activité du facteur d'échange eIF2B. La traduction au sein des cellules infectées est par conséquent inhibée.

Les virus résistants aux IFN ont développé indépendamment plusieurs stratégies pour contourner les multiples actions antivirales de PKR (pour une revue, [147]). Celles-ci sont résumées dans la figure 5.

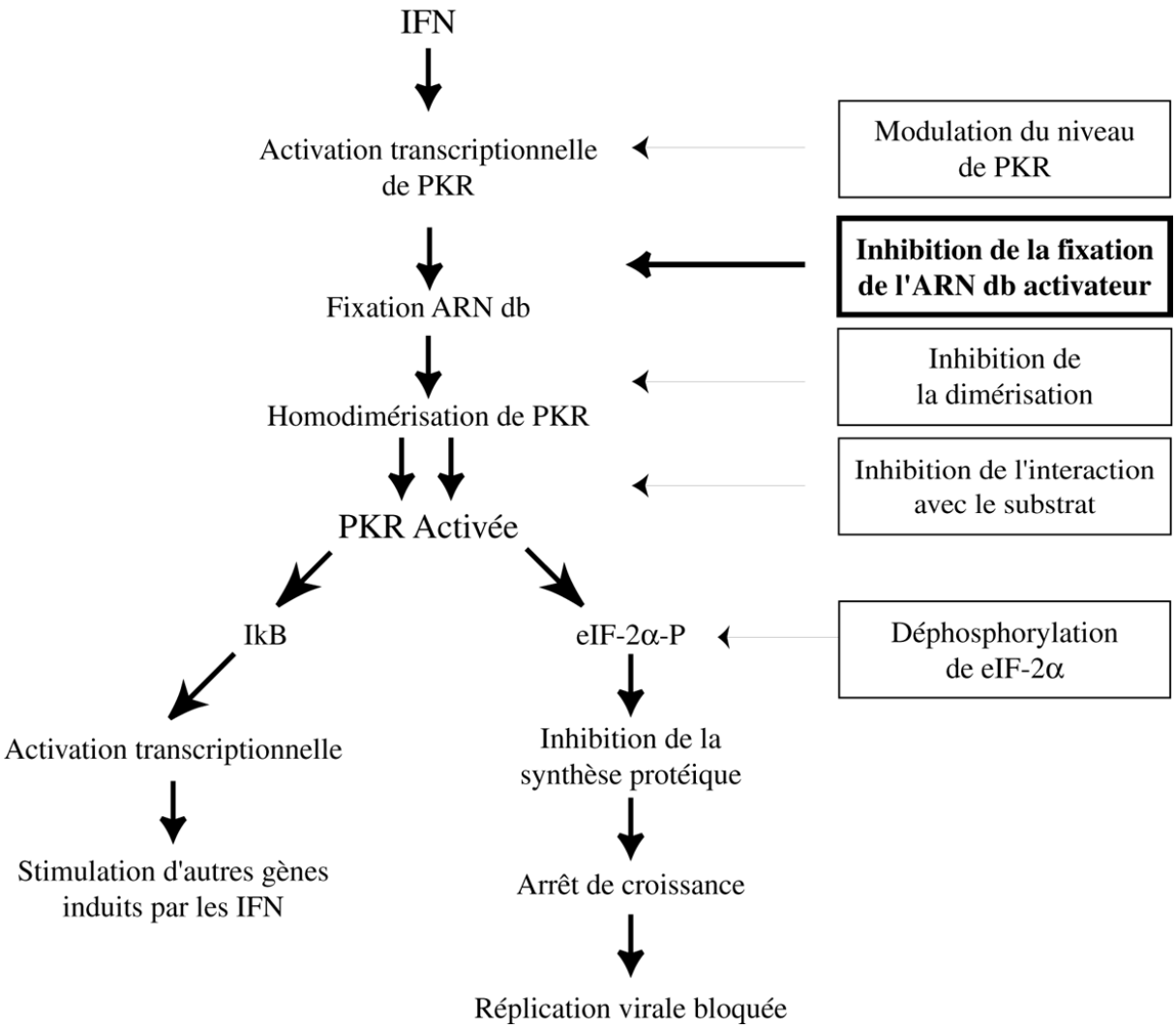

Figure 5 : ***Mécanisme de maturation de la protéine kinase PKR et sites de régulation virale***. PKR, induite au niveau transcriptionnel par les interférons (IFN), est activée par fixation d'ARNdb et par homodimérisation. Une fois activée, PKR peut phosphoryler ses substrats dont eIF-2α (voie de droite) et le facteur de transcription IκB (voie de gauche). Au sein d'une cellule infectée, la phosphorylation de eIF-2α par PKR inhibe la synthèse protéique, la croissance cellulaire et la réplication virale. Les flèches à droite indiquent les étapes du mécanisme d'activation de PKR pouvant être ciblées par les virus pour inhiber la réponse des cellules aux IFN. Le mécanisme d'inhibition décrit dans ce paragraphe est indiquée en gras.

Parmi les mécanismes viraux d'inactivation de PKR, un des plus intéressant est l'inhibition de la fixation des ARN activateurs double-brin. L'activation de PKR par de l'ARNdb est optimale pour une double hélice d'environ 70 pb, même si 40 pb permettent déjà une activation partielle [148]. Un domaine dsRBM isolé couvre 16 pb d'un ARNdb [149] et les deux dsRBM de PKR sont nécessaires pour son activation. La distance de 70 pb requise pour l'activation de la kinase reflète probablement un changement conformationnel de PKR. L'activation nécessiterait le besoin d'arranger précisément les deux dsRBM et la région charnière les connectant, ce qui ne pourrait être accompli qu'avec une taille minimale d'ARNdb. Si des fragments trop courts d'ARN bicaténaires ou des ARN hautement structurés

mais imparfaitement appariés ne peuvent pas activer PKR, ils peuvent néanmoins interagir avec les dsRBM [148]. Ainsi, certains virus ont tiré profit de cette propriété pour inhiber de façon compétitive la fixation d'ARN double-brin par PKR en synthétisant de petits ARN non codants qui bloquent l'activité de la kinase. Les ARN viraux les plus étudiés inhibant la reconnaissance PKR-ARNdb sont les ARN VAI des adénovirus. Des cellules infectées par des souches d'adénovirus mutantes délétées du gène *vai* possèdent une activité PKR élevée, présentent un haut degré de phosphorylation de eIF-2α, et sont défectives en traduction. Les ARN VAI des adénovirus comportent environ 260 nts. Ils s'accumulent à hautes concentrations dans le cytoplasme des cellules infectées où ils abolissent la fonction inhibitrice de la traduction conférée par PKR [126]. L'interaction PKR-VAI est directe et le domaine régulateur N-terminal de l'enzyme, contenant les deux dsRBM, est suffisant pour cette association [150]. Des études structurales combinées à des expériences de mutagenèse dirigée ont révélé les régions de l'ARN VAI importantes pour sa fonction [151,152]. Ainsi, le domaine central contenant les 4 appariements centraux hyperconservés GGGU :ACCC (Figure 6) est crucial pour inhiber spécifiquement PKR et est nécessaire à la réplication virale. Ces appariements constituent le cœur structural de l'ARN en maintenant les hélices II et III dans une orientation favorable à la reconnaissance par PKR. Des études comparatives entre différentes souches d'adénovirus ont révélé que les séquences des ARN VAI n'étaient que peu conservées. En revanche, l'établissement des modèles de repliements bidimensionnels déduits par approche phylogénétique ou par cartographie en solution montrent une topologie globale conservée entre les différents isolats [152]. Ces ARN doivent donc contenir une région bicaténaire suffisamment longue pour garantir la spécificité d'interaction avec les dsRBM, mais assez courte pour ne pas activer la fonction catalytique de PKR. Enfin, des études structurales similaires effectuées sur les ARN EBER du virus d'Epstein-Barr ont révélé des caractéristiques comparables aux ARN VAI [127]. L'ARN EBER-1 peut aussi complémenter la déficience en ARN VAI d'une souche d'adénovirus mutante [153] et inhiber l'activation de PKR *in vitro* [150]. Finalement, un tel mode d'action anti-PKR a été postulé pour la séquence TAR du virus de l'immuno-déficience humaine de type I (VIH-I) [154], mais est encore controversé à l'heure actuelle.

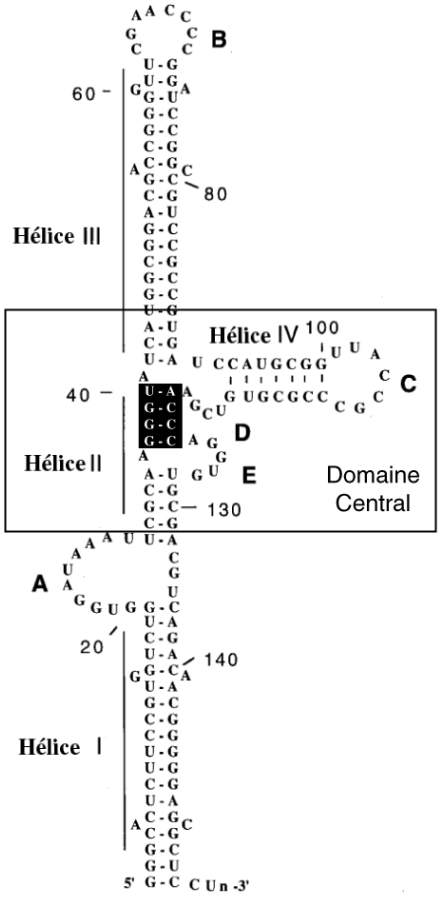

Figure 6 : ***Modèle de structure secondaire consensus des ARN VAI d'adénovirus***. Seuls les 157 premiers nucléotides sont représentés. Les hélices I à IV et les boucles A à E sont indiquées. Le domaine central est encadré et les quatre paires de bases hyperconservées sont représentées en caractères blancs sur fond noir.

Il est surprenant de constater à quel point ce mécanisme de résistance aux interférons est conservé chez les virus de mammifères. Pour le virus, il est primordial que la cascade de défense cellulaire en réponse aux IFN soit inhibée de façon très précoce dans le cycle viral. Ceci n'est permis qu'avec un nombre minimal de macromolécules pouvant être synthétisées dans un temps restreint. Un ARN est donc un régulateur de choix pour ce type de fonction.

IV. Les riborégulateurs multifonctionnels

Dans certains cas bien précis, il peut y avoir nécessité pour une cellule de réguler de façon coordonnée l'expression d'un ensemble de gènes. C'est un moyen efficace et rapide pour adapter la réponse des cellules à un stress ou plus généralement lors de conditions de croissance particulières. La synthèse d'une seule et unique molécule régulatrice aux cibles multiples peut permettre de remplir une telle fonction. Chez les procaryotes, plusieurs exemples d'ARN régulateurs multifonctionnels ont été décrits. Ceux-ci possèdent des domaines fonctionnels distincts par lesquels ils interagissent avec des cibles variées.

Deux systèmes faisant intervenir de tels ARN multifonctionnels ont été particulièrement étudiés. Le contrôle de l'expression des gènes codant pour les facteurs de virulence chez *S. aureus* étant l'objet d'une étude effectuée lors de cette thèse, il sera décrit en détail plus loin. Le système de régulation de l'expression du facteur de transcription alternatif σ^S chez *E. coli* fait intervenir au moins deux ARN multirégulateurs. Les ARN DsrA et OxyS jouent des rôles antagonistes dans ce contrôle en répondant à divers stress.

A. Implication d'ARN régulateurs dans le contrôle de l'expression du facteur de transcription alternatif σ^S

1. *Le facteur de transcription alternatif σ^S*

Comme tous les micro-organismes unicellulaires, *E. coli* a à faire face à de constantes variations des conditions environnementales. La sous-unité σ^S de l'ARN polymérase bactérienne a été caractérisée par sa capacité à induire l'expression de plus d'une cinquantaine de gènes lors de l'entrée des cellules en phase stationnaire de croissance. Depuis, le facteur σ^S, codé par le gène *rpoS*, a également été impliqué dans la réponse des cellules au stress oxydatif, aux basses températures, à une déplétion du milieu en nutriments et à un choc acide (pour une revue : [155,156]). σ^S intervient au centre de cascades d'intégration de signaux extracellulaires et est associé de façon générale avec la réponse des cellules induite par différents stress. En réponse à ces stimuli, l'expression de *rpoS* est modulée au niveau transcriptionnel par différents signaux dont le ppGpp, l'homosérine lactone, l'AMPc et l'UDP-glucose. Des régulateurs ont également été caractérisés comme modulant l'expression de *rpoS* au niveau traductionnel ou au niveau de la stabilité de la protéine [157]. La figure 7 résume les différentes voies de contrôle de l'expression de *rpoS*.

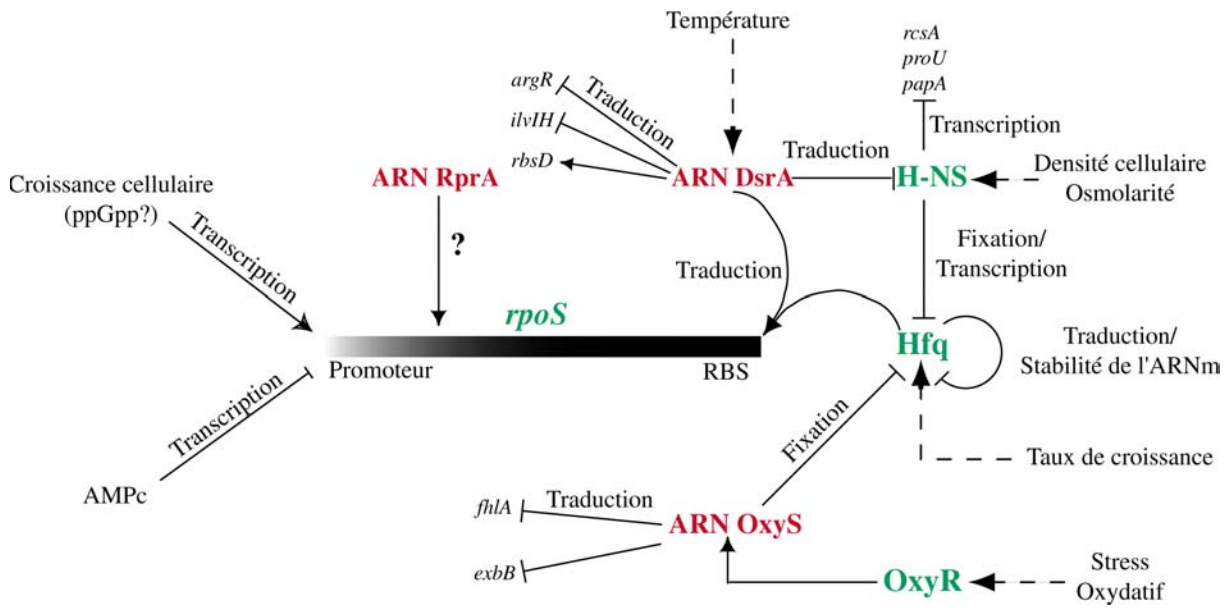

Figure 7 : ***Différentes voix de régulation du gène* rpoS *(d'après [155])***. La région 5' non traduite du gène *rpoS* est représentée. Les différents contrôles sont indiqués par des flèches pour les régulations positives et par des lignes terminées par une barre pour les régulations négatives. Les flèches en pointillés indiquent les gènes qui répondent à des signaux environnementaux. Les niveaux de l'expression de *rpoS* auxquels les différents contrôles se font sont indiqués. Les autres cibles potentielles ou caractérisées de DsrA et OxyS sont indiquées. Les facteurs globaux de régulation sont indiqués en vert. Les riborégulateurs sont indiqués en rouge.

Lors de certains stress, le taux d'ARNm *rpoS* ne varie que faiblement [157,158]. En revanche, la quantité de facteur σ^S fluctue de façon significative. Ceci laisse suggérer que l'essentiel du contrôle de l'expression de *rpoS* repose sur des mécanismes de contrôle au niveau post-transcriptionnel. Deux facteurs protéiques influencent la traduction de *rpoS* de manière divergente. Il s'agit de Hfq et de H-NS. Parallèlement, trois ARN contrôlent l'expression de *rpoS*. Il s'agit des ARN OxyS, DsrA et RprA.

2. *La protéine Hfq*

La délétion du gène *hfq* chez *E. coli* cause des phénotypes sévères et pléiotropiques parmi lesquels une diminution du taux de croissance et une augmentation de la susceptibilité des cellules au stress osmotique et à la lumière ultraviolette. Ces effets ressemblent en partie à ceux préalablement observés lors de l'inactivation de *rpoS*. En faveur d'une relation entre *hfq* et *rpoS*, il a été montré que la protéine Hfq était essentielle à la traduction de *rpoS* [159]. Des mutations diminuant ou éliminant l'effet de Hfq sur la traduction de *rpoS* ont été identifiées chez *Salmonella typhimurium* [160]. Elles sont toutes localisées dans la structure en tige-boucle présente en amont de la phase codante de *rpoS (*Figure 8) et qui contient le site de

fixation du ribosome. Hfq serait nécessaire pour reconnaître et déstabiliser cette structure, activant la traduction de *rpoS* [160].

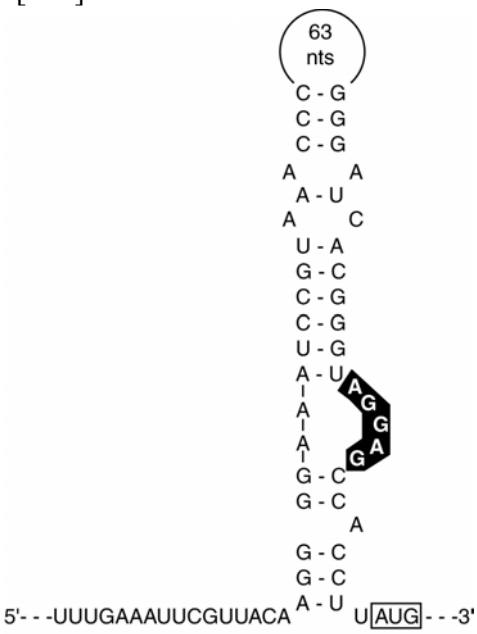

Figure 8 : ***Modèle de repliement secondaire de la région 5' non traduite de l'ARNm de* rpoS.** La séquence Shine et Dalgarno est représentée en caractères blancs sur fond noir et le codon d'initiation de la traduction est encadré.

3. *La protéine H-NS*

H-NS est un facteur de transcription qui inhibe spécifiquement l'expression d'un grand nombre de gènes. Un défaut d'expression de *hns* conduit à une augmentation du taux de σ^S dans les cellules. Même s'il a été montré que H-NS est une protéine se fixant à l'ADN, l'effet sur l'expression de *rpoS* est post-transcriptionnel puisque des mutants *hns* présentent une traduction de *rpoS* augmentée [161] et une stabilité accrue de la protéine σ^S [162]. La protéine H-NS s'associe à Hfq *in vitro* [163], suggérant que H-NS séquestre Hfq *in vivo* en l'empêchant ainsi de remplir son rôle activateur de la traduction de *rpoS*.

4. *Contrôle de l'expression de* rpoS *par l'ARN DsrA*

a. DsrA répresseur de *hns*

L'ARN DsrA a initialement été identifié par son habilité à accroître la synthèse des polysaccharides de la capsule chez *E. coli* lorsqu'il était surexprimé [164]. La synthèse des polysaccharides repose sur différents facteurs dont l'expression est sous la dépendance de la protéine régulatrice de transcription RcsA. Le gène *rcsA* est lui-même réprimé au niveau

transcriptionnel par la protéine H-NS et la surproduction de DsrA mime les effets observés par une délétion de *hns*. Le fait que la surproduction de DsrA induise l'expression des gènes *proU* et *papA*, également connus comme étant réprimé par H-NS, suggère que DsrA augmente la transcription de *rcsA* de manière indirecte en inhibant l'activité de H-NS.

DsrA est un ARN abondant et stable de 87 nts contenant trois structures en tige-boucle dont la troisième est un signal de terminaison de transcription ρ-indépendant (Figure 9). Le contrôle de l'expression de *hns* par DsrA se fait au niveau post-transcriptionnel. Des expériences de mutagenèse dirigée et de restauration par mutations compensatoires ont montré que DsrA agissait par appariement de type ARN-ARN pour remplir sa fonction régulatrice [165]. Il contient dans sa deuxième structure en tige-boucle deux régions de complémentarité à l'ARNm *hns* (Figure 9). La première couvre 13 pb, 4 nucléotides en aval du codon initiateur de la traduction de *hns* et la seconde couvre 9 pb immédiatement en amont du codon de terminaison de la traduction [49,165] (Figure 9A). Cette interaction bipartite entre DsrA et l'ARNm H-NS a été confirmée par cartographie en solution *in vitro* [49]. Le contrôle de l'expression de *hns* par DsrA se fait donc par un mécanisme antisens. Comme cela est représenté dans la figure 9B, l'interaction DsrA-*hns* génère un empilement coaxial des deux hélices intermoléculaires, circularisant ainsi l'ARNm *hns* et exposant sa partie centrale. L'effet de DsrA sur *hns* se situe donc à deux niveaux : d'une part en empêchant la reconnaissance de l'ARN messager par le ribosome et d'autre part en favorisant sa dégradation [49]. Il est à noter que ce dernier effet peut être imputé à l'absence de protection par les ribosomes en cours de traduction. DsrA a donc un effet positif indirect sur l'expression de *rpoS* en contrôlant directement le taux du répresseur H-NS au niveau traductionnel.

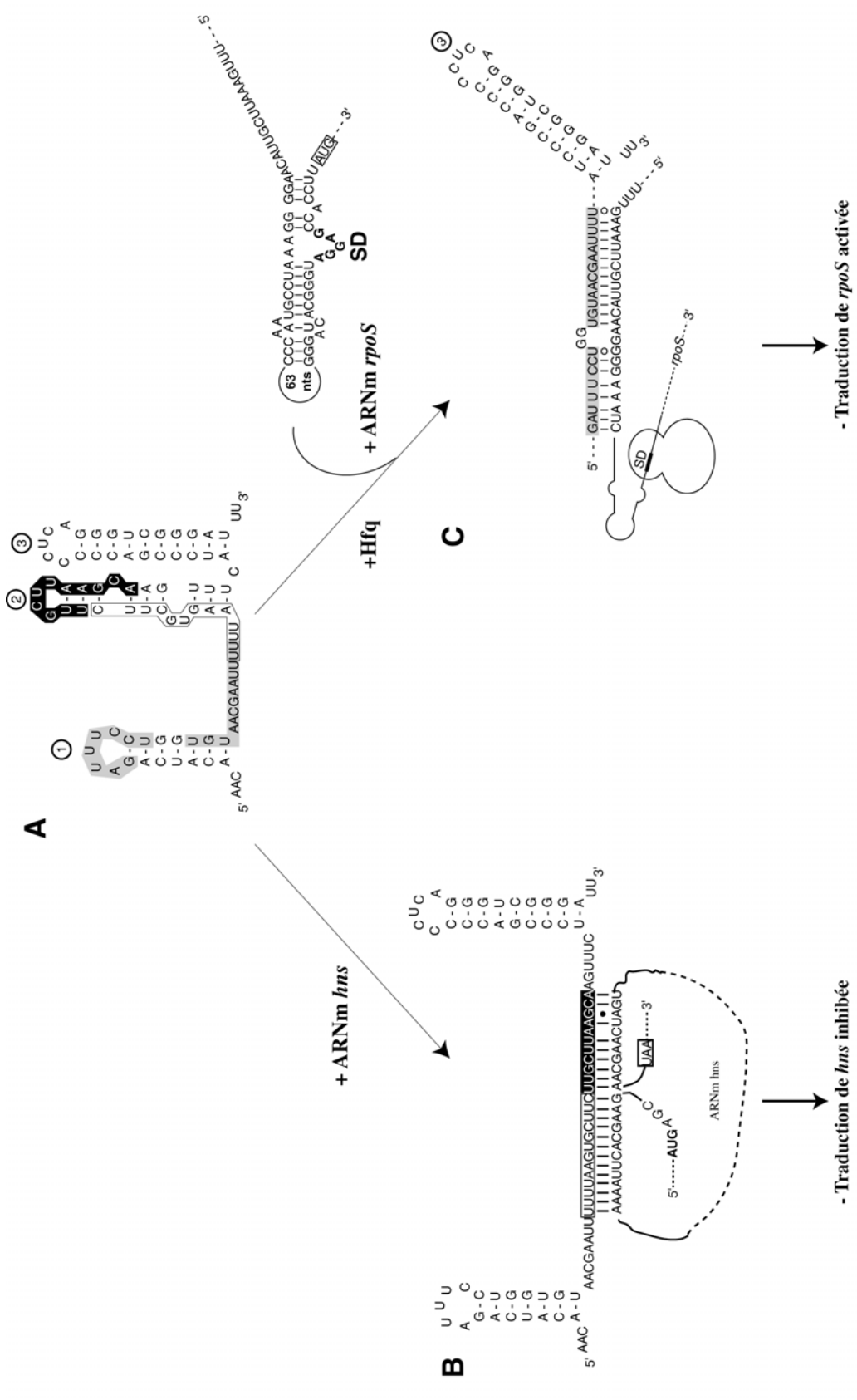

Figure 9 : *Modèle de stucture secondaire et mécanismes d'action de DsrA*. (A) Modèle de structure secondaire de l'ARN DsrA. Les trois tiges-boucles sont numérotées. Les régions complémentaires à *rpoS* et complémentaires aux deux séquences de *hns* sont indiquées en grisées, en noir et encadrées, respectivement. (B) Modèle de l'interaction entre DsrA et l'ARNm *hns*. Les codons initiateur et de terminaison de *hns* sont respectivement en gras et encadrés. (C) Modèle de l'interaction entre DsrA et l'ARNm *rpoS*. La séquence Shine et Dalgarno (SD) est indiquée en gras et le codon initiateur AUG est encadré. Le ribosome reconnaissant *rpoS* est représenté.

b. DsrA active directement l'expression de *rpoS*

La quantité de σ^S est normalement peu élevée dans les cellules en phase exponentielle de croissance à 37°C. En revanche, σ^S s'accumule en phase exponentielle à 20°C, ce qui coïncide avec une expression optimale de DsrA [166]. A faible température, DsrA est essentiel à l'activation de la traduction de *rpoS* et ceci de façon indépendante de son rôle inhibiteur sur celle de *hns* [166]. Il a été proposé qu'à faible température chez *Salmonella typhimurium*, l'ARNm *rpoS* adoptait une structure secondaire stable séquestrant les signaux d'initiation de la traduction [160] (Figure 8). Des analyses informatiques ont révélé une complémentarité entre une séquence de DsrA et une séquence de la structure inhibitrice de *rpoS* [165] (Figure 9). Des études génétiques ont validé cette interaction [165,167]. L'appariement DsrA-ARNm *rpoS* libère la région contenant la séquence Shine et Dalgarno de ce dernier et stimule sa traduction par un mécanisme anti-antisens (Figure 9C). On peut néanmoins s'interroger sur les raisons pour lesquelles dans ces conditions, la protéine Hfq est nécessaire [168]. En fait, Hfq interagit avec l'ARN *rpoS* mais également avec DsrA [169]. Hfq pourrait agir comme cofacteur pour les activités régulatrices de DsrA soit en jouant sur la structure des ARN, soit en formant un complexe ribonucléoprotéique actif.

Ainsi, DsrA régule *rpoS* d'une part directement par un mécanisme anti-antisens et d'autre part de façon indirecte en inhibant la synthèse de H-NS. Ces actions conduisent à une stimulation amplifiée de la synthèse de σ^S. Les effets répresseurs et activateurs de DsrA sur la traduction de respectivement *hns* et *rpoS*, peuvent être génétiquement séparés [165,167]. L'effet direct de DsrA sur *rpoS* repose uniquement sur la première structure en tige-boucle du régulateur [167], tandis que la seconde tige-boucle est déterminante pour l'inhibition de la traduction de *hns*. DsrA contient donc des domaines fonctionnels séparés jouant des rôles spécifiques dans le contrôle de l'expression de gènes cibles, ceci par interactions de type ARN-ARN. Les interactions entre DsrA et ses cibles, qu'elles soient établies ou hypothétiques, reposent toutes sur la première ou la seconde structure en tige-boucle. On peut également s'interroger sur l'importance de ces régions d'un point de vue structural comme cela a été démontré pour d'autres riborégulateurs, ou si seule la complémentarité entre DsrA et ses cibles est nécessaire.

5. *L'ARN RprA*

Récemment, un court ARN de 106 nts a été identifié par sa capacité à activer l'expression de *rpoS in vivo* en absence de DsrA [170]. L'ARN RprA est conservé entre différentes souches de Salmonella et Klebsiella, suggérant qu'il remplit une fonction

régulatrice importante. Il est non essentiel, mais sa délétion interfère avec la réponse des cellules aux chocs osmotiques lorsque DsrA est absent. Tout comme DsrA, il contient trois structures en tige-boucle. L'implication de ces structures sur sa fonction n'est cependant pas connue et reste à déterminer. Enfin, RprA et DsrA agissent probablement sur l'expression de *rpoS* par des mécanismes distincts. Contrairement à DsrA, RprA ne contient pas de régions complémentaires à la région 5'-UTR de l'ARNm *rpoS*. La découverte récente de RprA soulève un point intéressant. En effet, l'existence d'autres riborégulateurs positifs de l'expression de *rpoS* est à envisager, chacun pouvant répondre à un autre type de stress en absence de la synthèse de DsrA. Finalement, tout comme DsrA, RprA régule peut-être l'expression de plusieurs gènes cibles. Ceux-ci restent cependant à caractériser.

6. *L'ARN OxyS répresseur de* rpoS

L'ARN OxyS a été identifié lors de l'exposition de cellules d'*E. coli* à un stress oxydatif et protège les cellules contre des mutations spontanées ou induites [171]. La transcription d'OxyS est sous la dépendance du facteur protéique OxyR. L'ARN est stable et abondant et ne code pour aucune protéine. Chez *E. coli,* il contrôle positivement ou négativement l'expression de plus d'un quarantaine de gènes cibles [171]. L'ARN compte 109 nts et sa structure secondaire a été déterminée *in vivo.* Il contient trois structures en tige-boucle et une grande séquence connectrice entre la seconde et la troisième tige-boucle (Figure 10 [171]). La tige-boucle C est un terminateur de transcription ρ-indépendant et est impliqué dans la stabilité de l'ARN. Un rôle similaire peut également être envisagé pour la première structure en tige-boucle mais n'est pas démontré à l'heure actuelle.

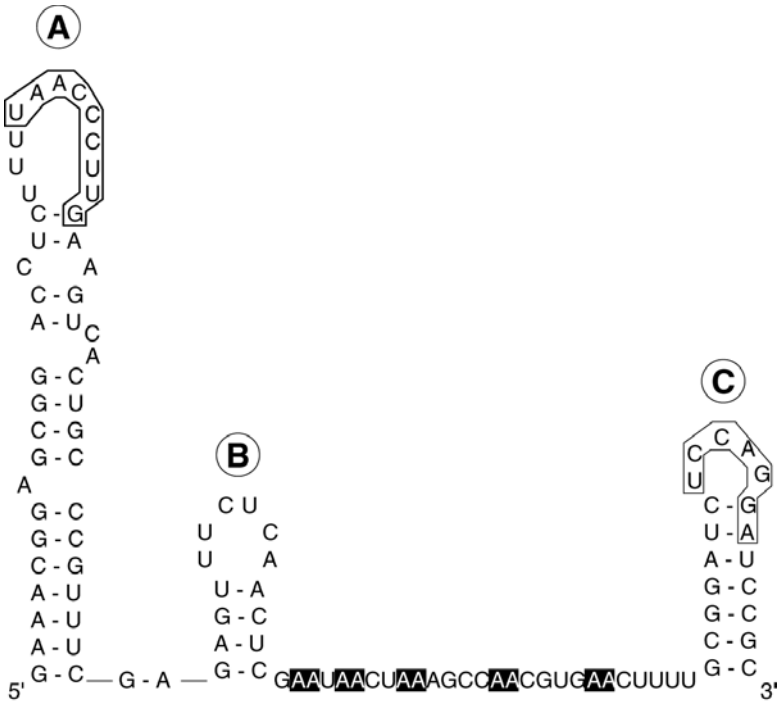

Figure 10 : ***Modèle de structure secondaire de l'ARN OxyS***. Les trois structures en tige-boucle sont indiquées A à C. Les régions de complémentarité à l'ARNm *fhlA* sont encadrées et les séquences dinucléotidiques AA sont indiquées en caractères blancs sur fond noir.

Comme précisé plus haut, la traduction de *rpoS* est sous la dépendance stricte de la protéine Hfq. L'ARN OxyS inhibe l'expression de *rpoS* de façon post-transcriptionnelle et cette inhibition dépend de la protéine Hfq [172]. OxyS interagit avec la protéine Hfq *in vivo* et *in vitro* suggérant qu'OxyS la séquestre et l'empêche de remplir son rôle activateur de la traduction de *rpoS*. L'interaction OxyS-Hfq n'empêche cependant pas la reconnaissance de l'ARNm *rpoS* par Hfq ce qui conduisit Storz et collaborateurs à émettre l'hypothèse que le complexe ternaire OxyS-Hfq-*rpoS* inactiverait la protéine Hfq [172]. La protéine Hfq fixe préférentiellement des régions simple-brin riches en A [173]. De manière intéressante, la région connectrice d'OxyS contient cinq répétitions du dinucléotide AA (Figure 10) et leur mutation en dinucléotide GG diminue la fonction inhibitrice d'OxyS [172]. Comme seuls des ARN OxyS tronqués possédant cette région conservent la capacité d'inhiber la traduction de *rpoS*, il est proposé que la région de connexion confère la spécificité de l'interaction OxyS-Hfq [43]. Finalement, l'intérêt de réprimer l'expression de *rpoS* lors d'un stress oxydatif peut paraître obscur, sachant que σ^S est impliqué dans l'activation d'un grand nombre de gènes de défense contre ce type de stress. Une explication serait qu'OxyS étant exprimé quand le facteur de transcription OxyR est activé, OxyS permettrait d'éliminer une activation inutile des gènes à la fois sous le contrôle d'OxyR et de σ^S.

B. Rôle central de DsrA dans le contrôle de l'expression génétique

Le facteur de transcription H-NS contrôle l'expression de plus d'une trentaine de gènes chez *E. coli* (pour revue : [174]). L'action inhibitrice de DsrA sur H-NS a donc un effet pléiotropique puisqu'en plus d'affecter positivement l'expression de *rpoS*, elle module celle de l'ensemble des cibles de H-NS. La recherche de complémentarité entre DsrA et d'autres ARNm par une approche informatique a permis de caractériser d'autres cibles potentielles pour DsrA [165]. Les ARNm des gènes *argR* et *ilvIH* peuvent tout deux être reconnus par DsrA de façon similaire à l'ARNm *hns*. Ils conservent effectivement la possibilité de former une interaction bipartite avec DsrA à proximité des signaux de traduction. L'ARNm du gène *rbsD* quant à lui possède une région de complémentarité à la première tige-boucle de DsrA. *ArgR* code pour une protéine répresseur qui contrôle la synthèse des pyrimidines, de l'arginine et des polyamines chez *E. coli* [175]. *IlvIH* code pour une enzyme impliquée dans le contrôle de la biosynthèse de la valine, de la leucine et de l'isoleucine [176]. De manière intéressante, H-NS active indirectement l'expression de *ilvIH* en réprimant la synthèse d'un répresseur spécifique [177]. *RbsD* a une fonction inconnue mais est codé au sein d'un opéron impliqué dans le transport du ribose [178]. Ainsi, le riborégulateur DsrA agit sur plusieurs cascades d'événements en co-régulant l'expression de ses gènes cibles négativement ou positivement (Figure 7). Finalement, DsrA est conservé chez *E. coli*, *S. typhimurium* et *Klebsiella pneumoniae* et adopte des structures secondaires semblables [49]. Sa structure doit par conséquent être cruciale pour son activité régulatrice et éventuellement pour sa stabilité. Ceci reste néanmoins à prouver par des approches génétiques et biochimiques. DsrA joue donc apparemment un rôle dans des cascades spécifiques d'intégration de signaux divers dans les organismes procaryotiques.

C. Plusieurs cibles pour OxyS

Le gène *fhlA* code pour un activateur de transcription dont l'expression est régulée négativement par OxyS [171]. Ce contrôle se fait au niveau traductionnel par un mécanisme distinct de l'inhibition de la traduction de *rpoS* [43]. Une recherche de complémentarité entre OxyS et l'ARNm *fhlA* a révélé que les boucles A et C d'OxyS contiennent chacune 7 nts complémentaires à deux régions de l'ARNm dont l'une contient la séquence Shine et Dalgarno. (Figure 10). L'implication de ces nucléotides et des positions complémentaires dans *fhlA* a été confirmée par mutagenèse dirigée et restauration par mutations compensatoires [43,179]. La structure de l'hélice C est importante d'un point de vue thermodynamique mais n'est pas importante d'un point de vue séquence. Un modèle pour le

mode de contrôle OxyS-dépendant de *fhlA* suggère qu'OxyS intervient comme ARN antisens et masque le site de fixation du ribosome de l'ARNm. En effet, il a été montré *in vitro* que l'appariement OxyS-*fhlA* inhibait la fixation des sous-unités ribosomiques 30S à l'ARNm [43].

Tout comme l'ARN DsrA, OxyS régule l'expression de deux facteurs de transcription. Ceci suggère qu'OxyS permet d'intégrer la réponse au stress oxydatif au sein d'autres réseaux régulateurs. Il a ainsi un rôle beaucoup plus global que celui d'un simple répresseur puisqu'il agit indirectement sur l'ensemble des cibles de *rpoS* et de *fhlA*. OxyS contrôle l'expression de ces deux gènes cibles par des mécanismes distincts qui impliquent des régions différentes de l'ARN. OxyS a probablement d'autres cibles directes que celles identifiées jusqu'à présent. Il comporte par exemple une séquence de 11 nucléotides complémentaire à la région contenant la séquence Shine et Dalgarno de l'ARNm *exbB*. ExbB code pour un facteur facilitant l'entrée du fer dans les cellules [180]. Les ions Fe^{3+} génèrent des radicaux hydroxyles hautement réactifs, particulièrement contre l'ADN. OxyS pourrait ainsi diminuer la concentration intracellulaire en fer en réprimant la synthèse de *exbB* lors d'un stress oxydatif, limitant les lésions de l'ADN. Ceci est en accord avec le rôle antimutateur attribué à OxyS lors d'un stress oxydatif. Cette fonction régulatrice d'OxyS n'est que spéculative et l'interaction ARN-ARN reste à confirmer dans ce cas.

D. Caractéristiques générales des ARN multirégulateurs

Les deux riborégulateurs multifonctionnels décrits dans ce chapitre intègrent différents réseaux régulateurs dans une réponse physiologique précise. OxyS et DsrA interviennent grâce à des régions distinctes, par plusieurs mécanismes, sur des cibles variées. Tout deux contrôlent l'expression de *rpoS* de façon Hfq-dépendante, mais avec des effets antagonistes. Ces deux ARN remplissent leurs fonctions soit partiellement (OxyS) soit entièrement (DsrA) par des mécanismes antisens. Contrairement à la plupart des ARN antisens décrits plus loin qui sont souvent codés à partir du même locus que leur gène cible, DsrA et OxyS sont codés à des loci distincts. La complémentarité avec leur ARN cible n'est donc que partielle. Ceci ne restreint pas ces deux riborégulateurs à une cible unique et permet que différentes régions puissent interagir avec des cibles différentes en augmentant la liberté évolutive des ARN malgré leur courte taille. De plus, ils remplissent leur rôle d'antisens grâce à de courtes régions de complémentarité avec leur cible et l'interaction se fait *via* une reconnaissance boucle-boucle ou boucle-région simple-brin. Ces caractéristiques rendent la reconnaissance spécifique et optimale. Ces deux ARN ont également la propriété d'être particulièrement

stables. Par conséquent, lors d'un stress, leur concentration est suffisante pour protéger les cellules sur une durée plus longue.

Une caractéristique générale, et non des moindres, de ces ARN est qu'ils interviennent au cœur de cascades d'intégration de signaux et que leurs cibles sont souvent des facteurs de transcription. La synthèse d'une seule courte molécule d'ARN à multiples domaines permet d'avoir des effets pléiotropiques sur l'expression des gènes d'un organisme, et ceci de façon coordonnée. Il est peu probable que de tels ARN se limitent à OxyS et DsrA. Un autre riborégulateur de cette catégorie sera décrit plus loin. Il s'agit de l'ARN-III de *S. aureus* impliqué dans le contrôle de l'expression des gènes de virulence chez ce pathogène. Dans ce cas précis, l'ARN est plus grand (plus de 500 nts), suggérant qu'il possède des domaines fonctionnels plus nombreux interagissant avec un plus grand nombre de cibles.

V. Les ARN antisens

A. Les ARN antisens eucaryotiques

Même si de nombreux ARN antisens présentant une complémentarité à un ARN cible ont été caractérisés chez les eucaryotes (résumés dans le tableau 3), leur intervention en tant que régulateurs de l'expression d'un gène n'a que rarement été prouvée. Ils ont été caractérisés dans de nombreux organismes tels que chez les insectes, les amphibiens, les amibes, les oiseaux et les mammifères, dont l'homme. Certains d'entre eux sont conservés entre différents organismes, ce qui suggère qu'ils remplissent des rôles régulateurs essentiels. Les modes d'action des ARN antisens chez les eucaryotes sont variés. Ils ont été postulés comme agissant à toutes les étapes de l'expression des gènes et beaucoup sont supposés agir dans le noyau. En général, et contrairement aux systèmes bactériens, peu de données expérimentales sont disponibles pour les mécanismes de contrôle au niveau transcriptionnel. Plusieurs d'entre eux agissent par un mécanisme antisens classique, où un ARN interagit avec sa cible en modulant son expression. Au contraire, certains exemples ont été décrits où ce n'est pas le cas. Dans ces cas, ce sont la transcription ou la traduction de l'ARN antisens elles-mêmes qui modulent l'expression de son gène cible. Ne seront décrits ici que les exemples dans lesquels un appariement ARN-ARN a été caractérisé comme étant la clé du processus du contrôle. Dans ces derniers cas, les ARN antisens régulent l'expression de leur gène cible au niveau post-transcriptionnel. Ils peuvent être regroupés en deux groupes : ceux régulant la traduction d'un ARNm et ceux régulant sa stabilité.

ARN sens	ARN antisens	Taille antisens	ORF	Organisme	Fonction ORF sens	Niveau de Régulation	Références
lin-14 / lin-28	***lin-4***	22 nts	-	*C. elegans*	Contrôle développement larvaire	Traduction	[181,182]
lin-41	***let-7***	21 nts	-	*C. elegans*	Contrôle développement larvaire	Traduction ?	[41]
EB4-SV		1,8 kb	-	*D. discoideum*	Formation des préspores	Stabilité	[39]
bFGF	***gfg***	1,5 kb	+	*X. laevis* *H. sapiens* *Rat*	Facteur mitotique	Désamination/ Dégradation	[183,184]
Hcb.12	*Hca.12*	475 nts	-	*Bombyx mori*	Formation du chorion	Désamination/ Dégradation	[185]
eIF2 α	*Inr*	*ND*	-	*H. sapiens*	Facteur de traduction	Transcription ?	[186]
ICP	*γ 34.5*	1,2 kb	+	HSV-1	Réplication virale	Transcription ?	[187]
α1(1) Collagen		> 10 kb	-	Poulet	Structure cellulaire	Transcription ?	[188]
c-erbA α	*Rev-erb*	*ND*	+	Rat	Réponse hormonale	Epissage ?	[189]
mbp		*ND*	-	*M. musculus*	Myélination	Transport ?	[190]
p53	*inRNA*	1,3 kb	-	*M. musculus*	Facteur de transcription	P. T. *ND*	[191]
bzlf1	*ebna*	2,5 kb	+	EBV	Transactivation virale	P. T. *ND*	[192]
ICPO	*lat*	1,4 kb	+	HSV-1	Réplication virale	*ND*	[193]
Env	*AS01*	2,3 kb	+	HIV-1	Enveloppe virale	*ND*	[194]
Ddc		2,2 kb	-	*D. melanogaster*	Biogenèse de la dopamine	*ND*	[195]
micropia		1,0 kb	-	*D. melanogaster*	Rétro - transposition	*ND*	[196]
Hoxd-3	*Dxhoh-3*	1,5 – 2,8 kb	+	*M. musculus*	Développement	*ND*	[197]
c-myc	*Plusieurs*	0,3 - 1 kb	-	*M. musculus* *H. sapiens*	Prolifération cellulaire	*ND*	[198,199]
ear-7	*ear-1*	2,9 kb	+	*H. sapiens*	Réponse hormonale	*ND*	[200]
N-myc	*N-cym*	*ND*	+	*H. sapiens*	Facteur de transcription	*ND*	[201]
CD3 ζ/η/θ	*Oct1*	> 12 kb	+	*H. sapiens*	Réponse immunitaire	*ND*	[202]
bcma		*ND*	+	*H. sapiens*	Réponse immunitaire	*ND*	[203]
WT1	*WIT1*	7–10 kb	-	*H. sapiens*	Contrôle du développement	*ND*	[204]
MCH	*AROM*	*ND*	+	*H. sapiens*	Réponse hormonale	*ND*	[205]
TS	*rTS*	*ND*	+	*H. sapiens*	Prolifération cellulaire	*ND*	[206]
SC35	*c-myb*	*ND*	+	*H. sapiens*	Epissage	*ND*	[207]
Gn-RH	*SH*	*ND*	+	*H. sapiens*	Réponse hormonale	*ND*	[208]
bcl-2	*bcl-2/IgH*	*ND*	-	*H. sapiens*	Contrôle de l'apoptose	*ND*	[209]

Tableau 3: ***ARN antisens caractérisés chez les organismes eucaryotes***. La présence ou l'absence d'une phase ouverte de lecture (ORF) dans les antisens sont matérialisées respectivement par un + ou un -. Les organismes dans lesquels ces ARN ont été identifiés sont indiqués, de même que la fonction de l'ARN cible potentiel. Le niveau de régulation de l'expression du gène cible est mentionné s'il est connu. P.T. : action post-transcriptionnelle. Les systèmes qui seront décrits ici sont indiqués en rouge. *ND* : Non déterminé.

1. *Régulation traductionnelle : l'exemple du contrôle du développement larvaire chez* C. elegans

Le développement larvaire de *C. elegans* compte 4 stades (L1 à L4). Un ensemble de gènes hétérochroniques est important pour spécifier la durée de chacun de ces stades et contrôler le passage de l'un à l'autre. Des mutations dans ces gènes peuvent générer soit une apparition précoce des stades tardifs, soit une réitération des stades précoces du développement. Parmi ces gènes, *lin-14* est important pour empêcher un passage précoce du stade L1 au stade L2 et *lin-28* du stade L2 au stade L3. Le gène *lin-29* est quant à lui requis pour assurer la transition du dernier stade larvaire L4 au stade adulte. L'expression de ces gènes doit être régulée de façon fine afin d'assurer un développement correct à l'animal.

a. L'ARN Lin-4

Le gène hétérochronique *lin-4* est un régulateur négatif de *lin-14*. Une mutation dans *lin-4* a un phénotype semblable à celui observé pour une mutation de gain de fonction [gf] dans *lin-14*, c'est-à-dire une réitération des stade larvaires précoces [210]. Les produits de *lin-4* sont deux courts ARN non traduits de 22 et 61 nts [211], le plus petit étant le régulateur majeur. La maturation de ce dernier à partir de son précurseur nécessite les gènes impliqués dans l'ARNi [121]. L'abondance des ARN Lin-4 est régulée dans le temps, avec un début d'expression en fin de stade L1, au moment où la quantité de protéine LIN-14 commence à décroître. La régulation de l'expression de *lin-14* se fait de manière post-transcriptionnelle et est dépendante de la région 3'-UTR de l'ARNm [212]. Celle-ci contient sept copies d'une séquence de 14 à 19 nucléotides complémentaires à une portion de l'ARN Lin-4. Ceci suggère que Lin-4 régule négativement l'expression de *lin-14* par la formation de plusieurs complexes ARN-ARN dans la partie 3'-UTR de l'ARNm. Comme la quantité d'ARNm Lin-14 ne varie pas de façon significative au cours du développement, le contrôle doit se faire au niveau post-transcriptionnel. La fixation de Lin-4 n'affecte pas la stabilité de l'ARNm. L'hypothèse avancée est que les complexes Lin-4-Lin-14 empêcheraient la polyadénylation et ainsi la traduction de l'ARNm. De façon alternative, les complexes pourraient interférer avec la formation du complexe d'initiation de la traduction en empêchant l'interaction 5'/3' au sein de l'ARNm.

Parmi les sept duplexes Lin-4-Lin-14, quatre présentent une cytosine non appariée dans la structure de Lin-4 (Figure 11) [181]. Ces cytosines non appariées sont conservées chez *Caenorhabtidis briggsae*, suggérant qu'elles ont une importance fonctionnelle. La présence d'un seul duplexe présentant une cytosine en bulge dans la partie 3'-UTR d'un gène

rapporteur est suffisante pour conférer un contrôle par Lin-4 *in vivo*. En revanche, la présence de duplexes sans cytosine non appariée ou présentant une autre base à cette même position ne permet pas l'inhibition. Curieusement, Lin-4 se fixe avec une plus grande efficacité *in vitro* à une séquence où cette cytosine est appariée. Ceci suggère qu'un facteur protéique pourrait intervenir *in vivo* pour reconnaître et stabiliser les duplexes imparfaits par une reconnaissance spécifique de la cytosine non appariée. L'intervention d'un tel facteur additionnel reste néanmoins à prouver.

La région 3' non traduite de l'ARNm du gène *lin-28* contient elle aussi un site de fixation potentiel de Lin-4 [182] (Figure 11). La déletion de ce site entraîne une expression de *lin-28* à des stades plus tardifs et un phénotype de type [*gf*]. Une délétion du gène *lin-4* conduit aux mêmes phénotypes. L'ARN Lin-4 a donc plus d'une cible dans *C. elegans* et régule l'expression de *lin-14* et de *lin-28* par des mécanismes semblables. Les deux gènes présentent cependant des différences. En effet, la région 3'-UTR de l'ARNm Lin-28 est plus courte que celle de Lin-14. Elle ne compte qu'un seul site de complémentarité avec Lin-4, contre sept pour Lin-14. Le duplexe Lin-4-Lin-28 potentiel ne génère pas de cytosine non appariée (Figure 11). Il doit par conséquent y avoir des variations subtiles dans les modes de régulation Lin-4-dépendants de *lin-14* et de *lin-28*. Si un seul duplexe sans cytosine non appariée est suffisant pour réguler *lin-28*, ce n'est pas le cas pour réguler *lin-14*. Une hypothèse serait que ces variations dans la structure des complexes permettent de contrôler l'expression de ces deux gènes par des mécanismes distincts à différents stades du développement embryonnaire, ceci avec une molécule d'ARN unique.

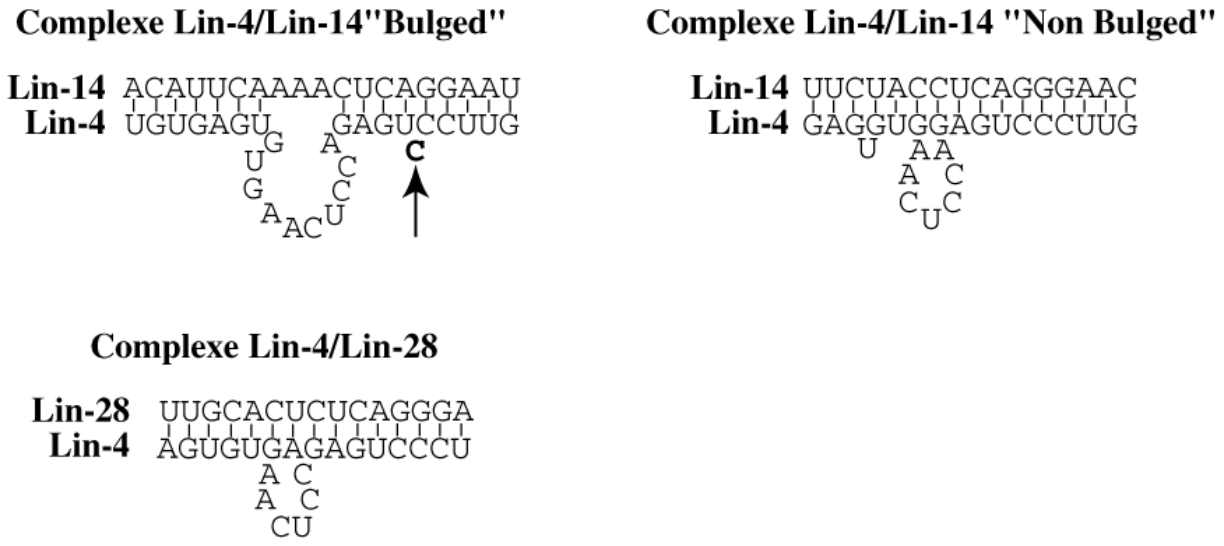

Figure 11 : ***Modèles des interactions entre Lin-4 et les régions 3'-UTR des ARN Lin-14 et Lin-28***. Dans le complexe Lin-4-Lin-14 présentant une cytosine non appariée (complexe n°2/7) celle-ci est représentée en gras et indiquée par une flèche. Le complexe Lin-4-Lin-14 avec la cytosine appariée est le complexe n°5 parmi les sept potentiels.

b. L'ARN let-7

Le gène *let-7* a été identifié par recherche de nouveaux gènes hétérochroniques chez *C. elegans* [41]. Une mutation dans ce gène entraîne un défaut dans la transition du dernier stade larvaire L4 au stade adulte, avec une répétition du stade L4. Au contraire, une surproduction de *let-7* conduit à une transition trop précoce, en fin de stade L3. *Let-7* a par conséquent été proposé comme étant l'interrupteur moléculaire nécessaire à la transition L4 --> adulte. Le gène hétérochronique *lin-29* est nécessaire à l'établissement du stade adulte et code pour un facteur de transcription à domaine « doigt de zinc » [213]. Dans une souche *let-7(-)*, la quantité de protéine Lin-29 est sévèrement diminuée [41]. *let-7* est requis pour l'activation de l'expression de *lin-29* et son expression est régulée dans le temps. *let-7*, non exprimé dans l'embryon et les stades larvaires L1 et L2, commence à être exprimé au stade L3 pour atteindre une expression maximale aux stades L4 et adulte. Ce profil d'expression coïncide avec celui de la protéine Lin-29.

Des expériences de northern-blot ont montré que le produit du gène *let-7* est un petit ARN non traduit de 21 nts, maturé à partir d'un précurseur [41] (Figure 12). Cette maturation nécessite une endoribonucléase de *C. elegans* orthologue à la protéine Dicer de *D. melanogaster,* impliquée dans le phénomène d'ARNi [121]. Des séquences complémentaires à let-7 ont été identifiées dans les régions 3'-UTR des ARNm Lin-14, Lin-28, Lin-41 et Lin-42 [41]. Des mutations dans *lin-41* et *lin-42* sont épistatiques à celles observées dans *let-7* [41,214], suggérant que let-7 régule négativement ces gènes. Cette régulation est directe et la région 3'-UTR de *lin-41* contenant deux séquences complémentaires à let-7 est suffisante pour réguler l'expression d'un gène rapporteur dans le temps de façon let-7 dépendante [41]. Comme Lin-29 est synthétisée de façon précoce dans un mutant *lin-41*, il est proposé que Lin-41 réprime l'expression de *lin-29*. La présence d'un domaine de reconnaissance à l'ARN de type RBCC dans Lin-41 suggère que la protéine pourrait inhiber la traduction de l'ARNm *lin-29*. Un modèle d'action de let-7 serait qu'il puisse réguler négativement la traduction d'ARNm cibles, dont l'ARNm *lin-41*, de manière similaire à Lin-4. Dans ce cas précis, l'inhibition de l'expression de *lin-41* permettrait la dérepression de celle de *lin-29*. La transition entre le stade larvaire L4 et le stade adulte serait alors activée. Dans ce modèle, l'interaction entre let-7 et la région 3'-UTR de Lin-41 reste encore à être démontrée. Le mécanisme d'action de let-7 reste également inconnu. L'implication de l'enzyme Dicer dans la maturation de let-7 laisse supposer que let-7 puisse moduler la stabilité de ces ARNm cibles par un mécanisme de type ARNi. L'ARN antisens de 21 nts s'hybriderait dans la région 3'-UTR des ARNm cibles et la structure double-brin formée serait un substrat des enzymes impliqués dans l'ARNi. Il est donc envisageable que les facteurs impliqués dans l'ARNi

puissent aussi intervenir dans le contrôle de l'expression génétique par des ARN antisens naturels dans les organismes eucaryotiques. Ceci reste néanmoins à montrer.

De manière intéressante le gène *let-7* est conservé et l'ARN est exprimé chez les vertébrés (dont l'homme), les ascidiens, les hémichordates, les mollusques, les annélides et les arthropodes [215] (Figure 12). Il est néanmoins absent chez *S. cerevisiae*, *E. coli* et *A. thaliana*. Le gène de *C. briggsae* a une séquence conservée et complémente une souche *let-7(-)* de *C. elegans*, suggérant une fonction identique. Le gène *lin-41* possède lui aussi des orthologues chez les vertébrés et chez *D. melanogaster* [214]. De plus, tous ces gènes possèdent des régions de complémentarité à let-7 dans leur région 3'-UTR. Il a été montré que l'expression de *let-7* chez *D. melanogaster* est également sujette à une régulation temporelle [215]. Ces observations indiquent que la cascade de gènes hétérochroniques pourrait être maintenue dans ces espèces. let-7 pourrait avoir une fonction et un mode d'action conservés dans tous ces organismes. Cela est surprenant dans la mesure où dans tous ces organismes, l'embryogenèse est radicalement différente. L'implication des ARN de type let-7 dans le développement de ces organismes devra être testée.

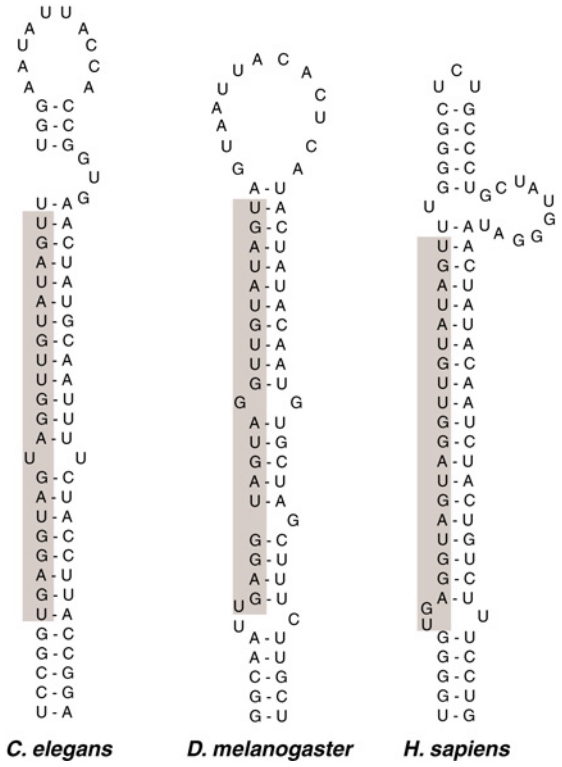

Figure 12 : ***Modèle de structure secondaire des ARN précurseurs de let-7 dans* C. elegans, D. melanogaster et H. sapiens.** La région de 21 nts correspondant aux ARN let-7 matures dans les organismes indiqués est grisée (d'après [215]).

2. *Régulation de la stabilité d'un ARNm*

a. Contrôle de la formation des préspores chez *Dictyostelium discoideum*

Le gène *EB4-SV* de *Dictyostelium discoideum* code pour une protéine associée avec la membrane des préspores. Lors de l'étape d'agrégation du développement de *D. discoideum*, l'ARNm EB4-SV s'accumule. En revanche, il ne peut pas être détecté lors de la croissance végétative ou dans l'étape de pré-agrégation [216]. EB4-SV est transcrit de façon constitutive, ce qui suggère que son accumulation différentielle résulte d'un contrôle post-transcriptionnel de son expression.

Des expériences de northern-blot en utilisant différentes sondes sens ou antisens sur la région EB4-SV ont révélé la présence de deux transcrits [39]. Un transcrit de 2,2 kb correspond à l'ARNm EB4-SV tandis qu'un second de 1,8 kb correspond à un ARN antisens. Ce dernier ne contient pas de phase ouverte de lecture. Il couvre les trois premiers exons de l'ARNm ainsi que la région 5'-UTR et une partie de la séquence promotrice du gène. Les deux transcrits s'accumulent de façon différentielle au cours du développement. L'ARN antisens est détecté en cours de croissance végétative et lors de la pré-agrégation tandis qu'il est indétectable lors de l'étape d'agrégation. La synthèse de l'ARN antisens est sous la dépendance d'un contrôle transcriptionnel permettant d'intégrer différents signaux indiquant l'état de différentiation de l'organisme. Au contraire de l'ARNm, l'ARN antisens est intrinsèquement instable ce qui permet de moduler sa concentration de manière précise lors des différentes étapes du développement de *D. discoideum*. L'ARN antisens déstabilise spécifiquement l'ARNm grâce à l'intervention d'une RNase spécifique des ARNdb [39]. La, ou les, région(s) responsables de l'interaction entre l'ARN antisens et l'ARNm n'a (n'ont) cependant pas encore été identifiée(s). Il est cependant fort probable que de manière similaire aux cas décrits pour les antisens procaryotiques (voir §V. C), des déterminants structuraux des deux ARN soient impliqués spécifiquement dans cette reconnaissance. Il est également envisageable que la dégradation de l'ARNm se fasse par un mécanisme de type ARNi.

b. Modification post-transcriptionnelle d'un ARNm : l'exemple de *bFGF*

Le gène *bFGF* code pour un facteur mitotique (basic Fibroblast Growth Factor) extrêmement conservé des batraciens à l'homme. Le locus *bFGF* code également pour l'ARN antisens *gfg* de 1,5 kb, tout aussi conservé et contenant une phase ouverte de lecture [183,184]. Les ARN sens et antisens se chevauchent sur une région d'environ 900 nts recouvrant l'exon 3 de b*FGF*. Dans les oocytes de xénope, les ARN sens et antisens sont

exprimés de façon constitutive jusqu'à la maturation. Lors de la fertilisation, les transcrits *bFGF* sont rapidement dégradés [183] tandis que la quantité de transcrit *gfg* demeure constante. Un séquençage de la région de complémentarité entre l'antisens et l'ARNm indique une forte proportion (~ 50%) de désamination d'adénosines en inosines dans *bFGF*. Cette modification est effectuée par une adénosine-désaminase (ADAR) présente dans le noyau et libérée dans le cytoplasme lors de la maturation de l'oocyte [217]. L'antisens étant vingt fois en excès par rapport à l'ARNm, la totalité de ce dernier serait complexé et par conséquent modifié par l'ADAR. La présence d'inosines dans le complexe déstabiliserait ce dernier ce qui rendrait ultérieurement l'ARNm susceptible à l'attaque par la I-RNase, spécifique des ARN contenant des inosines et caractérisée dans des extraits de cellules de mammifères [218].

Ce mode de régulation ne paraît pas restreint au contrôle de l'expression du gène *bFGF* chez les vertébrés et semble un mécanisme beaucoup plus général. En effet, il a récemment été montré que la désamination et la déstabilisation d'un ARNm d'un polyomavirus murin étaient dépendantes d'un ARN antisens viral [219]. De même, un tel mécanisme a également été proposé pour le contrôle de la formation du chorion chez *Bombyx mori* [185].

3. *Conclusions*

Une fois encore, comme dans le cas des riborégulateurs agissant *via* un facteur protéique ou des riborégulateurs multifonctionnels, les rares ARN antisens eucaryotiques bien étudiés contrôlent des fonctions biologiques nécessitant une réponse cellulaire rapide. Dans certains cas, le mécanisme de régulation est conservé pour contrôler une fonction essentielle, à savoir le développement embryonnaire. Il est également intéressant de noter que l'analyse de ces systèmes révèle un rôle fonctionnel des facteurs impliqués dans l'ARNi.

L'étude de systèmes procaryotiques a montré que la structure des antisens jouait un rôle déterminant dans l'appariement avec leur cible et que peu de facteurs protéiques sont impliqués dans ces systèmes de régulation (voir § V.B). Dans les cas d'antisens eucaryotiques qui ont été décrits ici, très peu d'informations structurales sont disponibles. Il est fort possible que la structure de ces antisens soit aussi essentielle à leur activité et cet aspect devra être approfondi. Il est également envisageable que des facteurs protéiques intervenant en *trans* puissent jouer un rôle essentiel dans la reconnaissance entre l'ARN antisens et sa cible et le mécanisme de contrôle chez les eucaryotes. L'ARN antisens agirait alors au cœur d'une particule ribonucléoprotéique. Malgré l'identification d'un grand nombre d'ARN antisens eucaryotiques, peu d'entre eux ont été étudiés dans le détail. Des études fonctionnelles et

structurales seront nécessaires pour la plupart des systèmes énoncés dans le tableau 3 afin de confirmer le rôle régulateur des ARN antisens et d'augmenter notre compréhension des mécanismes de contrôle de l'expression génétique chez ces organismes.

B. Les ARN antisens procaryotiques

Les ARN antisens procaryotiques ont été les premiers caractérisés. Ils sont courts et hautement structurés. Dans tous les systèmes étudiés en détail, des structures bien définies des ARN antisens et/ou de leur cibles sont essentielles pour leur assurer leur fonction (pour une revue, [65]). Curieusement, la plupart des antisens procaryotiques ont des caractéristiques structurales communes. Ils contiennent une ou plusieurs structures en tige-boucle requises pour l'interaction ARN-ARN. Les tiges contiennent souvent des nucléotides non appariés importants pour leur fonction et dans plusieurs cas les boucles partagent une séquence optimisant l'interaction. En effet, l'étape de nucléation de l'appariement fait généralement intervenir des paires de bases G≡C. Cette étape initiale étant souvent la seule à être réversible, ces appariements plus stables déplacent l'équilibre vers la formation du complexe. Ainsi, les antisens procaryotiques étant impliqués dans le contrôle de fonctions biologiques nécessitant un temps de réponse très court, ces structures garantissent une interaction optimale et rapide entre l'antisens et son ARN cible.

Parmi les ARN antisens procaryotiques peuvent être distingués deux classes. Les ARN antisens codés par le génome bactérien interviennent le plus souvent dans des mécanismes de réponse des cellules aux conditions environnementales et sont tous synthétisés en *trans* par rapport à leur gène cible. La complémentarité entre les deux ARN n'est donc que partielle. La seconde classe d'ARN antisens correspond à ceux exprimés à partir des éléments génétiques mobiles (plasmides, bactériophages et transposons). Dans ces systèmes, ils contrôlent des fonctions vitales pour l'élément extra-chromosomique et sont synthétisés en *cis* par rapport à leur gène cible. La complémentarité entre les deux ARN est alors totale. Dans ce paragraphe, des exemples caractéristiques de ces deux classes seront décrits. Un des objets de cette thèse étant l'étude d'un ARN antisens plasmidique, cette sous-classe d'ARN antisens sera décrite en détail dans le § VI.

1. *Les ARN antisens procaryotiques chromosomiques*

Deux ARN multifonctionnels et agissant en partie comme ARN antisens chez *E. coli* ont déjà été décrits au § IV.A. Il s'agit des ARN DsrA et OxyS codés par des gènes chromosomiques. Parallèlement à ces deux ARN agissant de façon pléiotropique et ceci sur plusieurs cibles, de nombreux autres ARN ont été caractérisés chez les bactéries et agissent exclusivement comme ARN antisens. Tout comme DsrA et OxyS, ils contrôlent l'expression de gènes cibles en réponse à des signaux extracellulaires. Les exemples connus sont répertoriés dans le tableau 4.

Organisme	**ARN antisens**	**ARN cible**	**Taille de l'antisens (nts)**	**Fonction régulée**	**Mode de contrôle**	**Référence**
E. coli	OxyS	*rpoS* & *fhlA*	109	réponse au stress oxydatif	Séquestration de Hfq & Inhibition de la traduction	[43,172]
E. coli	DsrA	*rpoS* & *hns*	87	réponse aux stress	activation & inhibition de la traduction	[166]
E. coli	MicF	*OmpF*	97	réponse aux stress	Inhibition de la traduction et dégradation de l'ARNm	[220,221]
E. coli	DicF	*ftsZ*	53	Division cellulaire	Inhibition de la traduction	[222]
V. anguillarum	ARNα	*fatA* & *fatB*	650	Internalisation du Fer	Dégradation de l'ARNm	[223,224]
E. coli	Sof	*gef*		Lyse cellulaire	Inhibition de la traduction	[225]
S. aureus	ARN-III	*hla*	517	Virulence	Activation de la traduction	[50]
B. subtilis	ARNt	ARNm des aa-ARNt synthétases	73	synthèse des aaRS	Antiterminaison de la transcription	[226]
Clostridium acetobutylicum	p3	*glnA*		Synthèse de la glutamine synthétase	*ND*	[227]

Tableau 4 : ***ARN antisens codés par des loci chromosomiques bactériens***. *ND* : mode d'action non déterminé.

a. Le contrôle du transport du Fer chez *Vibrio anguillarum*

Les souches les plus virulentes du pathogène *Vibrio anguillarum,* dont les hôtes sont les poissons, possèdent un système hautement efficace d'internalisation du Fer codé par un plasmide. Le plasmide pJM1 code pour les gènes *fatA*, *fatB*, *fatC* et *fatD* impliqués dans le transport du Fer vers l'intérieur de la cellule [228]. FatA est un récepteur de complexes ferriques et FatB, FatC et FatD sont des protéines transmembranaires dont les deux dernières possèdent un large domaine cytoplasmique. Les ions ferriques Fe^{3+} étant impliqués dans la formation de radicaux libres, ils sont dangereux pour la cellule. Leur internalisation doit par conséquent être contrôlée de manière rigoureuse. Elle est régulée négativement par des facteurs codés par le chromosome mais aussi par le plasmide pJM1. La protéine Fur codée par un gène chromosomique inhibe la transcription du locus *fatDCBA* [229] et un ARN antisens (ARNα), codé par le plasmide pJM1, réprime l'expression des gènes *fatA* et *fatB* au niveau post-transcriptionnel [224]. L'ARNα est distingué des ARN antisens plasmidiques qui seront décrits plus loin et qui contrôlent une fonction nécessaire au maintien du plasmide au sein des cellules (voir § VI) ; en effet, l'ARNα contrôle une fonction cellulaire. L'ARNα compte 650 nucléotides et est exprimé à partir de la région *fatB*, en orientation opposée au transcrit polycistronique *fatDCBA*. La transcription de cet ARN est induite à haute concentration en Fer et nécessite la protéine Fur [228]. A haute concentration en Fer, la concentration de l'ARNα augmente et les ARN *fatB* [224] et *fatA* [223] sont déstabilisés de façon concomitante. Des évidences indirectes suggèrent une interaction stœchiométrique entre l'ARNα et l'ARNm *fatB* [224]. Même si l'ARNα est complémentaire à l'ARNm *fatB*, le complexe ARN-ARN n'a jamais été caractérisé, ni *in vitro*, ni *in vivo*. Il serait par conséquent intéressant de caractériser l'interaction entre l'ARNα et les transcrit *fatDCBA* ou *fatB* et les structures impliquées dans cette reconnaissance. Il est fort probable que l'interaction ARN-ARN soit optimale pour que la répression de l'expression des transporteurs des complexes ferriques soit rapide à haute concentration extracellulaire en Fer. C'est probablement pour cette dernière raison que les cellules utilisent un double contrôle, transcriptionnel et post-transcriptionnel, de l'expression du locus *fatDCBA*. Finalement, un parallèle peut être fait entre l'intervention d'un ARN dans ce mécanisme de protection contre la formation de radicaux libres et l'intervention d'OxyS dans la protection des cellules contre les stress oxydatifs.

b. Contrôle de l'expression de la porine OmpF chez *Escherichia coli*

La porine OmpF est présente dans la membrane externe chez *E. coli* et est une des porines majeures. Son expression est contrôlée en réponse à des changements de température [221], de pression osmotique [230] et suite à un stress oxydatif [231]. Le contrôle de l'expression d'OmpF est effectuée au niveau post-transcriptionnel par l'ARN antisens MicF (Figure 13).

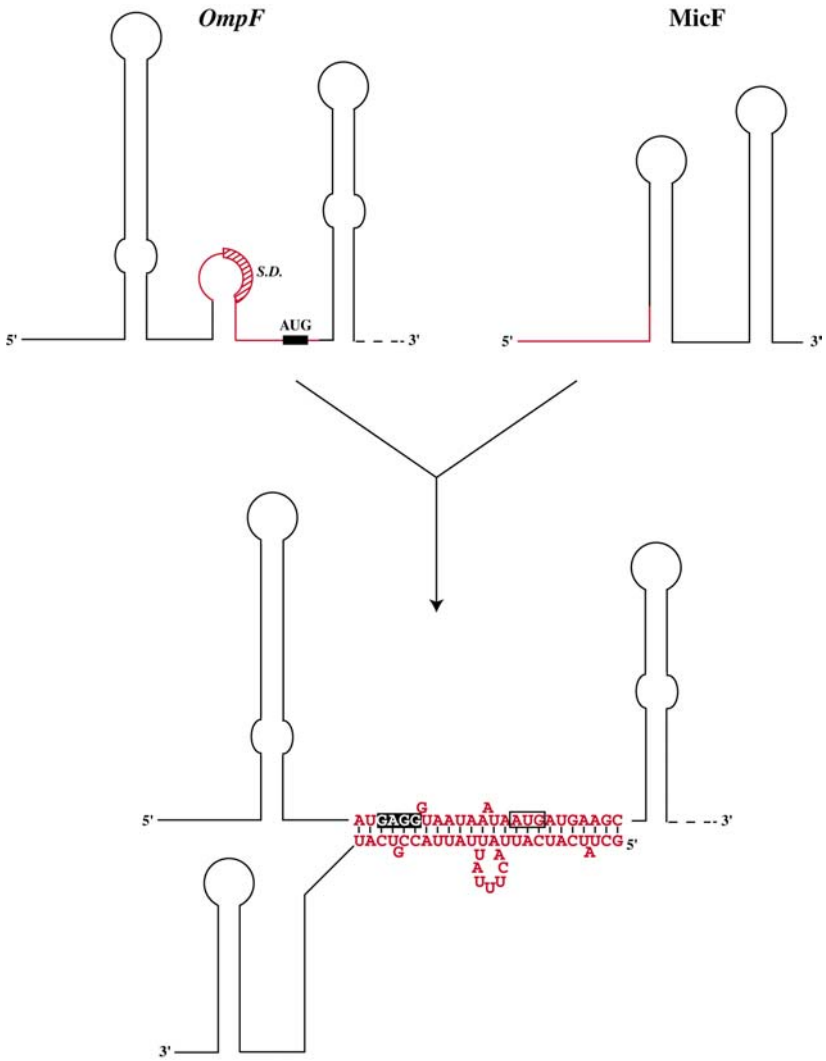

Figure 13 : ***Séquences et structures de la région 5'-UTR d'*OmpF*, de l'ARN antisens MicF et du complexe ARN-ARN formé***. Les régions complémentaires entre les deux ARN sont dessinées en rouge. La séquence Shine et Dalgarno est hachurée dans la région 5'-UTR de l'ARN *OmpF*. Le codon AUG initiateur de la traduction est indiqué par une boîte noire. Dans le complexe formé, ces mêmes séquences sont représentées en caractères blancs sur fond noir et encadrées, respectivement (d'après [220]).

MicF (94 nts) inhibe la traduction de l'ARNm *OmpF* [220] et favorise sa dégradation [221]. MicF est nécessaire pour cette régulation négative mais n'est pas suffisant. En effet, un

facteur protéique fixe MicF et intervient dans le contrôle de l'expression d'o*mpF*, mais par un mécanisme inconnu [45,221]. L'interaction entre MicF et sa séquence cible dans la partie 5'-UTR de l'ARNm *OmpF* a été caractérisée [45]. L'ARN MicF étant synthétisé à partir d'un locus différent, sa complémentarité avec *OmpF* n'est que partielle. Des études de cartographie en solution ont permis de déterminer les régions des deux ARN impliquées dans la formation du complexe [220]. Ce dernier implique les 33 premiers résidus de MicF et les nucléotides 95 à 120 de l'ARNm *OmpF* (Figure 13). Le complexe MicF-*OmpF* séquestre la séquence Shine et Dalgarno et le codon initiateur de l'ARNm, réprimant ainsi la traduction d'*OmpF*. Le mode d'action de la protéine fixant MicF n'est pas connu. On peut toutefois envisager qu'elle stimule la reconnaissance entre les deux partenaires ou qu'elle augmente la stabilité du complexe.

Le gène *micF* est hautement conservé chez plusieurs bactéries Gram-négatives [232]. Ceci souligne le rôle fondamental de l'ARN MicF comme régulateur. Par ailleurs, l'observation récente que l'expression de *micF* est sous le contrôle du facteur global de régulation H-NS décrit au § IV.A.3 supporte l'idée que MicF est requis pour la réponse des cellules à divers stress [233].

c. Contrôle de la division cellulaire chez *Escherichia coli*

La protéine FtsZ est nécessaire à la division cellulaire chez *E. coli* et semble être impliquée dans le couplage entre la séparation des nucléoïdes et la division cellulaire [234]. Son expression est sous la dépendance des produits des gènes *dicB* et *dicF*, tout deux codés par un même prophage défectif [235], et du facteur de transcription alternatif σ^S [236]. Alors que *dicB* code pour un inhibiteur protéique de la division cellulaire modulant l'activité de FtsZ, *dicF* code pour un court ARN de 53 nts de long (Figure 14A) [235,237], maturé à partir d'un précurseur, dépourvu de phase codante et particulièrement stable (Figure 14A). L'ARN DicF agit négativement comme ARN antisens sur l'expression du gène *ftsZ* [222]. DicF bloque la division cellulaire en empêchant la traduction de l'ARNm de *ftsZ*, ceci indépendamment de DicB. Les deux ARN ne sont pas strictement complémentaires mais une comparaison de séquences suggère une interaction entre la séquence Shine et Dalgarno de l'ARNm et DicF (Figure 14B, [222]). Celle-ci n'a pas été confirmée expérimentalement, mais la conservation de l'appariement potentiel dans différentes espèces eubactériennes suggère qu'elle soit une réalité biologique [238].

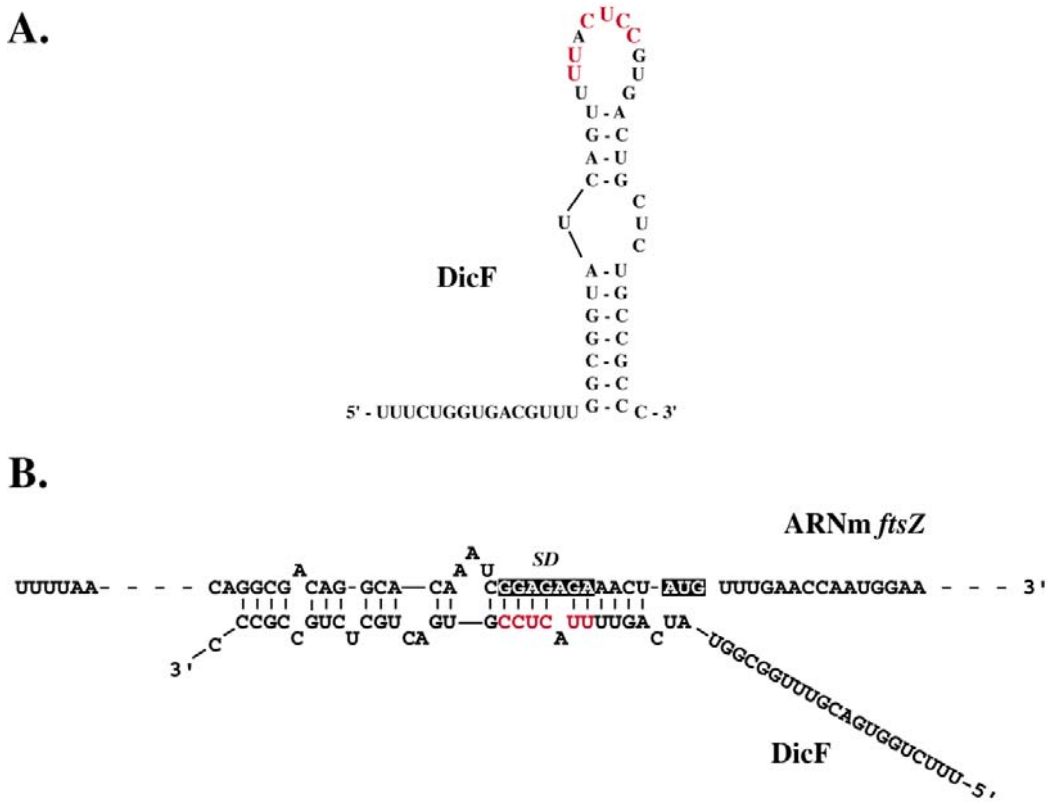

Figure 14 : ***Modèles de structures secondaires de l'ARN DicF (A) et du complexe DicF-*fstZ *(B)***. Seule une région de la 5'-UTR de l'ARNm *ftsZ* contenant le site de reconnaissance à DicF est représentée. La séquence de DicF complémentaire à la séquence Shine et Dalgarno de *ftsZ* est représentée en rouge. La séquence SD de *ftsZ* et le codon AUG initiateur de la traduction sont représentés en caractères blancs sur fond noir.

Il est intéressant de noter que les positions de l'ARN antisens DicF impliqués dans l'interaction avec la séquence Shine et Dalgarno de l'ARNm sont exposés dans la boucle apicale du motif en tige-boucle (Figure 14A). L'initiation de l'interaction pourrait se faire *via* une reconnaissance boucle-région simple-brin, comme cela a été montré dans de nombreux autres systèmes (§ V.B et VI). L'implication de cette séquence dans les contacts initiaux pourrait favoriser l'interaction et optimiser la répression de la traduction de *ftsZ*.

2. *Les ARN antisens codés par les éléments génétiques accessoires*

Les ARN antisens régulant diverses fonctions des éléments accessoires procaryotiques furent les premiers identifiés et à l'heure actuelle les mieux étudiés. Leurs caractéristiques ont d'ailleurs souvent servi à la conception de molécules antisens artificielles.

a. La régulation de la transposition de l'élément *IS10*

L'exemple de la régulation de la transposition de l'élément IS10 est un des exemples qui démontre l'importance de la structure de l'ARN antisens pour sa fonction. Il présente des caractéristiques réminiscentes à de nombreux ARN antisens, leur garantissant leur efficacité d'interaction avec leur ARN cible.

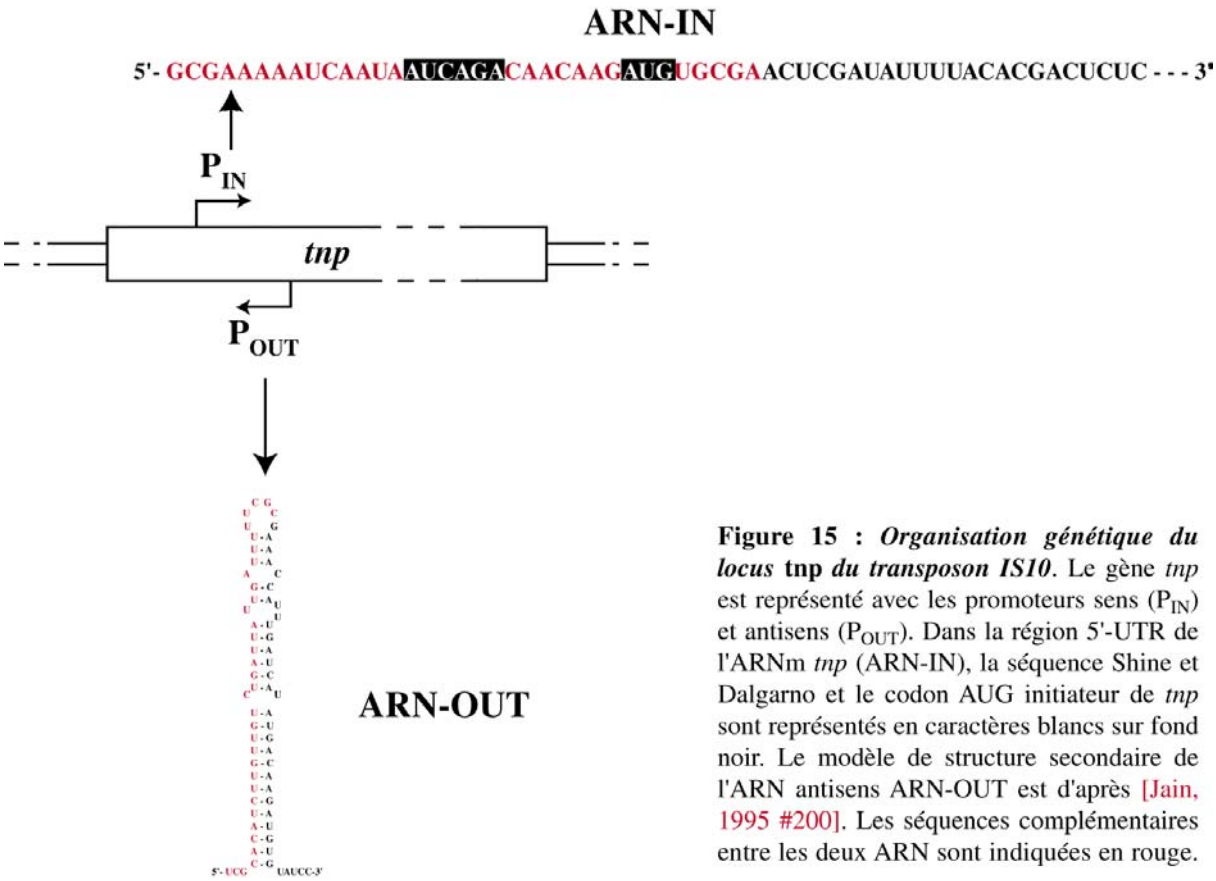

Figure 15 : *Organisation génétique du locus* tnp *du transposon IS10*. Le gène *tnp* est représenté avec les promoteurs sens (P_{IN}) et antisens (P_{OUT}). Dans la région 5'-UTR de l'ARNm *tnp* (ARN-IN), la séquence Shine et Dalgarno et le codon AUG initiateur de *tnp* sont représentés en caractères blancs sur fond noir. Le modèle de structure secondaire de l'ARN antisens ARN-OUT est d'après [Jain, 1995 #200]. Les séquences complémentaires entre les deux ARN sont indiquées en rouge.

Deux transcrits majeurs sont exprimés à partir de l'élément transposable IS10 chez *E. coli* (Figure 15). Le premier est l'ARNm *tnp* ou ARN-IN, produit à partir du promoteur faible P_{IN}, codant pour la transposase. Le second est l'ARN-OUT de 69 nts, produit à partir du promoteur fort P_{OUT} et en orientation inverse du précédent. Les deux transcrits se chevauchent sur une distance de 35 nucléotides englobant le site de fixation du ribosome de l'ARNm ARN-IN dans la partie 5' terminale de ce dernier (Figure 15A, [239]). Une transposition non contrôlée entraînerait des dommages au sein du génome de l'organisme hôte. C'est pourquoi la transposition est inhibée lorsque plus d'une copie du transposon est présente par cellule [240]. Ce contrôle est effectué d'une part par méthylation de l'ADN [241] et d'autre part majoritairement par l'action de l'ARN antisens ARN-OUT [240]. Cet ARN réprime l'expression de la transposase au niveau post-transcriptionnel en s'appariant à la région 5' de

l'ARN-IN sur toute la région de complémentarité entre les deux ARN [242]. Cette interaction empêche la fixation du ribosome *in vitro* et inhibe la traduction *in vivo* [243]. Par ailleurs, la formation du complexe ARN-IN-ARN-OUT crée un site de coupure par la RNase III *in vivo* [244], même si celle-ci ne contribue pas significativement au contrôle.

Le contrôle efficace *in vivo* est directement corrélé à la vitesse d'association des deux molécules *in vitro* [242,245]. Le processus d'appariement se fait en deux étapes. L'interaction est initiée par un petit nombre de nucléotides de la boucle de l'ARN-OUT et les nucléotides 5' terminaux de l'ARN-IN, formant trois paires de bases G≡C (Figure 16). Cette interaction se propage ensuite dans la direction 3' de l'ARN-IN et 5' de l'ARN-OUT afin de former un duplexe étendu. L'efficacité du système repose sur la structure de l'ARN antisens. La progression de l'interaction nécessite une fusion de la tige de l'ARN-OUT. Cette étape requiert des nucléotides non appariés dans l'hélice [245]. Ils pourraient servir d'éléments déstabilisateurs et faciliteraient la formation d'une hélice intermoléculaire régulière aux dépens des appariements intramoléculaires imparfaits. Finalement, une répression optimale de l'expression de la transposase nécessite que l'ARN antisens soit particulièrement stable. Cette stabilité de l'ARN-OUT ($t_{1/2} \approx 60$ min) résulte de la présence de nucléotides non appariés dans l'hélice qui le protègent contre l'attaque par la RNase III [246]. La stabilité intrinsèque de la partie inférieure de la tige inhibe aussi l'attaque par des exoribonucléases 3'--> 5'.

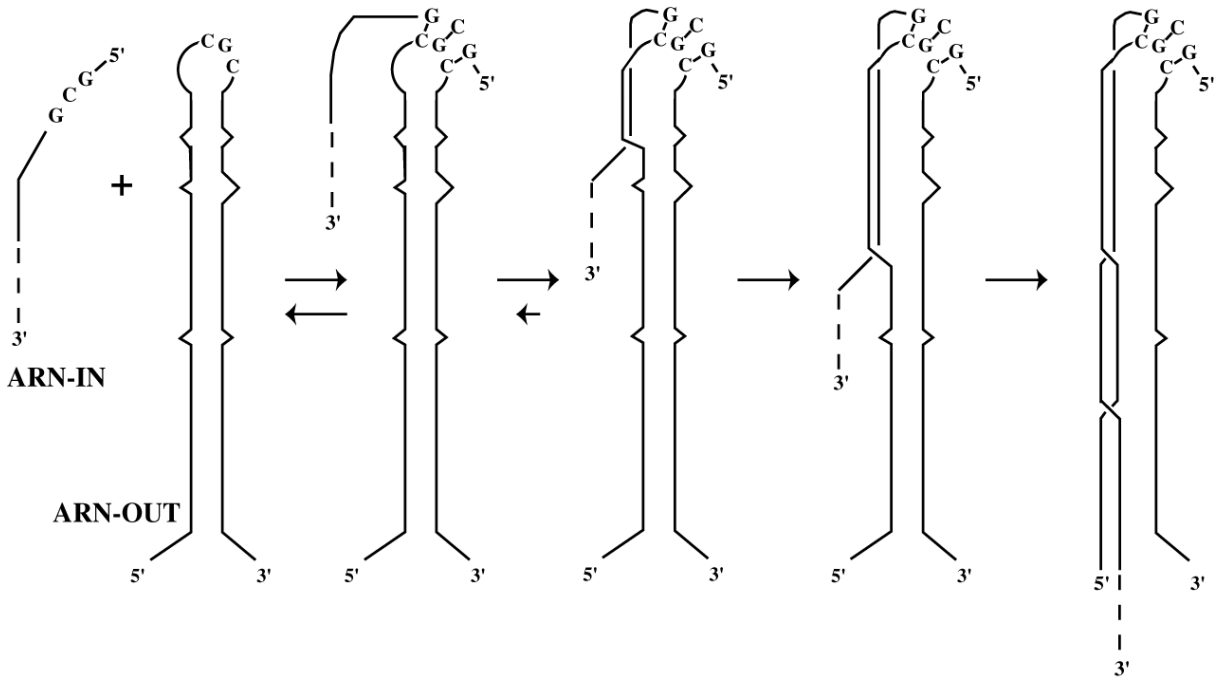

Figure 16 : ***Modèle du mécanisme d'appariement entre les ARN ARN-IN et ARN-OUT*** (d'après [242]) (Voir le texte pour le détail).

b. Les ARN antisens chez les bactériophages

Le tableau 5 répertorie tous les ARN antisens qui ont été caractérisés chez les bactériophages.

Bactériophage	Organisme hôte	ARN antisens	ARN cible	Taille de l'antisens (nts)	Fonction régulée	Mode de contrôle	Référence
P1/P7	*E. coli*	C4	*ant*	77	synthèse de l'antirépresseur inhibée	Terminaison de la transcription	[247]
P1	*E. coli*	P53as	*kilA*	180	inhibition d'un gène nécessaire à la réplication lytique	Traduction ?	[248]
P4	*E. coli*	C4	*CI*	69	Contrôle de la surinfection par un autre prophage	Terminaison de la transcription	[249]
P22	*Salmonella typhimurium*	Sas	*sieb/ esc*	105	Contrôle de la surinfection par un autre prophage	Inhibition de la Traduction	[250]
P22	*Salmonella typhimurium*	Sar	*ant*	68	synthèse de l'antirépresseur inhibée	Inhibition de la traduction Dégradation de la cible	[251]
ϕH	*Halobacterium salinarium*	Tant	*T1*	201	Inhibition de la phase lytique	Dégradation de la cible	[38]
λ	*E. coli*	Oop	C_{II}	77	Inhibition de la phase lytique	Dégradation de la cible	[54]

Tableau 5 : ***ARN antisens caractérisés chez les bactériophages***.

Tous les ARN antisens caractérisés chez les bactériophages contrôlent leur passage de la phase lytique à la phase lysogène dans les cellules hôtes. Ce contrôle s'effectue néanmoins de diverses manières et principalement au niveau post-transcriptionnel. Cette observation est en accord avec la nécessité d'une inhibition rapide du gène impliqué dans l'entrée en phase lytique. Ceci implique une fois de plus que ces ARN antisens doivent agir dans un intervalle de temps très court. Une fuite dans le mécanisme de régulation entraînerait la mort de la cellule hôte sans que le bactériophage n'aie eu le temps de se répliquer.

La plupart de ces ARN antisens (Tableau 5) sont synthétisés à partir du même locus que leur ARN cible, en orientation inverse, à l'exception de l'ARN C4 régulant la synthèse de l'antirépresseur *ant* chez les bactériophages P1 et P7 [247]. Pour les ARN antisens synthétisés en *cis*, les régions de complémentarité avec leur cible ont clairement été définies dans tous les cas. Cependant, dans un seul de ces systèmes (Sar/*ant* du bactériophage P22) les structures des deux ARN importantes pour l'efficacité du contrôle ont été caractérisées. La

reconnaissance initiale Sar/*ant* implique aussi une interaction boucle-boucle. Celle-ci étant instable, elle doit être rapidement stabilisée [251]. Finalement, dans le cas du système P1/P7, un nouveau mode de synthèse de l'ARN antisens a été caractérisé. L'ARN antisens C4 est transcrit à partir du même promoteur que son ARNm cible [252] et maturé à partir d'un précurseur polycistronique [253]. La traduction d'*ant* est couplée à celle de la courte phase ouverte de lecture *orfX*. Deux régions de l'antisens C4 sont complémentaires au transcrit *orfx-ant* sur des longueurs de 7 et 8 nts, respectivement. La dernière de ces régions peut s'apparier à la région Shine et Dalgarno de l'ARNm *orfx*, inhibant sa traduction. L'absence de protection par les ribosomes induit un changement de conformation du transcrit qui adopte une topologie contenant un terminateur de transcription ρ-indépendant [247]. La transcription d'*ant* est donc inhibée indirectement *via* l'inhibition de la traduction d'*orfx*.

VI. Cas particuliers des contrôles de fonctions plasmidiques

Les plasmides bactériens présentent un terrain d'étude fabuleux des ARN antisens. Ces derniers sont extrêmement efficaces, la vitesse de formation des complexes avec leurs cibles étant optimale ($\approx 10^6$ $M^{-1}.s^{-1}$) et dépendante de leur structure. Ils régulent une grande variété de fonctions vitales pour les plasmides (Tableau 6). Par ailleurs, le mode de régulation antisens-dépendant d'une fonction donnée peut différer entre deux plasmides. Encore une fois, ces points mettent en avant la grande versatilité des molécules d'ARN pour la régulation de l'expression des gènes.

Plasmide	ARN antisens	ARN cible	Taille de l'antisens (nts)	Fonction contrôlée	Mode de contrôle	Référence
R1 (IncFII)	CopA	ARNm *repA*	91	Réplication plasmidique	Inhibition de la traduction/dégradation	[254]
Col1b-P9 (IncIα) et dérivés	Inc	ARNm *repZ*	67	Réplication plasmidique	Inhibition de la traduction et de la formation d'une structure activatrice	[255]
ColE1 et dérivés	ARN-I	ARN-II	108	Réplication plasmidique	Inhibition de la maturation de l'amorce	[256]
pT181/pIP501 /pAMβ1	ARN-I	ARNm *repC*	83	Réplication plasmidique	Atténuation transcriptionnelle	[57,257,258]
ColE2	ARN-I	ARNm *rep*	115	Réplication plasmidique	Inhibition de la traduction	[259]
pLS1 et dérivés	ARN-II	ARNm *cop-rep*	50	Réplication plasmidique	Inhibition de la traduction	[260]
R1162	Ct	ARNm *repI*	75	Réplication plasmidique	Inhibition de la traduction	[261]
R1 (IncFII)	Sok	ARNm *mok/hok*	64	Lyse Cellulaire	Inhibition de la traduction/dégradation	[262]
R1 (IncFII) et groupe IncFI	FinP	ARNm *traJ*	79	Conjugaison	Inhibition de la traduction/dégradation	[263]

Tableau 6 : ***ARN antisens plasmidiques bactériens***. Seuls les systèmes les mieux étudiés sont indiqués dans chaque groupe. Les groupes d'incompatibilité (si connus) auxquels les plasmides appartiennent sont indiqués entre parenthèses.

A. Biologie des plasmides bactériens

Les plasmides sont des entités génétiques extrachromosomiques avec un nombre de copies par cellule qui leur est propre. Ils ont été caractérisés chez les archaebactéries, les eubactéries et chez les eucaryotes (pour une revue, [264]). Les plasmides bactériens confèrent généralement des avantages phénotypiques à l'hôte, notamment de nombreuses résistances envers des antibiotiques. Leur étude présente pour cette raison un intérêt médical puisqu'ils sont responsables de nombreuses surinfections en milieu hospitalier. Leur utilisation courante en laboratoire présente quant à elle de nombreux intérêts biotechnologiques.

Si les plasmides apportent de nombreux bénéfices à la bactérie hôte, la présence de nombreux exemplaires dans une cellule constitue néanmoins une charge métabolique. La réplication plasmidique doit par conséquent être contrôlée de manière à maintenir un équilibre entre coût énergétique et avantage phénotypique. Ce contrôle permet de limiter la réplication plasmidique à un événement par génération cellulaire. Parallèlement, la plupart des plasmides ont également développé des mécanismes pour contrôler leur répartition équitable dans les cellules filles lors de la division cellulaire. Enfin, la troisième fonction essentielle au maintien d'un plasmide au sein d'un population bactérienne est le transfert conjugatif. Tout comme la réplication et la ségrégation, celui-ci doit être contrôlé de façon précise pour ne pas tuer l'organisme hôte.

Réplication, ségrégation et transfert conjugatif sont les trois fonctions les plus essentielles au maintien stable d'un plasmide dans une population bactérienne. De manière surprenante, elles sont toutes trois contrôlées par des ARN antisens dans la plupart des systèmes plasmidiques caractérisés. La manière dont le contrôle s'effectue varie cependant d'un cas à l'autre. Dans cette partie, nous décrirons les différentes manières de contrôler ces trois fonctions à l'aide d'un ARN antisens.

B. Contrôle du transfert conjugatif

La conjugaison bactérienne est permise par des plasmides qui codent pour toutes les fonctions nécessaires au transfert d'ADN d'une cellule à une autre. Le système de transfert conjugatif le plus étudié est celui des plasmides de type F. Parmi eux, le plasmide R1 appartient au groupe d'incompatibilité IncFII. Le transfert conjugatif de ces plasmides nécessite la synthèse de pili sexuels à la surface de la cellule donneuse qui lui permettent de contacter la cellule receveuse [265]. La trentaine de gènes nécessaires à la formation des pili et au contrôle du transfert sont codés par l'opéron *tra* présent sur les plasmides de type F, et long de 33 kb. La transcription de l'opéron *tra* est contrôlée positivement par l'activateur

transcriptionnel TraJ [266]. La synthèse de TraJ est elle même régulée négativement par le système FinOP (pour Fertility Inhibition) dans lequel FinP est un ARN antisens de 79 nts (Figure 17A) [263] et FinO est une protéine de 22 kDa [267]. FinO est FinP agissent de concert pour réguler la fertilité plasmidique et sont nécessaires au contrôle [267]. FinP est synthétisé à partir du même locus que le gène *traJ*, mais en orientation inverse. FinP est par conséquent complémentaire à la partie 5'-UTR de l'ARNm *traJ*, englobant le site de fixation du ribosome et les deux premiers codons [263] (Figure 17B). FinP contrôle négativement l'expression de *traJ* à trois niveaux. Dans un premier temps, la transcription constitutive de FinP inhibe la transcription convergente de *traJ* [268]. L'essentiel du contrôle se fait cependant au niveau post-transcriptionnel. FinP interagit avec l'ARNm *traJ in vitro* [269] et inhibe sa traduction *in vivo* [268]. Enfin, le complexe FinP-*traJ* est rapidement dégradé par la RNase III *in vivo* [53].

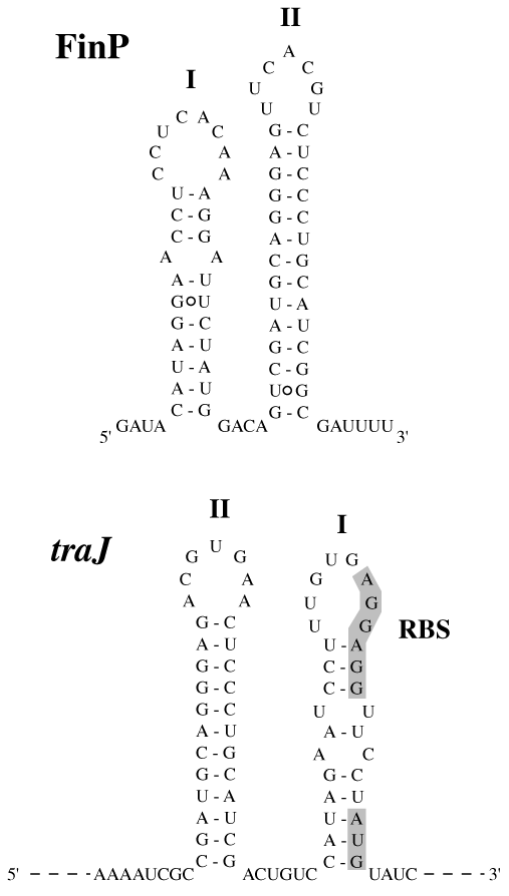

Figure 17 : ***Modèles de structures secondaires de l'ARN antisens FinP et de la région complémentaire dans la partie 5'-UTR de l'ARNm* traJ**. Les tiges boucles I et II sont indiquées dans les deux cas. Le site de fixation du ribosome (RBS) et le codon AUG initiateur de la traduction sont représentés en grisé.

L'efficacité de l'interaction FinP-*traJ* repose sur la structure des deux ARN [269]. Tout deux contiennent deux motifs en tiges-boucles (Figure 17), conservées entre différents

plasmides et essentielles à une interaction efficace *in vitro* et au contrôle optimal *in vivo* [263,270]. La séquence Shine et Dalgarno de *traJ* est partiellement contenue dans la boucle I (Figure 17B). Deux indications vont dans le sens qu'une interaction entre les boucles II des deux ARN est l'étape initiale et limitante de la formation du complexe. En effet, les mutations qui dérépriment le plus sévèrement le taux de conjugaison du plasmide R1 sont toutes localisées dans la boucle II de FinP [263,270]. Il a également été montré que les séquences des ARN FinP sont très conservées entre différents allèles plasmidiques mais qu'elles divergent au niveau de la boucle II. Cette observation suggère que cette région est responsable de la spécificité d'interaction entre l'ARN antisens et sa cible. Il ne peut donc pas y avoir d'interactions croisées entre les ARN de plasmides différents. Ce mode d'interaction de type boucle-boucle apparaît comme une caractéristique redondante de nombreux systèmes antisens.

Une particularité du contrôle de l'expression de *traJ* par l'ARN antisens FinP est l'intervention d'une protéine pour l'efficacité du système [267]. La protéine FinO n'est pas spécifique du plasmide et protège FinP contre l'attaque par la RNase E *in vivo* [53,271]. La concentration d'ARN antisens présent dans les cellules est donc élevée, garantissant une forte répression de la conjugaison. De plus, FinO reconnaît les secondes tiges-boucles de FinP et de *traJ* indépendamment de leur séquence et augmente la vitesse de l'interaction ARN-ARN de cinq fois *in vitro* [44]. Les régions requises pour la reconnaissance des deux ARN par FinO ont été caractérisées et résident dans les hélices II [270]. De même manière, les régions de la protéine impliquées dans la reconnaissance ont également été caractérisées [272], mais le mécanisme par lequel FinO facilite l'interaction demeure spéculatif.

C. Contrôle de la ségrégation plasmidique

Pour qu'un plasmide soit maintenu de façon stable dans une population bactérienne en absence de pression de sélection, chaque cellule fille doit en recevoir au moins une copie lors de la division cellulaire. Pour les plasmides à haut nombre de copies (> vingt copies par cellules) chaque cellule fille hérite statistiquement d'au moins un exemplaire du plasmide. En revanche, pour les plasmides à faible nombre de copies (< cinq copies par cellules) la ségrégation doit être contrôlée activement. Ceux-ci ont développé différents mécanismes pour assurer ce contrôle (pour une revue, [273]). Le contrôle de la ségrégation se fait grâce à une « mémoire » du plasmide puisque seules les cellules filles n'héritant pas du plasmide sont détruites. Dans ce but, les plasmides synthétisent un ARNm extrêmement stable (la mémoire) codant pour une toxine, ainsi qu'un ARN antisens antidote beaucoup plus labile. Les cellules contenant un plasmide survivent grâce à l'expression des deux gènes. En revanche, dans les

cellules qui en sont dépourvues l'antisens est rapidement dégradé tandis que l'ARNm est encore présent. Le système *hok/*Sok caractérisé chez le plasmide R1 est très représentatif de ces mécanismes de contrôle de la ségrégation plasmidique. La toxine Hok est une porine déstabilisant la membrane cellulaire et est synthétisée à partir de l'ARNm *hok* [274]. L'ARN antisens Sok représente quant à lui l'antidote [Thisted, 1994 #229]. *hok* et Sok sont synthétisés à partir du même locus génétique sur R1 mais en orientation inverse (Figure 18A).

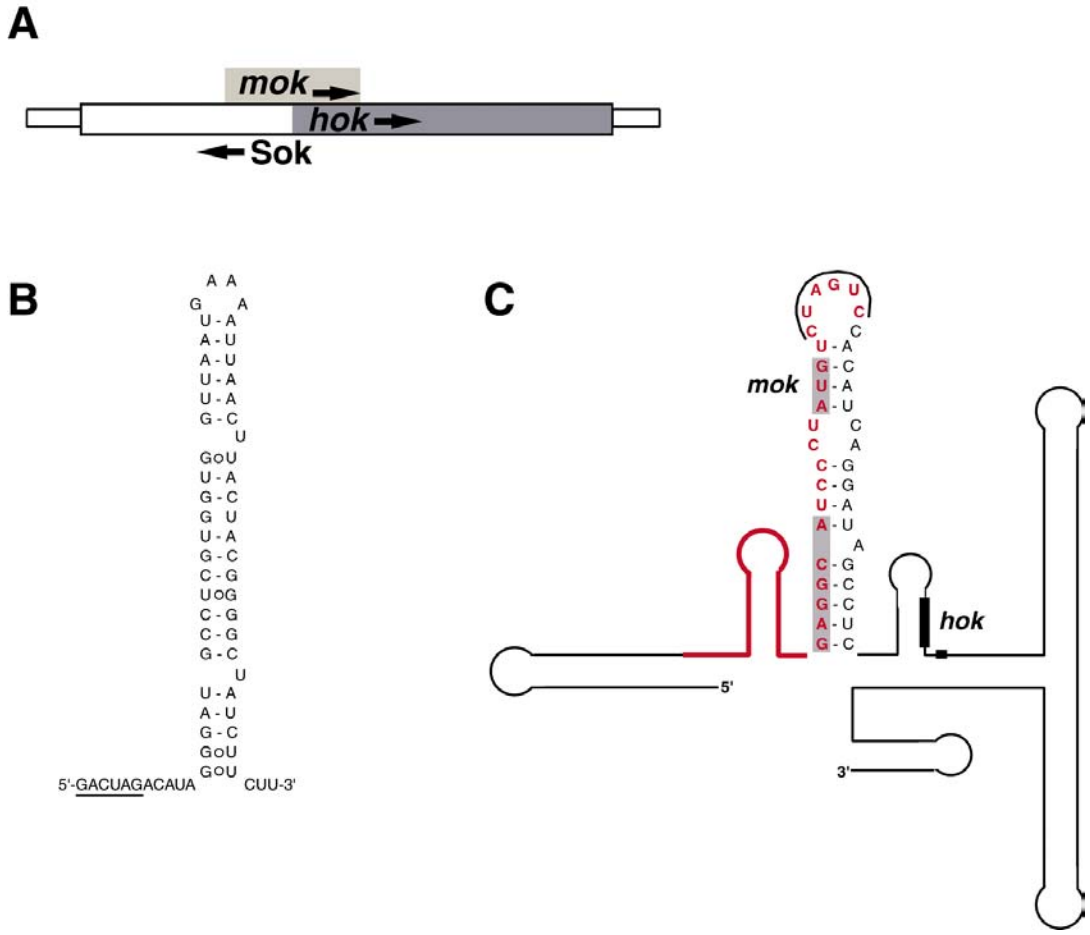

Figure 18 : ***Le locus génétique* hok/*Sok et modèles des structures secondaires des deux ARN***. (A) Organisation génétique du locus *hok/Sok*. Le gène *hok* et le gène chevauchant *mok* sont indiqués, ainsi que le sens de la transcription. L'ARN antisens Sok est indiqué en dessous. (B) modèle de structure secondaire de l'ARN antisens Sok. Les nucléotides de Sok et *hok,* impliqués dans l'interaction initiale, sont soulignés. (C) modèle de structure secondaire de l'ARNm *hok* maturé. La région de complémentarité avec Sok est indiquée en rouge. La séquence Shine et Dalgarno et le codon initiateur de la traduction de *mok* sont indiqués sur fonds grisés et ceux de *hok* par des boîtes noires.

Afin de ne pas tuer les cellules contenant le plasmide, la traduction de l'ARNm « mémoire » *hok* doit être contrôlée de manière très fine. En cours de transcription, l'ARNm *hok* passe par une succession d'intermédiaires de repliement qui bloquent l'accès des ribosomes [275,276]. Cette forme de *hok* ne peut également pas être reconnue par l'ARN antisens Sok. C'est une maturation exonucléolytique par son extrémité 3' qui permet un

changement conformationnel. Suivant le cas (plasmide présent ou non), cette maturation active soit la traduction, soit la reconnaissance par l'ARN antisens Sok [277]

Sok comporte 64 nts et est entièrement complémentaire à la région 5'-UTR de *hok*. Il est rapidement dégradé à l'intérieur des cellules avec un temps de demi-vie de l'ordre de 30 secondes [278]. La traduction de *hok* nécessite celle du peptide activateur *mok* par couplage traductionnel (Figures 18A & C) [279]. Sok interagit avec sa séquence complémentaire en amont de *mok* et inhibe directement sa traduction. L'inhibition de la traduction de *hok* par Sok se fait donc indirectement *via* l'inhibition de la traduction de *mok*. Finalement le complexe Sok-*hok* est rapidement clivé par la RNase III *in vivo* [52]. Comme dans d'autres systèmes, ce contrôle négatif est particulièrement efficace car il intervient de façon post-transcriptionnelle, d'une part par inhibition de la traduction et d'autre part par dégradation de l'ARN messager.

Pour la survie des cellules, l'interaction Sok-*hok* doit être optimale. A nouveau, l'efficacité de l'interaction ARN-ARN repose sur la structure des deux molécules. Dans l'ARNm *hok*, le site de fixation à l'antisens est présenté dans une structure en tige-boucle (Figure 18C). Parmi les sept nucléotides de la boucle, six sont complémentaires aux premiers nucléotides en 5' de Sok se trouvant eux dans la région non structurée. L'appariement entre la boucle de *hok* et la région 5' de Sok est l'étape limitante de la formation du complexe [280]. Le motif 5'-YUNR-3' de la boucle de *hok* adopte une structure particulière de type U-turn [66] nécessaire à une interaction *hok*-Sok optimale *in vitro* [66]. Par ailleurs, des nucléotides non appariés dans la tige de reconnaissance de l'antisens de *hok* sont nécessaires pour garantir une interaction optimale *in vitro* et un contrôle efficace *in vivo* [276]. Cette caractéristique est également redondante parmi les ARN antisens procaryotiques. Ainsi, ces déterminants structuraux permettent à *hok* et Sok d'interagir avec une constante de vitesse de $2x10^6$ $M^{-1}.s^{-1}$ [66]. Cet ordre de grandeur est caractéristique des systèmes naturels les plus efficaces puisqu'aucun couple antisens-cible n'interagit avec une constante de vitesse supérieure.

Des systèmes homologues ont été caractérisés dans de nombreux plasmides, y compris chez les bactéries Gram positives [281], ainsi que dans le chromosome d'*E. coli* (pour une revue, [262]). La présence de ces copies chromosomiques n'est à l'heure actuelle pas expliquée.

D. Mécanismes de contrôle de la réplication plasmidique

La réplication plasmidique doit être régulée de manière à ne pas détourner la machinerie de réplication de l'hôte. Les plasmides bactériens ont donc développé différents mécanismes qui impliquent généralement un ARN antisens. Ces derniers ont tous deux propriétés fondamentales. Ils sont synthétisés de manière constitutive à partir d'un promoteur plasmidique fort et sont hautement instables à l'intérieur des cellules ($t_{1/2}$ < 1 minute), avec une seule exception (voir plus bas au § VI.D.2).

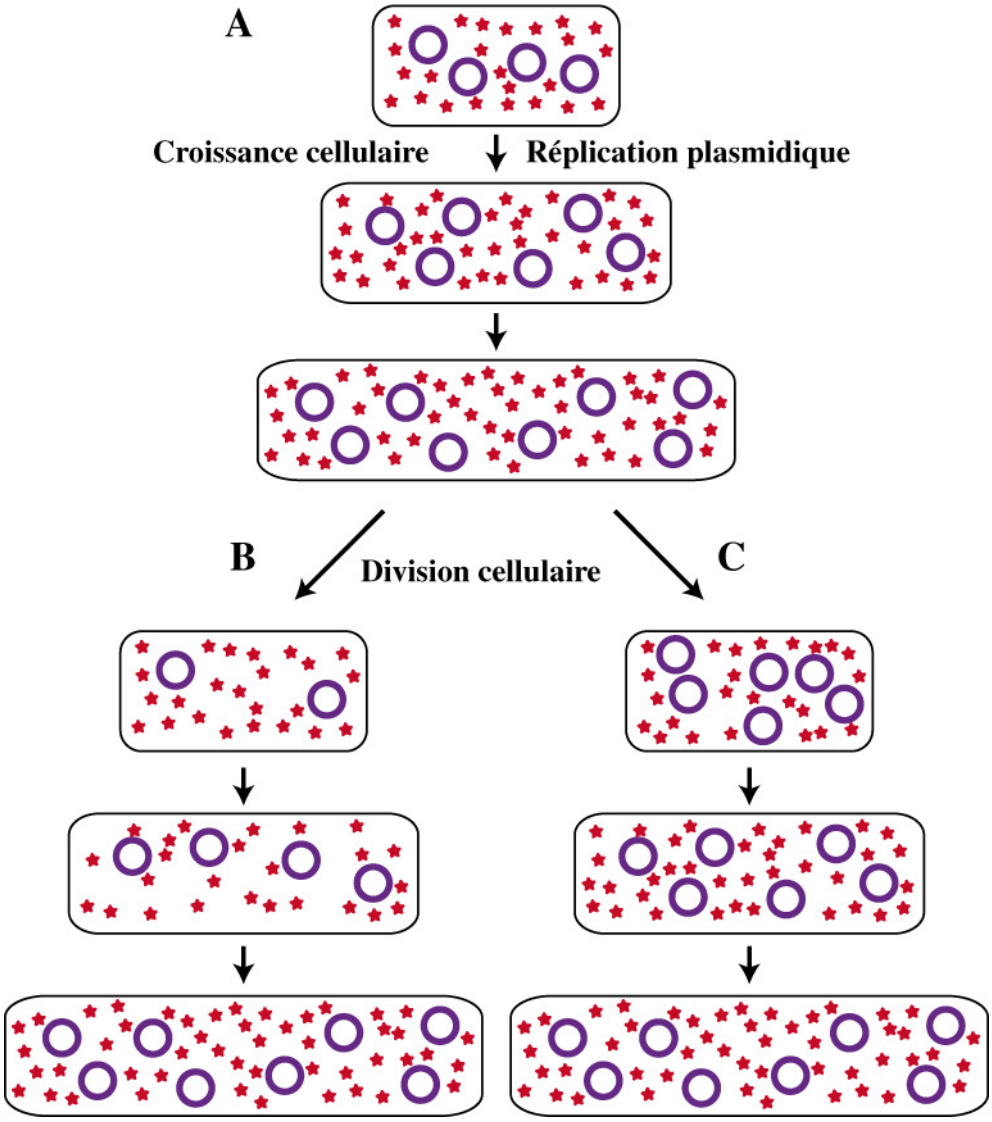

Figure 19 : ***Principe de dilution de Pritchard***. (A) Une bactérie en court de croissance est représentée. Les plasmides bactériens sont représentés par des cercles bleus et l'inhibiteur de la réplication plasmidique par une étoile rouge. (B) et (C) Deux cas de figures pouvant être rencontrés lors de la division cellulaire (Voir le texte pour le détail).

Basé sur le principe de dilution de Pritchard (Figure 19, [282]), ces deux propriétés leur permettent de mesurer en permanence la concentration intracellulaire en plasmide. Prenons l'exemple d'un plasmide présent de manière stable à raison de quatre copies par cellule. Sa réplication est réprimée grâce à la synthèse constitutive d'un ARN antisens inhibiteur instable. En cours de croissance cellulaire (Figure 19A), le volume des cellules augmente, diluant ainsi l'inhibiteur. La réplication reprend et la concentration en antisens augmente avec le nombre de copies du plasmide. Un nouvel équilibre est atteint : si le volume des cellules double avant la division, le nombre de copies double également. Lors de la division cellulaire, deux cas peuvent être distingués. Si une bactérie fille reçoit moins de la moitié des copies du plasmide (Figure 19B), l'inhibiteur instable est rapidement dégradé. La réplication plasmidique est déréprimée et continue jusqu'à atteindre un nombre de copie à partir duquel la concentration en antisens synthétisé est à nouveau inhibitrice. Dans le cas contraire (Figure 19C), si une cellule fille reçoit plus de la moitié des copies du plasmide, la réplication est immédiatement réprimée. Elle ne peut reprendre que lors de la croissance de la cellule avant sa division, lorsque l'inhibiteur est à nouveau dilué.

Les ARN antisens vont agir principalement sur l'étape limitante de la réplication plasmidique et la régulent par des mécanismes variés (Tableau 6). Dans le système ColEI, l'antisens inhibe la formation d'un complexe amorce/matrice productif lors de l'initiation de la réplication plasmidique. Dans les systèmes pT181 et pIP501, l'antisens inhibe la transcription du gène codant pour la protéine d'initiation de la réplication. Enfin, la traduction de l'ARNm codant pour la protéine d'initiation de la réplication peut être réprimée soit en empêchant la formation d'une structure activatrice (plasmides appartenant aux groupes d'incompatibilité IncIα, IncB, IncZ, IncL/M et IncK) soit en inhibant la traduction d'un peptide leader (plasmides appartenant aux groupes d'incompatibilité IncFII, IncFI, IncIα, IncB, IncZ, IncL/M et IncK).

1. *Contrôle de la formation de l'amorce de réplication*

Plusieurs plasmides assurent leur réplication en utilisant une amorce ARN. Parmi eux se trouvent les plasmides RSF1030, p15A, CloDF13 et ColE1 [256]. Ce dernier est le plus étudié d'entre eux et sert de modèle à la compréhension du contrôle de la réplication dans cette classe de plasmides. La transcription de l'amorce ARN-II débute 555 nts en amont de l'origine de réplication (Figure 20 : étape 1) [35]. L'ARN-II élongué ne reste associé à la matrice ADN qu'à l'endroit où a lieu la transcription (étape 2). Cependant, grâce à un changement conformationnel, l'ARN-II peut former un hybride persistant et étendu avec l'ADN lorsque le complexe de transcription arrive à proximité de l'origine de réplication

(étape 3). Ce complexe persistant devient cible de la RNase H au niveau de l'origine (étape 4) et l'extrémité 3' libre générée sert d'amorce à la synthèse d'ADN par l'ADN polymérase I (étape 5).

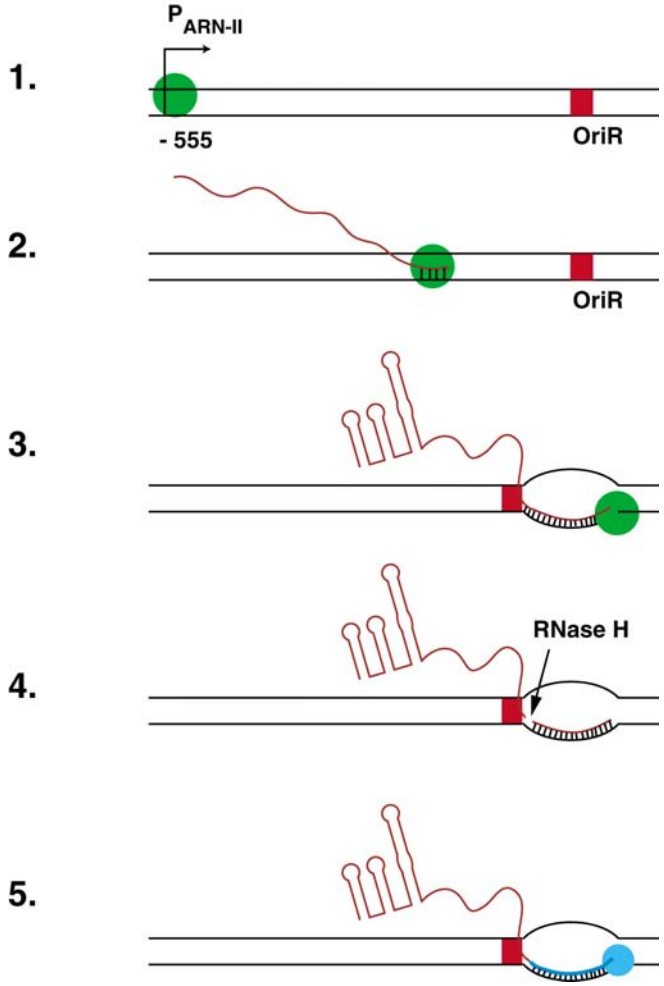

Figure 20 : ***Mécanisme d'initiation de la réplication chez les plasmides de type ColE1***. La région contenant l'origine de réplication (OriR : boîte rouge) est représentée. La synthèse de l'amorce ARN-II utilisée pour l'initiation de la réplication débute 555 nts en amont de l'origine. L'ARN-polymérase ADN dépendante est représentée par un rond vert et l'ARN naissant pas une ligne rouge. Le site d'action de la ribonucléase H est indiqué. L'ADN-polymérase ADN-dépendante est représentée par un rond bleu la molécule d'ADN naissante par une ligne bleue. Le détail de chaque étape est décrit dans le texte.

Un ARN antisens modulant la réplication plasmidique de ColE1 a été caractérisé [283]. Celui-ci est transcrit à partir de la région codant pour l'ARN-II, en orientation inverse. Des mutations dans cette région affectent le taux de réplication du plasmide *in vivo*. L'antisens ARN-I comporte 108 nts et est complémentaire aux 108 nucléotides 5' terminaux de l'ARN-II (Figure 21).

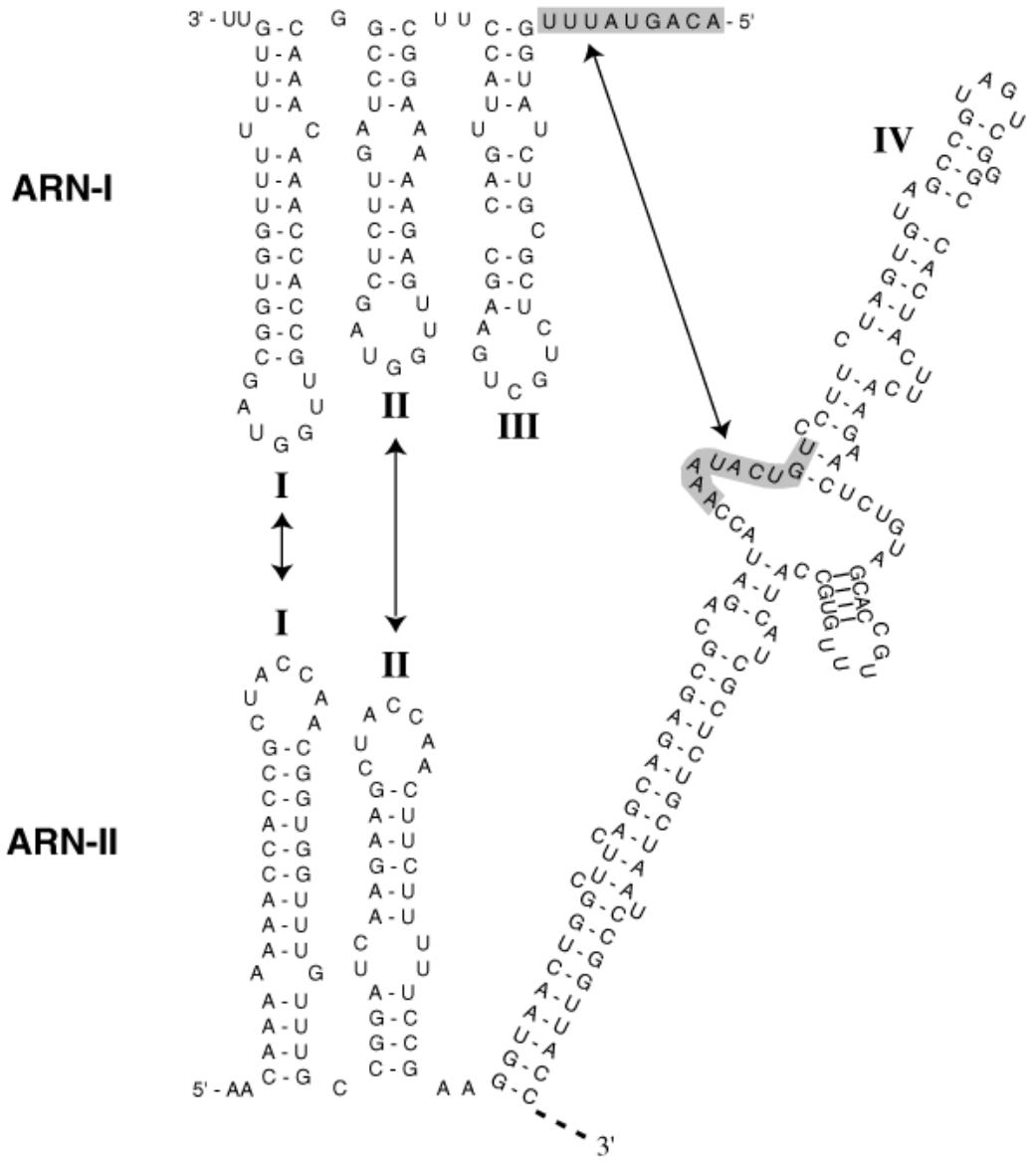

Figure 21 : ***Modèles de structures secondaires de l'ARN antisens ARN-I et de l'amorce ARN-II de la réplication plasmidique***. Les différentes tiges-boucles des deux ARN sont numérotées ainsi que le domaine IV de l'ARN-II. Les interactions initiales impliquant les tiges-boucles I et II des deux ARN sont représentées par des flèches, de même que l'interaction entre la région 5' de l'ARN-I et le domaine IV de l'ARN-II. Les régions des deux ARN impliquées dans cette dernière interaction sont grisées.

En cours de transcription, l'ARN-II passe par une succession d'intermédiaires conformationnels [284]. Chaque domaine structural a une implication fonctionnelle et est important pour assurer la formation de l'hybride stable. Parmi eux, le domaine IV (Figure 21) doit être correctement structuré pour permettre le repliement du reste de la molécule. *In vitro*, l'ARN-I forme un complexe stable avec l'ARN-II [285]. Lorsque l'ARN-I interagit avec son site cible sur l'ARN-II, le domaine IV est séquestré par le complexe ARN-ARN. L'ARN-II ne peut plus adopter la conformation nécessaire à la formation de l'hybride stable

amorce/matrice au niveau de l'origine de réplication et n'est plus reconnu par la RNase H *in vivo* [256]. La réplication plasmidique est alors inhibée.

L'ARN-I est synthétisé de façon constitutive et est rapidement dégradé par l'action concertée de la RNase E et d'exonucléases 3'--> 5' suite à sa polyadénylation [59,286,287]. La fenêtre temporelle pendant laquelle l'ARN-I peut agir efficacement est relativement courte. Il faut en effet que l'interaction avec l'ARN-II se fasse avant que la transcription de ce dernier atteigne l'origine de réplication de ColE1. L'efficacité de l'interaction ARN-I-ARN-II repose sur la structure des deux ARN. L'ARN-I contient trois structures en tige-boucle et une extension 5' simple-brin (Figure 21). L'appariement est initié par au moins deux interactions de type boucle-boucle [288,289]. Ces appariements initiaux sont suivis par des contacts additionnels impliquant la région 5' de l'ARN-I et le domaine IV de l'ARN-II (Figure 21).

Les interactions boucle-boucle initiales entre l'ARN-I et l'ARN-II sont cruciales pour garantir une reconnaissance optimale entre les deux ARN [290]. Elles impliquent chacune un appariement complet des boucles I et II des deux ARN sur sept positions [291]. Cependant, ce complexe est instable et doit être rapidement stabilisé par des interactions additionnelles. Cette stabilisation repose essentiellement sur la formation d'une hélice intermoléculaire entre la région 5' simple-brin de l'ARN-I et le domaine IV de l'ARN-II (Figure 21). Une succession d'intermédiaires de stabilités croissantes a été caractérisée dans le processus d'appariement [288]. La formation d'un duplexe étendu entre les deux ARN n'est cependant pas requise pour l'efficacité de l'inhibition. De manière intéressante, l'exemple de ColE1 est le seul système plasmidique avec FinOP (§ VI.B) nécessitant l'intervention d'une protéine pour l'action de l'antisens. Elles agissent néanmoins par des mécanismes distincts. La protéine Rom sous forme dimérique reconnaît spécifiquement la structure formée par l'interaction boucle-boucle initiale indépendamment de la séquence [289]. Contrairement à FinO, elle ne facilite pas l'interaction ARN-I-ARN-II mais stabilise le premier intermédiaire réactionnel. La première étape étant la seule à être réversible, Rom permet de déplacer le processus d'appariement vers le complexe stable [290,292]. Pour cette raison, Rom n'a qu'un effet limité sur le contrôle de la réplication de ColE1.

2. *Contrôle transcriptionnel du gène de la protéine initiatrice de la réplication*

A l'heure actuelle, ce mode de contrôle n'a été observé que chez des plasmides de bactéries Gram-positives. Un ARN antisens inhibe la transcription du gène de la protéine de réplication des plasmides pT181 chez *S. aureus* [57,293], pIP501 chez les Streptocoques [257] et pAMβ1 chez *Enterococcus faecalis* [258], ceci par un mécanisme d'atténuation. La

figure 22 représente le mécanisme tel qu'il a été décrit pour pT181. Dans les cas de pIP501 et pAMβ1, il est absolument similaire.

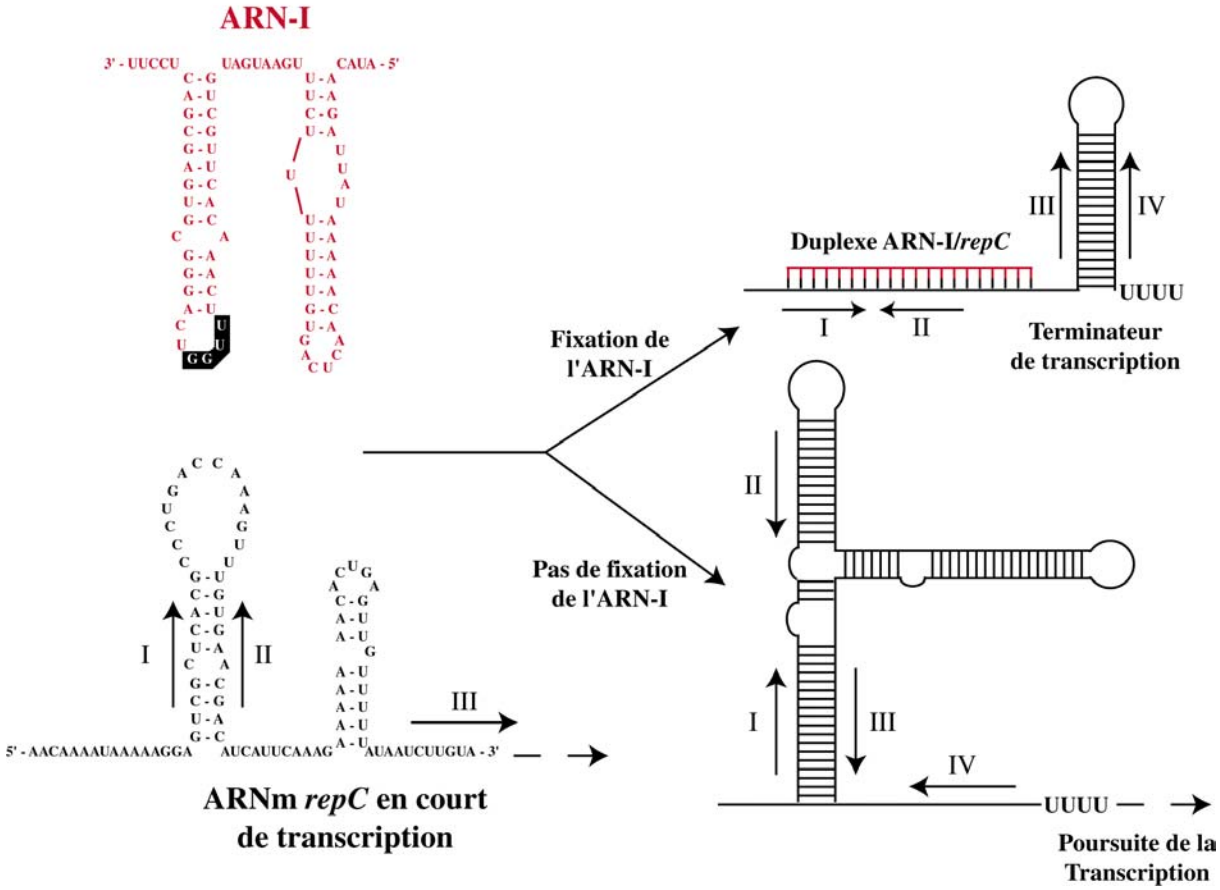

Figure 22 : ***Mécanisme d'atténuation de la transcription et modèles de structures secondaires de l'ARN antisens ARN-I et de la région 5'-UTR de l'ARNm* repC *(d'après [294])***. Seules les régions de l'ARN-I et de *repC* nécessaires pour une interaction optimale sont représentées. L'ARN-I est dessiné en rouge. Les nucléotides de la boucle II de l'ARN-I pouvant former un motif de type U-turn sont représentés en caractères blancs sur fond noir. Les séquences I, II, III et IV dans *repC* sont indiquées par des flèches. Suivant que l'ARN-I se fixe ou non à *repC*, deux cas peuvent être distingués (Voir le texte pour le détail).

Dans des conditions où la réplication doit être réprimée, la concentration intracellulaire de l'ARN antisens ARN-I est élevée. Il peut par conséquent interagir avec sa séquence cible sur l'ARNm *repC* avec une forte probabilité. L'interaction ARN-I-*repC* dans la partie 5'-UTR de l'ARNm séquestre la séquence I qui est complémentaire à la séquence III (Figure 22). La séquence III peut alors interagir avec la séquence IV. La structure ainsi formée est suivie d'une séquence poly-U et constitue alors un signal de terminaison de la transcription ρ-indépendant [295]. Dans des conditions où la réplication doit avoir lieu, la concentration intracellulaire en ARN antisens est faible et donc l'interaction avec l'ARNm *repC* n'a pas lieu. En cours de transcription, la séquence I interagit alors avec la séquence III

l'empêchant d'interagir avec la séquence IV. La transcription peut continuer et la protéine RepC est synthétisée.

La formation du complexe ARN-ARN se fait avec une vitesse de $1x10^6$ $M^{-1}.s^{-1}$. Les structures secondaires de l'ARN-I et de la séquence cible de l'ARNm *repC* ont été déterminées [57,294]. Les deux ARN sont très structurés et contiennent deux motifs en tige-boucle (Figure 22). Dans l'ARN-I toutes deux sont nécessaires pour assurer une interaction optimale *in vitro* [294] et une atténuation efficace de la transcription *in vitro* et *in vivo* [57,294]. L'interaction ARN-ARN est probablement initiée par la deuxième tige-boucle de l'ARN-I et suivie par l'interaction de la boucle I avec sa séquence complémentaire de l'ARNm *repC* (Figure 22). L'étude du système homologue pIP501 a conduit aux mêmes observations [296,297]. Il est intéressant de noter que dans ces systèmes également, la boucle II de l'ARN-I contient un motif de type 5'-YUNR-3'. Ces constatations suggèrent que de manière similaire à de nombreux antisens, une interaction boucle-boucle entre les deux ARN est la première étape du processus d'appariement.

Une particularité du système pIP501 est que l'ARN antisens est particulièrement stable [297]. C'est le seul cas décrit pour un antisens régulant le taux de réplication d'un plasmide. En fait, pIP501 utilise un deuxième mécanisme de contrôle impliquant la protéine CopR. CopR est un répresseur transcriptionnel qui agit sur le promoteur du gène *repR* [298]. A un faible nombre de copie et donc à une faible concentration de CopR, la synthèse de RepR est fortement activée. Cette activation a un double effet : une grande quantité d'ARNm est synthétisée et la transcription convergente inhibe la synthèse de l'ARN antisens [299]. Le rapport ARNm÷ARN antisens augmente ainsi jusqu'à une valeur de vingt. Ce mécanisme de régulation est permis car les promoteurs de *repR* et de l'antisens sont forts et faibles, respectivement, ce qui est également une particularité parmi les systèmes impliquant des ARN antisens synthétisés en *cis*.

Par rapport aux mécanismes contrôlant la synthèse de la protéine initiatrice de la réplication à des niveaux post-transcriptionnels, le mécanisme d'atténuation est particulièrement intéressant d'un point de vue énergétique car l'inhibition est effectuée de façon très précoce. On peut par conséquent s'interroger sur la raison pour laquelle tous les plasmides ne régulent pas leur réplication de cette manière. D'une manière intéressante, ce mode de contrôle est largement utilisé chez les bactéries Gram-positives. Par exemple, chez *Bacillus subtilis*, l'expression de la plupart des aminoacyl-ARNt synthétases est régulée par un mécanisme de type antisens impliquant l'ARNt correspondant [300,301]. La région 5' non codante des ARNm des aminoacyl-ARNt synthétases peut alterner entre deux conformations, l'une induisant la terminaison de la transcription et la seconde permettant la translecture. Lors d'un manque en un acide aminé donné, l'ARNt libre homologue interagit *via* son anticodon et

sa séquence NCCA 3'-terminale avec la région 5' non traduite de l'ARNm. Cette interaction stabilise la conformation permettant la translecture et favorise la synthèse de l'aminoacyl-ARNt synthétase.

3. *Mécanismes de contrôles post-transcriptionnels de la réplication plasmidique*

Chez certains plasmides, l'ARN antisens inhibe directement la traduction de l'ARNm codant pour la protéine initiatrice de la réplication en s'appariant à la région de fixation du ribosome [259,261]. Ces exemples sont cependant rares et peu étudiés. Dans la plupart des cas, la traduction de la protéine de réplication Rep est sous la dépendance de la traduction d'un peptide leader, impliquant un couplage traductionnel. Ce mode de traduction a été observé chez de nombreux plasmides parmi lesquels ceux des groupes IncFI, IncFII, IncIα, IncB, IncZ, IncK et IncL/M (pour une revue, [302]). Dans ces systèmes, un ARN antisens inhibe stériquement la fixation du ribosome au RBS du peptide leader et par conséquent, réprime la traduction du gène *rep* de façon indirecte. Deux classes de plasmides peuvent néanmoins être distinguées. Chez Col1b-P9 et ses dérivés, l'ARN antisens inhibe la traduction du peptide leader mais empêche également la formation d'une structure activatrice de la traduction de l'ARNm de la protéine de réplication. Chez les plasmide des groupes IncFI et IncFII, l'ARN antisens inhibe la traduction du peptide leader et le complexe ARN-ARN formé est cible de la RNase III.

a. Mécanisme de traduction de la protéine de réplication : l'exemple de Col1b-P9

La figure 23 représente la région 5'-UTR de l'ARNm codant pour la protéine RepZ initiatrice de la réplication chez le plasmide Col1b-P9 du groupe IncIα. L'organisation de cette région est similaire chez tous les plasmides des groupes mentionnés ci-dessus.

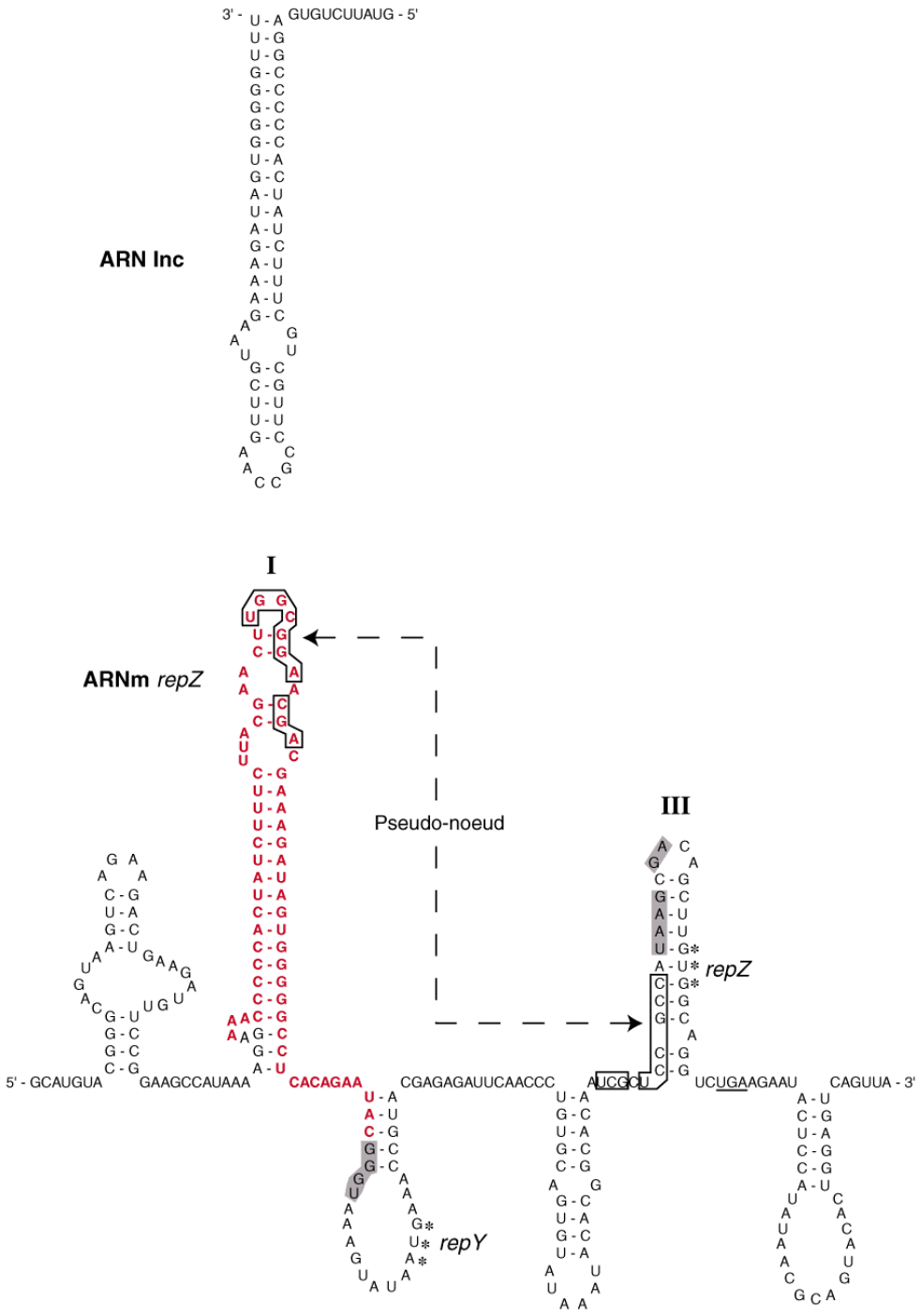

Figure 23 : ***Séquences et modèles de structures secondaires de la région 5'-UTR de l'ARNm* repZ *(d'après [303]) et de l'ARN antisens Inc du plasmide Col1b-P9***. La séquence de *repZ* complémentaire à Inc est dessinée en rouge. Les séquences Shine et Dalgarno de *repY* et de *repZ* sont indiquées par des boîtes grises et les codons AUG ou GUG initiateurs par des *. Le codon de terminaison de *repY* est souligné. Les séquences de la structure I et de la structure III impliquées dans la formation du pseudo-noeud sont encadrées et leur interaction est matérialisée par une ligne pointillée.

Les signaux d'initiation de la traduction de *repZ* sont séquestrés dans la structure en tige-boucle III (Figure 23). Pour que la traduction de *repZ* puisse avoir lieu, cette structure doit être déroulée. Ceci est permis grâce à la présence d'une phase ouverte de lecture (*repY*)

en amont de *repZ*. Les ribosomes traduisant *repY* déstabilisent la structure III et réinitient la traduction à la séquence Shine et Dalgarno de *repZ*. La position du codon de terminaison de *repY*, 10 nts en aval du codon d'initiation de *repZ*, est importante et optimale pour la ré-initiation, ceci dans tous les systèmes étudiés [304,305]. *repY* code pour un court peptide de 29 acides aminés dont la séquence peut être changée sans affecter l'expression de *repZ*. Ainsi, c'est la traduction de *repY* qui est importante pour la synthèse de RepZ, et non pas la protéine elle-même. Ceci n'est pas sans rappeler l'implication fréquente de petites ORF dans le contrôle traductionnel, tant chez les procaryotes que chez les eucaryotes (pour une revue, [306]).

Pour que la synthèse de RepZ aie lieu, la traduction de *repY* n'est pas suffisante. De manière intéressante, Asano et collègues ont montré que des mutations dans la boucle I diminuaient la traduction de *repZ in vivo* [307]. Le même effet a été observé en mutant une séquence dans la tige-boucle III et complémentaire à la séquence de la boucle I (Figure 23). Le rétablissement des appariements entre ces deux séquences par insertion de mutations compensatoires permet de restaurer la traduction de *repZ* et la réplication plasmidique [304,307]. Le même phénomène a été observé chez pMU720 du groupe IncB [308] et pMU604 du groupe IncL/M [48]. Ces approches génétiques ont permis de proposer la formation d'une structure en pseudo-noeud entre la séquence de la boucle I et celle de la tige-boucle III lors du couplage traductionnel (Figures 23 et 24). Sa formation a ultérieurement été confirmée par des expériences de cartographie en solution *in vitro* [47].

Lorsque le peptide leader est traduit, la structure III est déstabilisée, permettant la formation du pseudo-noeud [47]. Néanmoins, si la structure III est déstabilisée par mutagenèse, le pseudo-noeud est tout de même requis pour garantir la traduction de *repZ* [304]. Ce résultat montre que le rôle du pseudo-noeud n'est pas uniquement de maintenir la structure III ouverte mais qu'il a également un rôle activateur de la traduction de *repZ*. Cette activation est nécessaire du fait du caractère faible du RBS de *repZ*. Si celui-ci est rendu fort par la création d'une séquence SD consensus, d'un espacement optimal entre la séquence SD et le codon d'initiation, et par le remplacement du codon GUG initiateur par un codon AUG, alors le pseudo-noeud n'est plus nécessaire à la traduction de *repZ* [304]. On peut imaginer que l'interaction directe entre la sous-unité ribosomique 30S et le pseudo-noeud soit responsable de l'activation de la traduction du gène *rep* dans ces systèmes. Cette hypothèse reste cependant à prouver.

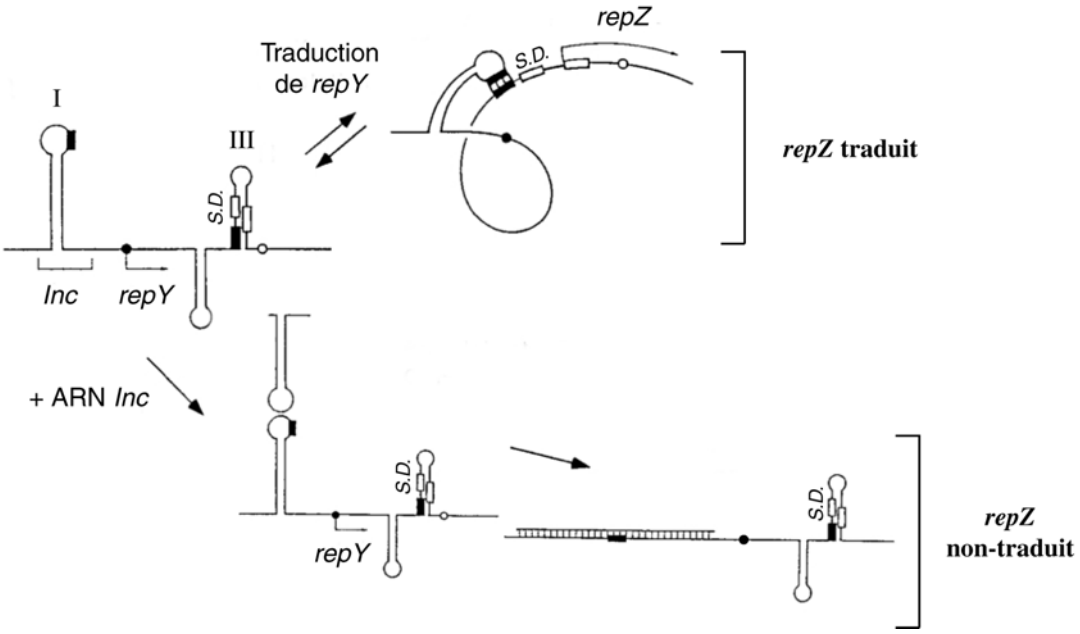

Figure 24 ***: Modèle du contrôle de l'expression de*** **repZ** ***par un pseudo-noeud (d'après [303])***. Les tiges-boucles I et III de l'ARNm *repZ* sont indiquées en chiffres romains. La région de complémentarité à l'ARN antisens Inc est délimitée et appelée 'Inc'. Les rectangles blancs représentent la séquence SD et le codon initiateur de *repZ*. Les rectangles noirs représentent les régions impliquées dans la formation du pseudo-noeud. Les ronds noirs et blancs représentent respectivement les codons d'initiation et de terminaison de la traduction de *repY*. Suivant que l'ARN Inc interagisse ou non avec l'ARNm, deux cas peuvent être distingués, ne permettant pas (bas) ou permettant (haut) la synthèse de RepZ, respectivement. Le complexe Inc/ *repZ* est schématisé de façon simplifiée par un duplexe étendu. Celui-ci n'est cependant pas nécessaire (Voir le texte pour le détail).

b. Contrôle de la traduction de *repZ* par un ARN antisens

Dans tous les groupes d'incompatibilité mentionnés plus haut (§VI.D.3), un ARN antisens inhibe la synthèse du peptide leader [254,255,309]. L'ARN antisens est synthétisé à partir du locus plasmidique codant pour la protéine initiatrice de la réplication, mais en orientation inverse. Chez Col1b-P9, l'ARN antisens Inc est complémentaire à la région 5'-UTR de *repZ* (Figure 23). Comme cela est schématisé, la région de complémentarité à Inc est adjacente à la séquence Shine et Dalgarno de *repY* dans l'ARN messager. L'interaction Inc-ARNm inhibe stériquement la traduction de *repY* et par conséquent celle de *repZ* [255].

L'interaction initiale est permise par une reconnaissance boucle-boucle (« kissing complex ») entre les tiges-boucles majeures de l'antisens et de la cible. Elle ne subsiste pas et progresse rapidement à travers les nucléotides des tiges [64,310]. Cette conversion est facilitée par la présence de nucléotides non appariés dans les deux ARN et conduit à la formation d'un complexe stable [61,311]. La formation d'un duplexe étendu n'est cependant pas nécessaire pour l'activité de l'ARN antisens [64]. Grâce à l'étude du complexe antisens-cible régulant la réplication du plasmide R1 du groupe IncFII, nous expliquerons dans cette

thèse en quoi sa structure est importante pour l'efficacité du contrôle (Chapitre 1 de la partie *Résultats & Discussion*).

La séquence proximale impliquée dans la formation du pseudo-noeud (5'-GGCG-3' dans la boucle I) contient les nucléotides par lesquels l'interaction avec l'ARN Inc est initiée [61]. L'interaction de la structure I avec l'antisens Inc et son implication dans la formation du pseudo-noeud sont donc mutuellement exclusives [61]. Ainsi, l'ARN antisens Inc agit à deux niveaux. Inc inhibe la traduction de *repZ* indirectement en inhibant celle de *repY* et directement en séquestrant les bases nécessaires à la formation de la structure activatrice de sa traduction. Ce double contrôle est retrouvé dans tous les plasmides des groupes IncIα, IncB, IncZ, IncK et IncL/M [47,48,308], mais pas dans les plasmides des groupes IncFI et IncFII (dont fait partie le plasmide R1). Il permet de réprimer efficacement la synthèse de la protéine Rep et par conséquent d'inhiber la réplication plasmidique, ceci à l'aide d'un ARN antisens unique.

VII. Le contrôle de la réplication du plasmide R1 : Etat des lieux

A. La biologie du plasmide R1

Le plasmide R1, initialement caractérisé chez *Salmonella typhimurium*, peut être maintenu de manière stable chez *E. coli* [312]. C'est un plasmide du groupe d'incompatibilité IncFII, d'environ 90 kpb (Figure 25A) et présent à un faible nombre de copie (entre 1 et 2) par cellule hôte (pour une revue, [313]). Il porte plusieurs gènes de résistance à des antibiotiques et pour cette raison présente un grand intérêt médical. Le plasmide R1 est auto-transmissible et porte tous les gènes nécessaires à sa réplication, à son transfert conjugatif et à sa maintenance stable lors de la division cellulaire. Ces trois fonctions étant contrôlées par des ARN antisens (décrits dans les § VI.B, VI.C et celui-ci) [254,263,279], le plasmide R1 est un système de choix pour l'étude des mécanismes de contrôle de l'expression génétique.

Tous les gènes nécessaires à la réplication du plasmide R1 et à son contrôle sont localisés sur un fragment d'environ 2.5 kpb appelé le réplicon (Figure 25A). La réplication débute à l'origine *ori*R1 et nécessite l'intervention en *cis* de la protéine RepA (Figure 25B) [314]. Cette étape est l'étape limitante de la réplication plasmidique. Par conséquent, le contrôle de la réplication plasmidique se fait en modulant le taux d'expression de *repA*. L'expression de *repA* est contrôlée au niveau transcriptionnel par le produit du gène *copB* et au niveau post-transcriptionnel par le produit du gène *copA* (Figure 25B).

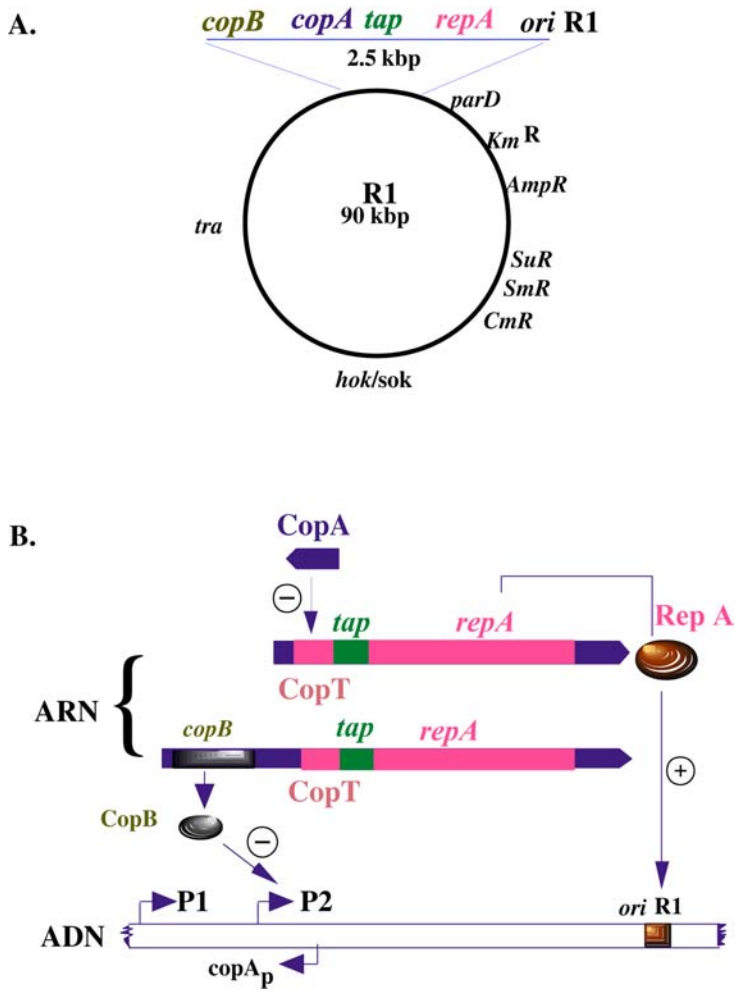

Figure 25 : ***Représentation simplifiée du plasmide R1 (A) et organisation génétique de la région de l'origine de réplication (B)***. (A) Les gènes de résistance à des antibiotiques portés par R1 sont indiqués. Les loci *hok/sok* et *tra*, impliqués respectivement dans le contrôle de la ségrégation et de la conjugaison sont également indiqués. L'organisation génétique du réplicon est représentée au-dessus du plasmide. Il contient les gènes de contrôle *copB* et *copA*, le peptide leader *tap* et le gène *repA* codant pour la protéine initiatrice de la réplication. «*ori* R1» indique la position de l'origine de réplication. (B) Les différents promoteurs sens (P1 et P2) et antisens ($copA_p$) sont indiqués sur l'ADN, de même que la position de l'origine de réplication *ori* R1. Les ARN transcrits à partir de cette région sont représentés. La répression de la transcription à partir de P2 par le produit du gène *copB* est indiquée par un - , de même que la répression de la traduction de *tap* par l'ARN antisens CopA lors de sa fixation à la séquence CopT. L'action activatrice de la réplication au niveau de l'origine *ori* R1 et conférée par la protéine RepA est indiquée par un +.

B. Contrôle transcriptionnel de l'expression de *repA*

Le contrôle transcriptionnel de l'expression de *repA* par la protéine CopB ne joue qu'un rôle mineur [315]. RepA peut être synthétisée à partir de deux transcrits différents (P1 et P2 : Figure 25B) et le plus long d'entre eux code pour le répresseur transcriptionnel CopB. A l'équilibre, CopB est présent en quantité saturante et intervient sous forme de tétramère pour réprimer la transcription à partir du promoteur P2 [316]. RepA est alors synthétisée à partir d'un transcrit polycistronique *copB-tap-repA* (Figure 25B). Cependant, lorsque la

concentration intracellulaire en plasmide chute, comme par exemple lors de la division cellulaire, la quantité de CopB présente n'est pas suffisante pour réprimer le promoteur P2. RepA est alors également synthétisée à partir d'un transcrit *tap-repA*. Par conséquent, la transcription de l'ARNm *repA* augmente ce qui conduit à une augmentation de la réplication plasmidique. La dérepression du promoteur P2 doit probablement jouer un rôle lors de l'entrée d'un plasmide dans une bactérie.

C. Contrôles post-transcriptionnels de l'expression de *repA*

1. *Inhibition de la traduction du peptide leader* tap

Comme dans le cas de Col1b-P9 décrit au § VI.D.3., la traduction de l'ARNm *repA* nécessite la traduction de la phase ouverte de lecture *tap* (pour translational activator peptide) en amont de *repA* qui code pour un court peptide (tap) de 24 acides aminés [42,317,318]. Dans ce cas également, seule la traduction de *tap* est importante car la séquence en acides aminés de la protéine n'est pas essentielle. La traduction de *tap* s'achève deux nucléotides en aval du codon initiateur de *repA* [42] (Figure 26). Le couplage traductionnel induit la déstabilisation de la structure en tige-boucle séquestrant les signaux d'initiation de la traduction de *repA*, permettant la réinitiation du ribosome [319] (Figure 26). Contrairement au système Col1b-P9, le codon GUG initiateur de la traduction de *repA* est plus proche du codon stop de *tap* (2 nts en amont contre 10) et la séquence Shine et Dalgarno est plus forte. Par conséquent, la traduction de *repA* ne nécessite pas la formation d'une structure activatrice en pseudo-nœud.

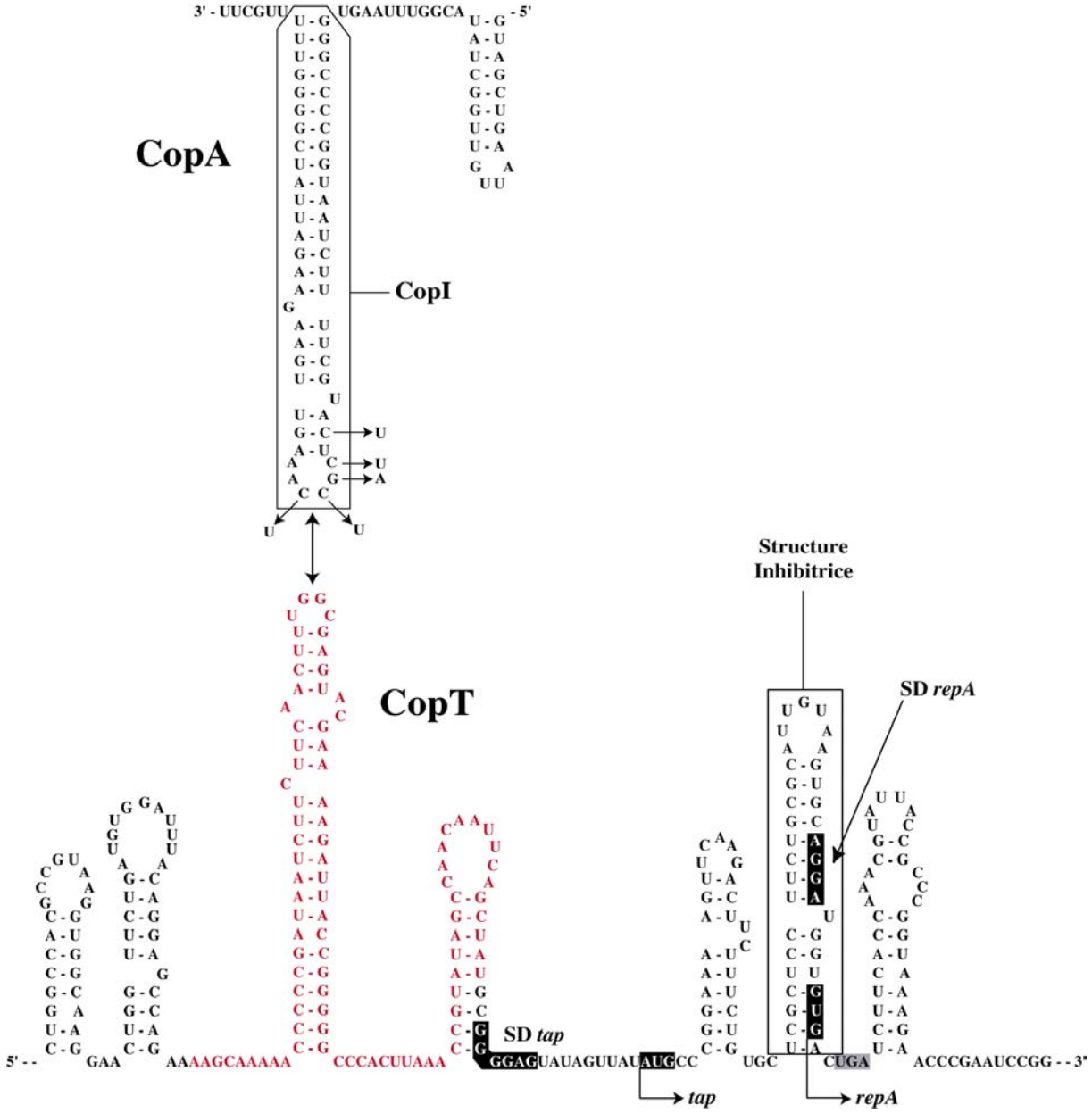

Figure 26 : ***Séquences et modèles de structures secondaires de la région 5'-UTR de l'ARNm* repA *et de l'ARN antisens CopA du plasmide R1***. La séquence CopT, complémentaire à CopA, est dessinée en rouge. Les séquences Shine et Dalgarno (SD) et les codons initiateurs de la traduction de *tap* et de *repA* sont indiquées par des caractères blancs sur fond noir. Le codon de terminaison de *tap* est indiqué par une boîte grisée. La structure en tige-boucle séquestrant le RBS de *repA* est indiquée. L'interaction boucle-boucle initiale entre CopA et CopT est matérialisée par une double flèche. Les mutations dans la boucle de CopA affectant le contrôle *in vivo* sont indiquées. L'ARN antisens tronqué CopI est encadré.

Le locus *repA* code pour l'ARN antisens CopA, long de 91 nts et complémentaire à la séquence CopT dans la région 5'-UTR de l'ARNm *repA* [254] (Figures 25B & 26). L'extrémité 3' de la région CopT est localisée deux nucléotides en amont de la séquence Shine et Dalgarno (SD) de *tap* (Figure 26). La fixation de CopA à CopT réprime l'expression de *repA* au niveau post-transcriptionnel en inhibant la traduction de *tap*. *In vitro*, cette interaction empêche de façon stérique la fixation des sous-unités ribosomiques 30S à la

séquence SD [320]. L'efficacité de ce contrôle repose sur une vitesse d'association optimale ($\approx 10^6$ $M^{-1}.s^{-1}$) des deux ARN *in vitro* [321]. CopA et CopT contiennent tout deux une structure majeure en tige-boucle (Figure 26) [322]. Des études cinétiques ont permis de proposer un mécanisme d'interaction à deux étapes. L'association est initiée par la séquence 5'-CGCC-3' de la boucle de CopA et la séquence complémentaire de CopT. Cette interaction boucle-boucle initiale (« kissing complex ») est néanmoins réversible et doit être stabilisée par des interactions additionnelles [321,323]. L'interaction boucle-boucle initiale est essentielle à l'efficacité du contrôle. En effet, des mutations affectant soit la séquence de la boucle de CopA (Figure 26, [324]), soit sa taille [325] affectent le nombre de copies plasmidiques. Ainsi, les boucles des deux ARN doivent avoir une séquence précise et une taille comprise entre 5 et 7 nts. Par ailleurs, la présence de nucléotides non appariés dans la tige des deux ARN est requise à la fois pour garantir une interaction efficace *in vitro* et assurer un contrôle optimal *in vivo* [62]. Ces caractéristiques sont redondantes dans la plupart des ARN antisens naturels procaryotiques décrits jusqu'à présent (pour une revue, [65]).

Le cumul des résultats biochimiques et génétiques sur le système CopA/CopT nous avaient permis de proposer un mécanisme préliminaire pour l'interaction entre les deux ARN [326]. Celui-ci est représenté dans la figure 27. Le « kissing complex » initial (Figure 27B) ne subsiste pas dans la produit final de l'interaction *in vitro* [326]. Celui-ci progresse vers une interaction plus étendue impliquant les nucléotides de la partie supérieure des tiges de CopA et de CopT et est stabilisé par l'interaction des régions 5' de CopA et 3' de CopT formant une longue hélice intermoléculaire d'environ 30 paires de bases (Figure 27C). Malgré la parfaite complémentarité entre CopA et CopT, la formation d'un duplexe étendu (Figure 27D) est très lente *in vitro* [320] et n'est pas essentielle à la fonction de l'antisens *in vivo* [327]. Cependant, la structure du complexe stable et fonctionnel (Figure 27C) n'était pas connue et le mécanisme de formation restait très imprécis.

Il est intéressant de noter qu'un ARN antisens artificiel tronqué (CopI, Figure 26) contrôle efficacement la traduction de *repA in vivo* et inhibe transitoirement la fixation des sous-unités ribosomiques 30S à la séquence SD de *tap in vitro* [320,327]. CopI, qui ne contient que la tige-boucle II de CopA, n'est pas capable de former un duplexe étendu avec CopT *in vitro*. Des expériences préliminaires de cartographie en solution suggèrent une topologie originale du complexe CopI-CopT qui doit subsister dans le complexe stable CopA-CopT [326]. Par ailleurs, les changements de structure de l'ARNm induits par la fixation de CopI sont restreints à la région de complémentarité. Ceci suggère fortement qu'une structure particulière du complexe CopI-CopT est responsable des propriétés inhibitrices de l'ARN antisens tronqué.

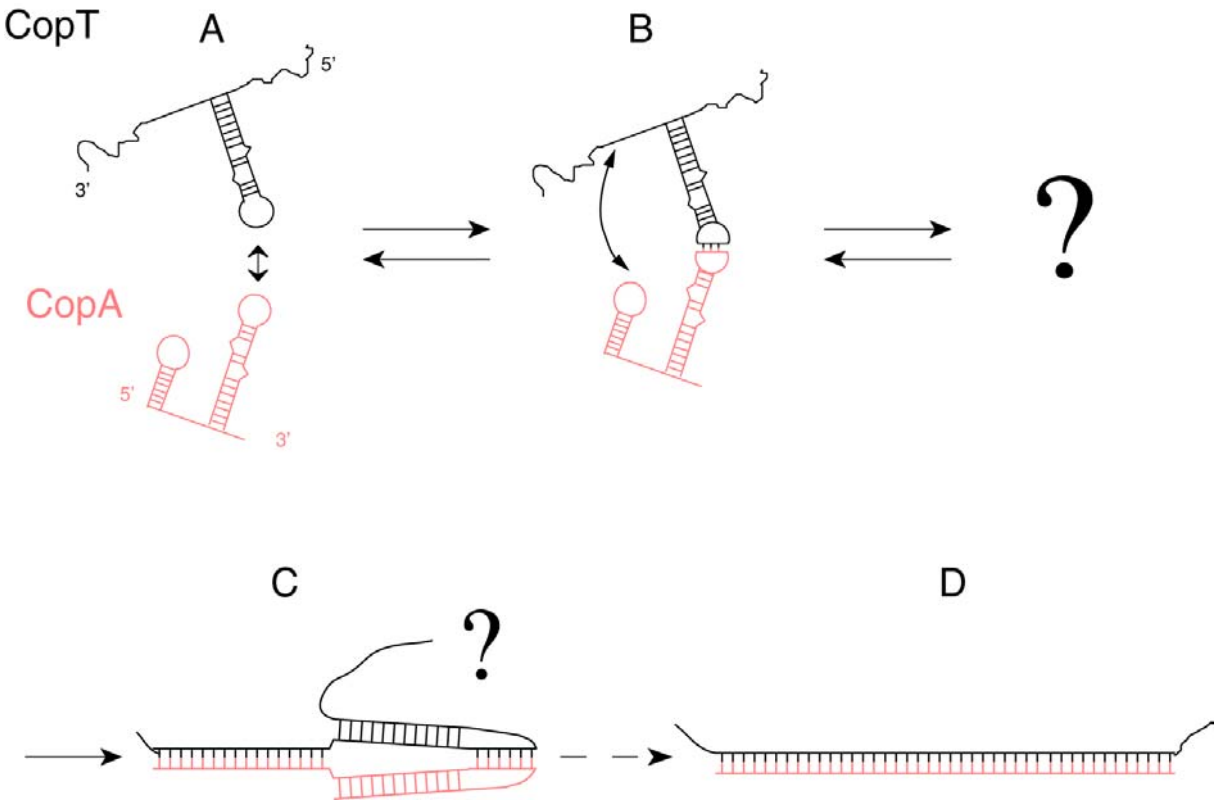

Figure 27 : ***Mécanisme d'interaction simplifié entre CopA et CopT (d'après [326])***. (A) CopA et CopT interagissent *via* un faible nombre de nucléotides présents dans les boucles de leur tige-boucle majeure. Le complexe (B) est réversible et doit être stabilisé par un mécanisme inconnu (« ? ») pour former une hélice d'environ 30 paires de bases entre la région 5' de CopA et la région 3' de CopT (C). La topologie du complexe C reste purement hypothétique, comme l'indique le symbole « ? ». La formation d'un duplexe étendu entre CopA et CopT (D) est très lente *in vitro* et est schématisée par une flèche pointillée.

2. *Dégradation de l'ARN messager* repA

En plus d'inhiber la traduction de *tap* et donc indirectement celle de *repA*, l'ARN antisens CopA favorise également la dégradation de l'ARNm *repA* d'une manière RNase III dépendante, à la fois *in vitro* et *in vivo* [51]. Ce clivage n'est cependant pas essentiel au contrôle et ne réprime l'expression de *repA* que de dix fois. Un dimère de RNase III reconnaît deux tours d'hélice de forme A et induit une coupure dans les deux brins d'ARN de façon symétrique [328]. La RNase III doit probablement agir avant la formation d'un duplexe étendu entre CopA et CopT, qui est un processus très lent *in vitro*. Des sites uniques de coupures ont été localisés *in vivo* dans CopA et dans CopT (17 nts en aval de l'extrémité 5' de CopA et de l'extrémité 3' de CopT), indiquant que l'hélice intermoléculaire entre ces deux régions est suffisante pour l'action de la RNase III. Il est cependant difficile de conclure si le duplexe étendu ou un complexe stable (comme celui schématisé dans la figure 27C) est la forme prédominante *in vivo*. En effet, les coupures observées *in vivo* peuvent refléter un intermédiaire stable issu de la dégradation d'un duplexe étendu par la RNase III et résistant

aux exoribonucléases 3’ --> 5’ cellulaires. Dans le système col1b-P9, la taille réduite de la région simple-brin en 5’ de l’ARN antisens Inc empêche le complexe Inc-*repZ* d’être reconnu par la RNase III au niveau de l’hélice intermoléculaire.

VIII. Motivation du travail de thèse

Cette introduction illustre l’importance des ARN dans le contrôle d’un grand nombre de fonctions biologiques, tant chez les procaryotes que chez les eucaryotes. Les mécanismes de régulation font preuve d’une grande diversité. Cependant, très peu d’études détaillées reliant la structure à la fonction des riborégulateurs sont disponibles à ce jour. L’objectif de cette thèse était d’établir cette relation pour deux ARN régulateurs bactériens et d’en extraire des règles générales. Nous avons focalisé notre attention sur un ARN antisens et sur un ARN multirégulateur. L’ARN antisens étudié est l’ARN CopA, impliqué dans le contrôle de la réplication du plasmide R1, et l’ARN multirégulateur est l’ARN-III, impliqué dans le contrôle de l’expression des gènes de virulence chez *S. aureus*.

PRESENTATION DU TRAVAIL DE THESE

I. Chapitre 1 : Relation structure-fonction du complexe ARN antisens – ARNm inhibiteur responsable du contrôle de la réplication du plasmide R1

Même si CopA était l'un des ARN antisens bactériens les plus étudiés avant que je ne débute ma thèse, son mode d'action n'était pas encore clairement compris. Un modèle de reconnaissance entre CopA et CopT avait été proposé au laboratoire, mais celui-ci restait encore incomplet. En particulier, la topologie du complexe fonctionnel et inhibiteur ainsi que les déterminants structuraux responsables de sa formation étaient inconnus. Dans un premier temps, nous avons focalisé notre attention sur la structure du complexe irréversible formé *in vitro* et capable d'inhiber la fixation du ribosome à *tap*. Cette première étude nous a permis de proposer un modèle de structure tridimensionnelle pour le complexe CopA-CopT (Article I). Nous montrons que malgré la parfaite complémentarité des deux ARN, celui-ci n'est pas un duplexe étendu mais adopte une topologie très particulière ; les interactions boucle-boucle initiales sont rompues au profit de nouveaux appariements intermoléculaires impliquant les nucléotides des tiges des deux ARN. Nous montrons dans l'article II que cette conversion nécessite des déterminants structuraux présents dans les hélices intramoléculaires de CopA et de CopT. La conversion est essentielle pour permettre la formation de l'hélice stabilisatrice qui rend le système irréversible. Nous avons alors voulu relier ces résultats structuraux aux propriétés fonctionnelles de CopA. Dans notre troisième étude, nous avons combiné des approches *in vitro* (cartographie en solution, mesure de la vitesse d'association des ARN), la mutagenèse dirigée et une approche *in vivo* (mesure de la fonction inhibitrice d'ARN antisens mutants) pour proposer un mécanisme détaillé de la formation de la structure inhibitrice. Nous avons pu révéler un nouvel intermédiaire réactionnel et nous avons caractérisé les étapes du processus essentielles au contrôle. Enfin, dans de nombreux systèmes bactériens, malgré la parfaite complémentarité de l'ARN antisens et de sa cible, la formation d'un duplexe étendu n'est pas nécessaire à l'efficacité du contrôle. Cette propriété est maintenant devenue une règle plutôt qu'une exception. Plusieurs plasmides régulent leur taux de réplication de manière similaire au plasmide R1 (voir § VI. D de l'introduction). Chez tous ces plasmides, les ARN antisens et leurs cibles partagent de nombreuses caractéristiques. Nous montrons que leurs mécanismes de reconnaissance et la topologie des complexes formés sont semblables à

ce que nous avions établi dans le système CopA-CopT (Article IV). Enfin, dans le but de déterminer à l'échelle atomique la structure du complexe formé entre CopA et CopT, nous avons initié un projet de cristallisation dont les résultats préliminaires obtenus seront présentés.

II. Chapitre 2 : Rôle de l'ARN-III dans la régulation de l'expression des gènes de virulence chez *Staphylococcus aureus*

Le locus *agr* est impliqué dans le contrôle de l'expression des gènes de virulence chez *S. aureus*. Il réprime la synthèse des protéines de surface et stimule la synthèse des exotoxines en fin de phase exponentielle de croissance. Il code pour tous les gènes qui composent un mécanisme de mesure de densité cellulaire dont la molécule effectrice est un ARN régulateur de 531 nts. Ce système de contrôle de l'expression des toxines est conservé chez de nombreuses souches de Staphylocoques. Ceux-ci représentent un enjeu majeur en terme de santé publique. Ils sont responsables de 30% des infections nosocomiales et d'une proportion équivalente d'infections communautaires. Par ailleurs, de nombreuses souches résistantes aux antibiotiques apparaissent. Afin de concevoir de nouveaux traitements efficaces et de définir de nouvelles cibles thérapeutiques potentielles, il est important de comprendre les bases moléculaires de la virulence des Staphylocoques.

L'ARN-III régule l'expression de plus d'une vingtaine de gènes chez *S. aureus*. Il joue donc un rôle pléiotropique dans le contrôle de l'expression génétique. Comme cela a été décrit dans le § IV de l'introduction pour d'autres ARN multirégulateurs, les différentes propriétés régulatrices de cet ARN reposent probablement sur des domaines structuraux distincts. Lorsque j'ai débuté mon travail de thèse, aucune étude structurale de l'ARN-III n'avait été réalisée. Ainsi, la structure de cet ARN a été étudiée au moyen de diverses sondes de structure *in vitro* et *in vivo*. Une fois la structure établie, nous avons testé l'implication de domaines structuraux indépendants dans le contrôle de l'expression de gènes cibles (Article V). Enfin, l'ARN-III remplit probablement certaines de ses fonctions de régulateur par des interactions spécifiques avec d'autres facteurs. Des résultats préliminaires de caractérisation de protéines fixant l'ARN-III seront présentés. Les différents résultats obtenus pour cette étude sont présentés dans le deuxième chapitre de la partie *Résultats et Discussion*.

RESULTATS & DISCUSSION

1er Chapitre

Structure et Fonction du Complexe Inhibiteur

CopA-CopT responsable du contrôle de la réplication du plasmide R1

I. Article I : Etude structurale du complexe inhibiteur CopA-CopT

A. Objectifs

Plusieurs observations suggéraient que le complexe final formé entre CopA et CopT *in vitro* adopte une structure inattendue. D'une part les nucléotides des boucles, impliqués dans la reconnaissance initiale, sont en conformation simple-brin dans le produit stable et d'autre part la formation d'un duplexe étendu entre les deux ARN est un processus très lent [326]. Par ailleurs, l'ARN antisens tronqué CopI, incapable de former une hélice intermoléculaire avec CopT en amont de la séquence Shine et Dalgarno de *tap*, conserve néanmoins les propriétés inhibitrices de CopA (Figure 26, [320]). Dans la présente étude, nous avons voulu déterminer la topologie globale des complexes formés *in vitro* entre CopA et CopT mais également entre CopI et CopT et la relier à leurs propriétés fonctionnelles.

B. Stratégie expérimentale

Afin d'étudier la structure du complexe CopA-CopT irréversible formé en solution, nous avons eu recours à une combinaison d'approches. Des expériences de cartographie en solution à l'aide de différentes sondes de structure chimiques et enzymatiques ont permis de définir les régions accessibles des deux ARN dans le complexe stable. Ces résultats ont été comparés à ceux obtenus en parallèle sur le complexe CopI-CopT et sur les duplexes étendus correspondant formés artificiellement. Cette approche n'est cependant pas suffisante pour définir les nucléotides impliqués dans des contacts intermoléculaires. Pour délimiter les appariements intermoléculaires, des inversions de paires de bases ont été effectuées à diverses positions de l'hélice intramoléculaire de CopA. L'influence de ces mutations sur la structure des complexes hétérologues formés avec un ARN CopT sauvage a ensuite été définie. Le faisceau d'évidences expérimentales que nous avons obtenu nous a permis de proposer une topologie particulière du complexe qui a ensuite été testée par modélisation graphique. Cet outil intègre les contraintes stéréochimiques et géométriques propres aux ARN [329] et permet un conception plus précise des reconnaissances intermoléculaires dans l'espace. Il nous a permis de sélectionner un modèle tridimensionnel du complexe qui tient au mieux compte de toutes les données disponibles.

L'utilisation de sondes de structure étant fréquente dans les études qui vont suivre, les spécificités de celles qui ont été utilisées dans ce travail sont indiquées dans la figure 28 (pour une revue, [330-332]). Le diméthylsulfate (DMS) méthyle spécifiquement les positions N3 des cytosines et N1 des adénines si celles ci ne sont pas appariées. Il méthyle également la

position N7 des guanines. Le diéthylpyrocarbonate est spécifique de la position N7 des adénines et le complexe NiCR modifie spécifiquement la position N7 de guanines. Ces deux sondes sont très sensibles aux phénomènes d'empilement dans une hélice d'ARN. La position N7 des purines n'est donc accessibles que si le grand sillon est élargi. Nous avons également utilisé les ions Pb^{2+} qui se fixent par liaison de coordination au niveau de sites de fixation de cations divalents et induisent un clivage au niveau des riboses à des endroits précis de la chaîne d'ARN. Ils reconnaissent aussi aspécifiquement les régions simple-brin et sont très sensibles à de subtiles variations de conformation. Enfin, les radicaux hydroxyles qui induisent des coupures au niveau des riboses permettent de définir les régions non accessibles au solvant [333]. En tant que sondes enzymatiques, nous avons utilisé la ribonucléase (RNase) T1 (spécifique des guanines non appariées), la RNase T2 (spécifique des régions non appariées avec une préférence pour des adénines) et la RNase V1 (spécifique des nucléotides en conformation de type A, qu'ils soient appariés ou non).

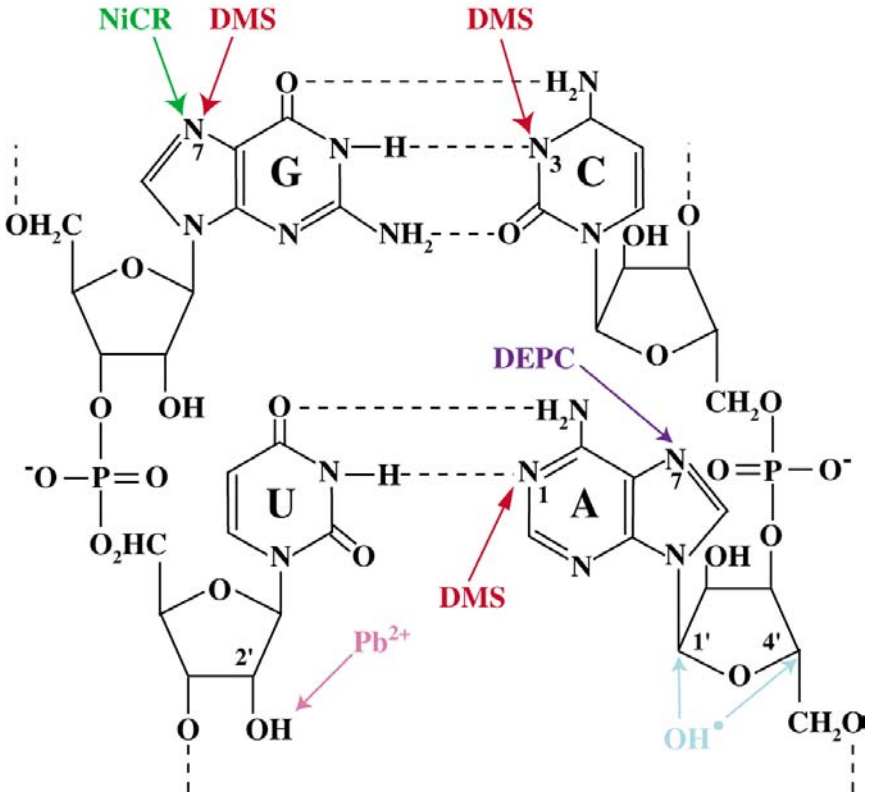

Figure 28 : ***Représentation schématique des paires de bases canoniques dans une hélice d'ARN et réactivités vis-à-vis de différentes sondes chimiques de structure.***

C. Résultats

Les signaux obtenus dans les expériences de cartographie en solution effectuées sur les complexes formés en conditions natives ou sur les duplexes étendus formés artificiellement

sont différents. Cela démontre que les duplexes étendus ne sont pas les produits finaux de l'interaction *in vitro*. La cartographie en solution nous a permis d'établir que les boucles de CopA (et CopI) et de CopT sont accessibles aux sondes spécifiques de régions simple-brin dans les complexes natifs formés *in vitro*. Bien que ces nucléotides soient impliqués dans la reconnaissance initiale (§ VII.C.1), le kissing-complex est détruit au profit de la formation d'interactions plus étendues. De plus, l'apparition de nouvelles coupures par la RNase V1 dans CopA et dans CopT suggère la formation de deux hélices intermoléculaires qui impliquent les nucléotides des tiges des deux ARN (Figure 29 : hélices B et B'). Cette structure est également formée dans le complexe CopI-CopT mais dans le complexe CopA-CopT, elle est ultérieurement stabilisée par la formation d'une troisième hélice intermoléculaire (hélice C) entre les trente premiers résidus de CopA et les nucléotides complémentaires de CopT. Sa formation est déduite de la protection de ces positions vis-à-vis des sondes spécifiques des régions simple-brin.

La mutagenèse dirigée a été utilisée pour définir la longueur des deux hélices intermoléculaires B et B' formées dans la partie supérieure du complexe. Nous avons sélectivement inversé plusieurs paires de bases le long des tiges majeures de CopA et de CopT (Figure 29 : mutants H1, H2 et H3). Si les positions mutées sont impliquées dans des interactions intramoléculaires dans le complexe CopA-CopT alors les mutations n'auront pas d'incidence sur la formation des complexes hétérologues avec un ARN CopT sauvage. En revanche, si ces positions sont impliquées dans des contacts intermoléculaires, alors les mutations auront un effet. Enfin, la formation d'un complexe entre deux ARN contenant des inversions compensatoires ne devrait pas être affectée. Ainsi, l'étude de la structure des complexes hétérologues ou homologues formés entre différents ARN nous a permis de définir la frontière entre les positions impliquées dans des contacts intermoléculaires et les interactions intramoléculaires. Ainsi, les nucléotides positionnés au-dessus des résidus non appariés dans les deux ARN (positions H1 et H2) sont impliqués dans l'interaction tandis que ceux situés en-dessous (positions H3) ne le sont pas (Figure 29). Ces résultats sont confirmés par l'observation que l'inversion de toute la partie inférieure de la tige de CopA n'affecte pas la structure du complexe (Voir § V de ce chapitre).

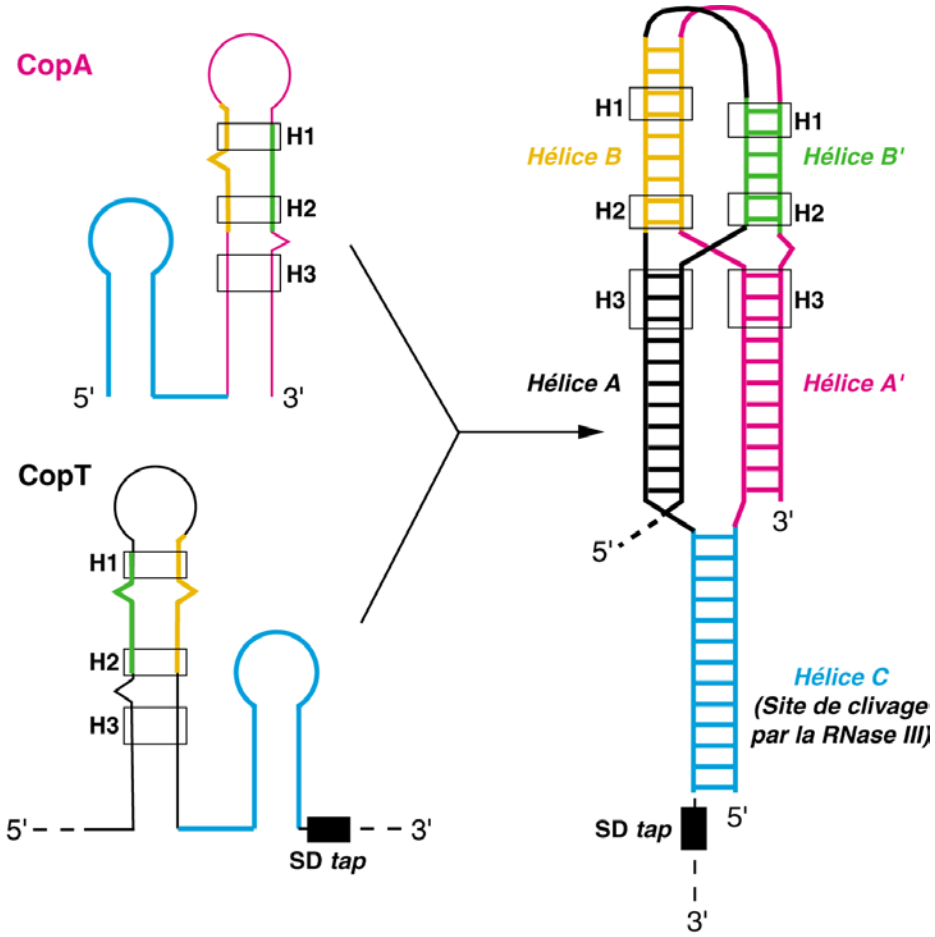

Figure 29 : ***Modèle bidimensionnel du complexe irréversible et inhibiteur formé entre CopA et CopT* in vitro**. Les régions non impliquées dans des appariements intermoléculaires sont indiquées en rose pour CopA et en noir pour CopT. Celles impliquées dans la formation des hélices intermoléculaires B, B' et C sont indiquées respectivement en jaune, vert et bleu. Les mutations (inversion de paires de bases) introduites dans les deux ARN au cours de cette étude sont indiquées (H1, H2 et H3). La séquence Shine et Dalgarno de *tap* est schématisée par une boîte noire.

D. Conclusions

L'ensemble de nos résultats converge vers un modèle de structure contenant une jonction à quatre hélices qui implique la formation de deux longs segments hélicoïdaux juxtaposés, chacun généré par l'empilement coaxial d'une hélice intermoléculaire et d'une hélice intramoléculaire. La topologie de ce complexe est tout à fait particulière et deux types d'empilements des hélices sont possibles (B-A/B'-A' ou B-A'/B'-A). Basée sur la stéréochimie des ARN, la modélisation graphique a permis d'exclure un de ces empilements. Elle a également contribué à définir le nombre d'appariements formés dans les hélices intermoléculaires B et B'. En raison de contraintes topologiques révélées par modélisation graphique, les appariements ne peuvent plus progresser vers un duplexe étendu à partir de la jonction à quatre hélices. Ceci est en accord avec les expériences de mutagenèse. Par ailleurs, la stabilité des hélices intramoléculaires A et A' conférée par leur richesse en paires de bases

de type G≡C requiert trop d'énergie pour fondre leur structure et permettre la progression de l'hélice C. La topologie décrite ci-dessus existe dans le complexe CopI-CopT. La structure générée est volumineuse dans l'espace et peut expliquer les propriétés inhibitrices de CopI. Le complexe CopA-CopT stable caractérisé dans cette étude est capable de bloquer de façon irréversible la fixation du ribosome à *tap in vitro*. Par ailleurs, nous avons récemment montré que le complexe CopA-CopT stable est reconnu et clivé efficacement par la RNase III *in vitro* (Huntzinger E., Kolb F.A. et Romby P., résultats non publiés). La protéine reconnaît exclusivement l'hélice stabilisatrice C alors que la jonction à quatre hélices est résistante à l'action de l'enzyme. Il est intéressant de signaler que les coupures dans l'hélice C sont les seules qui ont été identifiées *in vivo* [327].

Ainsi, le complexe CopA-CopT stable est nécessaire et suffisant pour inhiber de façon irréversible la synthèse de RepA et représente donc la forme biologique. La topologie de la jonction inhibe la fixation du ribosome probablement par un encombrement stérique et l'hélice C a un double rôle : elle stabilise irréversiblement le complexe et génère un site de clivage à la RNase III.

RNA (2000), *6*:311–324. Cambridge University Press. Printed in the USA.
Copyright © 2000 RNA Society.

An unusual structure formed by antisense-target RNA binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment

FABRICE A. KOLB,[1] CHARLOTTA MALMGREN,[3] ERIC WESTHOF,[1] CHANTAL EHRESMANN,[1] BERNARD EHRESMANN,[1] E. GERHART H. WAGNER,[2] and PASCALE ROMBY[1]

[1]UPR 9002 du Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 15 rue R. Descartes, Strasbourg cedex, France
[2]Department of Microbiology, SLU (Swedish University of Agricultural Sciences), Genetikvägen 5, S-75007 Uppsala, Sweden
[3]Department of Microbiology, Biomedical Center, Uppsala University, S-751 23 Uppsala, Sweden

ABSTRACT

The antisense RNA CopA binds to the leader region of the *repA* mRNA (target: CopT). Previous studies on CopA–CopT pairing in vitro showed that the dominant product of antisense RNA–mRNA binding is not a full RNA duplex. We have studied here the structure of CopA–CopT complex, combining chemical and enzymatic probing and computer graphic modeling. CopI, a truncated derivative of CopA unable to bind CopT stably, was also analyzed. We show here that after initial loop–loop interaction (kissing), helix propagation resulted in an extended kissing complex that involves the formation of two intermolecular helices. By introducing mutations (base-pair inversions) into the upper stem regions of CopA and CopT, the boundaries of the two newly formed intermolecular helices were delimited. The resulting extended kissing complex represents a new type of four-way junction structure that adopts an asymmetrical X-shaped conformation formed by two helical domains, each one generated by coaxial stacking of two helices. This structure motif induces a side-by-side alignment of two long intramolecular helices that, in turn, facilitates the formation of an additional intermolecular helix that greatly stabilizes the inhibitory CopA–CopT RNA complex. This stabilizer helix cannot form in CopI–CopT complexes due to absence of the sequences involved. The functional significance of the three-dimensional models of the extended kissing complex (CopI–CopT) and the stable complex (CopA–CopT) are discussed.

Keywords: antisense RNA; chemical mapping; RNA modeling; RNA–RNA interaction; RNA structure

INTRODUCTION

Many untranslated RNAs exert regulatory functions in both prokaryotes and eukaryotes. A subclass of these regulators, called antisense RNAs, affects target RNA function via binding to complementary sequences. Most antisense RNAs have been identified in prokaryotic cells, mainly in their plasmids, transposons, and bacteriophages (reviewed by Wagner & Simons, 1994; Wagner & Brantl, 1998; Zeiler & Simons, 1998). Plasmid R1 belongs to the IncFII group of plasmids whose initiation frequency is controlled by an antisense RNA, CopA. Synthesis of the replication initiator protein RepA requires translation of a short leader peptide (*tap*), located upstream of *repA*. CopA binds to the leader region of the *repA* mRNA (CopT), located about 80 nt upstream of the *repA* start codon (Fig. 1). Binding prevents *tap* translation and thereby *repA* expression (Blomberg et al., 1992, 1994; Malmgren et al., 1996). The CopA–CopT binding process is viewed as a series of reactions leading to progressively more stable complexes (Persson et al., 1988, 1990a, 1990b; Malmgren et al., 1997). CopA and CopT are fully complementary and both RNAs contain a major stem-loop structure (II/II′ in Fig. 1) that is essential for high pairing rates

Reprint requests to: Pascale Romby, UPR 9002 du Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 15 rue R. Descartes, Strasbourg cedex, France; e-mail: P.Romby@ibmc.U-strasbg.fr; or E. Gerhart H. Wagner, Department of Microbiology, SLU (Swedish University of Agricultural Sciences), Box 7025, Genetikvägen 5, S-75007 Uppsala, Sweden; e-mail: gerhart.wagner@mikrob.slu.se.

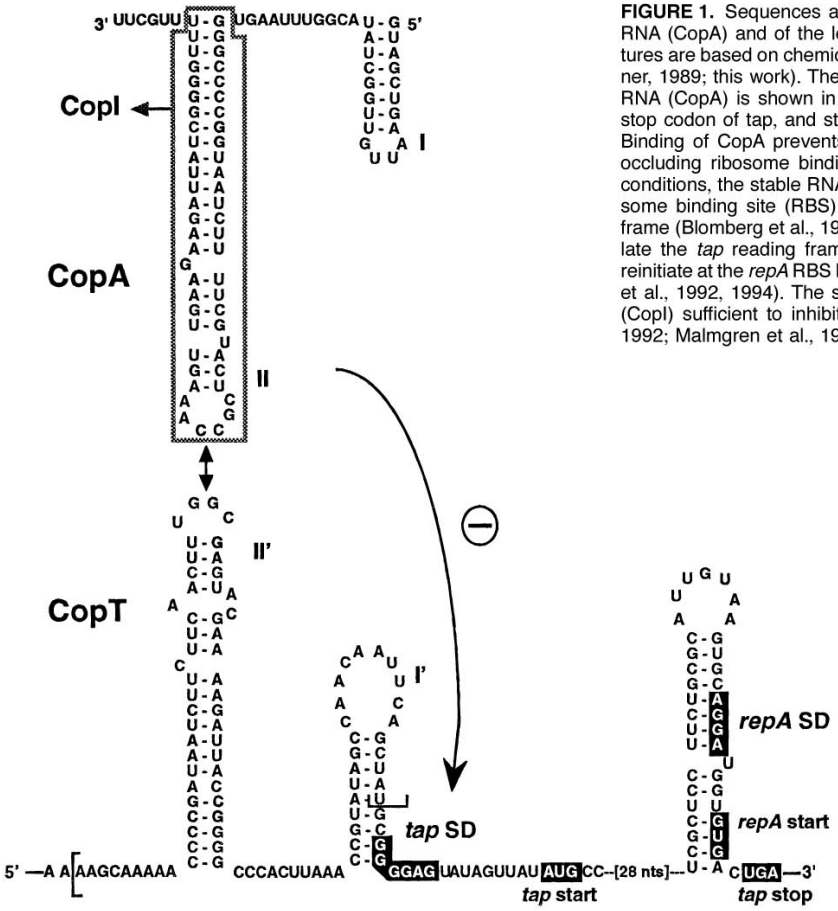

FIGURE 1. Sequences and secondary structures of the antisense RNA (CopA) and of the leader segment of the *repA* mRNA. Structures are based on chemical and enzymatic probing (Öhman & Wagner, 1989; this work). The target sequence (CopT) of the antisense RNA (CopA) is shown in brackets. The Shine and Dalgarno (SD), stop codon of tap, and start codons of *tap* and *repA* are indicated. Binding of CopA prevents translation of the *tap* reading frame by occluding ribosome binding at *tap* initiation site (−). Under these conditions, the stable RNA stem-loop that sequesters the *repA* ribosome binding site (RBS) prevents translation of the *repA* reading frame (Blomberg et al., 1994). If CopA fails to bind, ribosomes translate the *tap* reading frame, terminate at the *tap* stop codon, and reinitiate at the *repA* RBS by a direct translational coupling (Blomberg et al., 1992, 1994). The sequence of the truncated antisense RNA (CopI) sufficient to inhibit *tap* translation is boxed (Wagner et al., 1992; Malmgren et al., 1996).

and control (Öhman & Wagner, 1989; Hjalt & Wagner, 1992, 1995). The initial step involves a transient loop–loop interaction (kissing complex) between the complementary hairpin loops (Persson et al., 1990a, 1990b). Indeed, a truncated CopA (CopI, Fig. 1), lacking the 5′ proximal 30 nt and consisting only of the major stem-loop, does not form stable duplexes with CopT, but is capable of competing with CopA for binding (Persson et al., 1990b). It was recently shown that in both CopI–CopT and CopA–CopT complexes, initial kissing is rapidly followed by more extended intermolecular interactions (Malmgren et al., 1997). Subsequently, the single-stranded region in the 5′ tail of CopA pairs with its complement in CopT to yield the stable, inhibitory CopA–CopT complex. This complex is the dominant product of binding in vitro (Malmgren et al., 1996, 1997). Complete duplex formation is very slow and has been proposed to be irrelevant for control (Malmgren et al., 1996, 1997; Wagner & Brantl, 1998).

Different pairing pathways that result in rapid formation of stable antisense-target RNA complexes have been described (Kittle et al., 1989; Persson et al., 1990b; Tomizawa, 1990; Siemering et al., 1994; Thisted et al., 1994). A common feature is the use of a restricted single-stranded region in each interacting RNA for the initial step. In most cases, binding initiates between two loops, in some cases between a loop and a single-stranded RNA segment. Subsequently, more stable complexes are either formed by the pairing of distal RNA segments or, in the latter case, by extension of the first helix.

In the plasmids of the ColE1-family, control of replication is mediated by an antisense RNA, RNAI, that interacts with the preprimer, RNAII, via initial and transient base pairing between complementary loops (Eguchi et al., 1991; Eguchi & Tomizawa, 1991). NMR studies were performed on two RNA hairpins carrying seven-membered complementary loops derived from RNAI/RNAII (Marino et al., 1995; Lee & Crothers, 1998). These studies indicated that all seven loop bases were paired in the loop–loop helix, and continuous stacking of the loop nucleotides on the 3′ side of their respective stems was observed. Loop–loop interactions in plasmid R1 (IncFII plasmid; Persson et al., 1990a, 1990b; Malmgren et al., 1997), pMU720 (IncB plasmid; Siemering et al., 1994), and ColIb-P9 (IncIα plasmid; Asano et al., 1998)

are clearly similar. In all these systems, it was suggested that an initial loop–loop interaction is rapidly converted to an extended kissing complex. This requires partial melting of the upper stem regions, most probably facilitated by the presence of bulged residues (Siemering et al., 1994; Hjalt & Wagner, 1995). Interestingly, the extended kissing complex suffices for inhibition in vivo (Wagner et al., 1992; Wilson et al., 1993). In the case of plasmid R1, the extended kissing complex is also capable of blocking ribosome binding at the *tap* translation initiation site in vitro (Malmgren et al., 1996), suggesting the existence of a bulky structure.

In the present work, the conformation of the stable complex formed by CopA and its target, CopT, was studied using chemical and enzymatic probing (Ehresmann et al., 1987). For comparison, the extended kissing complex formed by the truncated antisense RNA CopI and CopT was tested in parallel. Based on extensive probing results, a model of the CopA–CopT complex was built using computer-graphic modeling. The proposed structure presents as its most unique feature an asymmetric cruciform-like structure whose formation requires extensive breakage of intramolecular base pairing in the upper stems and the formation of intermolecular base-pairing interactions. Base-pair inversions in CopA and compensatory mutations in CopT were introduced in different positions of the upper stems to define the boundaries of the newly formed intermolecular helices. The significance of these structural features is further discussed.

RESULTS

Enzymatic and chemical probing of the antisense and target RNAs

We showed previously that full CopA–CopT duplexes are only slowly formed in vitro, and identified a stable complex using double-strand-specific enzyme RNase III- and Pb^{2+}-catalyzed cleavages (Malmgren et al., 1997). To obtain detailed structural information about the CopA–CopT complex, we used a range of enzymes and chemical reagents. To study the extended kissing complex, CopA was replaced by the truncated CopA variant CopI, which contains only the major stem-loop structure (Persson et al., 1990a). Antisense or target RNAs were end labeled and incubated with complementary, unlabeled counterparts for complex formation under native conditions (see Materials and Methods). RNAs, free or complexed, were subjected to enzyme hydrolysis or chemical modifications under conditions where statistically less than one cleavage/modification per molecule takes place. Enzymes used were RNase T1 (specific for single-stranded guanines), RNase T2 (specific for single-stranded regions), and RNase V1 (specific for double-stranded regions). Dimethylsulfate (DMS; alkylates [N3]C and [N7]G), diethylpyrocarbonate (DEPC; modifies [N7]A), nickel complex (NiCR; modifies [N7]G), and hydroxyl radicals (cleavage at ribose moieties) were used as chemicals. Hydroxyl radicals were only used on end-labeled CopA free or bound to CopT. NiCR (Chen et al., 1993) and DEPC (Weeks & Crothers, 1993) are known to be very sensitive to stacking of base rings. Therefore, position N7 of purines within a helix is not reactive unless the deep groove is widened. RNAs were also subjected to Pb^{2+}-induced cleavage, which has proven exquisitely sensitive to subtle structural variations (Malmgren et al., 1997). Experiments are shown in Figures 2–5, and the determined reactivities are indicated schematically in Figure 6. For convenience, analogous nucleotides in CopA and CopI are referred to by their positions in CopA.

In general, most cleavages/modifications of the free RNAs correlate well with the proposed secondary structures of both RNAs (Fig. 6). The main cleavages by RNases T1 and T2 were located in the external loops and the single-stranded regions, whereas RNase V1 cleavages were only observed in helical regions of both RNAs (Figs. 2 and 4). All cytosines reactive towards DMS were located in loops and single-stranded regions of both RNAs (Fig. 3). The results indicated the existence of a particular structure for loop II of CopT: the absence of a RNase T1 cut at G114 (Fig. 4) and the absence of reactivity at N3 of U109 and N1 of G114 towards chemical probes (Malmgren et al., 1996) were indicative of a U–G base pair closing the loop. This base pair has recently been proposed to occur in a similar loop of *repZ* mRNA (ColIb-P9 plasmid; Asano et al., 1998).

Structure probing of the extended kissing CopI–CopT complex

Binding of CopI induces several reactivity changes in CopT, specifically in loop II′ and the upper stem II′ (Figs. 4 and 6B). Unexpectedly, the major RNase T1 cleavages at G111 and G112 were only weakly decreased (Fig. 4). The RNase T2 cleavages in loop II′ were still observed in the extended kissing complex, whereas significant protection of the weak cleavages occurred at C104, A105, and A118. Upon RNase V1 digestion, new cleavages appeared at C104 and A105, and A106 and C107 became protected (Fig. 4). CopI also induced strong protections at positions [N7]A118 and [N3]C119, and moderate protections at [N3]C101, [N3]C113, [N7]A105, and [N7]A106 (Fig. 6B). These data are in good agreement with the Pb^{2+}-induced cleavages, because strong CopI binding-induced protections were previously observed on both sides of the upper stem II′ of CopT (U102–C107 and U117–A123; Malmgren et al., 1997). In contrast, the entire region of complementarity was rendered resistant to single-strand-specific RNases when full duplexes were formed artificially (Fig. 4).

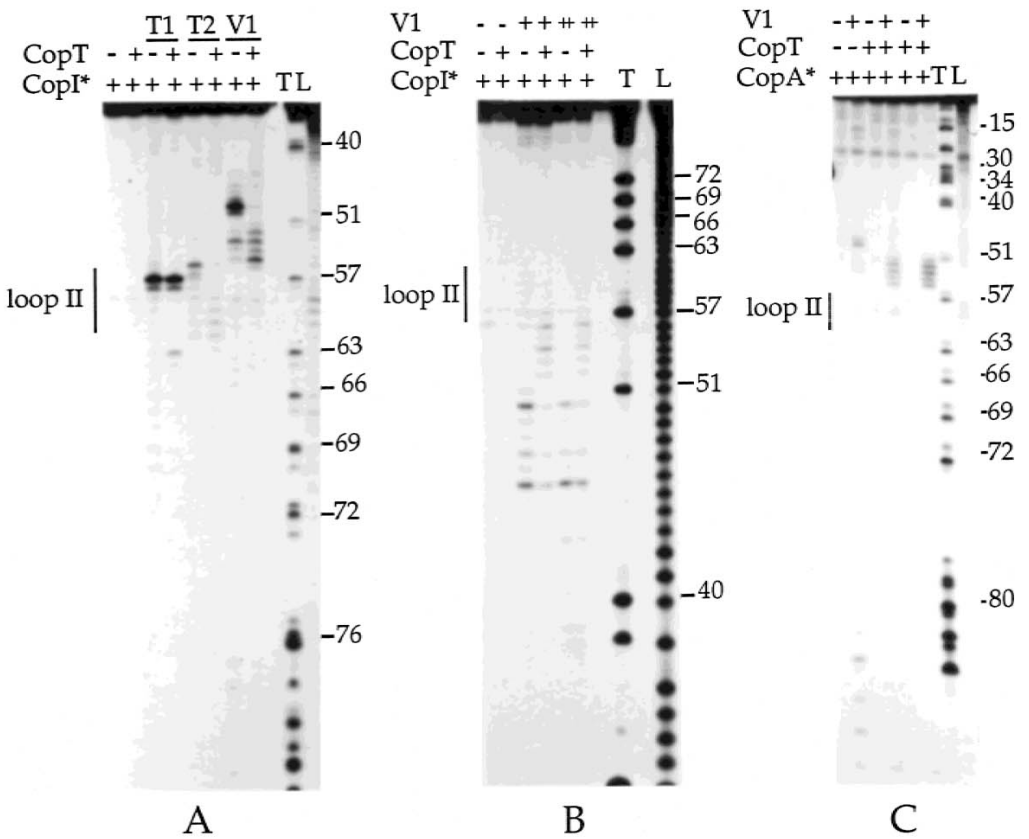

FIGURE 2. Enzymatic hydrolysis of CopI and CopA free or in complex with CopT. Enzymatic hydrolysis on 3′- (**A**) or 5′- (**B**) end-labeled CopI (CopI*), and (**C**) on 3′-end-labeled CopA (CopA*), free (− CopT) or in the presence of an excess of CopT (+ CopT). Complex formation was performed at 37 °C for 5 min. Enzymatic reactions were done at 37 °C for 5 min. T1: RNase T1, 0.001 unit; T2: RNase T2, 0.01 unit; V1: RNase V1, 0.05 (+) or 0.1 (++) U. Lanes T, L: RNase T1 and alkaline ladders, respectively.

Binding of CopT induced similarly located reactivity changes in loop II and the upper stem II of CopI (Figs. 2 and 3). The major RNase T1 cut at G57 remained unchanged or was only weakly protected, and a new weak cut appears at G63 (Fig. 2A). The strong RNase T2 cleavage at C56 became protected upon CopT binding whereas the cleavages at positions 60–62 were slightly enhanced (Fig. 2A). Again, these data were in accordance with the susceptibility of loop II of CopI to Pb^{2+}-catalyzed hydrolysis in the extended kissing complex (Malmgren et al., 1997). CopT also induced significant protection at [N3]C56 whereas cytosines 58–59 remained reactive at N3 (Fig. 3B), and adenines 60–61 at N7. RNase V1 hydrolysis was decreased at positions 48–50, concomitant with the appearance of new cleavages at positions 52–56 (Fig. 2). These new RNase V1 cleavages were not observed in the full CopA–CopT duplexes formed artificially (data not shown).

Thus, the accessibility of nt 57–63 in CopI and 107–113 in CopT to single-strand-specific probes and the RNase V1 cuts at positions 52–56 in CopI and 104–107 in CopT indicate that formation of the extended kissing complex requires intermolecular interactions in the upper stem regions at the expense of intramolecular base pairs.

Structure probing of the stable CopA–CopT complex

CopA–CopT complexes were formed at 37 °C for 2–15 min and subjected to enzymatic cleavages or chemical modifications. We previously showed that within 15 min of incubation, at given concentrations, most RNAs were in the form of stable complexes, but only a minor fraction was converted to a full duplex (Malmgren et al., 1997).

CopA and CopI induced virtually indistinguishable reactivity changes in the stem-loop II′ of CopT (Fig. 6B). In particular, residues in loop II′ remained accessible to the single-strand-specific RNases and Pb^{2+} hydrolysis when CopA–CopT complexes were analyzed (Fig. 4). Furthermore, enhanced RNase V1 cuts also occurred

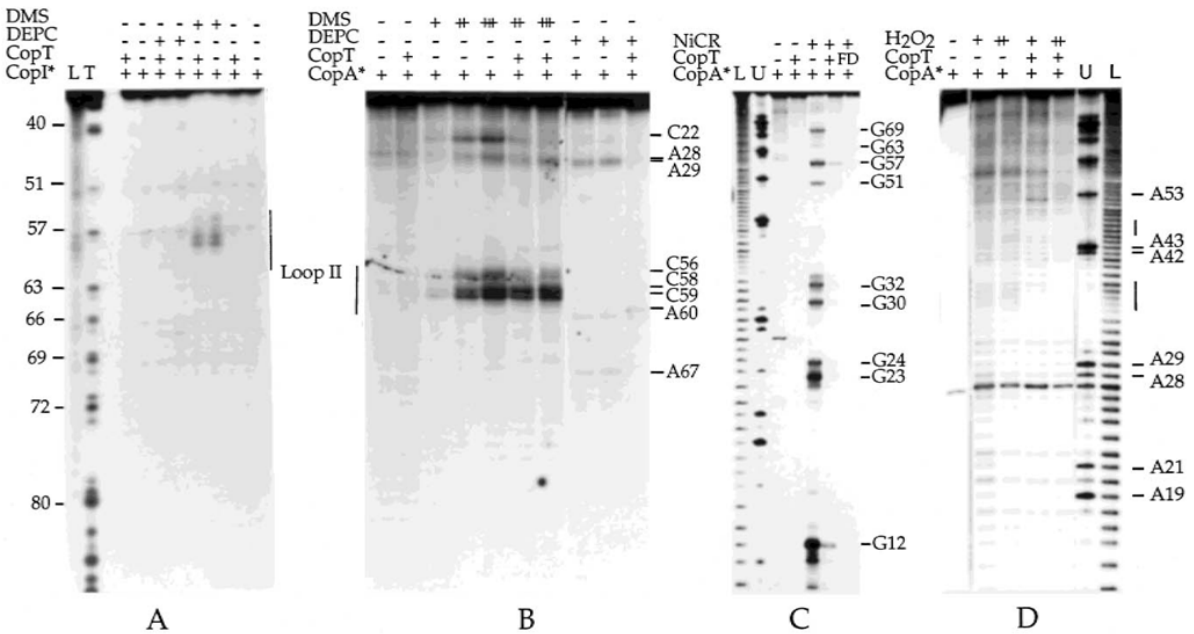

FIGURE 3. Chemical probing of CopA and CopI, free or in complex with CopT. **A**,**B**: DMS (N3C) or DEPC (N7A) modifications on (**A**) 3′-end-labeled CopI (CopI*) or (**B**) CopA (CopA*), free (− CopT) or bound to CopT (+ CopT). Complex formation with CopT was performed at 37 °C for 5 min. DMS, native conditions: 5 min (+), 10 min (++), and 15 min (+++) at 37 °C; DEPC, native conditions: 20 min at 37 °C. (**C**) NiCR modification on 5′-end-labeled CopA: free (− CopT), bound to CopT under native conditions (+ CopT), and full CopA–CopT duplex (FD); (**D**) Hydroxyl radical hydrolysis on 5′-end-labeled CopA free (− CopT) or bound to CopT (+ CopT). Reactions were performed either in the presence of H_2O_2 1% (+) or 10% (++). Lanes L, T, U: Alkaline, RNase T1, and RNase U2 ladders, respectively.

at C104 and A105. Additional strong protections were observed in the 140–169 region (T2, T1, and Pb^{2+} cleavages) and, concomitantly, new RNase V1 cuts appeared at U162–C163 and U167–A168 (Fig. 4). In contrast, formation of a full duplex between CopA and CopT conferred complete protection of the CopA binding site (positions 80–169) from single-strand-specific probes but induced new RNase V1 cuts at U162–C163 and U167–A168 (Fig. 4). In the case of the native complex, no major structural rearrangement of CopT RNA occurred in the noncomplementary region (Fig. 4). Minor changes were represented by enhanced RNase T1 cleavages at G172 and G173 in the Shine–Dalgarno sequence of *tap*, indicating melting of stem I′.

Conversely, CopT induced the same reactivity changes in stem-loop II of CopI and CopA (see Figs. 2 and 3). This is well illustrated by the new RNase V1 cuts at positions U52–C56 and the accessibility of residues 57–61 towards chemicals and RNase T2 (Figs. 2 and 3). Furthermore, most of the N7 position of guanines, with the exception of G57, showed protection towards NiCR (Fig. 3C). In addition, strong protections were detected in the 5′ tail of CopA encompassing nt 1–30 (T2 and T1), and [N3]C22 and [N7]/(A9, G23, G24, A28, and A29) were rendered unreactive in the complex (Figs. 2 and 3B). Finally, significant protections were induced by CopT binding at several riboses of CopA, at positions 35–38 and 46–48 (Fig. 3D).

These results further confirm that a full duplex between CopA and CopT is not rapidly formed in vitro. Instead, the extended intermolecular interactions present in the CopI–CopT complex are maintained in the CopA–CopT complex. These data suggest the formation of two intermolecular helices, B and B′, thus generating a four-helix junction. Additionally, the extended kissing complex is stabilized by an intermolecular helix C formed between the 5′ most 30 nt of CopA and the complementary sequences of CopT.

Base-pair inversions in stems II/II′ of CopA and CopT alter the formation of helices B and B′

Structure mapping of the complexes formed by CopI–CopT and CopA–CopT supported the presence of two intermolecular helices, B and B′, as modules in the four-helix junction structure (Fig. 6). To delimitate the number of base pairs in the two intermolecular helices, we used site-directed mutagenesis of the *copA/copT* gene sequence. Base-pair inversions were introduced at three positions within stem II. These mutations affect neither the stability nor the structure of the RNAs in their free state (data not shown), but were expected to affect formation of helices B and B′ (Fig. 6). Polymerase chain reaction (PCR)-generated templates were used for transcription of mutated CopI, CopA, and CopT

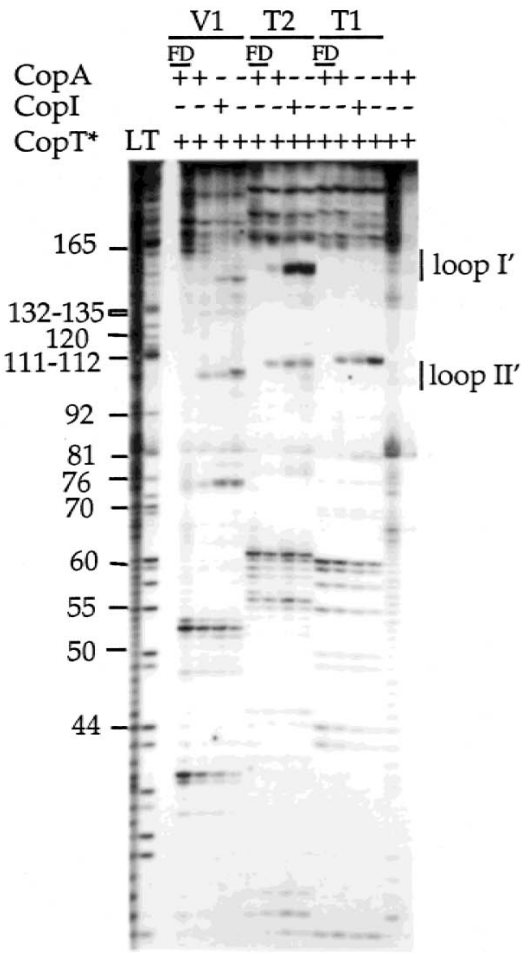

FIGURE 4. Enzymatic probing on CopT, free or in complex with antisense RNAs. Enzymatic hydrolysis on 5′-end-labeled CopT (CopT*). Complex formation with CopI (+ CopI), or with CopA (+ CopA) was performed at 37 °C for 5 min under native conditions. Full duplex with CopA (FD) was formed by denaturation/annealing treatment. Enzymatic hydrolysis were performed with RNase T1, 0.001 U (T1); RNase T2, 0.01 U (T2); RNase V1, 0.1 U (V1).

RNAs. The effect of the mutations on the structure of homologous or heterologous wild-type and mutant CopI–CopT or CopA–CopT pairs were analyzed by enzymatic mapping. Autoradiograms of such experiments are shown in Figure 5.

Cleavages by RNase V1 was used to obtain a signature of the extended kissing complex: new or enhanced cleavages were induced in CopA or CopI at consecutive positions U52–C56, and in CopT at C104–A105. Furthermore, in the CopA–CopT complex, the formation of the intermolecular helix C that greatly enhances complex stability was characterized by several V1 cuts in CopT at U162–C163 and U167–A168, as well as by the disappearance of V1 cleavages in CopA at positions G7–A9 and G16–A19, resulting from melting of helix I. Therefore, we used RNase V1 to probe the structure of the homologous and heterologous complexes formed with the H1, H2, and H3 RNA variants (Fig. 5). In all these experiments, we used conditions (RNA concentrations, time of incubation, buffers) under which the stable CopA–CopT complex is almost quantitatively formed.

RNase V1 probing of end-labeled CopA RNA variants indicated almost identical cleavage patterns of free mutant and wild-type RNAs. Hence, no major structural changes were caused by the mutations. The three homologous CopA/CopT and the heterologous CopA-H3/CopT-wt complexes showed the characteristic RNase V1 patterns, that is, the appearance of cleavages at positions U52–C56 in CopA (Fig. 5A). Interestingly, the heterologous complexes formed between CopT-wt and either CopA-H1 or CopA-H2 were different: significant protections were observed at U49–C50 of CopA-H1 and CopA-H2 (Fig. 5A), whereas positions U52–C56 remained uncleaved, that is, a pattern resembling that of free RNAs. The same experiments were also performed using end-labeled CopT variants (Fig. 5B). Again, no major structural changes occurred in CopT mutant RNAs, and identical cleavage patterns were observed for all homologous CopA–CopT complexes, as well as in the heterologous CopA-wt/CopT-H3 complex. Cleavages at positions C104–A105 in the different CopT species were enhanced after complex formation, and protections occurred at A106–C107 (Fig. 5B). In contrast, CopA-wt/CopT-H1 and CopA-wt/CopT-H2 complexes were not susceptible to cleavage at these positions (Fig. 5B). Identical data were obtained by using the mutated CopI RNAs which cannot form helix C (data not shown).

Thus, the base-pair inversions located above the lower bulge in CopA or CopT (H1 and H2) affected the overall topology of the four-way junction structure, whereas a wild-type-like cleavage pattern was restored by forming complexes between fully complementary RNAs. This supports the presence of canonical intermolecular base pairs in helices B and B′ that are required for RNase V1 hydrolysis in these two regions. In contrast, base-pair inversions below the lower bulged residues (H3) in each of the RNAs did not affect the global structure of the CopA–CopT complex, demonstrating that the corresponding nucleotides are not involved in intermolecular base pairing between CopA and CopT.

A three-dimensional model for the antisense RNA–target RNA complex

A secondary structure model consistent with the probing data is proposed for the CopA–CopT complex and shown in Figure 6C. The intramolecular and intermolecular helices are entirely formed by Watson–Crick base pairs. Our data argue for the formation of two intermolecular helices (B and B′) connected by single-

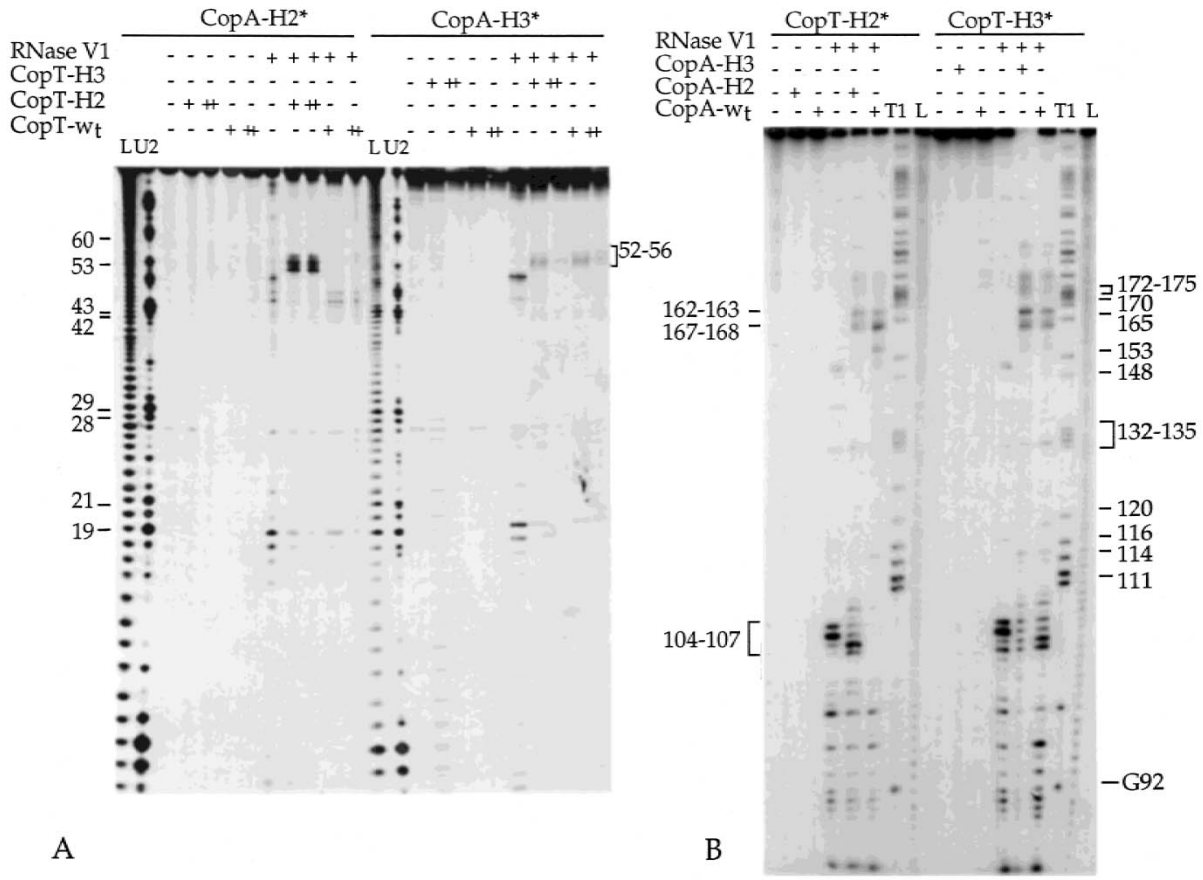

FIGURE 5. RNase V1 probing of homologous or heterologous CopA–CopT complexes. Enzymatic hydrolysis was performed on 5′-end-labeled CopA-H2 or CopA-H3, alone or in the presence of an excess of wild-type or mutant CopT (**A**) and on 5′-end-labeled CopT-H2, or CopT-H3, free or bound to either wild-type or mutant CopA (**B**). Complex formation was performed at 37 °C for 15 min in TMN buffer. Lanes T1, U2, L: RNases T1, U2, and alkaline ladders, respectively.

stranded regions (G57–A62 in CopA, U108–C113 in CopT). These two connecting regions were accessible to single-strand-specific probes (Fig. 6). Helix B contains 9 bp formed between U48–C56 of CopA and G114–A122 of CopT. This is supported by the fact that [N3]C56 and [N7]G51 in CopA, as well as [N3]C119 in CopT, were rendered unreactive, and by the presence of several RNase V1 cuts in CopA. Helix B′ is proposed to consist of 6 bp (G63–A68 in CopA, U102–C107 in CopT), and is supported by several RNase V1 cleavages in CopT and by the unreactivity of guanines 63, 66, and 69 at their N7 position towards NiCR. The number of base pairs in helices B and B′ was also determined by site-directed mutagenesis. Indeed, only base-pair inversions located above the lower bulge in CopA or CopT (H1 and H2) affected the formation of both intermolecular helices. Probing data also showed that the intramolecular helices A′ (35–47/70–83) in CopA and A (88–100/123–135) in CopT remain unchanged in the complex.

As base stacking is a major factor in the stabilization of RNA structures, the four (two intra- and two intermolecular) helices are proposed to stack forming two quasicolinear duplexes, centered around a four-helix junction, as a result of strand exchange between the two complementary RNAs (Fig. 6C). In principle, two types of junctions were possible, as the two helical domains, generated by coaxial stacking, could be constituted by either B–A and B′–A′ (junction 1 in Fig. 6C) or by B′–A and B–A′ (junction 2, not shown). The choice between the two junctions was made with the help of molecular modeling. A three-dimensional model of the CopA–CopT complex was assembled based on the experimental data and stereochemical constraints. Using this approach, the configuration of junction 2 could be rejected due to topological constraints that prevented connection of helices B and B′ via the single-stranded regions. The overall form of the resulting model shown in Figure 6 is that of an asymmetrical cruciform structure. The helical arms stack in pairs (A–B and A′–B′) to generate quasicontinuous, coaxial helices. These two helical domains were slightly rotated to adopt an asymmetrical X-shape required for the accommodation of single-stranded residues in grooves (Fig. 7A,B). This

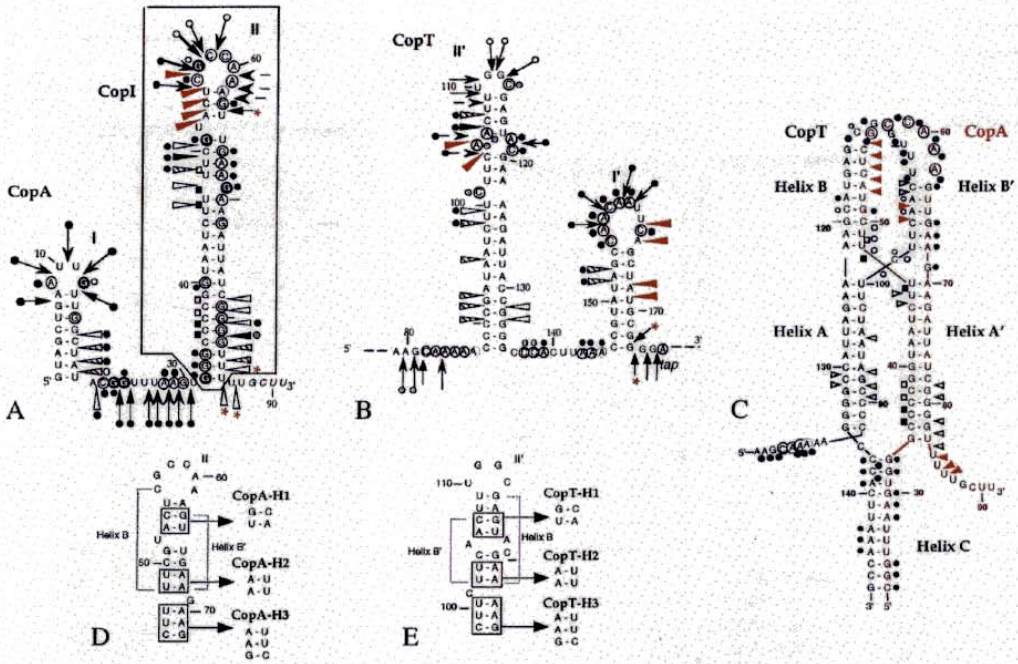

FIGURE 6. Enzymatic and chemical probing on free and complexed RNAs. A summary of the probing results is represented on the secondary structure of CopA (**A**), and CopT (**B**). The sequence of CopI is framed. The reactivity changes in CopI and CopA induced by CopT binding are strictly identical. **A**: Enzymatic cleavages: RNase T1 (o—►) T2 (o—►), RNase V1: (—◄) low, (—◄) moderate, (—◄) strong. Chemical modifications: cytosines at position N3, adenines at position N7, and guanines at position N7 reactive under native conditions are circled. The reactivity changes in CopA and CopT in the complex are indicated as follows: strong (filled circle), moderate (grey circle), and weak (open circle) protection; enhancements are represented by asterisks proportional to the intensity. New RNase V1 cleavages are shown by red arrows. Protected riboses in CopA are shown by squares. **C**: Summary of the enzymatic and chemical probing on the secondary structure model of the CopA–CopT complex. Pb^{2+}-cleavages are denoted by blue dots. Circled nucleotides: [N3]C, [N7]A reactive under native conditions. Positions ([N3]C, [N7]A, [N7]G) that became protected in the complex are denoted by black dots. RNase V1 cuts and protected riboses are indicated as described above. RNase T1 and T2 cleavages are not shown. Helix C comprises residues 1–33 of CopA and 137–169 of CopT. Nucleotides of CopA are in red. **D**,**E**: Localization of the mutated nucleotides (H1, H2, and H3) squared on the secondary structure model of CopA (**D**) and CopT (**E**). Nucleotides involved in the intermolecular helices B and B′ are indicated by bars.

was necessary to avoid steric interference between backbones and to permit the connections of B with A′ and B′ with A (Fig. 7C). The transition between crossed helices was accomplished by sharp turns of the backbone between U47 and U48 in CopA and U100 and C101 in CopT. It was difficult to build a canonical base pair between C101 in CopT and G69 in CopA. However, C101 and G69 were stacked within the helix, as their N3 and N7 positions, respectively, were unreactive in the complex (Fig. 6). The two connecting loops (residues 56–62 in CopA and 108–113 in CopT) were accessible to the solvent. Because of the limited number of constraints, a significant variation in local geometry for both loops may occur. Nevertheless, all the topological constraints resulting from the four-way junction are included in the molecular modeling. The unusual folding of the extended kissing complex favors a side-by-side alignment of stems A′ in CopA and A in CopT that facilitates formation of a third intermolecular helix between the 5′-most 30 nt of CopA and the complementary region of CopT (Fig. 7A,B). The close proximity of the phosphate backbone at this resulting three-way junction may explain the protections of several riboses from hydroxyl radicals, observed at position 35–38 in CopA (helix C, Figs. 3C and 6).

A putative divalent metal ion binding site in the CopA–CopT complex

The four-way junction imposes a close proximity of the helices that is well suited for divalent metal-ion binding. Computer modeling based on Brownian-dynamics sim-

Structure of antisense-target RNA complex

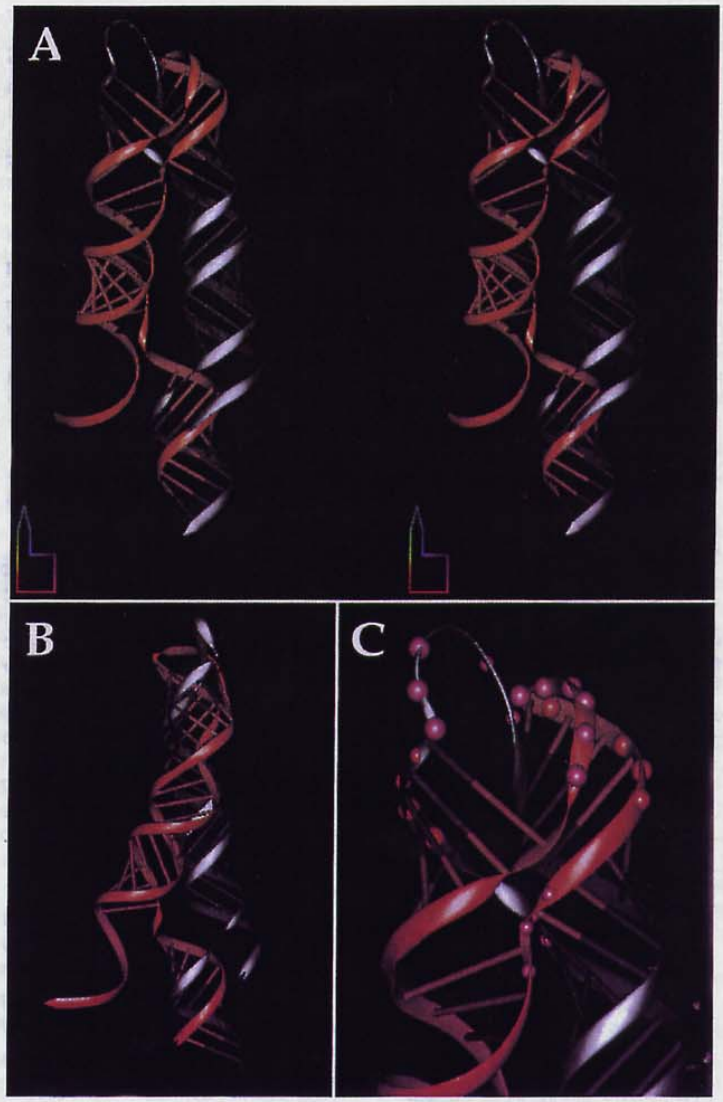

FIGURE 7. Models of the CopA–CopT complex. **A:** Stereoview of the three-dimensional model of the CopA–CopT complex. CopA [20–91] is in red, CopT [80–143] in white. **B:** The complex rotated by 45 degrees. **C:** An enlarged view of the four-way junction. Regions [U102–A122] in CopT and [U48–A68] in CopA are shown. RNase V1 cuts are represented by red dots, and Pb^{2+} cleavages by purple dots. The models were drawn with the program DRAWNA (Massire et al., 1994).

ulations for exploration of metal-ion binding sites in RNAs (Hermann & Westhof, 1998) predicted a magnesium binding site coordinated via the phosphate groups between the two adjacent helical domains (results not shown). With respect to this, the protection of riboses 46–48 towards hydroxyl radicals (Fig. 3D) might be well correlated with the presence of a Mg^{2+} binding pocket, as shown for tRNAs (Latham & Cech, 1989). Strong cleavage induced by Pb^{2+}-catalyzed hydrolysis at precise positions is also indicative of divalent cation

binding sites (for a review, see Pan et al., 1993). Interestingly, the inversion of 2 bp at positions 48–68 and 49–67 in CopA (mutant CopA-H2) and at positions 102–122 and 103–121 in CopT (CopT-H2) causes the appearance of a strong Pb^{2+}-cleavage site at U47 in CopA-H2 or CopI-H2 upon complex formation with CopT-H2 (Fig. 8A). These mutations do not affect the binding rate for the formation of the stable complex (data not shown). Also, the structure of the extended kissing complex is very similar to that of the wild-type CopI–CopT complex (Fig. 5), as indicated by a Pb^{2+}-cleavage pattern encompassing residues U52–G69 of CopI-H2 bound to CopT-H2 that is indistinguishable from that of the wild-type complex: the hairpin loop II of CopI in the complex was still cleaved at C59, A60, A61, and A62, and significant protections occurred on both sides of the upper helix II (Fig. 8B; Malmgren et al., 1997). Cleavage at U47 requires intermolecular base pairing between A48/A49 of the antisense RNA and U121/U122 of CopT, as no enhanced cleavage at U47 was detected in the heterologous complex CopI-H2/CopT-wt (data not shown). Finally, hydrolysis at U47 was strongly reduced by increasing concentrations of Mg^{2+} (Fig. 8C). Thus, the strong lead cleavage at U47 most likely results both from the presence of a divalent metal-ion binding site and from appropriate stereochemistry and mobility of the cleaved phosphodiester bond.

DISCUSSION

Conversion of a loop–loop interaction to a four-helix junction

Previous kinetic and mutational analyses showed that binding is initiated via a loop–loop interaction between the 5′-CGCC-3′ sequence in the hairpin loop of CopA and the complementary sequence of CopT (Persson et al., 1988, 1990b). However, these initial interactions are transient, as the present study of CopA–CopT and CopI–CopT complexes shows that most of the nucle-

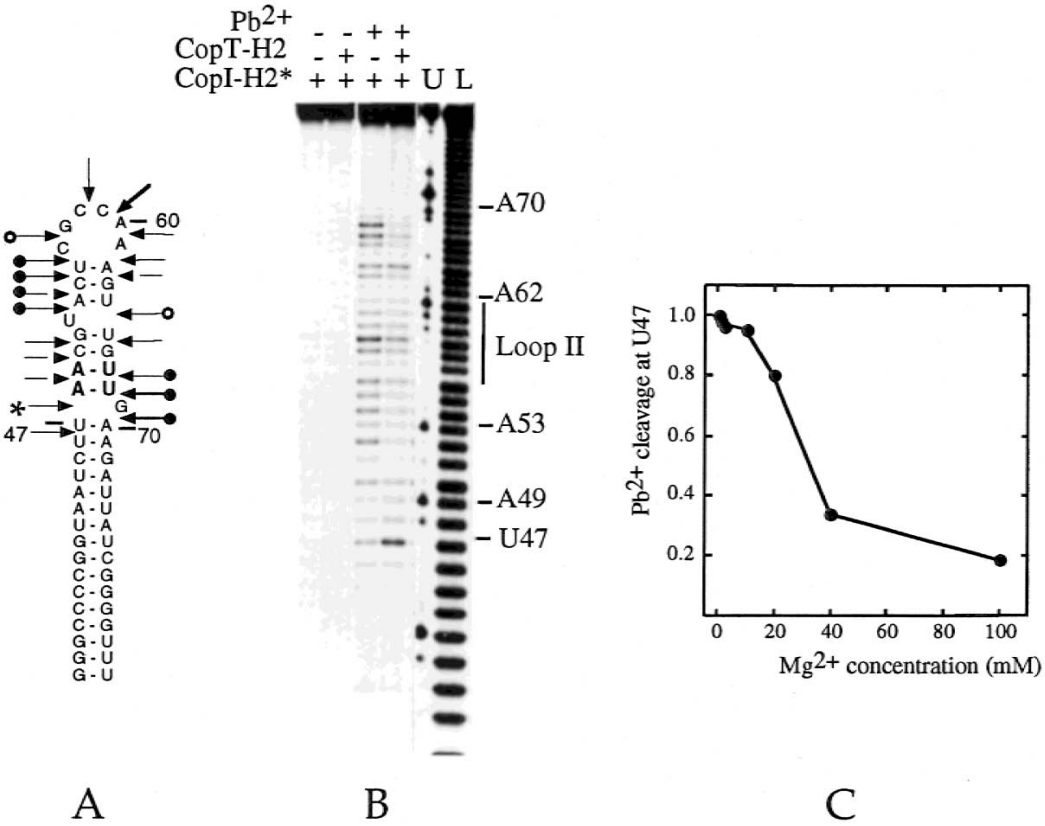

FIGURE 8. Pb(II) probing of CopI-H2 and the effect of binding of CopT-H2. **A**: Secondary structure of CopI-H2 showing the effect of CopT-H2 binding. The mutated nucleotides are shown in bold. Position of Pb^{2+} cleavages in the free RNAs are indicated by arrows. The protections induced by CopT-H2 are shown by spheres, and enhanced cleavage is indicated by an asterisk. **B**: Autoradiogram showing Pb^{2+}-hydrolysis (8 mM) on 5′-end-labeled CopI-H2 alone (− CopT-H2), or in the presence of CopT-H2 (+ CopT-H2). Lanes U, L: RNase U2 and alkaline ladders, respectively. **C**: Effect of magnesium concentration on Pb^{2+} hydrolysis. The intensity of the band corresponding to cleavage at U47 of CopI-H2 complexed to CopT-H2 is depicted as a function of magnesium concentration.

otides in loop II of the antisense RNA, as well as in the complementary target loop, are accessible to single-strand-specific probes. Thus, helix extension from this initial loop–loop complex generates helices B and B′, resulting in disruption of the initiating base pairs and formation of the extended kissing complex (Fig. 6C). The enzymatic and chemical probing data indicate that nucleotides on both sides of the upper stem II segments (48–56, 63–68) of CopA and CopI interact by interstrand pairing with the complementary nucleotides in CopT forming the two intermolecular helices B and B′ (Fig. 6C). Further support comes from the observation that mutations in helices B/B′ that disrupt the complementarity between CopA or CopI and CopT affect the structure of the extended kissing complex. On the other hand, mutations below the lower bulge had no effect on the formation of helices B and B′. Taken together, these results define the boundaries of helices B and B′ (Fig. 6C) and indicate that progression of the intermolecular helices B and B′ is arrested due to topological stress. Consequently, the intramolecular helices A and A′ remain identical in CopI–CopT and CopA–CopT complexes. We infer that conversion of the initial loop–loop interaction to the four-helix junction structure is facilitated by low stability of the upper stem regions, conferred by the bulged residues and the presence of several G–U and A–U base pairs in both RNAs. By independent experiments, it was previously shown that removal of bulges severely impaired binding rates and in vivo control (Hjalt & Wagner, 1995). The results presented here also indicate that complete pairing between CopA and CopT does not occur rapidly in vitro. It is likely that the formation of the peculiar four-helix junction structure results in kinetic or topological entrapment of the complex for extended periods of time (Malmgren et al., 1997; this work).

The extended kissing complex adopts an X-shaped structure with a side-by-side alignment of helical domains

Molecular modeling was used to deduce a global fold of the stable CopA–CopT complex. The validity of the overall architecture presented relies on a variety of experimental data (chemical and enzymatic probing; effect of mutations on CopA–CopT structure) that were used to assemble a structure with appropriate stereochemistry. We propose that the folding of the extended kissing complex at the four-way junction adopts an asymmetric stacked cruciform structure, formed by the colinear helices B–A and B′–A′ in a parallel configuration (Figs. 6C and 7). This model predicts a close contact between helices at the stacked cruciform junction (Fig. 7). Such a negatively charged cavity is likely to bind divalent ions with high affinity. Interestingly, computer modeling based on Brownian-dynamics simulations (Hermann & Westhof, 1998) predicted a magnesium-binding site coordinated via the phosphate groups between the two adjacent helical domains that may explain the protection of riboses 46–48 of CopA bound to CopT towards hydroxyl radicals (Fig. 3D). The presence of a divalent ion at the junction is also suggested by a strong Pb^{2+} cleavage at position 47 of CopA-H2 bound to CopT-H2 that was strongly reduced by increasing concentrations of magnesium (Fig. 8). Finally, we recently observed that magnesium ions are essential for optimal formation of the stable CopA–CopT complex (data not shown).

The overall topology structure of the extended kissing complex is similar to some other RNA four-way junctions (Krol et al., 1990; Walter et al., 1998a; Nowakowski et al., 1999) and DNA Holliday junctions (Duckett et al., 1992). In both DNA and RNA four-way junctions, divalent ions are required for the formation and stabilization of antiparallel X-shaped structures (Duckett et al., 1992; Walter et al., 1998b). The particularity of the proposed CopA–CopT structure is the crossing over of the strands at the junction under the constraints imposed by the two loops connecting intermolecular helices B and B′. This forces a side-by-side alignment of the two helical domains that brings the 5′ tail of CopA in close proximity to the complementary region of CopT (Fig. 7). The formation of intermolecular helix C, which clamps the two long helical domains, greatly enhances the stability of the complex (Persson et al., 1990a; Malmgren et al., 1997). Crystallographic analysis of a group I ribozyme domain revealed a similar organization (Cate et al., 1996): a sharp bend induced by an internal loop allows a side-by-side alignment of two helical domains that is additionally stabilized by metal- and ribose-mediated backbone contacts and two long-range tertiary interactions. A side-by-side configuration was also proposed for the hairpin ribozyme, here stabilized by interactions between two internal loops (Earnshaw et al., 1997).

The formation of a stable RNA–RNA complex is not unique to CopA–CopT, and is also a key feature in the replication control of plasmids belonging to the IncB and Inclα groups (Siemering et al., 1994; Asano & Mizobuchi, 1998; Asano et al., 1998) plasmids. In these systems, the antisense RNAs inhibit the formation of a pseudoknot structure that activates *rep* translation (Wilson et al., 1993; Asano & Mizobuchi, 1998). All these antisense and target RNAs are characterized by stable hairpins with identical loop sequences and bulged residues in the upper stem regions. Enzymatic probing performed on (antisense) RNAI in pMU720 plasmid bound to its target indicated that a full duplex was not rapidly formed in vitro. Instead, binding resulted in an extended kissing complex stabilized by 5′ tail interactions (Siemering et al., 1994). One may therefore speculate that, in all these systems, the final product of the binding reaction in vitro is characterized by an overall topology very similar to that reported here, except

that the lengths of helices B and B′, if formed in the IncB/IncIα cases, could be different.

Functional implications of the CopA–CopT structure for regulation

The role of CopA is to regulate the synthesis of RepA protein. Previous work showed that the stable CopA/CopT complex detected in vitro prevents ribosome from initiation complex formation at the *tap* RBS (Malmgren et al., 1996). CopI, the truncated antisense RNA unable to form fully paired duplexes, also repressed *repA* expression (Wagner et al., 1992), and the extended kissing complex (CopI–CopT) sufficed to transiently interfere with ribosome binding (Malmgren et al., 1996). We show here that the conversion of the loop–loop interaction to the four-helix junction occurs in both CopA–CopT and CopI–CopT complexes (Fig. 7). Therefore, association of the stabilizer segments (Helix C) of both CopA and CopT is not required for the formation of the extended kissing complex. Experiments with mismatched RNAs carrying mutations H1 and H2 indicate that rapid base pairing throughout helices B and B′ is essential for the formation of the stable CopA–CopT complex and in vivo control (F.A. Kolb, H. Engdhal, E.G.H. Wagner, & P. Romby, in prep.). This implicates the formation of helices B and, most likely B′, in promoting the side-by-side alignment required for formation of the stabilizer helix C and, thus, the inhibitory complex. Binding of CopI or CopA does not induce major conformational changes in the noncomplementary sequences of CopT, that is, the *tap* and *repA* RBS regions, in agreement with previous results (Öhman & Wagner, 1989; Malmgren et al., 1997). The model proposed here for the CopI–CopT complex is a bulky structure (Fig. 7) that might, by steric hindrance, prevent ribosome binding at *tap*. In full-length CopA, the presence of its 5′ extension has several functional implications. It stabilizes the extended kissing complex, providing approximately three helical turns of double-stranded RNA immediately 5′ of the *tap* SD. Thus, its main role is to promote complete and irreversible inhibition of *tap* translation. This intermolecular helix is also a substrate for RNase III cleavage, although destabilization of *repA* mRNA contributes little to control of *repA* expression (Blomberg et al., 1990). Finally, factors that alter CopA turnover will also affect plasmid copy number, because the degree of inhibition is correlated with the intracellular concentration of CopA. The 5′ tail of CopA carries a cleavage site for RNase E, the enzyme that initiates rapid turnover of CopA (Söderbom et al., 1997).

In many regulatory antisense systems, the formation of complete antisense-target RNA duplexes appears to be a slow process in vitro and often becomes arrested at the stage of a stable binding intermediate (Wagner & Brantl, 1998; Zeiler & Simons, 1998). For CopA–CopT, topological barriers are encountered during helix propagation, especially at the four-way junction of the extended kissing complex. Coaxial stacking with parallel packing of helices is known to be a general driving force towards RNA folding and probably contributes to the stabilization of the CopA–CopT binding intermediate. Moreover, unwinding of the stems of CopA and CopT from helix C, although topologically possible, requires overcoming important energy barriers. Thus, the CopA–CopT system appears to exploit binding intermediates as active key structures for the inhibitory step, rather than fully paired species.

So far, direct experimental evidence on structures of CopA–CopT complexes in bacterial cells is lacking. However, identically located RNase III-dependent cleavages occurring on both RNAs in vitro and in vivo (Blomberg et al., 1990; Malmgren et al., 1997) provide circumstantial evidence that, even in the cell, binding may be arrested at the stage of the extended kissing complex stabilized by the intermolecular helix C.

MATERIALS AND METHODS

Oligodeoxyribonucleotides

For RNA transcription in vitro, PCR fragments containing mutations H1, H2, or H3 were generated from the pGW177-III-L series plasmids using the following primer pairs: for CopT-H1: 5T-H1(5′ ACG TAC TTT AAA GCA AAA ACC CCG ATA ATC TTC TTC ATG TTT GGC GAC AAC GAA AAG ATT ACC G 3′) and SeqP-BamHI (5′ CGC GGA TCC CGG ATT CGG GTT CTT TA 3′); for CopT-H2: 5T-H2 (5′ ACG TAC TTT AAA GCA AAA ACC CCG ATA ATC TTC AAC AAC TTT GGC GAG TAC GTT AAG ATT ACC GGG GCC 3′) and SeqP-BamHI; for CopT-H3: 5T-H3 (5′ ACG TAC TTT AAA GCA AAA ACC CCG ATA ATG AAC TTC AAC TTT GGC GAG TAC GAA TTC ATT ACC GGG GCC CAC 3′) and SeqP-BamHI. The resulting PCR-fragments were cloned in the vector plasmid pUT7 (Serganov et al., 1997).

DNA templates and RNA synthesis

CopT, CopA, CopI, and the mutant RNAs (H1–H3) were synthesized by T7 RNA polymerase from PCR-generated DNA fragments as described (Hjalt & Wagner, 1992). PCR fragments were generated from plasmid pGW58 (Blomberg et al., 1990) carrying the wild-type *copA*/*copT* region. Transcription of CopT yields a run-off product of 302 nt initiated with GG instead of the GU sequence of the wild-type *repA* mRNA. The CopA RNA contains a 5′ terminal G instead of an A residue. Neither of these nucleotide changes affects structure or binding properties. In CopA or CopI mutants, 2 or 3 bp were inverted as shown in Figure 6. The complementary changes were also introduced in CopT mutants. Purification of RNAs was performed by fast protein liquid chromatography (FPLC, Pharmacia) on a Bio-Sil TSK250 column. The RNAs were eluted by 0.2 M sodium acetate, pH 6.5, containing 1% methanol, and precipitated.

5′-end labeling of dephosphorylated RNA was performed with T4 polynucleotide kinase and [γ-^{32}P]ATP (Maniatis et al.,

1982). 3′-end labeling of RNA was performed with T4 RNA ligase and [^{32}P]-pCp (England & Uhlenbeck, 1978). Labeled RNAs were purified by polyacrylamide-urea gel electrophoresis, eluted, and precipitated twice with ethanol. Before use, the RNAs were dissolved in water and renatured by incubation at 90 °C for 2 min, followed by slow cooling at 20 °C in TMN buffer (20 mM Tris-acetate, pH 7.5, 10 mM Mg-acetate, 100 mM sodium acetate).

Enzymatic probing

Antisense RNA binding was carried out at 37 °C for 1–15 min in TMN buffer with end-labeled CopT (3×10^{-8} M) and a fivefold excess of unlabeled CopA or CopI (1.5×10^{-7} M), or with end-labeled CopA or CopI (4×10^{-8} M) and a fivefold excess of unlabeled CopT (2×10^{-7} M). Full duplexes between CopT and antisense RNA species were formed by incubation at 90 °C for 2 min followed by slow cooling to 37 °C in TMN buffer. Enzymatic hydrolysis was performed in 10 μL of TMN, in the presence of 1 μg carrier tRNA at 37 °C for 5 min: RNase T1 (0.0025 U), RNase V1 (0.1 U), RNase T2 (0.05 U). Reactions were stopped by phenol/chloroform extraction, followed by ethanol precipitation, and washing with 80% ethanol. Incubation controls were done in parallel to detect nicks in the RNA.

Chemical probing

RNA–RNA complex formation was performed as described above. Chemical modifications of end-labeled RNA, free or in complex, were performed at 37 °C in 20 μL of TMN containing 2 μg of carrier tRNA. Alkylations of [N3]C and [N7]G positions were done with 1 μL DMS (diluted 1/8 in ethanol) for 5 and 10 min, and modifications at [N7]A positions were done with 5 μL of DEPC for 20 and 40 min. After ethanol precipitation, cleavage at modified C residues was induced by incubation with 10 μL of 50% hydrazine for 5 min on ice, and cleavage at modified G residues was performed in 10 μL of 1 M Tris-HCl, pH 8.3, and 10 μL of sodium borohydride (8 mg/mL) for 10 min on ice. All samples were then treated by aniline (Peattie & Gilbert, 1980). Reactions were stopped by ethanol precipitation in the presence of 0.3 M sodium acetate pH 6.0. RNA pellets were washed twice in 80% ethanol and vacuum dried.

Pb^{2+}-induced hydrolysis was done in 20 mM HEPES-NaOH, pH 7.5, 10 mM Mg-acetate, 100 mM sodium acetate, in the presence of 8 or 16 mM $PbOAc_2$ for 5 min. Reactions were stopped by addition of 50 mM EDTA, followed by ethanol precipitation and treatment as above.

Modification of [N7]G by NiCR was performed according to Chen et al. (1993) in 20 μL containing 25 mM phosphate-KOH, pH 7.0, 10 mM $MgCl_2$, 100 mM NaCl, 2 μg carrier tRNA, in the presence of 3 μM NiCR and 200 μM $KHSO_5$. Incubation was for 30 min at 20 °C. Reactions were quenched by addition of 180 μL of a solution containing 3 μg of carrier tRNA, 0.3 M NaOAc, 10 mM Tris-HCl, pH 7.5, 10 mM EDTA, and 0.5% (w/v) sodium dodecyl sulphate (SDS), followed by ethanol precipitation. Cleavages at modified G residues were induced by aniline treatment (Peattie & Gilbert, 1980). Reactions were stopped by ethanol precipitation and treated as above.

Fe(II)-EDTA reactions were done in 25 μL of TMN buffer in the presence of 1 μg carrier tRNA, 2 mM $Fe(SO_4)_2(NH_4)_2$, 4 mM EDTA, 1 mM DTE, 0.1% H_2O_2 for 10 min on ice. Reactions were stopped by ethanol precipitation and treated as above.

Identification of cleavage sites

End-labeled RNA fragments were dissolved in loading buffer (formamide 93%, 30 mM EDTA, xylene cyanol 0.05%, bromophenol blue 0.05%, SDS 0.5%), incubated at 90 °C for 3 min, and sized on 15% polyacrylamide/8 M urea gels. Cleavage positions were identified by running in parallel RNase T1, RNase U2, and alkaline ladders of the end-labeled RNA (Donis-Keller et al., 1977). Incubation controls were done to detect nicks in the RNA.

Graphic modeling

The three-dimensional model of CopT (nt A85–A150) interacting with CopA (nt U20–U91) was built using several algorithms (Westhof, 1993) incorporated in the program MANIP (Massire & Westhof, 1998). The generated model was subjected to restrained least-squares refinement using the programs NUCLIN and NUCLSQ (Westhof, 1993) to ensure geometry and stereochemistry with allowed distances between interacting atoms and to avoid steric conflicts. The color views were generated with the program DRAWNA (Massire et al., 1994). The model was then tested by comparing the theoretical accessibility of atoms with the observed experimental reactivities, as described earlier (Westhof et al., 1989).

ACKNOWLEDGMENTS

We thank Drs. B. Masquida and C. Massire for their help during graphic modeling and helpful discussions, S. Lodmell for stimulating discussions and critical reading of the manuscript, and T. Hermann for the exploration of metal-ion binding sites. We are also grateful to Dr. S. Rokita for providing us with the NiCR compound. C.M. was supported by a fellowship from the Swedish Institute. This work was supported by grants from Centre National de la Recherche Scientifique (CNRS), Ministère des affaires étrangères (MAE) through a PICS (Programme International de Coopération Scientifique, P.R.), and by the Swedish Natural Science Research Council (NFR, G.W.) and Swedish Research Council for Engineering Sciences (TFR, G.W.).

Received September 29, 1999; returned for revision November 15, 1999; revised manuscript received November 24, 1999

REFERENCES

Asano K, Mizobuchi K. 1998. An RNA pseudoknot as the molecular switch for translation of the *repZ* gene encoding the replication initiator of the Inclα plasmid CoIIb-P9. *J Biol Chem 273*:11815–11825.

Asano K, Niimi T, Yokoyama S, Mizobuchi K. 1998. Structural basis for binding of the plasmid CoIIb-P9 antisense Inc RNA to its target RNA with the 5′-rUUGGCG-3′ motif in the loop sequence. *J Biol Chem 273*:11826–11838.

Blomberg P, Engdahl HM, Malmgren C, Romby P, Wagner EGH. 1994. Replication control of plasmid R1: Disruption of an inhibitory RNA structure that sequesters the *repA* ribosome binding site permits *tap*-independent RepA synthesis. *Mol Microbiol 12*: 49–60.

Blomberg P, Nordström K, Wagner EGH. 1992. Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. *EMBO J 11*:2675–2683.

Blomberg P, Wagner EGH, Nordström K. 1990. Control of replication of plasmid R1: The duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically in vivo and in vitro by RNase III. *EMBO J 9*:2331–2340.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot GE, Cech TR, Doudna J. 1996. Crystal structure of a group 1 ribozyme domain: Principles of RNA packing. *Science 273*:1678–1685.

Chen X, Woodson SA, Burrows CJ, Rokita SE. 1993. A highly sensitive probe for guanine N7 in folded structures of RNA: Application to tRNA-Phe and *Tetrahymena* group I intron. *Biochemistry* 32:7610–7616.

Donis-Keller H, Maxam AM, Gilbert W. 1977. Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res 4*:2527–2538.

Duckett DR, Murchie AIH, Bhattacharyya A, Clegg RM, Diekmann S, Kitzing E, Lilley DMJ. 1992. The structure of DNA junctions and their interaction with enzymes. *Eur J Biochem 207*:285–295.

Earnshaw DJ, Masquida B, Müller S, Sigurdsson STh, Eckstein F, Westhof E, Gait M. 1997 Inter-domain cross-linking and molecular modeling of the hairpin ribozyme. *J Mol Biol 274*:197–212.

Eguchi Y, Itoh T, Tomizawa J. 1991. Antisense RNA. *Annu Rev Biochem* 60:631–652.

Eguchi Y, Tomizawa J. 1991. Complexes formed by complementary RNA stem-loops. Their formations, structures and interactions with ColE1 Rom protein. *J Mol Biol 220*:831–842.

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res 16*:9109–9128.

England TE, Uhlenbeck OC. 1978. 3′-terminal labeling of RNA with T4 RNA ligase. *Nature 275*:560–561.

Hermann T, Westhof E. 1998. Exploration of metal ion binding sites in RNA folds by Brownian-dynamics simulations. *Structure 15*: 1303–1314.

Hjalt T, Wagner EGH. 1992. The effect of loop size in antisense and target RNAs on the efficiency of antisense RNA control. *Nucleic Acids Res 20*:6723–6732.

Hjalt T, Wagner EGH. 1995. Bulged-out nucleotides in an antisense RNA are required for rapid target RNA binding in vitro and inhibition in vivo. *Nucleic Acids Res 23*:571–579.

Kittle JD, Simons RW, Lee Y, Kleckner N. 1989. Insertion sequence *IS*10 antisense pairing initiates by an interaction between the 5′ end of the target RNA and a loop in the antisense RNA. *J Mol Biol 210*:561–572.

Krol A, Westhof E, Bach M, Lührmann R, Ebel JP, Carbon P. 1990. Solution structure of human U1 snRNA: Derivation of a possible three-dimensional model. *Nucleic Acids Res 18*:3803–3811.

Latham JA, Cech TR. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science 245*:276–282.

Lee AJ, Crothers DM. 1998. The solution structure of an RNA loop–loop complex: The ColE1 inverted loop sequence. *Structure 6*:993–1005.

Malmgren C, Engdahl H, Romby P, Wagner EGH. 1996. An antisense/target RNA duplex or a strong intramolecular RNA structure 5′ of a translation initiation signal blocks ribosome binding: The case of plasmid R1. *RNA 2*:1022–1032.

Malmgren C, Wagner EGH, Ehresmann C, Ehresmann B, Romby P. 1997. Antisense RNA control of plasmid R1 replication: The dominant product of the antisense RNA–mRNA binding is not a full RNA duplex. *J Biol Chem 19*:12508–12512.

Maniatis T, Fritsch EF, Sambrook J. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Marino JP, Gregorian RS, Csankovski G, Crothers DM. 1995. Bent helix formation between RNA hairpins with complementary loops. *Science 268*:1448–1454.

Massire C, Gaspin C, Westhof E. 1994. A program for drawing schematic views of nucleic acids. *J Mol Graph 12*:201–206.

Massire C, Westhof E. 1998. MANIP: An interactive tool for the modeling of ribonucleic acids. *J Mol Graph 16*:197–205.

Nowakowski J, Shim PJ, Prasad GS, Stout CD, Joyce GF 1999. Crystal structure of an 82-nucleotide RNA-DNA complex formed by the 10–23 DNA enzyme. *Nat Struct Biol 6*:151–156.

Öhman M, Wagner EGH. 1989. Secondary structure analysis of the RepA mRNA leader transcript involved in control of plasmid R1. *Nucleic Acids Res 17*:2557–2579.

Pan T, Long DM, Uhlenbeck OC. 1993. Divalent metal ions in RNA folding and catalysis. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 271–302.

Peattie DA, Gilbert W. 1980. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci USA 77*:4679–4682.

Persson C, Wagner EGH, Nordström K. 1988. Control of replication of plasmid R1: Kinetics of in vitro interaction between the antisense RNA, CopA, and its target, CopT. *EMBO J 7*:3279–3288.

Persson C, Wagner EGH, Nordström K. 1990a. Control of replication of plasmid R1: Structures and sequences of the antisense RNA, CopA, required for its binding to the target RNA, CopT. *EMBO J 9*:3767–3775.

Persson C, Wagner EGH, Nordström K. 1990b. Control of replication of plasmid R1: Formation of an initial transient complex is rate-limiting for antisense RNA-target RNA pairing. *EMBO J 9*:3777–3785.

Serganov A, Rak A, Garber M, Reinbolt J, Ehresmann B, Ehresmann C, Grunberg-Manago M, Portier C. 1997. Ribosomal protein S15 from *Thermus thermophilus*: Cloning, sequencing, overexpression of the gene and RNA-binding properties of the protein. *Eur J Biochem 246*:291–300.

Siemering KR, Praszkier J, Pittard AJ. 1994. Mechanism of binding of the antisense and target RNAs involved in the regulation of IncB plasmid replication. *J Bacteriol 176*:2677–2688.

Söderbom F, Binnie U, Masters M, Wagner EGH. 1997. Regulation of plasmid R1 replication: PcnB and RNase E expedite the decay of the antisense RNA, CopA. *Mol Microbiol 26*:493–504.

Thisted T, Sørensen NS, Wagner EGH, Gerdes K. 1994. Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5′-end single-stranded leader and competes with the 3′ end of Hok mRNA for binding to the *mok* translational initiation region. *EMBO J 13*:1960–1968.

Tomizawa J. 1990. Control of ColE1 plasmid replication. Interaction of Rom protein with an unstable complex formed by RNA I and RNA II. *J Mol Biol 212*:695–708.

Wagner EGH, Blomberg P, Nordström K. 1992. Replication control in plasmid R1: Duplex formation between the antisense RNA, CopA, and its target, CopT, is not required for inhibition of RepA synthesis. *EMBO J 11*:1195–1203.

Wagner EGH, Brantl S. 1998. Kissing and RNA stability in antisense control of plasmid replication. *Trends Biochem Sci 23*:451–454.

Wagner EGH, Simons RW. 1994. Antisense RNA control in bacteria, phages, and plasmids. *Annu Rev Microbiol 48*:713–742.

Walter F, Murchie AIH, Lilley DMJ. 1998a. Folding of the four-way RNA junction of the hairpin ribozyme. *Biochemistry 37*:17629–17636.

Walter F, Murchie AIH, Thomson JB, Lilley DMJ. 1998b. Structure and activity of the hairpin ribozyme in its natural junction conformation: Effect of metal ions. *Biochemistry 37*:14195–14203.

Weeks KM, Crothers DM. 1993. Major groove accessibility of RNA. *Science 261*:1574–1577.

Westhof E. 1993. Modeling the three-dimensional structure of ribonucleic acids. *J Mol Struct 286*:203–211.

Westhof E, Romby P, Romaniuk PR, Ebel JP, Ehresmann C, Ehresmann B. 1989. Computer modeling from solution data of spinach chloroplast and *X. laevis* somatic and oocyte 5S rRNAs. *J Mol Biol 207*:417–431.

Wilson IW, Praszkier J, Pittard AJ. 1993. Mutations affecting pseudoknot control of the replication of B group plasmids. *J Bacteriol 175*:6476–6483.

Zeiler BN, Simons RW. 1998. Antisense RNA structure and function. In: Simons RW, Grunberg-Manago M, eds. *RNA structure and function*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 437–464.

II. Article II : Rôle des nucléotides non appariés des hélices de CopA et de CopT dans la formation du complexe irréversible

Jusqu'à présent, aucune protéine n'a été impliquée dans le processus d'appariement de CopA à sa cible CopT. Ces deux ARN contiennent donc toute l'information au niveau de leur séquence et de leur structure pour leur assurer une interaction rapide et garantir ainsi un contrôle optimal. Une caractéristique commune à de nombreux ARN antisens est la présence d'une structure en tige-boucle par laquelle l'interaction est initiée. Dans le système CopA/CopT, la séquence et la taille de la boucle sont importantes [324,325]. De même, la présence d'une extension simple brin dans l'ARN antisens est nécessaire pour la stabilisation du complexe [334]. Par ailleurs, CopA et CopT contiennent des nucléotides non appariés dans les tiges-boucles majeures qui sont cruciaux tant pour l'efficacité de l'association *in vitro* que pour celle du contrôle *in vivo* [62]. Leur position est également importante puisqu'ils doivent être situés à proximité des boucles apicales dans les deux ARN.

A. Objectifs

Nous avons montré que le processus de formation du complexe CopA-CopT stable impliquait une conversion du « kissing-complex » initial vers une jonction à quatre hélices (Article I, [335]). Cette progression se fait au détriment des appariements intramoléculaires des tiges des deux ARN. La présence de nucléotides non appariés pourrait faciliter l'ouverture des hélices pour induire la formation des appariements intermoléculaires. Cette conversion serait favorable d'un point de vue thermodynamique puisque des appariements canoniques remplaceraient des hélices imparfaites. Afin de tester cette hypothèse, nous avons créé des appariements canoniques à la place des nucléotides non appariés des tiges-boucles de CopA et de CopT. Dans ces ARN mutants, soit le nucléotide non apparié du haut a été éliminé, soit celui du bas, soit les deux à la fois. Nous avons alors réalisé une étude structurale et mécanistique des complexes mutants.

B. Stratégie expérimentale

Dans un premier temps, nous avons comparé l'accessibilité des bases des deux ARN dans les complexes mutants et sauvages par des expériences de cartographie en solution. Trois sondes (Pb^{2+}, NiCR et RNase V1) qui donnent une signature des trois hélices intermoléculaires B, B' et C ont été utilisées. Par ailleurs, l'interférence chimique a été utilisée pour analyser l'interaction initiale dans les complexes sauvages et mutants. Les résultats structuraux obtenus ont été reliés à une étude fonctionnelle de ces complexes. Nous

avons testé *in vitro* l'habilité des ARN antisens mutants à inhiber la fixation des sous-unités ribosomiques 30S à la région 5'-UTR de *repA*. Finalement, afin d'expliquer les caractéristiques topologiques responsables du défaut d'activité des ARN antisens mutés, un modèle tridimensionnel d'un complexe mutant a été construit et comparé au modèle du complexe sauvage.

C. Résultats

Jusqu'à présent, l'implication des nucléotides des boucles de CopA et de CopT dans l'interaction initiale avait été déduite d'expériences génétiques, celle-ci étant trop transitoire pour être détectée *in vitro*. C'est pourquoi nous avons utilisé des expériences d'interférence chimique à la N-éthyl-N-nitrosourée (ENU) pour identifier les phosphates des deux ARN dont la modification pourrait interférer avec la formation des complexes sauvages et mutants. Les nucléotides 5'-GGCG-3' et 5'-CGCC-3' des boucles de CopT et de CopA, respectivement, sont essentiels à la formation des complexes sauvages et mutants. Ainsi, la délétion des nucléotides non appariés des hélices de CopA et de CopT n'affecte pas l'interaction initiale entre les ARN mutants.

Le complexe CopA-CopT sauvage possède une signature caractéristique vis-à-vis de différentes sondes de structure. Les boucles de CopA et de CopT sont non appariées dans le complexe et la présence des hélices intermoléculaires B et B' génère de nouveaux sites de clivage par la RNase V1. Au contraire, dans les complexes mutants, nous montrons que les trois hélices B, B' et C ne sont pas formées. En revanche, les boucles de CopA et de CopT sont appariées sur toute leur longueur (Figure 30). La conversion de l'interaction boucle-boucle vers la jonction à quatre hélices est donc bloquée. Contrairement aux ARN sauvages, la cartographie des ARN mutants libres montre que les tiges-boucles sont stabilisées et que seules les boucles apicales restent accessibles aux sondes spécifiques de régions simple-brin. Ce résultat explique l'absence de conversion.

Ces données de cartographie en solution ont été incorporées dans un modèle tridimensionnel d'un complexe mutant. Celui-ci est radicalement différent du complexe sauvage (Figure 30 ; Article I, figure 7). En effet, l'interaction boucle-boucle crée une nouvelle hélice quasi-colinéaire aux deux longues hélices intramoléculaires de CopA et de CopT. Il en résulte que l'extrémité de l'ARN antisens est distante de plus de 120 Å du site de fixation de ribosome (RBS) de *tap*. En conséquence, l'hélice stabilisatrice C ne peut pas être formée (Figure 30). La conversion de l'interaction initiale est donc essentielle pour rapprocher dans l'espace les deux extrémités complémentaires de CopA et de CopT. Cet effet structural explique pourquoi l'absence de résidus non appariés affecte fortement la vitesse de formation du complexe irréversible [62].

L'élimination des nucléotides non appariés affecte fortement l'interaction *in vitro* et donc l'efficacité du contrôle *in vivo* [62]. Nous montrons ici que les complexes boucle-boucle formés entre les différents ARN mutants sont incapables d'empêcher la formation d'un complexe ternaire entre les sous-unités ribosomiques 30S, l'ARNt$_i^{fmet}$ et l'ARNm *repA* au RBS de *tap*. Ce résultat indique que la topologie des complexes CopA-CopT et CopI-CopT est essentielle pour empêcher de manière stérique la fixation du ribosome et donc pour inhiber la traduction.

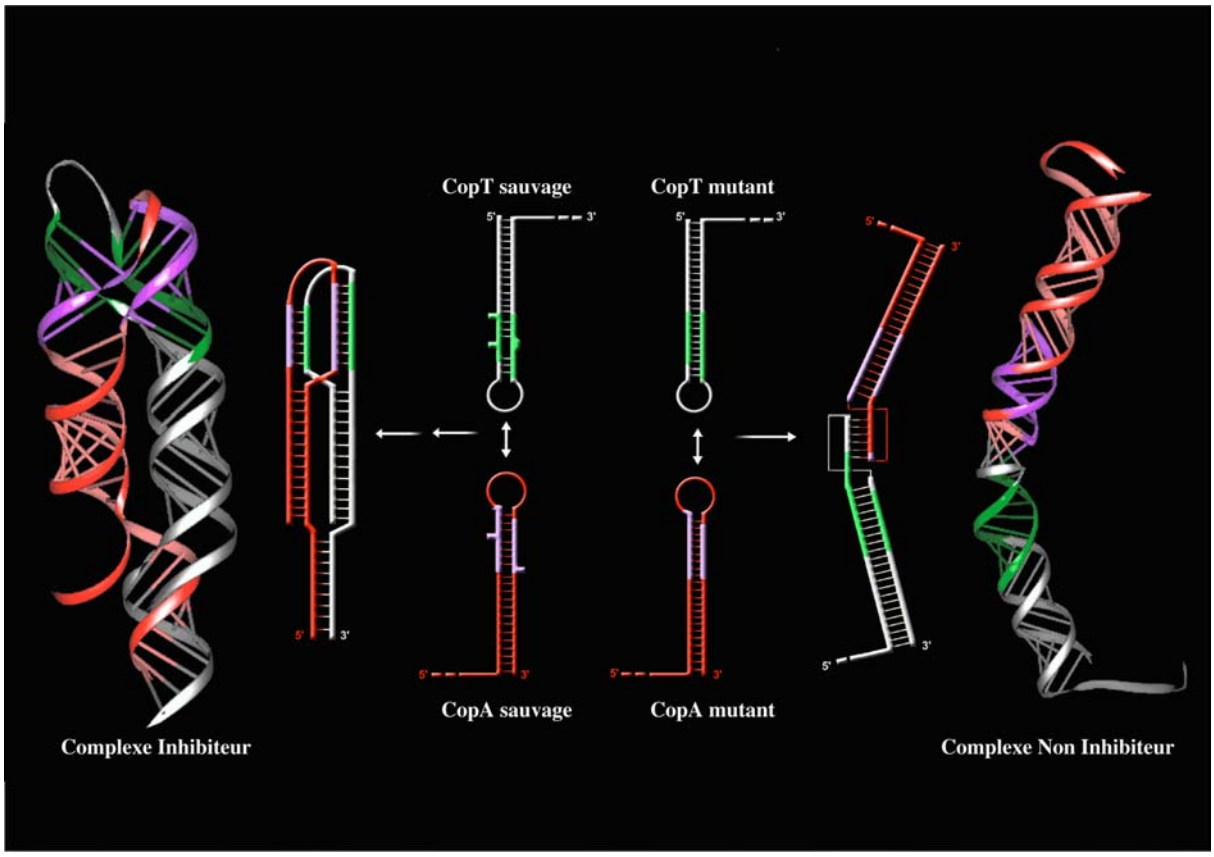

Figure 30 : ***Modèles de structures bi- et tri-dimensionnelles des complexes CopA-CopT sauvage (gauche) ou mutant (droite)***. Les ARN CopA et CopT sauvages sont indiqués, de même que les ARN mutants dépourvus de nucléotides non appariés dans la tige-boucle majeure. Les nucléotides impliquées dans la formation des hélices intermoléculaires B et B' dans le complexe sauvage sont indiquées en rose pour CopA et vert pour CopT. Ces mêmes positions ne sont pas impliquées dans des contacts intermoléculaires dans le complexe mutant. Dans ce dernier, la conversion de l'interaction boucle-boucle initiale est abolie.

D. Conclusions

Les résultats obtenus dans cette étude démontrent que les nucléotides non appariés dans CopA et dans CopT sont des éléments structuraux essentiels à la conversion de l'interaction boucle-boucle vers la jonction à quatre hélices. Lorsque ceux-ci sont éliminés, la progression est stoppée au niveau de la boucle. Dans un système dépourvu de protéine, les

nucléotides non appariés des tiges peuvent donc être assimilés à des ARN hélicases intramoléculaires.

© 2001 Oxford University Press

Nucleic Acids Research, 2001, Vol. 29, No. 15 **3145–3153**

Bulged residues promote the progression of a loop–loop interaction to a stable and inhibitory antisense–target RNA complex

Fabrice A. Kolb, Eric Westhof, Chantal Ehresmann, Bernard Ehresmann, E. Gerhart H. Wagner[1] and Pascale Romby*

UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue Rene Descartes, 67084 Strasbourg Cedex, France and [1]Institute of Cell and Molecular Biology, Biomedical Center, Uppsala University, Box 596, Husargatan 3, S-75124 Uppsala, Sweden

Received May 8, 2001; Revised and Accepted June 20, 2001

ABSTRACT

In several groups of bacterial plasmids, antisense RNAs regulate copy number through inhibition of replication initiator protein synthesis. These RNAs are characterized by a long hairpin structure interrupted by several unpaired residues or bulged loops. In plasmid R1, the inhibitory complex between the antisense RNA (CopA) and its target mRNA (CopT) is characterized by a four-way junction structure and a side-by-side helical alignment. This topology facilitates the formation of a stabilizer intermolecular helix between distal regions of both RNAs, essential for *in vivo* control. The bulged residues in CopA/CopT were shown to be required for high *in vitro* binding rate and *in vivo* activity. This study addresses the question of why removal of bulged nucleotides blocks stable complex formation. Structure mapping, modification interference, and molecular modeling of bulged-less mutant CopA–CopT complexes suggests that, subsequent to loop–loop contact, helix propagation is prevented. Instead, a fully base paired loop–loop interaction is formed, inducing a continuous stacking of three helices. Consequently, the stabilizer helix cannot be formed, and stable complex formation is blocked. In contrast to the four-way junction topology, the loop–loop interaction alone failed to prevent ribosome binding at its loading site and, thus, inhibition of RepA translation was alleviated.

INTRODUCTION

Many antisense RNAs are regulators of plasmid copy number (1). Plasmid R1 belongs to the IncFII family of plasmids which shares the overall genetic organization with respect to replication control functions. In plasmid R1, replication is negatively controlled at the translational level by binding of the antisense RNA (CopA) to its target site (CopT) in the leader region of the *repA* mRNA, ~80 nt upstream of the *repA* start codon (2). Synthesis of the replication initiator protein RepA requires translation of a short leader peptide (*tap*), encoded upstream of *repA*. Binding of CopA prevents *tap* translation by occluding ribosome binding at the *tap* ribosome binding site (RBS), and thereby *repA* expression is inhibited (3–5). Interestingly, several copy number control antisense RNAs which all inhibit translation of replication initiator proteins have structural properties similar to those of CopA. These RNAs are fully complementary to their target site and contain a long stem–loop structure. For proper control, the intracellular concentration of the antisense RNA must be a measure of the plasmid concentration: this is ensured by constitutive synthesis and a short half-life of these RNAs (6). Furthermore, efficiency of *in vivo* control is correlated with high rates of antisense RNA binding to its target site [in the order of 10^6 $M^{-1}s^{-1}$ (7)]. This class of antisense RNAs does not require *trans*-acting proteins to promote inhibition. Thus it can be speculated that the structure of these antisense RNAs has evolved to be optimized for the critical steps that determine their regulatory functions.

Kinetic studies have shown that the initial step of antisense RNA binding involves a loop–loop interaction mediated by a limited number of Watson–Crick base pairs, the so-called kissing complex [R1 (8); pMU720 (9); ColIb-P9 (10,11)]. In plasmid R1, we have recently shown that this loop–loop interaction is rapidly converted into a stable and inhibitory complex formed through a hierarchy of distinguishable intermediates, and that the formation of a full duplex *in vitro* is too slow to be of biological relevance (12–14). The major product of the binding reaction adopts an unusual structure characterized by a four-helix junction. Its formation involves extensive breakage of intramolecular base pairs to promote the formation of two intermolecular helices (see Fig. 1) (13). This conversion is essential for the formation of an irreversible complex and for the activity of the antisense RNA *in vivo* (14). We proposed previously that the four-way junction structure could promote a side-by-side helical alignment to allow formation of a third long intermolecular helix, involving the 5′ tail of CopA and the complementary region of CopT (13). Recent data indicated that a topology, strikingly similar to that of the stable CopA–CopT complex, may be present in complexes of all R1-related plasmids (11,15). This structure completely and irreversibly

*To whom correspondence should be addressed. Tel: +33 3 88 41 70 51; Fax: +33 3 88 60 22 18; Email: p.romby@ibmc.u-strasbg.fr

inhibits *repA* translation, mainly by occluding ribosome binding [e.g. pMU720 (16); R1 (5); Col1b-P9 (17)].

In addition to the recognition loops and the single-stranded regions required for stable complex formation, these antisense and target RNAs carry internal loops or unpaired residues in the major stem close to the hairpin loop. This is characteristic for antisense RNAs with long hairpins for which, unlike RNAI/RNAII of ColE1 (18,19), no protein is involved in regulation. The upper stem structure-disrupting elements can have several functional implications. They protect the RNAs against the double strand-specific RNase III since removal of bulged residues rendered CopA sensitive to cleavage *in vivo* and *in vitro* (20). Furthermore, their deletion strongly affected control *in vivo* and rapid formation of stable antisense/target RNA complexes *in vitro* [R1 (21); pUM720 (22); Col1b-P9 (23)]. Importantly, the positioning of these bulged residues close to the apical loop is critical since an internal loop placed at the bottom part of the stem did not restore inhibitory activity (21).

In the present paper, we show that removal of the bulged residues in CopA and CopT interferes with progression of the initial loop–loop helix towards the four-way junction. Instead, the loop–loop interaction formed by bulge-less CopA and CopT complex becomes arrested in a state that involves pairing of all loop bases, and does not convert to a stable complex. This complex fails to prevent ribosome binding at *tap* RBS.

MATERIALS AND METHODS

Oligodeoxyribonucleotides

Oligodeoxyribonucleotides were bought from IBA-NAPS. Primers for the generation of transcription templates were for $CopT_{302}$: T7GG, 5′-GAA ATT AAT ACG ACT CAC TAT AGG TTA AGG AAT TTT GTG GCT GG-3′ (T7 RNA polymerase promoter sequence underlined) and SeqP/II, 5′-CGG ATT CGG GTT CTT TA-3′; for $CopT_{74}$: GW46, 5′-GAA ATT AAT ACG ACT CAC TAT AGG CCC CGA TAA TCT TCT TCA ACT T-3′ and GW47, 5′-TGT TGG CTA TAC GGT TTA AGT-3′; for CopA: T7-SA, 5′-GAA ATT AAT ACG ACT CAC TAT AGT AGC TGA ATT GTT GGC TAT ACG-3′ and T7-EA, 5′-AAA GCA AAA ACC CCG ATA ATC TTC-3′. The primers for CopI were T7-SI, 5′-GAA ATT AAT ACG ACT CAC TAT AGG GCC CCG GTA ATC TTT TCG T-3′ and T7-EI, 5′-AAA CCC CGA TAA TCT TCT TCA-3′. SeqP/II was also used for primer extension in the toeprinting experiments.

DNA templates and RNA synthesis

CopT, CopA, CopI and mutant RNAs (Lo, Up, L/U) were synthesized by T7 RNA polymerase from PCR-generated DNA fragments as described (20). PCR fragments were generated from plasmid pGW58 carrying the wild-type *copA/copT* region (3), and from mutant plasmids (20). Transcription of CopT yields a run-off product of 302 nt initiated with GG instead of the GU sequence of the wild-type *repA* mRNA. A shorter $CopT_{74}$ corresponding to the complementary sequence of the antisense RNA was also produced, the RNA starts with GG instead of AG, i.e. CopA RNA contains a 5′ terminal G instead of an A residue. Neither of these nucleotide changes affects structure or binding properties. CopA and CopI mutants contained either a U52-A64:2 base pair inserted in the upper stem (Lo), or a C47:2-G69 base pair instead of a bulged guanine 69 (Up), or both (L/U) (Fig. 1) (20). The complementary changes were also introduced in CopT mutants. RNAs were purified from 8% polyacrylamide–urea gels.

5′ End-labeling of dephosphorylated RNA was performed with T4 polynucleotide kinase and [γ-^{32}P]ATP (24). Labeled RNAs were purified by polyacrylamide–urea gel electrophoresis, eluted and precipitated twice with ethanol. Before use, the RNAs were dissolved in water and renatured by incubation at 90°C for 2 min, followed by slow cooling at 20°C in TMN buffer (20 mM Tris-acetate pH 7.5, 10 mM Mg-acetate, 100 mM sodium acetate).

Enzymatic and chemical probing

Antisense RNA binding was carried out at 37°C for 15 min in TMN buffer with end-labeled CopT (3×10^{-8} M) and a 5-fold excess of unlabeled CopA or CopI (1.5×10^{-7} M), or with end-labeled CopA or CopI (4×10^{-8} M) and a 5-fold excess of unlabeled CopT (2×10^{-7} M). Full duplexes between CopT and CopA species were formed by incubation at 90°C for 2 min followed by slow cooling to 37°C in TMN buffer. RNase V1 hydrolysis, Pb^{2+}-induced hydrolysis and modification [N7]G by lead- and nickel-complex (NiCR) were done as previously described (13). Cleavage positions were identified by running, in parallel, RNase T1, RNase U2 and alkaline ladders of the end-labeled RNA (25). Incubation controls were done to detect nicks in the RNA.

Phosphate modification interference with ethylnitrosourea

5′ End-labeled RNA (10 nM) was first modified in 50 µl of buffer containing 50 mM sodium cacodylate pH 8.0, 1 mM EDTA with 16 µl of an ethylnitrosourea (ENU)-saturated ethanol solution for 1 min at 90°C. Reactions were stopped by ethanol precipitation with 0.3 M sodium acetate. Pellets were dissolved in TMN buffer, and complexes were formed with 1–5 nM of the unlabeled complementary RNA. An aliquot of labeled RNA was submitted to the same treatment except that ENU was omitted. Free and complexed RNAs were separated by electrophoresis at 4°C on 8% non-denaturing polyacrylamide gels in 45 mM Tris-borate pH 8.3. The corresponding bands were identified by autoradiography, cut out, and the RNA was eluted overnight at 4°C in 500 mM ammonium acetate pH 6, 0.1 mM EDTA, followed by precipitation. Ethylated phosphates were identified by incubating the modified RNA in 10 µl of 0.1 M Tris–HCl pH 9.0 at 50°C for 10 min. The control experiment was submitted to the same treatment. After precipitation, the RNA fragments were separated by denaturing gel electrophoresis on 15% polyacrylamide gels. Cleavage positions were identified by running in parallel RNase T1, and alkaline ladders of the end-labeled RNA.

Toeprinting assays

Toeprinting was performed essentially as previously described (5). The [^{32}P]5′ end-labeled primer SepP/II was first annealed to CopT RNA in 20 mM Tris-acetate pH 7.5, 60 mM NH_4Cl, 3 mM β-mercaptoethanol, heated at 90°C for 1 min and cooled on ice for 1 min. After the addition of 10 mM Mg-acetate, the RNA was incubated at 20°C for 20 min. Initiation complex formation was performed at 37°C from 30 s to 15 min in a reaction mixture containing 10 mM Tris-acetate pH 7.4, 60 mM

Nucleic Acids Research, 2001, Vol. 29, No. 15 **3147**

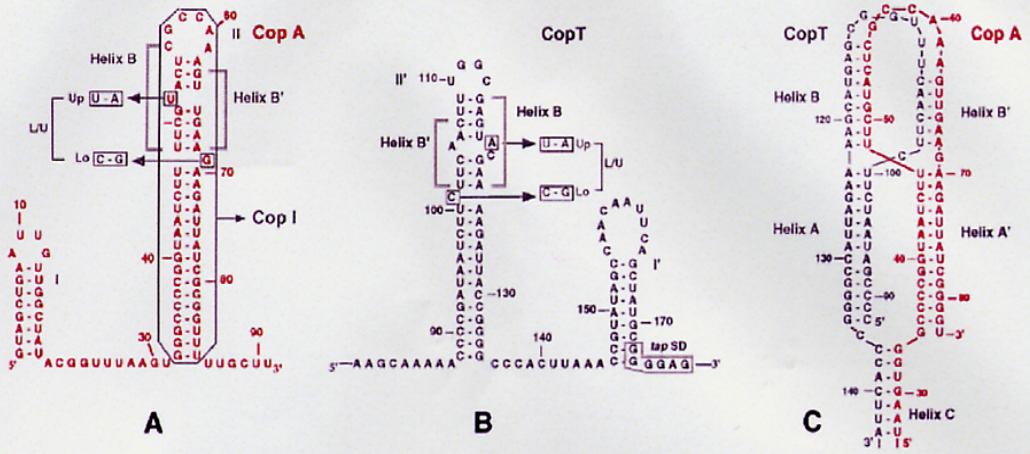

Figure 1. Secondary structures of the antisense RNA CopA (**A**), its target site CopT (**B**) and the stable CopA–CopT complex (**C**). The sequence of CopI is framed. Positions of mutations Lo, Up and L/U are shown on both RNAs. The Shine–Dalgarno sequence at *tap* is indicated. The secondary structure model for CopA–CopT is from Kolb *et al.* (13).

NH_4Cl, 10 mM Mg-acetate, 6 mM β-mercaptoethanol and, unless otherwise stated, 40 nM CopT RNA, 125 nM 30S subunits and 1 μM $tRNA_f^{Met}$. In the toeprinting inhibition experiments, antisense RNA/target RNA complexes were formed prior to ternary complex formation at 37°C for 15 min (as described above) in the presence of a 5-fold excess of CopA or CopI species, followed by the addition of 30S and tRNA. Concentrations of the antisense RNA was chosen to get an optimal formation of the kissing complex. Primer extension reactions were subsequently performed at 37°C for 15 min by adding 2 U of AMV reverse transcriptase (Life Sciences). Reactions were stopped by phenol extraction and ethanol precipitation.

Graphic modeling

The 3D model of CopT-L/U (nucleotides C88 to A178) interacting with CopA-L/U (nucleotides G1 to U87) was built using several algorithms (26) incorporated in the program MANIP (27). The generated model was subjected to restrained least-squares refinement using the programs NUCLIN and NUCLSQ (26) in order to ensure geometry and stereochemistry with allowed distances between interacting atoms and to avoid steric conflicts. The color views were generated with the program DRAWNA (28).

RESULTS

The loop–loop interaction is essential to initiate pairing in the wild-type and mutant complexes

Mutations in CopA and CopT which generated fully helical upper stems decreased the rate and stability of stable CopA–CopT complex formation by more than one order of magnitude (21). These effects suggested the topology of mutant and wild-type complexes to be different. Here, we used chemical modification interference experiments to probe the initial interaction in the wild-type CopA–CopT, CopI–CopT and mutant CopI-L/U–CopT-L/U complexes. CopI is a truncated CopA, unable to form the stable complex with CopT, but capable of forming the extended kissing complex (8,13). In the L/U mutant RNAs, the bulged nucleotides have been replaced by standard Watson–Crick base pairs (Fig. 1) (20,21). Phosphate groups were modified by treatment with *N*-ethyl-*N*-nitrosourea. The protocol used ensured ethylation of the non-bridging oxygens to, on average, less than one modification per molecule. Such an approach had been successfully used to probe the loop–loop interaction involved in the dimerization process of HIV-1 RNA (29). Modified antisense or target RNAs were allowed to form complexes with their counterparts. Free RNA and RNA–RNA complex, were separated by gel electrophoresis under native conditions and eluted from the gel. Sites of modifications in each fraction were then identified by alkaline treatment which induces RNA scission at the modified phosphate. Cleavages present in the free RNA species but absent from the ones in the complex indicate the positions at which modification prevents complex formation (negative interference). Consequently, the free RNA fraction is enriched for RNA species that carry modifications that interfere with complex formation, and thus the cleavages corresponding to the positions of negative interference are enhanced (Fig. 2, lanes F).

Strong inhibition of wild-type CopA–CopT complex formation was observed upon ethylation of phosphates 112–114 in CopT, and 58–59 in CopA (Fig. 2A and B). Interestingly, the strongest interference pinpointed a subset of nucleotides in both CopT and CopA loops (Fig. 2D) in which many copy number mutations had been mapped [5′-GGCG in CopT; e.g. (30)]. In addition, ethylation of phosphates 111, 115–120 in CopT, and phosphates 57 and 60 in CopA, resulted in partial interference. Identical results were obtained with the truncated antisense RNA, CopI (result not shown). The mutant complex formed between CopI-L/U and CopT-L/U was also analyzed (Fig. 2C). Compared to the wild-type case, a single and extended region of strong negative interference was observed, comprising phosphates 57–62 within the loop of CopI-L/U (Fig. 2D). In addition, partial interference was observed upon modification of phosphates 54–56 in the 5′ part of the CopI-L/U stem. Given the number of modifications that prevent complex

3148 Nucleic Acids Research, 2001, Vol. 29, No. 15

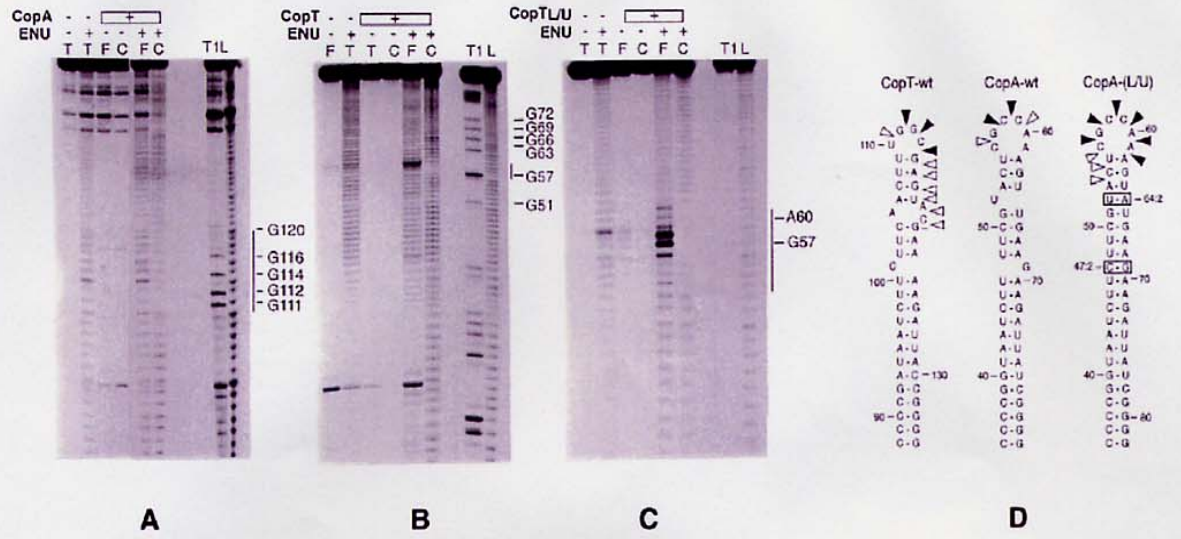

Figure 2. Effect of phosphate ethylation on complex formation. 5′ End-labeled CopT (**A**), CopA (**B**) and CopI-L/U (**C**) were modified with ENU and submitted to complex formation with either CopA (A), CopT (B) and CopT-L/U (C). F and C refer to RNA extracted from the bands containing the free RNA and the RNA–RNA complex, respectively; T corresponds to the total population of unmodified and ENU-modified CopA used for the interference experiment. (+) and (–) RNA are treated or untreated with ENU, respectively. Lanes T1 and L correspond to RNase T1 and alkaline ladders. Negative interference is indicated by vertical bars on the autoradiograms. In (B), lanes F, the cleavage at position U27 found in the free RNA population is not specific and may originate from a fortuitous RNase contamination. (**D**) Secondary structures of the RNAs showing the position of modified phosphates which interfere with complex formation. Filled and empty triangles indicate strong and weak interference, respectively. For clarity, the numbering of nucleotides in CopI-L/U was identical to that of wild-type CopI, except that the additional nucleotides were assigned as positions 47:2 and 62:2. Mutations are boxed.

formation, this negative interference may also reflect local perturbation of the CopA/I and CopT loop structures.

These experiments indicate that, in both the wild-type and L/U mutant cases, modification of phosphates within the recognition loops interfered with complex formation. This suggests that removal of bulges did not impair formation of the initial loop–loop interaction. However, the differences in the interference patterns observed between the wild-type and mutant complex suggest that these two complexes adopt different topologies.

Removal of the bulged residues prevents the formation of the intermolecular helices

We previously showed that the initial loop–loop interaction is converted to an unusual four-way junction structure (13). The initial base pairs formed are disrupted to permit helix progression into the upper stems, resulting in the formation of two intermolecular helices B and B′ (Fig. 1C). RNase V1 probing of CopA–CopT or CopI–CopT complexes provided a characteristic signature for helices B and B′ (Fig. 3A) (13): new or enhanced cleavages were induced in CopA or CopI at consecutive positions U52–C56 (helix B), and in CopT at C104–A105 (helix B′). In addition, the intermolecular helix C, which greatly enhances CopA–CopT complex stability, was characterized by the occurrence of several V1 cuts in the 3′ end of CopT, as well as by the disappearance of V1 cleavages in CopA at positions G7–A9 and G16–A19, resulting from melting of helix I (Fig. 3A). Thus, RNase V1 was used to probe the structure of the homologous antisense/target RNA complexes containing mutations of either the lower bulge (Lo), upper bulge (Up) or both (L/U) (Fig. 3B–D), as well as the heterologous complexes formed by wild-type CopA and mutant CopT RNAs (results not shown).

None of the mutant complexes presented the characteristic RNase V1 pattern observed for the wild-type complex. In the three homologous mutant complexes (CopA-Up/CopT-Up, CopA-Lo/CopT-Lo, CopA-L/U/CopT-L/U), helix C did not form properly. This is indicated by cleavages at positions 7–9 and 17–19 in CopA-Lo and CopA-L/U bound to their target CopT, which remained unchanged as compared to free mutant CopA (Fig. 3B and D), whereas CopT-Up induced only weak protections at positions 17–19 in CopA-Up (Fig. 3C). Several new cuts were detected in the hairpin loops of CopA-Up (C54–C56), of CopA-Lo (U55 and C56) and of CopA-L/U (U55–C56, A60 and A61) upon binding of CopT-Up, CopT-Lo and CopT-L/U, respectively (Fig. 3). In addition, enhanced cleavages were observed on the 5′ side of the major stem of CopA-Lo and CopA-L/U. Conversely, in all three homologous mutant complexes, binding of CopA variant did not change the pattern of RNase V1 cleavage in the corresponding CopT target. In particular, no enhanced cleavages at C104–A105 were detected in mutant CopT, indicating that helix B′ is not formed in the mutant complexes (results not shown).

Thus, replacing the bulged residues by Watson–Crick base pairs in the antisense and target RNAs induced dramatic changes in the overall topology of the CopA–CopT complex: the two intermolecular helices B and B′ were not formed, and thus formation of the third intermolecular helix C could not occur. These data are in good agreement with the dramatic effects of these three mutations on the rate of stable CopA–CopT complex formation (21).

Nucleic Acids Research, 2001, Vol. 29, No. 15 **3149**

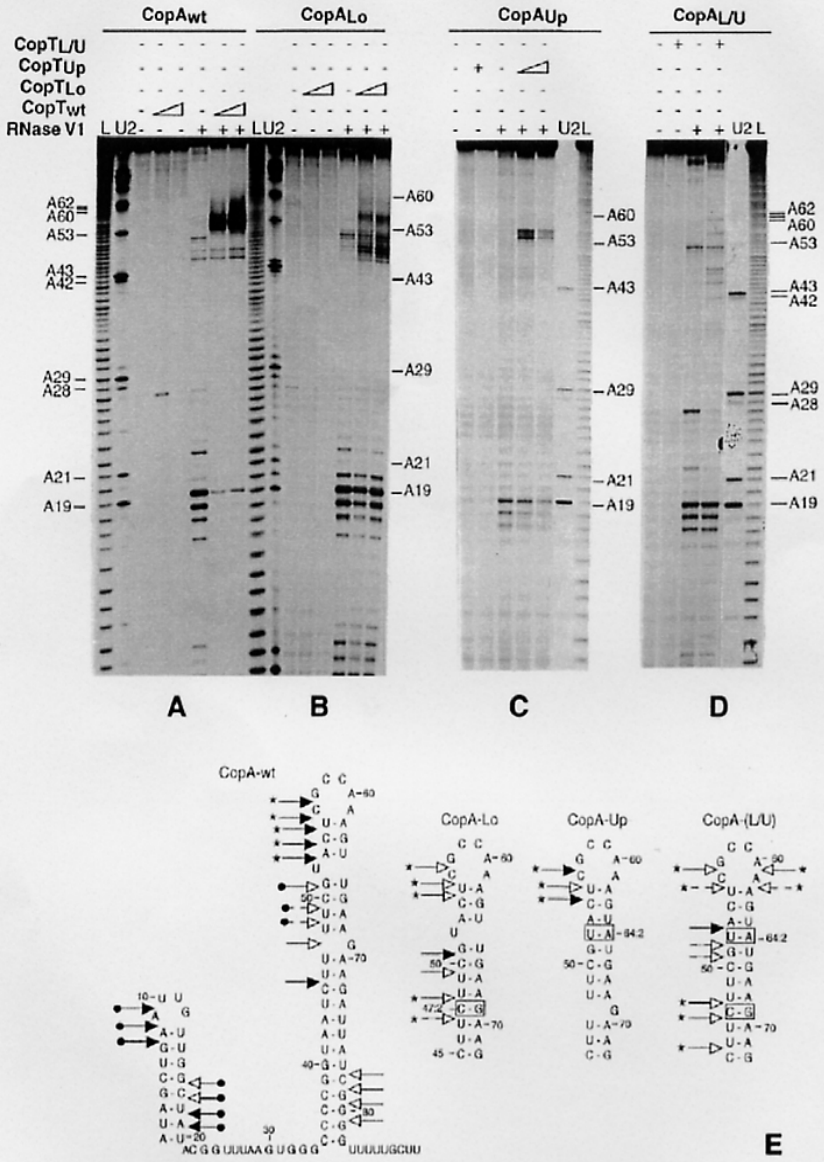

Figure 3. RNase V1 probing of homologous wild-type and mutated CopA–CopT complexes. Enzymatic hydrolysis was performed on 5′ end-labeled CopA-wt (**A**), CopA-Lo (**B**), CopA-Up (**C**) and CopA-L/U (**D**). Reactions were done in the absence or in the presence of an excess of wild-type or mutant CopT. Lanes U2 and L are RNase U2 and alkaline ladders, respectively. (**E**) Cleavages are shown on the secondary structure of the wild-type and mutant CopA by arrows. Only part of the secondary structure of the three mutant CopA, where changes induced by complementary CopT binding occurred, are shown. Effect of homologous CopT binding is as follows: asterisks, enhanced or new cut; filled circles, protection.

A fully base paired loop–loop interaction in the mutant CopA–CopT complexes

To obtain more detailed information on the accessibility of nucleotides within the CopA-L/U/CopT-L/U complex, NiCR-induced hydrolysis was used (13). Lead-induced cleavages occur mainly in single-stranded regions and in helices of low stability. NiCR modifies [N7]G and is very sensitive to stacking of base rings. Thus, N7 of purines within a helix are not reactive unless the deep groove is widened (31).

We showed previously that Pb^{2+}-cleavages in the CopA–CopT complex were restricted to unpaired residues located in the loops (C58–A61) connecting the two intermolecular helices B and B′ (13). In free CopA-L/U RNA, lead cleaved mainly in the hairpin loop and in the 5′ single-stranded region of the RNA (Fig. 4) (21). Binding of either CopT-L/U or CopT-wt induced strong protections at residues U55–A61 in the apical loop whereas minor protections were detected in the 5′ domain of CopA-L/U (Fig. 4A). This is supported by the NiCR probing experiment (Fig. 4B). In the major stem–loop of CopA-L/U, only the N7 position of G57 was modified whereas all guanines within the stem were unreactive. Notably, G69 which is base paired in CopA-L/U was not modified, whereas

3150 Nucleic Acids Research, 2001, Vol. 29, No. 15

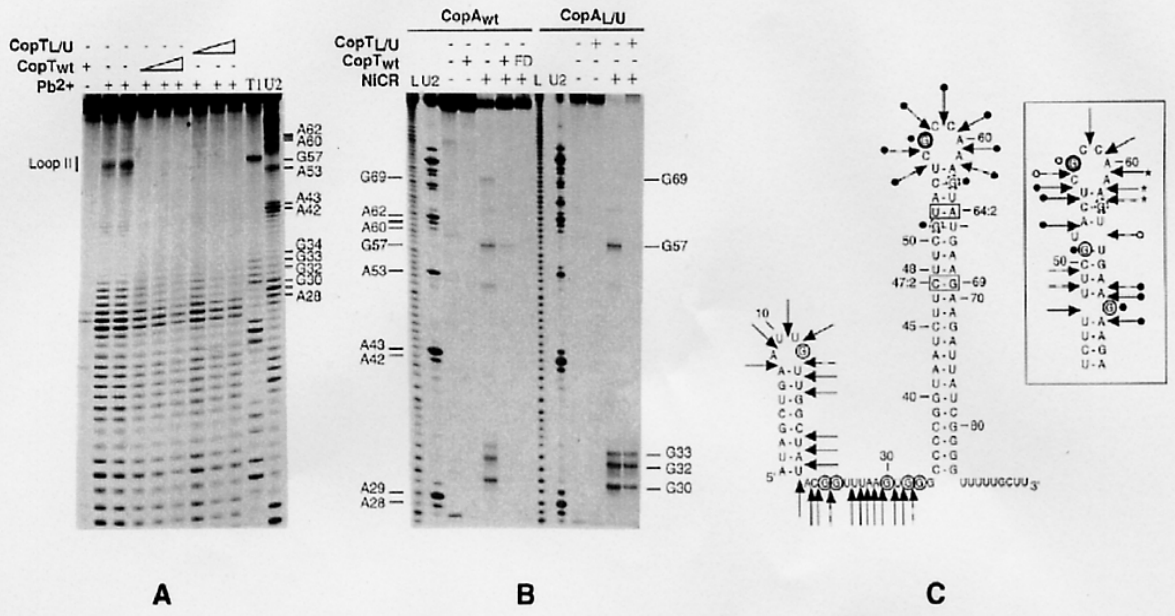

Figure 4. Chemical probing of CopA-L/U, free or in complex with CopT-L/U. (**A**) Pb^{2+}-induced hydrolysis was performed on 5′ end-labeled CopA-L/U alone (–) or in the presence of an excess of CopT-L/U or CopT-wt (+). Complex formation was performed at 37°C for 1, 5 and 10 min in TMN buffer. Lanes L, U2 and T1 are alkaline, RNase U2 and RNase T1 ladders, respectively. (**B**) NiCR modification on 5′ end-labeled CopA-wt or CopA-L/U: free (–CopT), bound to wild-type or mutant CopT under native conditions (+CopT/CopT-L/U). Full duplex with CopA (FD) was formed by denaturation/annealing treatment. (**C**) Secondary structure of CopA-L/U: G(N7) which are reactive under native conditions towards NiCR are circled, the line width of the symbols is proportional to the intensity of cleavage (strong, medium or weak). Pb^{2+}-cleavages are shown by arrows, dotted and full lines represent weak and moderate cleavages, respectively. Effect of CopT-L/U binding: strong and medium protections are shown by filled and empty circles, respectively; asterisks show enhanced cleavages. For comparison, the data obtained on CopA-wt bound to CopT-wt is shown in the insert.

it was reactive in CopA-wt (Fig. 4B). The only change induced by binding of CopT-L/U was restricted to the complete protection of G57 at N7 in CopA-L/U (Fig. 4B). In agreement with results from the RNase V1 probing, the stabilizer helix C was not formed in the CopA-L/U/CopT-L/U complex since position N7 of guanines 12, 23–24, 30, 32–33 in the 5′ domain of CopA-L/U were still reactive in the complex (Fig. 4B). Binding of CopT-Up also induced protection restricted to G57 at N7 in CopA-Up whereas G69 at N7 remained reactive towards NiCR (data not shown). By contrast, most of the N7 positions of guanines, with the notable exception of G57, showed protection towards NiCR in CopA-wt when bound to CopT-wt (Fig. 4B). In full duplexes (formed artificially), all the guanines at N7 of CopA-wt were fully protected by binding of CopT (Fig. 4B). Thus, these experiments allow us to distinguish between the full duplex, the extended kissing in CopA–CopT complex, and the loop–loop interaction in CopA-L/U/CopT-L/U.

Altogether these data indicate that removal of the bulged residues within the major stems of CopA and CopT resulted in a complex characterized by a fully base paired loop–loop helix whose formation was incompatible with subsequent formation of the stabilizer helix C.

A loop–loop interaction is insufficient to prevent ribosome binding

CopA-mediated inhibition works by steric interference: the CopA–CopT complex prevents binding of ribosomes at the *tap* translation initiation site (5). Even a truncated CopA, resembling CopI, inhibited RepA synthesis *in vivo*, although at somewhat reduced efficiency (32). CopI, although unable to form a stable complex with CopT, forms the four-way junction structure (8,13), and was capable of interfering transiently with ribosome binding *in vitro* (5). Since removal of the bulged residues inhibited the activity of the mutant antisense RNAs *in vivo* (21), we tested here whether the loop–loop interaction in the three mutant CopI–CopT complexes could interfere with ribosome binding.

Toeprinting analysis was used to monitor the formation of translation initiation complexes on mRNAs *in vitro* (33). Previous experiments showed that CopI-dependent inhibition of ribosome binding at the *tap* initiation site was only detectable after a short incubation time. Therefore, complexes between CopI and CopT carrying either mutation Lo, Up or L/U were pre-formed, at RNA concentrations allowing essentially quantitative formation of the kissing complex during pre-incubation. Thereafter, the ternary mRNA/30S/$tRNA_f^{Met}$ complex was formed by adding 30S subunits and $tRNA_f^{Met}$. Ternary complex formation was quantified over time by monitoring the signal of the *tap* toeprint (Fig. 5). For reference, the same experiments were carried out with the wild-type CopI/CopT and CopA/CopT complexes, respectively. As shown previously, inhibition by CopA is irreversible within the time interval assayed, whereas CopI-dependent inhibition was progressively lost with 50% inhibition after 2 min incubation (Fig. 5) (5). In contrast, the three homologous mutant CopI–CopT complexes were completely inactive in inhibition of initiation complex formation at the *tap* RBS (Fig. 5).

Nucleic Acids Research, 2001, Vol. 29, No. 15 **3151**

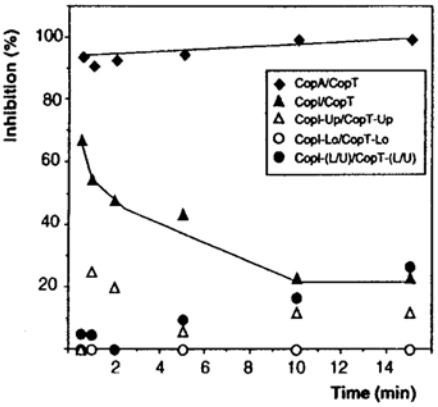

Figure 5. Kinetic analysis of antisense RNA-dependent inhibition of *tap* toeprints. Quantitations were done by Bioimager (BAS2000 Fuji) analysis of dried gels. Relative *tap* toeprint values were first calculated by relating the intensity of the toeprint to that of the run-off extension product. The yield of inhibition (%) was given by the signal of the relative *tap* toeprint in the presence of the antisense RNA divided by that of the relative toeprint in the absence of antisense RNA at a defined time. The different wild-type or mutated complexes are given in the insert. The values represent an average of three (wild-type) and two (mutant) independent experiments.

DISCUSSION

A common characteristic feature of most, if not all, prokaryotic antisense and target RNAs is the presence of well-defined stem–loop(s) as key structure elements. In many antisense regulatory systems, RNAs initiate binding by transient and reversible interactions between complementary loops and, depending on their particular topological constraints, use different pathways to proceed rapidly to more stable and inhibitory complexes (e.g. 9,11,34,35). Formation of sufficiently stable complexes is important to obtain maximal inhibition rates, and to achieve quasi-irreversible inhibition kinetics.

For CopA/CopT of plasmid R1, we have recently shown that the initial loop–loop interaction becomes disrupted when intermolecular base pairing is propagated into the upper stems, creating a four-way junction topology (Fig. 1C) (13). The same sequence of events is also supported for the truncated antisense RNA (CopI) which only contains the recognition stem–loop structure (Fig. 1). The four-way junction topology is essential to promote the formation of the additional intermolecular helix between the single-stranded regions of the two RNAs that results in the stable CopA–CopT complex (14). The question addressed here concerns the driving force for conversion of the initial loop–loop to the extended kissing complex. Both CopA and CopT carry, in the upper part of the recognition stem, several unpaired nucleotides. Replacement of these residues by Watson–Crick base pairs decreased binding rate *in vitro* and control *in vivo* (21). Given the magnitude of impairment, it could not be excluded that these mutant RNAs followed different binding pathways which did not initiate by a loop–loop interaction but instead by contacts within the 5′ single-stranded region of CopA. The experiments presented here indicate that removal of the bulged residues does not prevent the initial formation of a loop–loop helix, but abolishes the subsequent transition towards the four-way junction. Structure probing of the three homologous mutant RNA complexes indicates that helices B and B′ do not form (Figs 3 and 4). Instead, the data suggest that an intermolecular helix is formed between all 6 nucleotides in the complementary loop sequences, and that stem sequences do not participate in helix propagation. Interestingly, an extensively paired loop–loop helix was proposed for complexes between stem–loops of RNAI and RNAII of ColE1 (19,36). RNAI regulates ColE1 plasmid replication by inhibiting maturation of the pre-primer, RNAII. Binding is initiated by at least two loop–loop interactions followed by a series of reactions that progressively lead to the formation of a stable duplex (34,37). The NMR structures of two kissing complexes (6 and 7 bp loop–loop base pairs) have been solved (38–40). Both structures show full base pairing between the loops, and are characterized by a pronounced bend at the loop–loop helix towards the major groove. Interestingly, ethylation of any phosphate within the 6-nucleotide loop sequence of CopI-L/U prevents nucleation and initiation events for binding to CopT-L/U (Fig. 2). The dimerization of HIV-1 RNA which involves a loop–loop interaction between 6 complementary nucleotides in a 9-membered loop, is also inhibited by the ethylation of most of the phosphates of the loop nucleotides (29). However, in the wild-type CopA/CopT complex, where the initial interactions are transient, interference was restricted to the 5′ phosphates of the GCCA sequence of CopA and CopI, as well as to the complementary sequence of CopT. Weak interference observed at phosphates 114–120 of CopT implicates these residues in formation of helix B which is an early step required to commit the interacting RNAs to stable complex formation (14). Taken together, our study strongly argues for a striking difference in topology between complexes formed by the bulge-less mutant and wild-type RNAs.

In contrast to the ColE1 case where extensive loop–loop helices are involved in the inhibitory pathway, the loop–loop helix in the context of CopA-L/U/CopT-L/U is not productive since the long intermolecular stabilizer helix C cannot form (Fig. 3) (21). The suggested overall topology of the loop–loop helix tentatively explains this defect. Based on the known NMR loop–loop structures, a tertiary model for the CopA-L/U/CopT-L/U complex was built by graphic modeling, taking into account stereochemical constraints and probing data (Fig. 6). Despite the presence of a pronounced bend at the loop–loop helix, the two complementary regions that normally interact to give the stabilizer helix C are too far apart in space for efficient pairing (>100 Å; Fig. 6). In the ColE1 plasmid, the stem–loop structures of RNAI/II are shorter than in CopA/T, and bulged residues comparable to those in CopA/T are lacking (19). Also, bending at the loop–loop helix is further enhanced by Rom protein, most likely facilitating the distal contact between the 5′ tail of RNAI and its complement (38) which initiates full pairing. In plasmid R1 and related plasmids, no protein is known to play such a role. Instead, all these plasmids carry unpaired residues or bulged loops in the upper stem regions of antisense and target RNAs although their positions are not strictly conserved (15). These elements are essential for rapid *in vitro* binding and *in vivo* control [R1 (21); pMU720 (22); Col1b-P9 (23)]. In all these systems, the rapid conversion of the initial loop contacts to the extended kissing complexes is crucial for efficient inhibition. This step requires substantial melting of the upper stem regions of both antisense and target

3152 Nucleic Acids Research, 2001, Vol. 29, No. 15

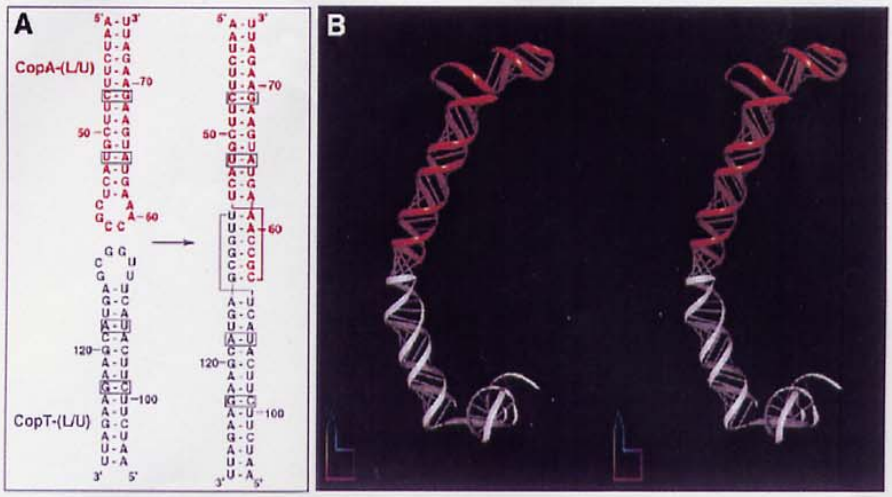

Figure 6. Models of the CopA-L/U/CopT-L/U complex. (**A**) Secondary structure representation and (**B**) stereoviews of the three-dimensional model of the CopA-L/U/CopT-L/U complex. Nucleotides in CopA-L/U [G1–U87] are shown in red, and in CopT-L/U [C88–A178] in white. The models were drawn with the program DRAWNA (28).

RNAs (9,11,13). We show here that, in the CopA/CopT system, this step is driven via the bulged residues; unpaired residues in the upper stems favor inter-strand helix progression over imperfect intra-strand base pairing. Bulged residues in long hairpin structures were also shown to be essential in other antisense regulated systems which initiate antisense/target binding via loop–linear RNA pairing (41–43).

Previous work showed that CopA inhibits the synthesis of RepA indirectly. Translation of a leader peptide, Tap, encoded immediately downstream of the CopA target site and in front of the *repA* reading frame is required for *repA* translation (4,44). CopA prevents ribosome binding at the *tap* ribosome loading site (5). Even CopI, although it only forms the four-way junction structure, transiently interferes with ribosome binding *in vitro* (Fig. 5) (5) and inhibits *repA* expression *in vivo* albeit at lower efficiency (32). Here we show that the CopI–CopT complexes carrying mutations Lo, Up and L/U do not inhibit ribosome binding at the *tap* RBS (Fig. 5). Since previous work has ruled out distal structure rearrangements in CopT as the mode of inhibition (12,13,45), the likely difference between inhibition (wild-type CopI–CopT) and lack of inhibition (bulge mutant CopI–CopT) should be due to their different structures. Even though CopI binds far upstream of the *tap* RBS, the bulky topology of the wild-type complex can sterically interfere with initiation, at least transiently. When full CopA is tested, the stabilizer helix C then renders inhibition irreversible. The suggested structure for the bulge mutant CopI–CopT complex (Fig. 6B) might be oriented so that ribosome access fails to be blocked. However, detailed structural analysis of the four-way junction topology will be required to elucidate the molecular mechanism by which a distal kissing complex can interfere with ribosome binding.

Intermolecular interactions between RNAs play fundamental roles in gene expression. Most of these are based on sequence complementarity, and loop–loop interactions are often implicated. *In vitro* selection approaches designed to select aptamers directed against RNA predominantly resulted in the isolation of sequences complementary to a hairpin loop in a defined structural context (46,47). It appears that, depending on the biological context of the RNA–RNA interaction, RNA structures have evolved either to 'freeze' a complex once formed or, alternatively, to convert initial interactions by propagating helices along topologically feasible pathways (reviewed in 48). In antisense regulation, formation of stable RNA–RNA complexes requires a propagation of the initial loop–loop helix to proceed rapidly to sufficiently stable and inhibitory complexes. The features that govern the high efficiency of natural antisense RNA are thus related to the secondary and tertiary structures of the interacting partners as well as of the complexes formed.

SUPPLEMENTARY MATERIAL

Coordinates of the molecular model of CopA-L/U/CopT-L/U complex are available as supplementary material at NAR Online.

ACKNOWLEDGEMENTS

We thank Hervé Moine for stimulating discussions and critical reading of the manuscript and G. Bec for advice. This work was supported by grants from the Centre National de la Recherche Scientifique (CNRS), and by the Swedish Natural Science Research Council (NFR, E.G.H.W.) and the Swedish Research Council for Engineering Sciences (TFR, E.G.H.W.).

REFERENCES

1. Wagner,E.G.H. and Simons,R.W. (1994) Antisense RNA control in bacteria, phages and plasmids. *Annu. Rev. Microbiol.*, **48**, 713–742.
2. Nordström,K., Molin,S. and Light,J. (1984) Control of replication of bacterial plasmids: genetics, molecular biology and physiology of the plasmid R1 system. *Plasmid*, **12**, 71–90.
3. Blomberg,P., Wagner,E.G.H. and Nordström,K. (1990) Control of replication of plamsid R1: the duplex between the antisense RNA, CopA and its target, CopT, is processed specifically *in vivo* and *in vitro* by RNase III. *EMBO J.*, **9**, 2331–2340.

4. Blomberg,P., Nordström,K. and Wagner,E.G.H. (1992) Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. *EMBO J.*, **11**, 2675–2683.
5. Malmgren,C., Engdhal,H.M., Romby,P. and Wagner,E.G.H. (1996) An antisense/target RNA duplex or a strong intramolecular RNA structure 5′ of a translation initiation signal blocks ribosome binding: the case of plasmid R1. *RNA*, **2**, 1022–1032.
6. Wagner,E.G.H. and Brantl,S. (1998) Kissing and RNA stability in antisense control of plasmid replication. *Trends Biochem. Sci.*, **23**, 451–454.
7. Zeiler,B.N. and Simons,R.W. (1998) Antisense RNA structure and function. In Simons,R.W. and Grunberg-Manago,M. (eds), *RNA Structure and Function*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 437–464.
8. Persson,C., Wagner,E. and Nordström,K. (1988) Control of replication of plasmid R1: kinetics of *in vitro* interaction between the antisense RNA, CopA and its target, CopT. *EMBO J.*, **7**, 3279–3288.
9. Siemering,K.R., Praszkier,J. and Pittard,A.J. (1994) Mechanism of binding of the antisense and target RNAs involved in the regulation of IncB plasmid replication. *J. Bacteriol.*, **176**, 2677–2688.
10. Asano,K., Niimi,T., Yokoyama,S. and Mizobuchi,K. (1998) Structural basis for binding of the plasmid ColIb-P9 antisense Inc RNA to its target RNA with the 5′-rUUGGCG-3′ motif in the loop sequence. *J. Biol. Chem.*, **273**, 11826–11838.
11. Asano,K. and Mizobuchi,K. (2000) Structural analysis of late intermediate complex formed between plasmid ColIb-P9 Inc RNA and its target RNA. How does a single antisense RNA repress translation of two genes at different rates? *J. Biol. Chem.*, **275**, 1269–1274.
12. Malmgren,C., Wagner,E.G.H., Ehresmann,C., Ehresmann,B. and Romby,P. (1997) Antisense RNA control of plasmid R1 replication. The dominant product of the antisense RNA-mRNA binding is not a full RNA duplex. *J. Biol. Chem.*, **272**, 12508–12512.
13. Kolb,F.A., Malmgren,C., Westhof,E., Ehresmann,C., Ehresmann,B., Wagner,E.G. and Romby,P. (2000) An unusual structure formed by antisense-target RNA binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment. *RNA*, **6**, 311–324.
14. Kolb,F.A., Engdahl,H.M., Slagter-Jager,J.G., Ehresmann,B., Ehresmann,C., Westhof,E., Wagner,E.G. and Romby,P. (2000) Progression of a loop–loop complex to a four-way junction is crucial for the activity of a regulatory antisense RNA. *EMBO J.*, **19**, 5905–5915.
15. Kolb,F.A., Westhof,E., Ehresmann,B., Ehresmann,C., Wagner,E.G.H. and Romby,P. (2001) Four-way junctions in antisense RNA-mRNA complexes involved in plasmid replication control: a common theme? *J. Mol. Biol.*, **309**, 605–614.
16. Wilson,I.W., Praszkier,J. and Pittard,A.J. (1993) Mutations affecting pseudoknot control of the replication of IncB group plasmids. *J. Bacteriol.*, **175**, 6476–6483.
17. Asano,K., Hama,C., Shin-ichi,I., Moriwaki,H. and Mizobuchi,K. (1999) The plasmid ColIb-P9 antisense Inc RNA controls expression of the RepZ replication protein and its positive regulator *repY* with different mechanisms. *J. Biol. Chem.*, **274**, 17924–17933.
18. Tomizawa,J. (1984) Control of ColE1 plasmid replication: the process of binding of RNAI and the primer transcript. *Cell*, **38**, 861–870.
19. Tomizawa,J. (1993) Evolution of functional structures of RNA. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 419–445.
20. Hjalt,T.A. and Wagner,E.G.H. (1995) Bulged-out nucleotides protect an antisense RNA from RNase III cleavage. *Nucleic Acids Res.*, **23**, 571–579.
21. Hjalt,T.A. and Wagner,E.G. (1995) Bulged-out nucleotides in an antisense RNA are required for rapid target RNA binding *in vitro* and inhibition *in vivo*. *Nucleic Acids Res.*, **23**, 580–587.
22. Siemering,K.R., Praszkier,J. and Pittard,A.J. (1993) Interaction between the antisense and target RNAs involved in the regulation of IncB plasmid replication. *J. Bacteriol.*, **175**, 2895–2906.
23. Asano,K., Kato,A., Moriwaki,H., Hama,C., Shiba,K. and Mizobuchi,K. (1991) Positive and negative regulations of plasmid ColIb-P9 *repZ* gene expression at the translational level. *J. Biol. Chem.*, **266**, 3774–3781.
24. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Donis-Keller,H., Maxam,A. and Gilbert,W. (1977) Mapping adenines, guanines and pyrimidines in RNA. *Nucleic Acids Res.*, **4**, 2527–2538.
26. Westhof,E. (1993) Modeling the three-dimensional structure of ribonucleic acids. *J. Mol. Struct.*, **286**, 203–211.
27. Massire,C. and Westhof,E. (1998) MANIP: an interactive tool for modelling RNA. *J. Mol. Graph. Model*, **16**, 197–205, 255–257.
28. Massire,C., Gaspin,C. and Westhof,E. (1994) A program for drawing schematic views of nucleic acids. *J. Mol. Graph. Model*, **12**, 201–206.
29. Jossinet,F., Paillart,J.C., Westhof,E., Hermann,T., Skripkin,E., Lodmell,S.J., Ehresmann,C., Ehresmann,B. and Marquet,R. (1999) Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. *RNA*, **5**, 1222–1234.
30. Givskov,M. and Molin,S. (1984) Copy mutants of plasmid R1: effects of base pair substitutions in the *copA* gene on the replication control system. *Mol. Gen. Genet.*, **194**, 286–292.
31. Chen,X., Woodson,S.A., Burrows,C.J. and Rokita,S.E. (1993) A highly sensitive probe for guanine N7 in folded structures of RNA: application to tRNA-Phe and Tetrahymena group I intron. *Biochemistry*, **32**, 7610–7616.
32. Wagner,E.G.H., Blomberg,P. and Nordström,K. (1992) Replication control in plasmid R1: duplex formation between the antisense RNA, CopA and its target, CopT, is not required for inhibition of RepA synthesis. *EMBO J.*, **11**, 1195–1203.
33. Hartz,D., McPheeters,D.S., Traut,R. and Gold,L. (1988) Extension inhibition analysis of translation initiation complexes. *Methods Enzymol.*, **164**, 419–425.
34. Tomizawa,J. (1990) Control of ColE1 plasmid replication: intermediates in the binding of RNAI and RNAII. *J. Mol. Biol.*, **212**, 683–694.
35. Persson,C., Wagner,E.G. and Nordstrom,K. (1990) Control of replication of plasmid R1: formation of an initial transient complex is rate-limiting for antisense RNA-target RNA pairing. *EMBO J.*, **9**, 3767–3775.
36. Eguchi,Y. and Tomizawa,J. (1991) Complexes formed by complementary RNA stem-loops: their formation, structures and interaction with ColE1 Rom protein. *J. Mol. Biol.*, **220**, 831–842.
37. Eguchi,Y. and Tomizawa,J. (1990) Complex formed by complementary RNA stem-loops and its stabilization by a protein: function of ColE1 Rom protein. *Cell*, **60**, 199–209.
38. Marino,J.P., Gregorian,R.S., Csankovszki,G. and Crothers,D.M. (1995) Bent helix formation between RNA hairpins with complementary loops. *Science*, **268**, 1448–1454.
39. Chang,K. and Tinoco,I.,Jr (1997) The structure of an RNA kissing hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.*, **269**, 52–66.
40. Comolli,L.R., Pelton,J.G. and Tinoco,I.,Jr (1998) Mapping of a protein-RNA kissing hairpin interface: Rom and Tar-Tar*. *Nucleic Acids Res.*, **26**, 4688–5502.
41. Thisted,T., Sorensen,N.S., Wagner,E.G. and Gerdes,K. (1994) Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5′-end single-stranded leader and competes with the 3′-end of Hok mRNA for binding to the mok translational initiation region. *EMBO J.*, **13**, 1960–1968.
42. Kittle,J.D., Simons,R.W., Lee,J. and Kleckner,N. (1989) Insertion sequence IS10 antisense pairing initiates by an interaction between the 5′ end of the target RNA and a loop in the antisense RNA. *J. Mol. Biol.*, **210**, 561–572.
43. Jain,C. (1995) IS10 antisense control *in vivo* is affected by mutations throughout the region of complementarity between the interacting RNAs. *J. Mol. Biol.*, **246**, 585–594.
44. Blomberg,P., Engdahl,H.M., Malmgren,C., Romby,P. and Wagner,E.G.H. (1994) Replication control of plasmid R1: disruption of an inhibitory RNA structure that sequesters the *repA* ribosome binding site permits tap-independent RepA synthesis. *Mol. Microbiol.*, **12**, 49–60.
45. Öhman,M. and Wagner,E.G.H. (1989) Secondary structure analysis of the RepA mRNA leader transcript involved in control of replication of plasmid R1. *Nucleic Acids Res.*, **17**, 2557–2579.
46. Scarabino,D., Crisari,A., Lorenzini,S., Williams,K. and Tochinni-Valentini,G.P. (1999) tRNA prefers to kiss. *EMBO J.*, **18**, 4571–4578.
47. Duconge,F., Di Primo,C. and Toulme,J. (2000) Is a closing "GA pair" a rule for stable loop–loop RNA complexes? *J. Biol. Chem.*, **275**, 21287–21294.
48. Grosjean,H., Houssier,C., Romby,P. and Marquet,R. (1998) Modulation role of modified nucleotides on RNA-RNA interaction. In Grosjean,H. and Benne,R. (eds), *Modification and Editing of RNA: The Alteration of RNA Structure and Function*. ASM press, WA, pp. 113–133.

III. Article III : Etude mécanistique de la formation du complexe irréversible CopA-CopT

Dans plusieurs systèmes antisens, une interaction boucle-boucle garantit la spécificité de l'interaction ARN-ARN. Pour l'efficacité du contrôle, celle-ci doit cependant être stabilisée, soit par des interactions additionnelles, soit grâce à l'intervention d'une protéine. Dans le système CopA-CopT, la formation d'un complexe stable et inhibiteur nécessite une conversion rapide du kissing-complex initial vers une jonction à quatre hélices (Article I : [335]). Cette conversion est facilitée par la présence de nucléotides non appariés dans chacun des deux ARN (Article II : [336]).

A. Objectifs

Le but de la présente étude est de démontrer l'importance de l'étape de conversion pour la rapidité de la formation du complexe CopA-CopT stable *in vitro* et pour l'efficacité du contrôle *in vivo*. Nous avons disséqué précisément le mécanisme de formation du complexe stable en caractérisant de nouveaux intermédiaires réactionnels entre le « kissing-complex » initial et la topologie de la jonction à quatre hélices.

B. Stratégie expérimentale

Dans le premier paragraphe (Article I) nous avons montré que les mutations H1 et H2 interféraient avec la formation de la jonction à quatre hélices entre deux ARN hétérologues. Cet effet délétère est dû à l'absence de complémentarité des deux ARN au niveau des positions mutées. Elles empêchent par conséquent la progression de l'interaction initiale vers les hélices B et B'. Ici, nous avons approfondi l'étude des mutants H1, H2 et H3 en analysant leur effet sur la vitesse de formation du complexe stable *in vitro*. Des mutants empêchant la progression de l'interaction boucle-boucle dans une direction précise ont également été construits et analysés (Figure 31 : mutants L1 et L2). Enfin, des mutations qui déstabilisent sélectivement l'hélice B (mutants H5 et H6) ou l'hélice B' (mutant H4), mais sans affecter la structure des ARN seuls, ont été construites. Parallèlement à des études cinétiques et de cartographie en solution *in vitro* sur différents complexes homologues ou hétérologues, nous avons également analysé l'effet des mutations sur l'efficacité du contrôle *in vivo*. Dans ce but, des fusions traductionnelles ont été construites entre la partie 5'-UTR de *repA*, contenant une séquence CopT sauvage ou mutée, et le gène *lacZ*. Les effets des mutations sur le contrôle ont été estimés en mesurant l'activité β-galactosidase des bactéries co-exprimant des ARN CopA et CopT homologues ou hétérologues.

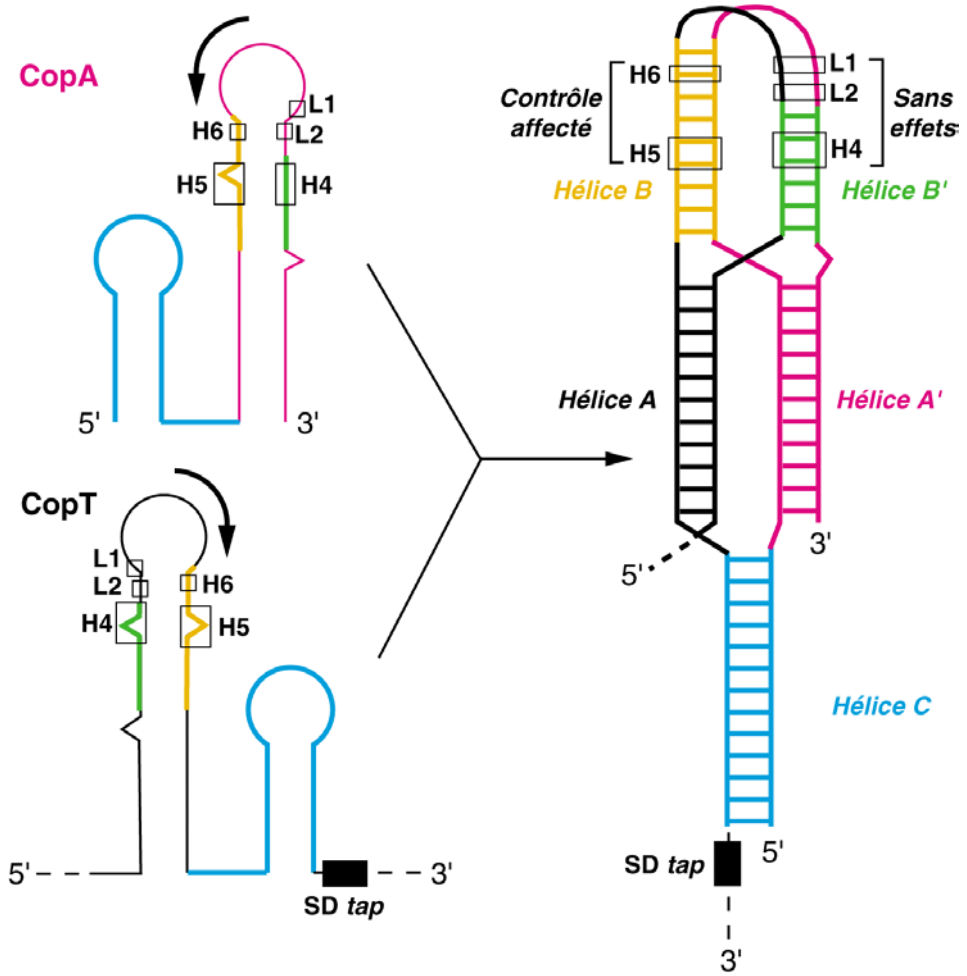

Figure 31 : ***Modèle bidimensionnel du complexe stable formé entre CopA et CopT* in vitro *et mutations introduites dans cette étude***. Les régions non impliquées dans des appariements intermoléculaires sont indiquées en rose pour CopA et en noir pour CopT. Celles impliquées dans la formation des hélices B, B' et C sont indiquées respectivement en jaune, vert et bleu,. Les mutations introduites dans les deux ARN au cours de cette étude sont indiquées (L1, L2, H4, H5 et H6). La séquence Shine et Dalgarno de *tap* est schématisée par une boîte noire. Le sens de progression de l'interaction boucle-boucle initiale est indiquée par une flèche au-dessus des boucles apicales de chacun des deux ARN.

C. Résultats

Tous les résultats obtenus *in vitro* et *in vivo* concordent, suggérant que le mécanisme de formation du complexe stable obéit aux mêmes règles dans les deux cas. Dès lors que la conversion est abolie dans les couples hétérologues H1-wt ou H2-wt, la formation d'un complexe stable est ralentie et le contrôle est moins efficace. De manière intéressante, nous montrons également que la disruption de l'hélice B par l'introduction des mutations H5 et H6 est délétère tant *in vitro* qu'*in vivo*. En revanche, la formation de l'hélice B' n'est pas essentielle mais dépend de celle de l'hélice B. Ces résultats montrent que la conversion de l'interaction boucle-boucle initiale est unidirectionnelle. Elle progresse du coté 5' de la boucle de CopA et 3' de la boucle de CopT formant alors l'hélice B. Ceci est renforcé par l'observation que les deux mutants L1 et L2 empêchant la progression dans le sens opposé

sont sans effets, tant *in vitro* qu'*in vivo*. Ces résultats ont finalement été validés par des expériences de compétition. Différents ARN complémentaires à CopT ont été testés sur leur habilité à interférer avec la formation d'un complexe CopA-CopT stable *in vitro*. Seuls ceux empêchant la propagation du coté 3' de la boucle de l'ARNm sont inhibiteurs.

D. Conclusions

Cette étude révèle que l'efficacité de l'interaction entre CopA et CopT *in vitro* et du contrôle *in vivo* repose sur deux étapes essentielles. L'interaction boucle-boucle est nécessaire à la reconnaissance initiale des deux ARN mais elle doit rapidement progresser vers la formation de l'hélice intermoléculaire B. De plus, une particularité de ce mécanisme est sa directionnalité et nous avons pu caractériser un nouvel intermédiaire réactionnel. Ce n'est que lorsque l'hélice B est formée que les hélices B' et C peuvent l'être à leur tour.

The EMBO Journal Vol. 19 No. 21 pp. 5905–5915, 2000

Progression of a loop–loop complex to a four-way junction is crucial for the activity of a regulatory antisense RNA

Fabrice A.Kolb, Hilde M.Engdahl[1], Jacoba G.Slagter-Jäger[1], Bernard Ehresmann, Chantal Ehresmann, Eric Westhof, E.Gerhart H.Wagner[1,2] and Pascale Romby[2]

UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue R. Descartes, Strasbourg cedex, France and [1]Department of Microbiology, SLU (Swedish University of Agricultural Sciences), Box 7025, Genetikvägen 5, S-75007 Uppsala, Sweden

[2]Corresponding authors
e-mail: P.Romby@ibmc.u-strasbg.fr or gerhart.wagner@mikrob.slu.se

The antisense RNA, CopA, regulates the replication frequency of plasmid R1 through inhibition of RepA translation by rapid and specific binding to its target RNA (CopT). The stable CopA–CopT complex is characterized by a four-way junction structure and a side-by-side alignment of two long intramolecular helices. The significance of this structure for binding *in vitro* and control *in vivo* was tested by mutations in both CopA and CopT. High rates of stable complex formation *in vitro* and efficient inhibition *in vivo* required initial loop–loop complexes to be rapidly converted to extended interactions. These interactions involve asymmetric helix progression and melting of the upper stems of both RNAs to promote the formation of two intermolecular helices. Data presented here delineate the boundaries of these helices and emphasize the need for unimpeded helix propagation. This process is directional, i.e. one of the two intermolecular helices (B) must form first to allow formation of the other (B′). A binding pathway, characterized by a hierarchy of intermediates leading to an irreversible and inhibitory RNA–RNA complex, is proposed.

Keywords: antisense RNA/loop–loop complex/plasmid/replication control/RNA–RNA interaction

Introduction

Naturally occurring antisense RNAs in prokaryotes are generally short, highly structured and untranslated (Wagner and Simons, 1994; Zeiler and Simons, 1998). Many of these antisense RNAs have been identified as plasmid copy-number regulators and are extremely efficient inhibitors; efficiency of *in vivo* control correlates with high *in vitro* binding rates to their respective target RNAs. Several *in vitro* studies indicated that binding rates are maximal with association rate constants in the range of 10^6 M^{-1} s^{-1} (reviewed in Wagner and Simons, 1994; Zeiler and Simons, 1998), and that the reaction is determined by the on-rate (e.g. Nordström and Wagner, 1994). By contrast, antisense RNAs used in artificial gene silencing strategies generally display much lower binding rates, often ~100-fold lower (e.g. Patzel and Sczakiel, 1998; E.G.H.Wagner, unpublished data). These results emphasize the important role of antisense and target RNA sequence and structure in promoting high pairing rates and high specificity of regulation. They also suggest that the structures of these RNAs have evolved to optimize their regulatory functions. Kinetic studies performed on antisense–target RNA pairs of plasmids ColE1 (Tomizawa, 1984), R1 (Persson *et al.*, 1988), pMU720 (Siemering *et al.*, 1994) and ColIb-P9 (Asano *et al.*, 1998; Asano and Mizobuchi, 2000) demonstrated that the initial step of binding involves a loop–loop interaction mediated by Watson–Crick base pairs, the so-called kissing complex, which is subsequently converted into more stable complexes. Loop–loop interactions play diverse roles in various biological systems, e.g. in dimerization of HIV-1 genomic RNA (for a review see Paillart *et al.*, 1996), in the interaction between human $tRNA^{Lys}$ and HIV-1 genomic RNA (Isel *et al.*, 1993), and in the localization of *Drosophila* bicoid mRNA (Ferrandon *et al.*, 1997). Furthermore, in phi29 RNA, a dimer was shown to be formed by a loop–loop interaction that constitutes a building block for the assembly of the hexamer that gears the DNA-translocating machinery of this bacteriophage (Chen *et al.*, 2000). Loop–loop interactions also participate in the architecture of large RNAs (for reviews see Pyle and Green, 1995; Brion and Westhof, 1997). However, the specific properties of antisense–target RNA systems are distinct from these latter cases: rapid kissing complex formation is a prerequisite for proper control, but rapid conversion to sufficiently stable complexes is essential to obtain maximal inhibition rates. The precise mechanism of formation of stable and inhibitory RNA–RNA complexes is still not understood in detail. Hence, it is important to delineate steps in the binding pathway as well as to identify structural features that permit unimpeded progression to the inhibitory complex.

Replication of plasmid R1 is controlled at the translational level by binding of the antisense RNA (CopA) to its target site (CopT) in the leader region of *repA* mRNA (Nordström *et al.*, 1984). Synthesis of the replication initiator protein RepA requires translation of a short leader peptide (*tap*), located upstream of *repA*. CopA binding prevents *tap* translation and thereby *repA* expression (Figure 1A; Blomberg *et al.*, 1992, 1994; Malmgren *et al.*, 1996). CopA and CopT are fully complementary, and both RNAs contain a major stem–loop structure as the main element responsible for high binding rate and control (Wagner and Nordström, 1986; Öhman and Wagner, 1989; Hjalt and Wagner, 1992, 1995). Binding initiates by an interaction between these complementary hairpin loops (Persson *et al.*, 1990a,b). Structure mapping of the CopA–CopT complex revealed that the initial loop–loop

© European Molecular Biology Organization

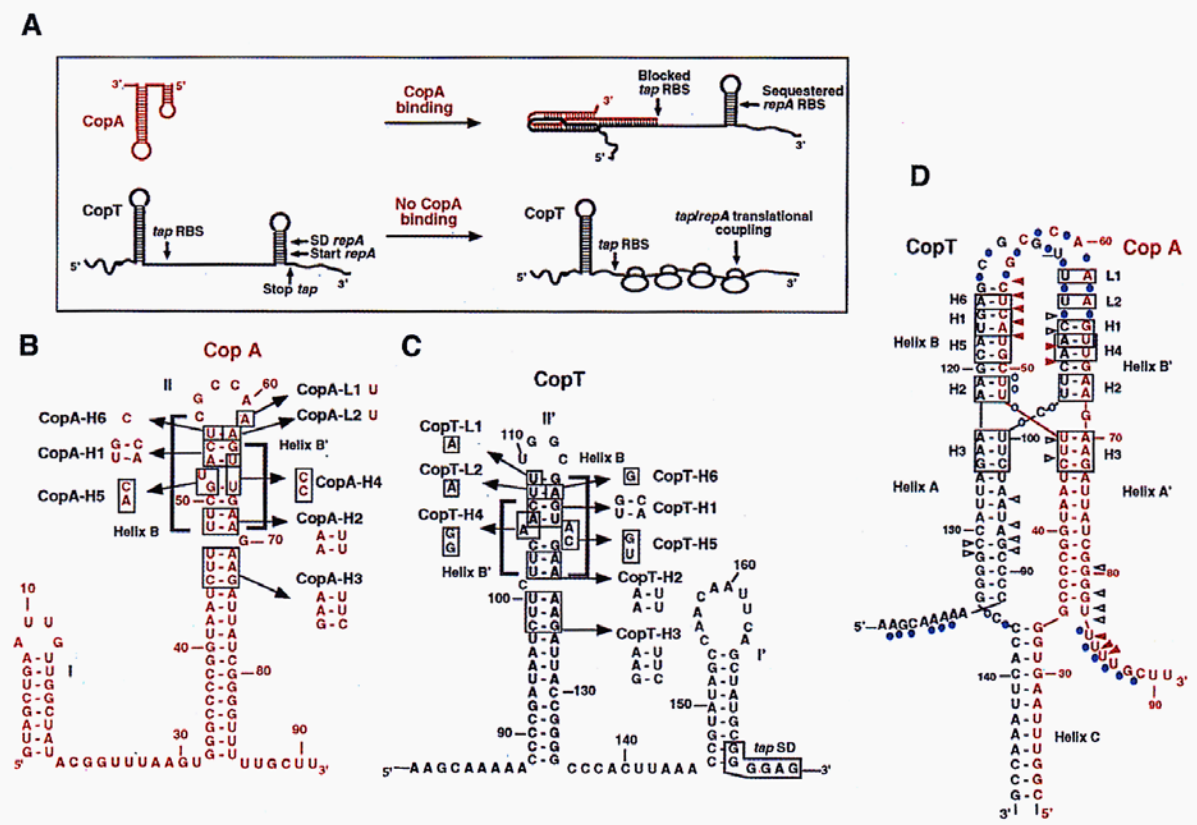

Fig. 1. Schematic model for antisense RNA control of RepA synthesis (**A**) and secondary structures of the antisense RNA CopA (**B**), its target site CopT (**C**) and the stable CopA–CopT complex (**D**). (A) Binding of CopA prevents ribosome binding at the *tap* ribosomal binding site (RBS), and the presence of a stem–loop structure sequesters the *repA* RBS so that translation of RepA is inhibited. In the absence of CopA binding, translation of *tap* permits ribosome entry at the *repA* loading site (translational coupling). SD, Shine–Dalgarno sequence. (B–D) The mutated nucleotides (H1–H6, L1 and L2) are boxed. The secondary structure model for the stable CopA–CopT complex (D) was derived from chemical and enzymatic probing (Kolb *et al.*, 2000). RNase V1 cleavages in CopA and CopT that occur in the CopA–CopT complex are shown by arrows: increased cleavages are shown by red arrows, unchanged cleavages are shown by black arrows; Pb^{2+}-induced cleavages in the CopA–CopT complex are shown by blue circles.

interaction (kissing complex) is rapidly followed by more extended interactions, and that full duplex formation is a slow process *in vitro* (Malmgren *et al.*, 1997; Kolb *et al.*, 2000). The major product of this binding reaction is an extended kissing complex that is characterized by a four-helix junction whose formation involves extensive breakage of intramolecular base pairs and the formation of two intermolecular helices (Figure 1; Kolb *et al.*, 2000). This bulky structure is further stabilized by the formation of a third long intermolecular helix involving the 5′ tail of CopA and the complementary region of CopT (Kolb *et al.*, 2000). In the absence of the 5′-most 30 nucleotides of CopA (truncated RNA variant denoted CopI; Persson *et al.*, 1990b; Malmgren *et al.*, 1997), the stabilizer helix cannot form, yet the complex is inhibitory (Wagner *et al.*, 1992; Malmgren *et al.*, 1996).

Previous experiments had shown that bulged residues located in the stems of both RNAs were required for rapid binding *in vitro* and control *in vivo* (Hjalt and Wagner, 1995). We speculated that these bulged residues would destabilize the upper stem to facilitate inter-strand helix propagation. Therefore, we have defined here the requirements for inter-strand base pairing and its importance for formation of the functional inhibitory complex *in vivo* and *in vitro*. By mutational analysis, we have identified a new binding intermediate and have demonstrated that the two intermolecular helices B and B′ (Figure 1D) are not of equivalent importance. This leads us to propose a binding pathway characterized by a hierarchy of distinguishable steps, ultimately resulting in the formation of an irreversible, inhibitory antisense–target RNA complex.

Results

Conversion of the initial kissing complex to extended interactions is required for rapid pairing in vitro

Structure probing of the stable CopA–CopT complex indicated that the primary loop–loop interaction is converted to extended interactions: intra-strand base pairs are melted and two intermolecular helices, B and B′, are formed (Kolb *et al.*, 2000). This extended kissing complex is further stabilized by the formation of an additional intermolecular helix C (shown schematically in Figure 1D). Base pair inversions were introduced at three positions within stem II of CopA and stem II′ of CopT to create

Table I. Binding rate constants of wild-type and mutant CopA–CopT pairs

CopA–CopT RNAs used[a]	k_{app} (M^{-1} s^{-1})[b]	Impairment relative to wt RNA pair[c]
wt–wt	1.6×10^6 [± 43%]	1.0×
H1–H1	4.0×10^5 [± 23%]	4×
H1–wt	**2.3×10^4 [± 37%]**	**69.5×**
wt–H1	**1.2×10^4 [± 0%]**	**133.3×**
H2–H2	2.2×10^5 [± 20%]	7.2×
H2–wt	**3.3×10^4 [± 0%]**	**48.5×**
wt–H2	**2.9×10^4 [± 0%]**	**55.2×**
H3–H3	3.8×10^5 [± 73%]	4.2×
H3–wt	8.7×10^5 [± 47%]	1.8×
wt–H3	2.3×10^5 [± 0%]	6.9×
H4–H4	2.5×10^5 [± 38%]	6.4×
H4–wt	5.9×10^5 [± 59%]	2.7×
wt–H4	2.2×10^5 [± 47%]	7.2×
H5–H5	6.5×10^5 [± 28%]	2.5×
H5–wt	**7.0×10^4 [± 26%]**	**22.8×**
wt–H5	**8.6×10^4 [± 27%]**	**18.6×**
H6–H6	1.1×10^6 [± 10%]	1.5×
H6–wt	**9.0×10^4 [± 22%]**	**17.8×**
wt–H6	**2.1×10^5 [± 26%]**	**7.6×**
L1–L1	1.4×10^6 [± 52%]	1.1×
L1–wt	2.1×10^6 [± 79%]	0.8×
wt–L1	8.7×10^5 [± 16%]	1.8×
L2–L2	2.0×10^6 [± 18%]	0.8×
L2–wt	1.2×10^6 [± 7%]	1.3×
wt–L2	2.1×10^6 [± 43%]	0.8×

[a]The RNAs produced *in vitro* are indicated as follows: wt, wild-type CopA or CopT; H1, H2, H3, H4, H5, H6, L1 and L2 mutated RNAs.
[b]k_{app} represents the rate constant of stable complex formation as determined from experiments such as in Figure 2. Values are averages of four independent experiments for the heterologous and homologous mutated RNA complexes, and of six for the wild-type pairs. Standard deviations are represented in brackets.
[c]The relative impairment is the (wt/wt) k_{app} value divided by the k_{app} value for the respective pair tested.

mismatches upon the formation of heterologous CopA–CopT complexes (mutations H1, H2 and H3; Figure 1). These mismatches should affect the binding rate if the formation of the helical segments is important. The rate of stable complex formation between homologous or heterologous CopA–CopT pairs was analysed by gel shift assays using denaturing polyacrylamide–urea gel (Persson *et al.*, 1988; Hjalt and Wagner, 1995). Second-order binding rate constants (k_{app}) were calculated according to Persson *et al.* (1988) and are summarized in Table I.

The values obtained for the homologous pairs CopA-H1–CopT-H1 and CopA-H3–CopT-H3 were only 4-fold lower compared with the wild-type (wt) RNA pair, whereas CopA-H2–CopT-H2 was decreased by 7-fold (Table I). Thus, base pair inversions have only relatively minor effects on the rate of formation of stable complexes. In agreement with this, structure mapping of the three homologous complexes revealed a topology of the extended kissing complex similar to that of the wild-type complex (Kolb *et al.*, 2000). By contrast, the heterologous CopA-H1–CopT-wt and CopA-H2–CopT-wt pairs were decreased by 70- and 49-fold, respectively, compared with the wild-type RNA pair. The reverse combinations showed similar decreases. Compared with the homologous mutated CopA-H2–CopT-H2 pairs, the effect of the H2 mutations on the rate constants in the heterologous complexes is reduced but still significant (~10-fold). Finally, the corresponding value for CopA-H3–CopT-wt was identical to that of the wild-type pair, and its reverse was only 6-fold lower (Table I).

The results in Table I thus show that mismatches in upper stem segments (H1, H2) decreased binding rates, whereas a mismatch below the lower bulge (H3) had no effect. This indicates that inter-strand base pairing throughout the upper two stem regions is required for high binding rates. Restored binding rates of the homologous mutant H1 and H2 RNA pairs additionally suggests that base pairing *per se*, and not the nucleotide sequence in this region, is the critical determinant for the rate of complex formation. These data are congruent with recent chemical probing results performed on the heterologous and homologous complexes (Kolb *et al.*, 2000).

Mutations that destabilize helix B decrease CopA–CopT complex formation rates

The previous data obtained on mutants H1–H3 cannot assess whether both intermolecular helices (B and B′; Figure 1) are equally important for stable complex formation. In order to test whether there was a preference in the directionality of helix progression subsequent to the first loop–loop interaction, the upper stems of CopA or CopT were mutated to alter only helix B′ (mutation H4) or helix B (H5, H6). Mutations H4 and H5 had also been designed so that the structures of both CopA and CopT were preserved. In addition, two point mutations were introduced: in the major hairpin loops of CopA (A61→U, CopA-L1) and CopT (U109→A, CopT-L1), and at the base pair that closed the hairpin loops of CopA (A62→U, CopA-L2) and CopT (U108→A, CopT-L2). These nucleotides were known to be unpaired in the stable CopA–CopT complex (Figure 1D). The k_{app} values of the homologous and heterologous wild-type and mutant CopA–CopT pairs are shown in Table I.

The binding rates of heterologous or homologous complexes carrying either mutations H4, L1 or L2 were similar to that of the wild-type pair (Table I). Only a 6-fold decrease in k_{app} was observed for the CopA-wt–CopT-H4 and CopA-H4–CopT-H4 pairs. This could be due to the formation of an additional G–C pair in the upper stem II′ of CopT-H4, resulting in increased stability of the stem and slight interference with helix propagation. By contrast, the binding rate constants of the heterologous CopA-H5–CopT-wt and CopA-H6–CopT-wt pairs were decreased by 23- and 18-fold, respectively, compared with the wild-type RNA pair or the homologous mutant pairs. A similar decrease in the k_{app} value was also obtained for the CopA-wt–CopT-H5 complex (18.6-fold), but only an 8-fold decrease was observed for the CopA-wt–CopT-H6 complex (Table I). Mutation H6 in CopT changed the intermolecular base pair U55(CopA)–A115(CopT) in helix B by a U(CopA)–G(CopT-H6) wobble pair, thus explaining the wild-type-like behaviour.

Thus, the effect of the H5 and H6 mutations indicates that inter-strand base pairing through helix B is required for high rates of formation of the stable CopA–CopT complex (which also contains helix C), whereas the formation of helix B′ (H4) does not appear to be essential. The silent phenotype of the L1 and L2 mutations in the

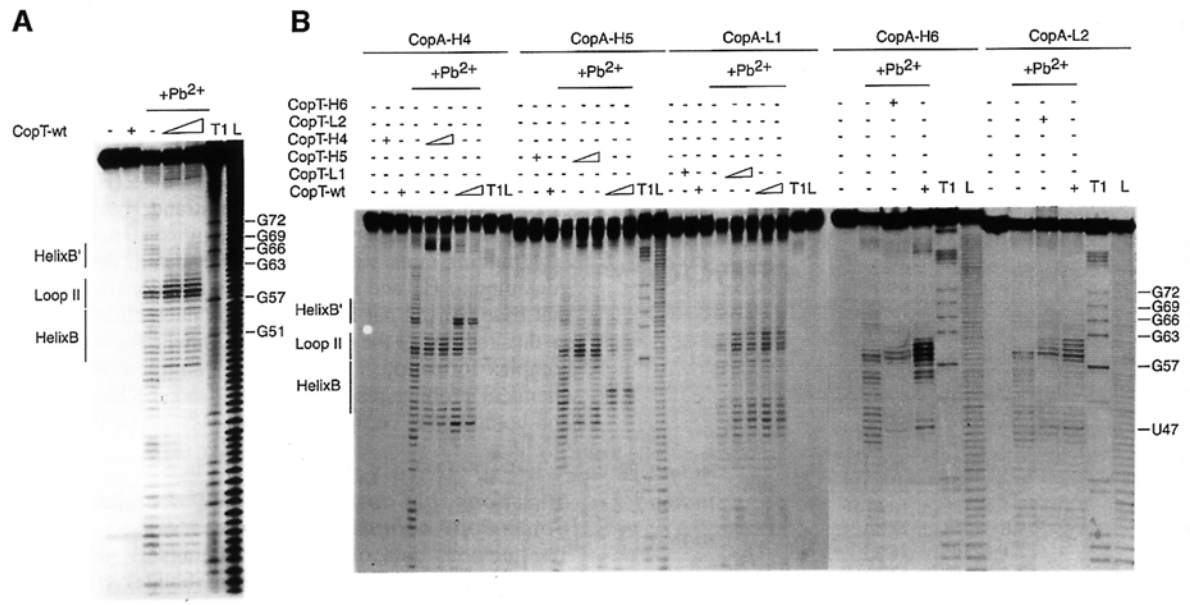

Fig. 2. Pb^{2+}-induced hydrolysis of homologous or heterologous CopA–CopT complexes using end-labelled CopA species. Hydrolysis was performed on 5′-end-labelled wild-type CopA (**A**) or CopA-H4, CopA-H5, CopA-L1, CopA-H6 and CopA-L2 (**B**), alone (–) or in the presence of an excess of wild-type or mutant CopT (+). Complex formation was performed at 37°C for 15 min in TMN buffer. Lanes T1 and L, RNases T1 and alkaline ladders, respectively.

in vitro binding assay also supports a directional mechanism for helix propagation.

Formation of helix B is a prerequisite for formation of helix B′

Pb^{2+}-induced hydrolysis occurs preferentially at unpaired residues and is very sensitive to subtle RNA rearrangements. This probe can be used to obtain a signature of the extended kissing complex (Malmgren *et al.*, 1997): strong cleavages occur in CopA at positions C56–A62 and U47–U49, and in CopT at U100–C101, C107–U110 and G112–C113. Furthermore, in the CopA–CopT complex, the formation of the intermolecular helix C, which greatly enhances complex stability, was characterized by strong protections in regions A8–G30 in CopA and A139–A164 in CopT (Kolb *et al.*, 2000). Here, Pb^{2+} cleavages were used on homologous and heterologous complexes formed with the H4, H5, H6, L1 and L2 RNA variants (Figures 2 and 3).

Structure probing of end-labelled CopA and CopT RNAs indicated almost identical cleavage patterns for free mutant and wild-type RNAs (Figures 2 and 3). Hence, no major structural changes were caused by the mutations. The homologous complexes of CopA–CopT (H4, H5, H6, L1 and L2), and the heterologous CopA-L1–CopT-wt and CopA-L2–CopT-wt complexes and their reverse, showed characteristic wild-type-like cleavage patterns (Figures 2 and 3). The heterologous CopA-H4–CopT-wt complex was almost identical to the wild-type complex except that strong cleavages were induced at C64 and C65 of CopA-H4 (Figure 2B). This shows that several inter-strand base pairs in helix B′ were not formed due to the mismatches introduced by mutation H4, whereas—as in the homologous complexes—the intermolecular helices B and C were supported by strong protections in regions C50–U55 and A8–G30, respectively (Figure 2B). Similarly, Pb^{2+} cleavages at positions U100–U108 in CopT-H4 (Figure 3B) in the reverse complex, CopA-wt–CopT-H4, also supported the finding that helix B′ was not formed. The wild-type-like cleavage pattern, which was restored in complexes formed between the fully complementary H4 mutant RNAs, further argues for the presence of helix B′ (Figures 2B and 3B). Mutation H6 is expected to disrupt the second intermolecular base pair in helix B (Figure 1C). Indeed, structure mapping of CopA-H6–CopT-wt complexes revealed a new Pb^{2+} cleavage at position C56 and an increased cleavage at U55 (Figure 2B). These data suggest a destabilization/disruption of the two first intermolecular base pairs in helix B. Conversely, structure mapping of CopT-H6–CopA-wt complexes supported a characteristic wild-type-like cleavage pattern. Thus, replacement of the intermolecular base pair U55(CopA)–A115(CopT) by U–G had no significant effect on the four-way junction topology. A strikingly different result was obtained with mutation H5, designed to alter two intermolecular base pairs in helix B. Here, the Pb^{2+} cleavage pattern of CopA-H5, when in complex with CopT-wt, was significantly different from that of a homologous pair. The only significant protections now occurred at nucleotides A53–C59, whereas protections at U49–C52 were no longer observed (Figure 2B). In the reverse pair (CopA-wt–CopT-H5), significant protections were restricted to residues G112–U117 in CopT-H5 (Figure 3B). This indicates that neither helix B nor B′ could form properly.

Thus, the difference in effects of mutations H4, H5 and H6 in binding experiments (Table I) was paralleled by effects on complex structure. In particular, disruption of

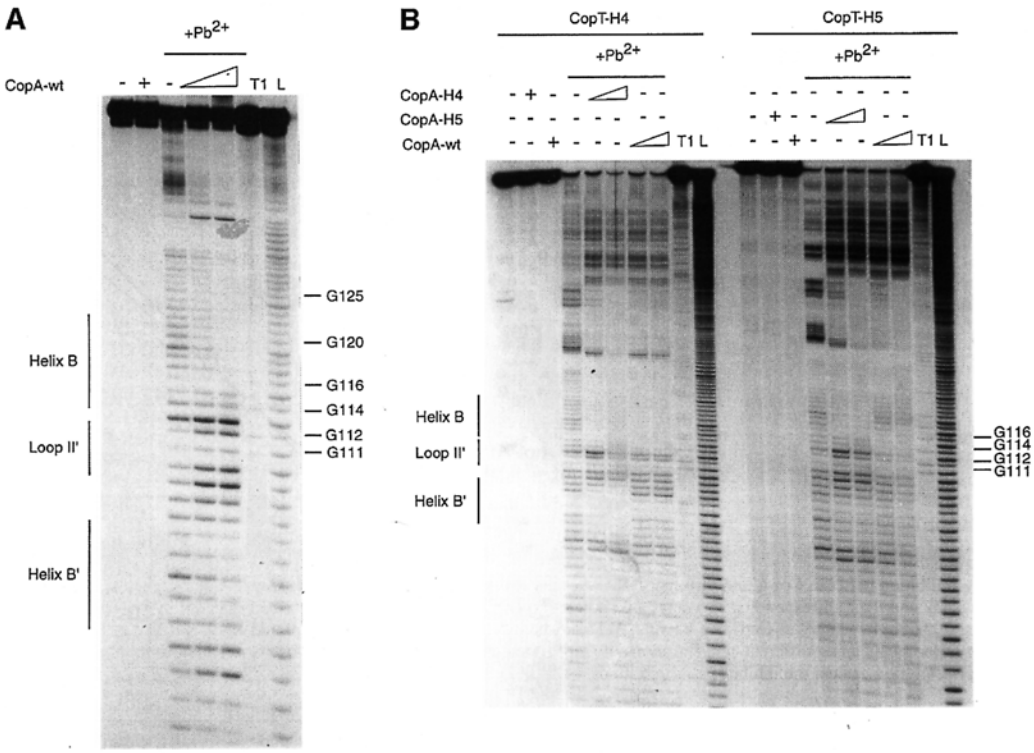

Fig. 3. Pb^{2+}-induced hydrolysis of homologous or heterologous CopA–CopT complexes using end-labelled CopT species. Hydrolysis was performed on 5′-end-labelled wild-type CopT (**A**) or CopT-H4 and CopT-H5 (**B**), alone (–) or in the presence of an excess of wild-type or mutant CopA (+). Lanes T1 and L, RNases T1 and alkaline ladders, respectively.

two Watson–Crick base pairs in helix B (H5) affected the conversion of the initial kissing complex to the four-way junction structure. Conversely, destabilization of helix B′ (H4) did not interfere with the formation of helices B and C. These data suggest that the formation of the intermolecular helix B′, but also of helix C, depends on prior formation of helix B. The fact that the L1 and L2 mutations did not show any effect on structure and on *in vitro* binding rate further supports a B→B′ hierarchy of helix formation.

Directionality of helix progression is supported by competition experiments

Competition assays can be used to obtain information about the rate of formation and stability of early intermediates in a binding pathway (Tomizawa, 1985; Asano *et al.*, 1998). In the experiments summarized in Figure 4, standard binding rate measurements were performed in the presence or absence of competitor RNAs. Initial rates of stable CopA–CopT complex formation were determined over a range of competitor concentrations, and a K_i value was calculated from graphs of the ratio of initial rates, v_o/v_i, against the concentration of inhibitor (Tomizawa, 1985). As expected, CopI, unable to form a stable complex but capable of forming the extended kissing complex, inhibited stable complex formation with a K_i value of 8 nM (Figure 4). Two RNA fragments (R3 and R4) were designed to mimic the formation of helices B′ and B, respectively. The 12mer R4 was equally efficient as CopI (K_i = 7 nM), whereas the putative helix B′ former R3 was >>100-fold less effective, as was CopI-H1 (inverted upper stem). The two shorter RNAs, R2, only consisting of the CopA six base loop sequence, and R1, which carries the CopA loop closed by a GC-rich stem (not complementary to target), were both unable to compete at submicromolar concentrations. The high K_i values of R1 and R2 may indicate that the stability of a putative initial loop–loop interaction is low.

From these experiments it is evident that R4 RNA and wild-type CopI were the only two efficient inhibitors of stable complex formation (Figure 4), and that their respective K_i values were practically identical to equilibrium dissociation constants determined previously (K_d = 7.4 nM; Hjalt and Wagner, 1995). The inability of R3 to compete successfully with CopA for binding to CopT, in spite of an expected helix stability similar to R4, clearly indicates that the formation of helix B must precede that of helix B′ (and helix C) to form the stable complex.

Mutations affecting in vitro binding also affect in vivo control

Mutations that affect antisense–target RNA binding rates should have proportionate effects on control *in vivo*. To assay the effects of the H1–H5 and L1 mutations, we used translational fusions between the *repA* and *lacZ* genes, preceded by wild-type or mutant *copA*/*copT* control regions. All plasmids of the pGW177-III-L series

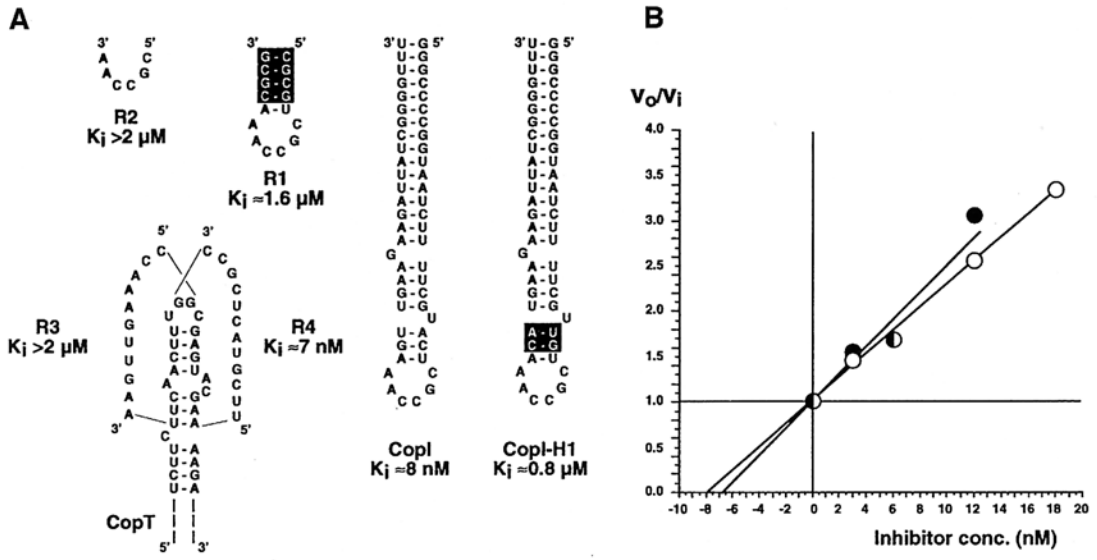

Fig. 4. Competitive inhibition between CopA and different RNA variants for the formation of the stable CopA–CopT complex. (**A**) The inhibitory RNAs used in this study and their corresponding K_i values are shown. Experimental conditions are described in detail in Materials and methods. Nucleotides that are not complementary to CopT are boxed. (**B**) Determination of the K_i values for CopI and RNA R4. The ratio v_o/v_i was calculated from the inhibition experiments using different concentrations of CopI (open circles) or of RNA R4 (closed circles) at time intervals between 1 and 6 min at 37°C, and are plotted against the inhibitor concentrations.

(Blomberg *et al.*, 1994) carry a *copA* promoter down-mutation to abolish CopA transcription and to permit supply of heterologous CopA from a second, compatible plasmid. RepA–LacZ fusion protein synthesis was measured in extracts of cells carrying appropriate combinations of plasmids. Assays of unrepressed *repA–lacZ* expression (no CopA *in trans*) indicated that neither of the tested mutations significantly affected target RNA stability or translatability [Table II: pGW177III-(H1–H5 or L1)-L].

Wild-type CopA supplied from a high-copy-number plasmid *in trans* decreased RepA–LacZ synthesis to background levels (pGW177-III-L + pGW643; Table II), and all mutant CopAs were equally effective on their cognate mutant CopT targets. When heterologous antisense–target RNA combinations were assayed, the CopA-H1–CopT-wt, CopA-H2–CopT-wt and CopA-H5–CopT-wt pairs, and their reverse, were clearly less inhibitory than any of the homologous combinations (0.23/0.40 relative activities for H1/wt; 0.04/0.07 for H2/wt; 0.2/0.31 for H5/wt; Table II). By contrast, the CopA-H3–CopT-wt, CopA-H4–CopT-wt and CopA-L1–CopT-wt pairs, and their reverse, were as efficient in inhibition as any homologous pair. For comparison, note that cells carrying pGW177-L alone give a value of 0.14. This plasmid supplies CopA from its own, intact gene. When CopA is supplied *in trans* (pGEM2 plasmids; Table II), it is present at ~20-fold higher concentration (due to a higher copy number of carrier plasmid and lack of convergent transcription; Stougaard *et al.*, 1982 and data not shown). This emphasizes that, for example, CopA-H1 is >20 times less active on a wild-type target [cf. pGW177-L and (pGEM2-H1 + pGW177-III-L); Table II]. The *in vivo* data correlate well with the *in vitro* results. All heterologous combinations of wild type and mutant that impaired inhibitory activity *in vivo* (H1, H2 and H5) also decreased *in vitro* binding rates (Tables I and II), whereas mutations H3, H4 and L1 induced no significant effect on stable complex formation *in vitro* and no phenotype *in vivo*. It has to be noted that no significant effect *in vivo* was observed for mutations that induced less than a 7-fold decrease in the binding rate constant.

In conclusion, only base pair inversions located above the lower bulged nucleotide in both RNAs (H1 and H2) were important for *in vivo* inhibition efficiency. The data also indicated that formation of helix B (H5), but not helix B′ (H4), is necessary and sufficient to initiate the inhibition of *repA* translation. We infer that the H1, H2 and H5 mutations impair CopA's inhibitory capacity by impeding the rapid progression through the early intermediates to the stable CopA–CopT complex.

Discussion

Rapid binding between antisense and target RNAs is a prerequisite for efficient control. This is particularly important for plasmids regulating their copy numbers, since antisense RNAs are the principal regulators that have to accomplish inhibition of their target RNAs within a very short time frame. The structural features that favour the rapid formation of inhibitory complexes are not well understood, and the complexity of binding pathways makes general conclusions difficult. Nevertheless, in several systems, the initial recognition involves formation of a reversible loop–loop interaction (kissing complex).

Based on the previous and present data, a tentative binding pathway between CopA (or CopI) and CopT is shown in Figure 5. The initial kissing complex occurs between a subset of loop bases in both RNAs (Figure 5B), and is defined by the location of copy number mutations (5′-GGCG in CopT; e.g. Givskov and Molin, 1984). This

Table II. β-galactosidase synthesis in cells carrying *repA–lacZ* translational fusion plasmids, alone or in the presence of CopA supplied *in trans*

Plasmids[a]	CopA/CopT pairs[b]	Specific β-gal activity[c]	Relative β-gal activity[d]
pGEM2	–/–	9 ± 4	<0.01
pGW177-L	[wt/wt][e]	459 ± 101	0.16
pGW177-III-L	–/wt	2940 ± 992	1.00
pGW177-H1-III-L	–/H1	3469 ± 198	1.18
pGW177-H2-III-L	–/H2	3065 ± 406	1.04
pGW177-H3-III-L	–/H3	3086 ± 278	1.05
pGW177-H4-III-L	–/H4	3223 ± 765	1.10
pGW177-H5-III-L	–/H5	3389 ± 885	1.15
pGW177-L1-III-L	–/L1	3239 ± 641	1.10
pGEM2-H1 + pGW177-H1-III-L	H1/H1	10 ± 2	<0.01
pGEM2-H2 + pGW177-H2-III-L	H2/H2	12 ± 3	<0.01
pGEM2-H3 + pGW177-H3-III-L	H3/H3	8 ± 0	<0.01
pGEM2-H4 + pGW177-H4-III-L	H4/H4	15 ± 5	<0.01
pGEM2-H5 + pGW177-H5-III-L	H5/H5	3 ± 2	<0.01
pGEM2-L1 + pGW177-L1-III-L	L1/L1	6 ± 2	<0.01
pGEM2-H1 + pGW177-III-L	**H1/wt**	**662 ± 300**	**0.23**
pGEM2-H2 + pGW177-III-L	**H2/wt**	**124 ± 62**	**0.04**
pGEM2-H3 + pGW177-III-L	H3/wt	9 ± 2	<0.01
pGEM2-H4 + pGW177-III-L	H4/wt	4 ± 1	<0.01
pGEM2-H5 + pGW177-III-L	**H5/wt**	**578 ± 24**	**0.20**
pGEM2-L1 + pGW177-III-L	L1/wt	5 ± 1	<0.01
pGW643 + pGW177-III-L	wt/wt	8 ± 3	<0.01
pGW643 + pGW177-H1-III-L	**wt/H1**	**1163 ± 411**	**0.40**
pGW643 + pGW177-H2-III-L	**wt/H2**	**216 ± 69**	**0.07**
pGW643 + pGW177-H3-III-L	wt/H3	11 ± 6	<0.01
pGW643 + pGW177-H4-III-L	wt/H4	16 ± 2	<0.01
pGW643 + pGW177-H5-III-L	**wt/H5**	**911 ± 342**	**0.31**
pGW643 + pGW177-L1-III-L	wt/L1	23 ± 5	<0.01

[a]Fusion plasmids and CopA-donor plasmids: pGW63 carries the *copA* gene in vector plasmid derivative from pSP64 (Persson *et al.*, 1990); pGW177-L plasmid carries the control region of plasmid R1 (Blomberg *et al.*, 1992); pGW177-III-L derivative of pGW177-L contains the mutation III in *copA* promoter (Blomberg *et al.*, 1994). The mutated plasmids (pGW177-III-L and pGEM2 series) were constructed as described in Materials and methods.
[b]The RNAs produced *in vivo* are indicated as follows: wt, wild-type CopA or CopT; H1, H2, H3, H4, H5 and L1 mutated RNAs; –, no RNA.
[c]β-galactosidase (β-gal) activity assays were performed as in Materials and methods, and represent averages of four independent determinations. Values are given in Miller units (Miller, 1972). Standard deviations are indicated.
[d]Miller units were converted to relative activities. The activity obtained with pGW177-III-L was set to unity.
[e]CopA encoded from fusion plasmid (the CopA concentration is lower than when encoded *in trans*).

interaction is fully reversible and most likely very unstable, as indicated by the inability of loop-only sequences to compete for CopA binding to CopT (R1 and R2; Figure 4). The results presented show that subsequent helix progression proceeds unidirectionally into the upper stem regions. In support of this, changes at residues located at the 3′ side of the CopA loop (L1 and L2) had no effect on antisense efficiency *in vivo* and *in vitro*. Furthermore, only the mutations that altered helix B (H5 and H6; Figure 1) but not helix B′ (H4) significantly affected the binding rate (Table I), the structure of heterologous CopA–CopT complexes (Figures 2 and 3) and control *in vivo* (Table II). Thus, these data suggest that helix B must be formed (structure C in Figure 5) to permit subsequent formation of helices B′ and C. Whether, in a wild-type context, one of the latter two helices (B′ and C) is formed earlier than the other, cannot be determined. Conversely, whether or not helix B′ can form, activity tests *in vitro* (Table I) and *in vivo* (Table II), as well as the structure of the complex (except for helix B′ itself; Figures 2 and 3), are wild-type like. Thus, the stable complex is only formed under conditions in which helix B is formed, whereas helix B′ is dispensable. In line with these data, the competition experiments in Figure 4 argue that helix B formation is an early step required to commit the interacting RNAs to stable complex formation, since the RNA fragment R4 gave a K_i value comparable to that obtained with the complete CopI stem–loop, whereas the R3 RNA (helix B′) was >100-fold less effective.

The asymmetric strand migration of helix B requires stem disruption above the position of the lower bulge (Figure 1C), and frees the opposing side of the stem for the second intermolecular helix B′ to form. The experiments presented here do not address the reason for the observed directionality of the step from the first loop–loop contact to structure C, and subsequently D (Figure 5). However, an interesting possibility is suggested by the recent finding that CopT, like antisense or target loops in most antisense RNA systems, carries a putative U-turn (π-turn) structure motif (YUNR, Y = pyrimidine, R = purine; Asano *et al.*, 1998; Franch *et al.*, 1999). Such a motif has recently been proposed in the target loop of the *rep* mRNA of ColIb-P9 (Asano *et al.*, 1998), and is also present in the CopT loop (5′-UUGGCG). In this motif, the sharp turn at the 3′-phosphate of the invariant U positions the bases downstream in a pre-formed A-helical conformation, stabilized by additional hydrogen bonds within the loop. This structural module, in the tRNA context, is crucial for rapid and specific decoding (Grosjean *et al.*, 1998; Ashraf *et al.*, 1999) and it has been suggested that its ubiquitous occurrence in antisense–target RNA systems might reflect

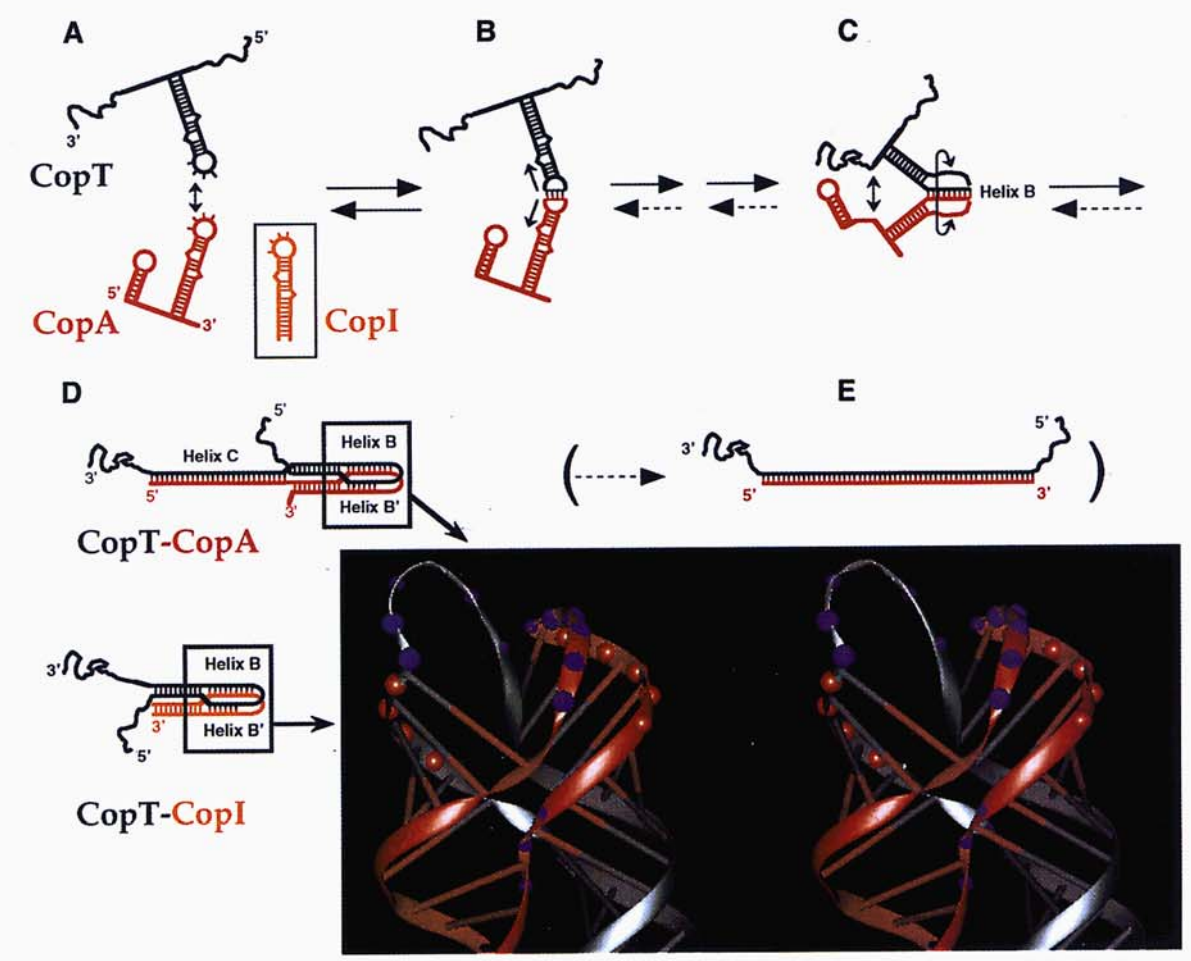

Fig. 5. The binding pathway of CopA–CopT. (**A–E**) refer to steps in the pathway as explained in the Discussion. The parentheses indicate that full duplex formation is slow and biologically irrelevant (step E). Dotted arrows denote very slow reactions. The structure model of the four-helix junction derived from computer modelling by Kolb *et al.* (2000) is shown. Red and purple circles represent the RNase V1 and Pb^{2+}-induced cleavages.

the requirement for binding specificity and rate (Franch and Gerdes, 2000). In CopT, the bases immediately following the invariant U are the ones at which binding initiates (5′-GGCG; Givskov and Molin, 1984; Nordström *et al.*, 1984). Extension of this early intermolecular helix (to give B) would not require a rearrangement of the U-turn structure in CopT loop; hence, the orientation of the U-turn might determine the preference for helix B over helix B′ formation.

Once helix B is formed, the interacting RNAs are committed to form the extended kissing complex shown as structure D in Figure 5. This four-helix junction structure is formed as either a stable CopA–CopT complex, or as the less stable CopI–CopT complex lacking helix C. Altogether, the results presented here support the model of the extended kissing complex, which was derived from computer modelling, structure probing data, site-directed mutagenesis and appropriate stereochemistry (Kolb *et al.*, 2000; Figure 5D). This structure shows two long helical domains, each one generated by co-axial stacking of two helices (Figure 1). The base pairs that initiate binding in CopA and CopT are disrupted upon conversion to the extended kissing complex, and serve as connecting loops between helices B and B′. Since mismatches between CopA and CopT below the bulged residues in stems II/II′ (mutation H3) show no phenotype *in vitro* and *in vivo* (Tables I and II), this supports the view that intermolecular base pairing proceeds to, and is arrested at, the position of the lower bulged nucleotides (CopA:G69/CopT:C101) to form the four-helix junction (Figure 1C). Progression of the intermolecular helices B and B′ is arrested, probably due to topological stress. This four-way junction adopts an asymmetrical X-shaped conformation as the result of strand exchange (Figure 5). This folding and the constraints imposed by the connecting loops force a side-by-side alignment of the helical segments that, in turn, facilitates the formation of the stabilizer intermolecular helix C in CopA–CopT (Kolb *et al.*, 2000). The stable CopA–CopT complex (Figure 5D) is sufficient to exert inhibition *in vitro* (Figure 1A; Malmgren *et al.*, 1996) and, under these conditions, full duplexes are neither formed nor required (Malmgren *et al.*, 1996, 1997). In conclusion, the pathway of binding suggested in Figure 5 represents a stepwise progression from free RNAs through a hierarchy

of distinguishable intermediates to the inhibitory stable complex.

Analysis of binding kinetics in the CopA–CopT (Persson *et al.*, 1990b) system had shown that overall rates of stable complex formation are very similar to the rate constants of the extended kissing complex formation, at ~10^6 M^{-1} s^{-1}. These data indicate that dissociation rarely occurs once the RNAs have encountered each other in a compatible orientation, and consequently, binding proceeds quasi-irreversibly. This process requires that subsequent steps, such as helix B formation, have to be unimpeded by topology. In accordance with this requirement, CopA and CopT carry upper stem bulges whose destabilizing effects are crucial for rapid binding and inhibition (Hjalt and Wagner, 1995). Thus, the characteristic high rates of stable CopA–CopT complex formation (Figure 5D) imply an almost instantaneous conversion of the reversible kissing complex (Figure 5B) to the extended kissing complex (four-way junction structure; Figure 5D).

Several other plasmids, belonging to incompatibility groups IncIα, B and Z, share a similar genetic organization. Their mechanism of regulation is different from that of plasmid R1; antisense RNAs primarily inhibit activator pseudoknot formation (e.g. Asano *et al.*, 1991; Wilson *et al.*, 1993). Analyses of pMU720 (IncB) and ColIb-P9 (IncIα) have indicated similarities to R1 in that: (i) the structural features of the antisense stem–loops are reminiscent of those of CopA (identical loop sequences, destabilized upper stems, single-stranded region 5′ of stem–loop); (ii) full duplex formation is not required for control; and (iii) a stable binding intermediate involves inter-strand pairing within the upper stems of antisense and target RNAs (Siemering *et al.*, 1993, 1994; Asano *et al.*, 1998). Based on enzymatic probing experiments, Asano and Mizobuchi (2000) recently reported on a late intermediate antisense–target RNA complex in ColIb-P9, of similar overall topology to CopA–CopT, but differing in the lengths of the intermolecular helices (analogous to B and B′). Asano and Mizobuchi (2000) have also postulated that the initial loop–loop propagates in both directions to promote formation of the four-helix junction. However, the results presented here indicate that the two helices in the CopA–CopT complex are formed in a hierarchical order. Whether or not these differences will stand further experimental tests, we suggest similar binding pathways for antisense and target RNAs of plasmids belonging to the incompatibility groups FII, B, Z and Iα.

A binding pathway has also been proposed for RNAI–RNAII of ColE1 (Tomizawa, 1984, 1990; Eguchi and Tomizawa, 1991). Binding of RNAI and RNAII is initiated by at least two loop–loop interactions followed by a series of reactions that progressively lead to the formation of a stable duplex (Tomizawa, 1990; Eguchi and Tomizawa, 1991). NMR studies were performed on a complex between two RNA hairpins carrying seven-membered complementary loops derived from a mutant pair of RNAI–RNAII (Marino *et al.*, 1995; Lee and Crothers, 1998). These studies indicated that all seven loop bases were paired in the loop–loop helix, and continuous stacking of the loop nucleotides on the 3′ side of their respective stems was observed. Thus, both the R1 and ColE1 systems employ recognition loop sequences for the first reversible interaction, but, depending on their particular topological constraints, use radically different pathways to proceed rapidly to more stable intermediates and ultimately to stable inhibitory complexes. In spite of this, these as well as most other naturally occurring antisense systems show comparable binding rate constants.

In conclusion, this study shows that structural features in both CopA and CopT determine the topologically possible and kinetically favoured pathway. The pathway proposed is characterized by a hierarchy of intermediates that leads to a stable and functional RNA–RNA complex (Figure 5D). The results presented, together with previous studies on the effects of loop size and bulge mutations (Hjalt and Wagner, 1992, 1995), also emphasize the requirement for unimpeded helix progression immediately subsequent to initial kissing, in order to sustain the high binding rates characteristic of efficient plasmid copy-number regulators. One can speculate that such a step-wise mechanism for RNA–RNA recognition may also play an important regulatory role in the folding of large RNA molecules and in many other cellular functions involving RNA–RNA interaction.

Materials and methods

Bacterial strains and plasmids

The *Escherichia coli* strain DH5α [F⁻ *endA1*, *hsdR17*, *supE44*, *thi-1*, *recA1*, *gyrA96*, *relA1*, Δ(*argF-lacZYA*)*U169*, Φ*80lacZ*ΔM15 (Hanahan, 1985)] was used for plasmid constructions and *in vivo repA–lacZ* expression assays. Plasmid pGW177-L carries the control region of plasmid R1 (Blomberg *et al.*, 1992). Plasmids of the pGW177-III-L series used for β-galactosidase measurements are derivatives of pGW177-L (Blomberg *et al.*, 1994). Furthermore, these plasmids carry the mutation III [promoter-down mutation in *copA*, silent in *tap* (Öhman and Wagner, 1991)], and thus CopA is not transcribed. To create plasmid pGW177-H1-III-L, two PCRs were performed, using primers HE27-bio/GW110 and GW109-bio/HE40, respectively, on pGW58-III (Blomberg *et al.*, 1994) template DNA. Each biotinylated PCR fragment was immobilized on streptavidin-coated Dynabeads® M (Dynal, Norway), and the complementary strands were separated by denaturation. The biotinylated (+) strand of the HE27-bio/GW110 fragment and the non-biotinylated (–) strand of the GW109-bio/HE40 fragment were annealed, followed by fill-in using the Klenow enzyme in the presence of 100 μM dNTPs. A further PCR using the external primers HE27 and HE40 was performed to obtain a fragment containing the desired mutation. This fragment was cleaved with the restriction enzymes *Bgl*II and *Sal*I and ligated to the larger cleavage product of plasmid pGW177-L, cleaved with the same enzymes. pGW177-H2-III-L and pGW177-H3-III-L were constructed using the same protocol, except that the primer combinations differed: HE27-bio/GW112 and GW111-bio/HE40 (mutation H2); HE27-bio/GW114 and GW113-bio/HE40 (mutation H3). Construction of pGW177-H4-III-L, pGW177-H5-III-L, pGW177-H6-III-L, pGW177-L1-III-L and pGW177-L2-III-L was achieved using the Stratagene Quickchange™ Site-Directed Mutagenesis Kit with plasmid pGW58-III and primers H4-FW/H4-Rev, H5-FW/H5-Rev, H6-FW/H6-Rev, L1-FW/L1-Rev and L2-FW/L2-Rev, respectively. The mutated plasmids were digested with *Sal*I and *Bgl*II, and fragments were inserted into pGW177-L cleaved with the same two enzymes. pGEM2 cloning vector was purchased from Promega. Plasmids pGEM2-H1, pGEM2-H2 and pGEM2-H3, carrying the mutant *copA* genes, were constructed by using the above-mentioned primer combinations on pGW58 template DNA (Blomberg *et al.*, 1990) for PCR, yielding H1, H2 and H3 fragments, which were then amplified with primers HE27 and HE40, cleaved with *Sma*I and *Sal*I, and ligated into *Sma*I–*Sal*I-cleaved pGEM2 DNA. Plasmids pGEM2-H4, pGEM2-H5 and pGEM2-L1 were obtained using the Stratagene Quickchange™ Site-Directed Mutagenesis Kit with pGW58 plasmid and the same set of primers as described above. The mutated plasmids were digested with *Xmn*I and *Sal*I, and the generated fragments were ligated into *Sma*I–*Sal*I-cleaved pGEM2 DNA. Purification of plasmid DNA, restriction enzyme cleavages and other DNA techniques were essentially according to Sambrook *et al.* (1989).

Oligodeoxyribonucleotides

Oligodeoxyribonucleotides were from Amersham Pharmacia Biotech, Interactiva Biotechnologie GmbH and NAPS Goettingen GmbH. For construction of plasmids containing mutations H1–H6, L1 and L2, the following primer pairs were used: for H1, HE27-bio (5′ biotin–GTGATCTTCCGTCACAGGTAT) and GW110 (5′-CGTTGTCGC-CAACATGAAGAAGATT), GW109 (5′-CATGTTTGGCGACAACGA-AAAGATTA) and HE40 (5′-GAATTTCGACCTCTAGACCAA); for H2, HE27-bio and GW112 (5′-TAACGTACTCGCCAAAGTTGTTGA-AGATT), GW111-bio (5′ biotin–CAACAACTTTGGCGAGTACGT-TAAGATTA) and HE40; for H3, HE27-bio and GW114 (5′-TGA-ATTCGTACTCGCCAAAGTTGAAGTTCATTATCGG), GW113-bio (5′ biotin–TGAACTTCAACTTTGGCGAGTACGAATTCATTACCG) and HE40; for H4, H4-FW (5′-CCGATAATCTTCTTCGGCTTTGGC-GAGTAC) and H4-Rev (5′-GTACTCGCCAAAGCCGAAGAAGAT-TATCGG); for H5, H5-FW (5′-TTCAACTTTGGCGAGTGTGAA-AAGATTACCGGG) and H5-Rev (5′-CCCGGTAATCTTTTCACA-CTCGCCAAAGTTGAA); for H6, H6-FW (5′-CGATAATCTTCTTCA-ACTTTGGCGGGTACGAAAAGATTACCGG) and H6-Rev (5′-CCG-GTAATCTTTTCGTACCCGCCAAAGTTGAAGAAGATTATCG); for L1, L1-FW (5′-TAATCTTCTTCAACTATGGCGAGTACGAAAA) and L1-Rev (5′-TTTTCGTACTCGCCATAGTTGAAGAAGATTA); for L2, L2-FW (5′-CGATAATCTTCTTCAACATTGGCGAGTACGAAAAG) and L2-Rev (5′-CTTTTCGTACTCGCCAATGTTGAAGAAGATTA-TCG).

For sequence determinations, PCR fragments were generated from the pGW177-III-L-series plasmids using primer pairs: GW58-bio (5′ biotin–CAGGCTCAGTTCGTTGAGAAAA) and GW59 (5′-CACCGCCTT-TTCCATCAGTTT). Primer GW60 (5′-GGATTCGGGTTCTTTAC) was used for sequencing. For the pGEM2 plasmid series, inserts were amplified using primer GW120-bio (5′ biotin–CATACGATTTAG-GTGACACTAT) and GW45 (5′-GAAATTAATACGACTCACTATA). GW45 was used as sequencing primer.

For *in vitro* RNA transcription, all CopT mutant PCR fragments were generated from the pGW58 plasmid using a 5′ primer that contains the appropriate mutations (e.g. CopT-H1: 5′-ACGTACTTTAAAGCAAAA-ACCATGTTTGGCGACAACGAAAAGATTACCG) and the 3′ primer (5′-CGCGGATCCCGGATTCGGGTTCTTTA). The resulting PCR fragments were cloned in the pUT7 vector under the control of T7 promoter (Serganov *et al.*, 1997). The wild-type CopT PCR fragment was synthesized with primers T7-G3 (5′-GAAATTAATACGACTCAC-TATAGGGTTAAGGAATTTTGTGGCTGG; T7 promoter sequence underlined) and SeqP/II (5′-CGGATTCGGGTTCTTTA). Wild-type CopI template was synthesized with primers T7SI (5′-GAAATT-AATACGACTCACTATAGGGCCCCGGTAATCTTTTCGT) and T7EI (5′-AAACCCCGATAATCTTCTTCA), and mutated CopI species were synthesized with 5′ primers containing the corresponding mutations. Finally, the wild-type and mutant (H1, H2, H4, H5, H6, L1 and L2) CopA templates were synthesized with primers T7SA (5′-GAAATTAATAC-GACTCACTATAGTAGCTGAATTGTTGGCTATACG) and T7EA (5′-AAAGCAAAAACCCCGATAATCTTC), and for CopA-H3 the 3′ primer was T7EA-short (5′-AAAGCAAAAACCCCGATAAT).

RNA preparation and labelling

Mutant CopT RNAs were synthesized with T7 RNA polymerase using *Bam*HI-linearized pUT7 plasmids. Wild-type CopT and wild-type or mutant CopA and CopI were transcribed from PCR-generated DNA fragments (Hjalt and Wagner, 1992). Purification of RNAs was performed either by fast protein liquid chromatography (FPLC; Pharmacia) on a Bio-Sil TSK250 column or by polyacrylamide–urea gel electrophoresis as described previously (Kolb *et al.*, 2000). Transcription of CopT yields a run-off product of 302 nucleotides, initiated with GG instead of the GU sequence of the wild-type *repA* mRNA. The CopA RNA contains a 5′ terminal G instead of an A residue. Neither of these nucleotide changes affect structure or binding properties (Kolb *et al.*, 2000).

5′-end labelling of dephosphorylated RNA was performed for 30 min at 37°C with T7 polynucleotide kinase and [γ-^{32}P]ATP (Sambrook *et al.*, 1989), and 3′-end labelling of RNA with T4 RNA ligase and [^{32}P]pCp (England and Uhlenbeck, 1978). For competition experiments, CopT RNA was uniformly labelled using [α-^{32}P]UTP in the transcription reaction. Labelled RNAs were purified by polyacrylamide–urea gel electrophoresis, eluted, and precipitated twice with ethanol. Before use, unlabelled or labelled RNAs were dissolved in RNase-free water and renaturated by incubation at 90°C for 2 min, followed by slow cooling at 20°C in TMN buffer (20 mM Tris–acetate pH 7.5, 10 mM magnesium acetate, 100 mM sodium acetate).

Determination of rate constants of stable CopA–CopT complex formation and competitive inhibition with different CopA variants

Binding rate constants of CopA–CopT pairs were measured as described previously (Persson *et al.*, 1988). RNAs R1–R4 used as competitors were purchased from Interactiva (Germany). Binding of ^{32}P-labelled CopT to an excess of unlabelled CopA (1.5 nM) was performed at 37°C in the presence of different concentrations of inhibitors (R1, R2 and R3: from 15 to 1000 nM; R4 and CopI: as indicated in Figure 4; CopI-H1: from 3 to 1000 nM). Samples were withdrawn at various time points (1–6 min), added to gel application buffer and immediately loaded onto a denaturing 5% polyacrylamide–urea gel. The gel was run at constant voltage (500 V) for 3 h and subsequently dried. Bands corresponding to the CopA–CopT complex and free CopT, respectively, were quantified using a PhosphorImager (Molecular Dynamics). The inhibition constant K_i was determined according to Tomizawa (1985), using the equation $v_o/v_i = 1 + i_o/K_i$, where v_o and v_i are the initial rates of hybridization in the absence and in the presence of inhibitory RNA, and i_o is the initial concentration of the inhibitory RNA.

Pb^{2+}-induced hydrolysis

Pb^{2+}-induced hydrolysis was carried out on end-labelled CopA or CopT, free or in complex, as described by Kolb *et al.* (2000). Hydrolysis was conducted in 20 mM HEPES–NaOH pH 7.5, 10 mM magnesium acetate, 100 mM sodium acetate, in the presence of 8 or 16 mM $Pb(OAc)_2$ for 5 min. Reactions were stopped by the addition of 50 mM EDTA, followed by ethanol precipitation. Identification of the cleavage sites was carried out according to Kolb *et al.* (2000).

In vivo repA–lacZ expression assays

RepA–LacZ fusion protein synthesis was measured in cell extracts from exponentially growing cultures. The protocol used for the experiments reported in Table II was essentially as in Berzal-Herranz *et al.* (1991), except that cultures were grown in M9 minimal medium supplied with 0.2% glucose, 0.2% casamino acids, ampicillin (100 µg/ml) and/or kanamycin (50 µg/ml), tryptophan (2 µg/ml) and thiamine (1 µg/ml), or in Luria–Bertani medium supplied with ampicillin (100 µg/ml) and/or kanamycin (50 µg/ml).

Acknowledgements

We thank Steve Lodmell, Hervé Moine and Christine Brunel for stimulating discussions and critical reading of the manuscript. F.A.K. was supported by a FEBS summer fellowship. This work was supported by grants from the Centre National de la Recherche Scientifique (CNRS), the Swedish Natural Science Research Council (NFR; E.G.H.W.) and the Swedish Research Council for Engineering Sciences (TFR; E.G.H.W.).

References

Asano,K. and Mizobuchi,K. (2000) Structural analysis of late intermediate complex formed between plasmid ColIb-P9 Inc RNA and its target RNA. *J. Biol. Chem.*, **275**, 1269–1274.

Asano,K., Kato,A., Morikawi,H., Hama,C., Shiba,K. and Mizobuchi,K. (1991) Positive and negative regulations of plasmid ColIb-P9 *repZ* gene expression at the translational level. *J. Biol. Chem.*, **266**, 3774–3781.

Asano,K., Niimi,T., Yokoyama,S. and Mizobuchi,K. (1998) Structural basis for binding of the plasmid ColIb-P9 antisense Inc RNA to its target RNA with the 5′-rUUGGCG-3′ motif in the loop sequence. *J. Biol. Chem.*, **273**, 11826–11838.

Ashraf,S.S., Ansari,G., Guenther,R., Sochacka,E., Malkiewicz,A. and Agris,P.F. (1999) The uridine in 'U-turn': Contributions to tRNA–ribosomal binding. *RNA*, **5**, 503–511.

Berzal-Herranz,A., Wagner,E.G.H. and Diaz-Orejas,R. (1991) Control of replication of plasmid R1: the intergenic region between *copA* and *repA* modulates the level of expression of *repA*. *Mol. Microbiol.*, **5**, 97–108.

Blomberg,P., Wagner,E.G.H. and Nordström,K. (1990) Control of replication of plasmid R1: the duplex between the antisense RNA, CopA and its target, CopT, is processed specifically *in vivo* and *in vitro* by RNase III. *EMBO J.*, **9**, 2331–2340.

Blomberg,P., Nordström,K. and Wagner,E.G.H. (1992) Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA

through inhibition of leader peptide translation. *EMBO J.*, **11**, 2675–2683.

Blomberg,P., Engdahl,H.M., Malmgren,C., Romby,P. and Wagner, E.G.H. (1994) Replication control of plasmid R1: disruption of an inhibitory RNA structure that sequesters the *repA* ribosome binding site permits *tap*-independent RepA synthesis. *Mol. Microbiol.*, **12**, 49–60.

Brion,P. and Westhof,E. (1997) Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137.

Chen,C., Sheng,S. and Guo,P. (2000) A dimer as a building block in assembling RNA. *J. Biol. Chem.*, **275**, 17510–17516.

Eguchi,Y. and Tomizawa,J. (1991) Complexes formed by complementary RNA stem–loops. Their formations, structures and interaction with ColE1 Rom protein. *J. Mol. Biol.*, **220**, 831–842.

England,T.E. and Uhlenbeck,O.C. (1978) 3′-terminal labeling of RNA with T4 RNA ligase. *Nature*, **275**, 560–561.

Ferrandon,D., Koch,I., Westhof,E. and Nüsslein-Volhard,C. (1997) RNA–RNA interaction is required for the formation of specific bicoid mRNA 3′ UTR Staufen ribonucleoprotein particles. *EMBO J.*, **16**, 1751–1758.

Franch,T. and Gerdes,K. (2000) U-turns and regulatory RNAs. *Curr. Opin. Microbiol.*, **3**, 159–164.

Franch,T., Petersen,M., Wagner,E.G.H., Jacobsen,J.P. and Gerdes,K. (1999) Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J. Mol. Biol.*, **294**, 1115–1125.

Givskov,M. and Molin,S. (1984) Copy mutants of plasmid R1: effects of base pair substitutions in the *copA* gene on the replication control system. *Mol. Gen. Genet.*, **194**, 286–292.

Grosjean,H., Houssier,C., Romby,P. and Marquet,R. (1998) Modulatory role of modified nucleotides in RNA loop–loop interaction. In Grosjean,H. and Benne,B. (eds), *Modification and Editing of RNA*. ASM Press, Washington, DC, pp. 113–133.

Hanahan,D. (1985) In Glover,D.M. (ed.), *DNA Cloning I*. IRL Press, Oxford, UK, pp. 109–135.

Hjalt,T. and Wagner,E.G.H. (1992) The effect of loop size in antisense and target RNAs on the efficiency of antisense RNA control. *Nucleic Acids Res.*, **20**, 6723–6732.

Hjalt,T. and Wagner,E.G.H. (1995) Bulged-out nucleotides in an antisense RNA are required for rapid target RNA binding *in vitro* and inhibition *in vivo*. *Nucleic Acids Res.*, **23**, 571–579.

Isel,C., Marquet,R., Keith,G., Ehresmann,C. and Ehresmann,B. (1993) Modified nucleotides of tRNA(3Lys) modulate primer/template loop–loop interaction in the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.*, **268**, 25269–25272.

Kolb,F.A., Malmgren,C., Westhof,E., Ehresmann,C., Ehresmann,B., Wagner,E.G.H. and Romby,P. (2000) An unusual structure formed by antisense–target RNA binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment. *RNA*, **6**, 311–324.

Lee,A.J. and Crothers,D.M. (1998) The solution structure of an RNA loop–loop complex: the ColE1 inverted loop sequence. *Structure*, **6**, 993–1005.

Malmgren,C., Engdahl,H.M., Romby,P. and Wagner,E.G.H. (1996) An antisense/target RNA duplex or a strong intramolecular RNA structure 5′ of a translation initiation signal blocks ribosome binding: the case of plasmid R1. *RNA*, **2**, 1022–1032.

Malmgren,C., Wagner,E.G.H., Ehresmann,C., Ehresmann,B. and Romby,P. (1997) Antisense RNA control of plasmid R1 replication: the dominant product of the antisense RNA–mRNA binding is not a full RNA duplex. *J. Biol. Chem.*, **272**, 12508–12512.

Marino,J.P., Gregorian,R.S., Csankovski,G. and Crothers,D.M. (1995) Bent helix formation between RNA hairpins with complementary loops. *Science*, **268**, 1448–1454.

Miller,J.H. (1972) In *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 352–355.

Nordström,K. and Wagner,E.G.H. (1994) Kinetics aspects of control of plasmid replication by antisense RNA. *Trends Biochem. Sci.*, **19**, 294–300.

Nordström,K., Molin,S. and Light,J. (1984) Control of replication of bacterial plasmids: genetics, molecular biology and physiology of the plasmid R1 system. *Plasmid*, **12**, 71–90.

Öhman,M. and Wagner,E.G.H. (1989) Secondary structure analysis of the RepA mRNA leader transcript involved in control of plasmid R1. *Nucleic Acids Res.*, **17**, 2557–2579.

Öhman,M. and Wagner,E.G.H. (1991) Regulation of replication of plasmid R1: an analysis of the intergenic region between *copA* and *repA*. *Mol. Gen. Genet.*, **230**, 321–328.

Paillart,J.C., Marquet,R., Skripkin,G., Ehresmann,C. and Ehresmann,B. (1996) Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie*, **78**, 639–653.

Patzel,V. and Sczakiel,G. (1998) Theoretical design of antisense RNA structures substantially improves annealing kinetics and efficacy in human cells. *Nature Biotechnol.*, **16**, 64–68.

Persson,C., Wagner,E.G.H. and Nordström,K. (1988) Control of replication of plasmid R1: kinetics of *in vitro* interaction between the antisense RNA, CopA and its target, CopT. *EMBO J.*, **7**, 3279–3288.

Persson,C., Wagner,E.G.H. and Nordström,K. (1990a) Control of replication of plasmid R1: structures and sequences of the antisense RNA, CopA, required for its binding to the target RNA, CopT. *EMBO J.*, **9**, 3767–3775.

Persson,C., Wagner,E.G.H. and Nordström,K. (1990b) Control of replication of plasmid R1: formation of an initial transient complex is rate-limiting for antisense RNA–target RNA pairing. *EMBO J.*, **9**, 3777–3785.

Pyle,A.M. and Green,J.B. (1995) RNA folding. *Curr. Opin. Struct. Biol.*, **5**, 303–310.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Serganov,A., Rak,A., Garber,M., Reinbolt,J., Ehresmann,B., Ehresmann,C., Grunberg-Manago,M. and Portier,C. (1997) Ribosomal protein S15 from *Thermus thermophilus*: Cloning, sequencing, overexpression of the gene and RNA-binding properties of the protein. *Eur. J. Biochem.*, **246**, 291–300.

Siemering,KR., Praszkier,J. and Pittard,A.J. (1993) Interaction between the antisense and target RNAs involved in the regulation of IncB plasmid replication. *J. Bacteriol.*, **175**, 2895–2906.

Siemering,K.R., Praszkier,J. and Pittard,A.J. (1994) Mechanism of binding of the antisense and target RNAs involved in the regulation of IncB plasmid replication. *J. Bacteriol.*, **176**, 2677–2688.

Stougaard,P., Light,J. and Molin,S. (1982) Convergent transcription interferes with expression of the copy number control gene, CopA, from plasmid R1. *EMBO J.*, **1**, 323–328.

Tomizawa,J. (1984) Control of ColE1 plasmid replication: The process of binding of RNA I to the primer transcript. *Cell*, **38**, 861–870.

Tomizawa,J. (1985) Control of ColE1 plasmid replication: initial interaction of RNAI and the primer transcript is reversible. *Cell*, **40**, 527–535.

Tomizawa,J. (1990) Control of ColE1 plasmid replication. Interaction of Rom protein with an unstable complex formed by RNA I and RNA II. *J. Mol. Biol.*, **212**, 695–708.

Wagner,E.G.H. and Nordström,K. (1986) Structural analysis of an RNA molecule involved in replication control of plasmid R1. *Nucleic Acids Res.*, **14**, 2523–2538.

Wagner,E.G.H. and Simons,R.W. (1994) Antisense RNA control in bacteria, phages and plasmids. *Annu. Rev. Microbiol.*, **48**, 713–742.

Wagner,E.G.H., Blomberg,P. and Nordström,K. (1992) Replication control in plasmid R1: duplex formation between the antisense RNA, CopA and its target, CopT, is not required for inhibition of RepA synthesis. *EMBO J.*, **11**, 1195–1203.

Wilson,I.W., Praszkier,J. and Pittard,A.J. (1993) Mutations affecting pseudoknot control of the replication of B group plasmids. *J. Bacteriol.*, **175**, 6476–6483.

Zeiler,B.N. and Simons,R.W. (1998) Antisense RNA structure and function. In Simons,R.W. and Grunberg-Manago,M. (eds), *RNA Structure and Function*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 437–464.

Received July 5, 2000; revised September 4, 2000; accepted September 8, 2000

IV. Article IV : Conservation de la jonction à quatre hélices dans les complexes impliqués dans le contrôle de la réplication plasmidique

De nombreux plasmides régulent leur taux de réplication par un mécanisme en partie similaire à celui décrit pour le plasmide R1. Les séquences cibles des ARNm codant pour la protéine de réplication sont présentés dans la figure 1 de l'article IV. Les boucles par lesquelles l'interaction avec l'ARN antisens est initiée présentent toutes une séquence hyper-conservée 5'-UGGC-3'. Par ailleurs toutes les tiges majeures sont longues et contiennent des nucléotides non appariés importants pour le contrôle [61,337], comme dans le cas de CopA-CopT.

A. Objectifs

Tous les déterminants nécessaires à la formation de la jonction à quatre hélices dans le système CopA-CopT sont présents dans les ARN de systèmes homologues. Ceci suggère que le mécanisme de reconnaissance et la topologie du complexe inhibiteur doivent être conservés. Récemment, Asano & Mizobuchi ont proposé la formation d'une jonction à quatre hélices dans le complexe stable formé entre l'ARN antisens Inc et l'ARNm *repZ* dans le système Col1b-P9 [310]. L'architecture qu'ils proposent diffère néanmoins de celle du complexe CopA-CopT. En effet, les deux hélices intermoléculaires comptent toutes deux 12 paires de bases et les boucles les connectant contiennent 3 résidus. La jonction est donc décalée de 3 paires de bases vers le bas par rapport à la topologie que nous proposons dans l'article I. Un critère de validité pour le modèle CopA-CopT est la conservation d'une telle topologie entre systèmes homologues. C'est pourquoi, nous avons étudié le complexe stable Inc-*repZ* à l'aide de deux sondes de structure qui donnent une signature caractéristique de la jonction à quatre hélices (Pb^{2+} et RNase V1). Les ions Pb^{2+} permettent de définir les régions simple-brin alors que la RNase V1 permet de visualiser les hélices B et B'. La modélisation graphique a été essentielle pour définir le modèle Inc-*repZ* qui soit possible au niveau topologique et qui tienne compte de l'ensemble des données expérimentales.

B. Résultats

Les réactivités du complexe Inc-*repZ* vis-à-vis de la RNase V1 et des ions Pb^{2+} sont semblables à celles observées pour le complexe CopA-CopT. Dans les deux cas, les boucles connectant les deux hélices intermoléculaires comportent six résidus, comme l'indique l'accessibilité des ces nucléotides aux ions Pb^{2+}. Cette sonde génère également une forte coupure dans chacun des deux ARN Inc et *repZ* et qui coïncident avec la position de la

jonction de CopA-CopT. Il reflète probablement un site de fixation pour les cations divalents au cœur de la jonction dont le rôle est de stabiliser cette dernière.

Le modèle de structure tridimensionnelle du complexe Inc-*repZ* construit à partir de nos résultats de cartographie et sur les résultats d'Asano & Mizobuchi est entièrement superposable au modèle du complexe CopA-CopT. En revanche, sa topologie diffère de celui construit à partir des résultats d'Asano & Mizobuchi. Nous montrons que ce dernier n'est pas topologiquement viable. En effet, la taille supérieure des hélices B et B' impose la formation d'un nœud entre les deux ARN, ce qui ne peut pas avoir de réalité biologique. Ce résultat met en avant l'importance de la modélisation graphique.

C. Conclusions

Nous avons réalisé un alignement entre les séquences des ARNm codant pour les protéines de réplication de nombreux plasmides homologues à R1. Tous partagent les caractéristiques de séquence (boucles conservées) et de structure (présence des nucléotides non appariés) nécessaires à la formation de la jonction à quatre hélices dans le système R1. Dans cette étude, nous avons montré que cette structure était formée pour deux de ces systèmes. Malgré des différences de séquence, les deux modèles tridimensionnels sont superposables. Des contraintes topologiques liées à la formation de l'hélice intermoléculaire B définissent probablement la position de la jonction à quatre hélices. Ainsi, il est vraisemblable que dans toutes ces familles de plasmides, les complexes formés entre l'ARN antisens et sa cible présentent tous la même topologie. De plus, le mécanisme de formation du complexe, tel que nous l'avons décrit dans l'article III, doit lui aussi être conservé. Cette étude met également en avant l'importance de l'étape de conversion de l'interaction initiale vers la jonction à quatre hélices. Les plasmides dont les ARNm codant pour les protéines de réplication sont alignés dans la figure 1 de l'article IV sont tous compatibles entre eux. Ceci signifie qu'un ARN antisens codé par un plasmide ne peut pas interagir de manière croisée avec la séquence de type CopT codée par un plasmide appartenant à un groupe d'incompatibilité différent. Comme les séquences des boucles apicales sont hyper-conservées entre tous ces plasmides, l'interaction initiale instable est maintenue entre des ARN hétérologues. En revanche, les séquences impliquées dans la formation des hélices B et B' divergent. Par conséquent, en plus de garantir une formation rapide du complexe stable et donc un contrôle optimal, l'étape de conversion est également responsable de la spécificité d'interaction entre les ARN.

Il est intéressant de noter que parmi ces plasmides, la manière dont l'ARN antisens contrôle le taux de synthèse de la protéine de réplication varie. En effet, chez les plasmides de type Col1b-P9, l'ARN antisens inhibe la formation d'un pseudo-noeud activateur de la

traduction (voir § VI.D.4 de l'Introduction). Dans ce cas, les séquences de *repZ* impliquées dans la formation du pseudo-noeud sont également celles impliquées dans la formation de l'hélice intermoléculaire B ([338], cette étude). Ces deux interactions sont donc mutuellement exclusives. Ainsi, les résultats présentés ici mettent en exergue l'importance de l'étape de conversion de l'interaction boucle-boucle initiale vers une topologie commune pour la fonction des ARN antisens de R1 et de Col1b-P9, malgré des modes d'action sensiblement différents. Des ARN pouvant former une telle structure ont dû être sélectionnés au cours de l'évolution parce qu'ils peuvent interagir de façon rapide et optimale. Ces caractéristiques sont requises pour un contrôle efficace de la réplication plasmidique.

doi:10.1006/jmbi.2001.4677 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **309**, 605–614

JMB

Four-way Junctions in Antisense RNA-mRNA Complexes Involved in Plasmid Replication Control: A Common Theme?

Fabrice A. Kolb[1], Eric Westhof[1], Bernard Ehresmann[1] Chantal Ehresmann[1], E. Gerhart H. Wagner[2] and Pascale Romby[1]*

[1]*UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 Rue R. Descartes, F-67084, Strasbourg Cedex France*

[2]*Institute of Cell and Molecular Biology, Department of Microbiology, Uppsala University, Box 596, Husargatan 3, S-75124 Uppsala, Sweden*

**Corresponding author*

In several groups of bacterial plasmids, antisense RNAs regulate copy number through inhibition of replication initiator protein synthesis. In plasmid R1, we have recently shown that the inhibitory complex between the antisense RNA (CopA) and its target mRNA (CopT) is characterized by the formation of two intermolecular helices, resulting in a four-way junction structure and a side-by-side helical alignment. Based on lead-induced cleavage and ribonuclease (RNase) V_1 probing combined with molecular modeling, a strikingly similar topology is supported for the complex formed between the antisense RNA (Inc) and mRNA (RepZ) of plasmid Col1b-P9. In particular, the position of the four-way junction and the location of divalent ion-binding site(s) indicate that the structural features of these two complexes are essentially the same in spite of sequence differences. Comparisons of several target and antisense RNAs in other plasmids further indicate that similar binding pathways are used to form the inhibitory antisense-target RNA complexes. Thus, in all these systems, the structural features of both antisense and target RNAs determine the topologically possible and kinetically favored pathway that is essential for efficient *in vivo* control.

© 2001 Academic Press

Keywords: antisense RNA; replication control; plasmid; RNA-RNA interaction; RNA modeling

Introduction

Most bacterial antisense RNAs are post-transcriptional regulators of processes that require rapid and adaptative responses.[1] Most of these have been identified as regulators of plasmid copy number. These RNAs are small and untranslated, but extremely efficient inhibitors. Usually, their exquisite efficiency with respect to *in vivo* control correlates with high *in vitro* binding rates to their respective target RNAs. Despite the fact that these antisense RNAs are fully complementary to their targets, formation of full duplexes is too slow to account for the observed biological effect. Instead, recent data indicate that many antisense RNAs mediate inhibition by forming complexes that involve limited numbers of base-pairs with their targets. In almost all of these cases, kinetic studies have shown that the initial step of binding involves a loop-loop interaction mediated by Watson-Crick base-pairs, the so-called kissing complex (ColE1,[2] R1,[3] pMU720,[4] and ColIb-P9[5,6]). Although all these systems use rapid kissing complex formation as a prerequisite for proper control, further steps are required to form a stable, inhibitory complex.

Plasmid R1 belongs to a family of plasmids that differ in primary sequence and, in some cases, also in the mode of regulation (see below), but share a similar overall genetic organization of their replication control. Replication of plasmid R1 is controlled at the translational level by binding of the antisense RNA (CopA) to its target site (CopT) in the leader region of the *repA* mRNA,[7] thereby inhibiting the synthesis of the replication initiator protein RepA.[8–10] We recently proposed a CopA-CopT binding pathway as a stepwise progression from free RNAs to a stable and inhibitory complex

Abbreviations used: CopA, antisence RNA of plasmid R1; CopT, target mRNA of CopA; RNase, ribonuclease; Inc, antisense RNA of plasmid Col1b-P9; RepZ, target mRNA of Inc.

E-mail address of the corresponding author: P.Romby@ibmc.u-strasbg.fr

0022-2836/01/030605–10 $35.00/0 © 2001 Academic Press

through a hierarchy of experimentally distinguishable intermediates.[11] The major product of this binding reaction adopts an unusual structural motif characterized by a four-helix junction, the formation of which involves extensive breakage of intramolecular base-pairs to promote the formation of two intermolecular helices.[12] This process is directional, since one of the two intermolecular helices must form first to allow formation of the other.[11] The resulting four-way junction structure induces a side-by-side helical alignment further stabilized by a third long intermolecular helix involving the 5′ tail of CopA and the complementary region of CopT.[12] This structure is the stable, inhibitory complex.[10] From studies of the CopA/CopT system we infer that the structural features of both RNAs determine the topologically possible and kinetically favored pathway.

Interestingly, several copy number control antisense RNAs that all inhibit translation of replication initiator proteins have overall structural features similar to those of CopA in plasmid R1. These antisense RNAs, and their targets, carry imperfectly base-paired major stems closed by conserved six-membered loops at which binding initiates (Figure 1 and references therein). Thus, the antisense-target RNA pairs can be expected to encounter the same topological constraints during stable complex formation and may use binding and folding pathways with characteristics like those described for CopA-CopT. Asano & Mizobuchi[6] recently reported on the structure of the late intermediate antisense-target RNA complex in ColIb-P9, based on enzymatic probing experiments. These authors found an overall topology similar to that of CopA-CopT; however, the lengths of the two intermolecular helices differed, such that the position of the four-way junction was shifted by three base-pairs. Since the 3D model of the CopA-CopT complex indicated that such a junction shift could not be accommodated, we have reinvestigated the structure of the inhibitory Inc-RepZ complex of the ColIb-P9 plasmid by lead-induced cleavages and the double-strand-specific ribonuclease (RNase) V_1. These two probes provided a characteristic signature for the extended kissing CopA-CopT complex (i.e. the four-way junction), which was supported by extensive site-directed mutagenesis.[11,12] We show here that the stable complexes of CopA-CopT and Inc-RepZ adopt very similar topologies and that their four-way junctions are located at equivalent positions. Furthermore, the data support divalent metal ion-binding sites at these junctions in both complexes. Thus, we propose that similar topological constraints in antisense-target RNA complexes of many related plasmids (see Figure 1) determine very similar binding pathways and structures of complexes, as exemplified by CopA-CopT and Inc-RepZ.

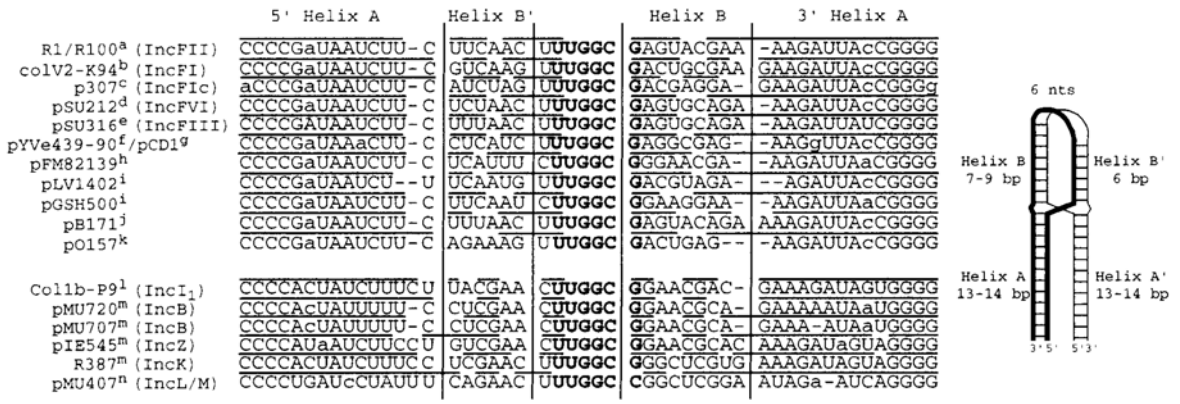

Figure 1. Structural alignment of the major stem-loop regions of the target RNAs of several closely related plasmids and schematic model for the antisense RNA-mRNA complex. Plasmids and their respective incompatibility groups when known are indicated in parenthesis. The target RNA sequences were divided in two parts (R1-related and Col1b-P9-related plasmids) according to their ability, or not, to form a pseudo-knot structure. Sequences are aligned relative to their 5′-UUGGCG-3′ loop sequences. Secondary structure elements for all target RNAs are shown as follows: the loop nucleotides are represented in bold characters and nucleotides involved in Watson-Crick pairing are indicated by horizontal bars. In helix A, potential non-canonical base-pairs are denoted by lower-case letters. Elements of the proposed four-way junctions: the two intermolecular helices B and B′, and the intramolecular helix A are delimited by vertical bars. These helices are shown on a schematic model derived from both CopA-CopT and Inc-RepZ complexes (right panel). The positions of the intermolecular helices in all plasmids have been inferred based on experimental data and geometrical constraints obtained on CopA-CopT and Inc-RepZ complexes. In most of these systems, helix B is predicted to contain seven to nine base-pairs, and helix B′ six base-pairs. References for plasmids are from (a) Ryder *et al.*,[38] (b) Weber *et al.*,[39] (c) Saadi *et al.*,[40] (d) EMBL Genbank X55895, (e) Lopez *et al.*,[41] (f) Vanootegham & Cornelis,[42] (g) Perry *et al.*,[43] (h) Rodriguez-Pena *et al.*,[44] (i) Osborn *et al.*,[45] (j) Tobe *et al.*,[46] (k) Burland *et al.*,[47] (l) Hama *et al.*,[48] (m) Praszkier *et al.*,[32] and (n) Athanasopoulos *et al.*[49]

Results and Discussion

Lead and RNase V_1 probing reveal the signature of the four-way junction in Inc-RepZ

We have previously shown that lead-induced hydrolysis and RNase V_1 probing of the CopA-CopT complex provided a characteristic signature of the four-way junction. These two probes gave complementary data, since Pb^{2+}-induced hydrolysis occurs preferentially at unpaired residues and is very sensitive to subtle RNA rearrangements, whereas RNase V_1 is specific for helical regions. Indeed Pb^{2+} cleavages were restricted to unpaired residues located in the loops connecting the two intermolecular helices B and B′, and to residues located at the cross-over sites of the four-way junction in the CopA-CopT complex.[12] Characteristic RNase V_1 cleavages were observed in helices B′ and B in the CopA-CopT complex (Figure 2(d)). Identical experiments were performed on the Inc/RepZ complex of plasmid ColIb-P9, using either end-labeled Inc or repZ RNAs. Autoradiograms of probing experiments are shown in Figure 3, and cleavage positions are shown schematically on the secondary structures of the RNAs (Figure 2(a)-(c)).

In free RepZ and Inc RNAs, the main lead-cleavages occurred at unpaired residues. They also extended into the upper stems that contain several bulged residues (Figures 2 and 3). Strikingly, the cleavage pattern obtained on the RepZ-Inc complex is fully comparable to that of the CopA-CopT complex.[13] Binding of Inc RNA induced strong protections in regions U318-A323 and A332-C337 of RepZ RNA, whereas residues C324-C329 were still susceptible to hydrolysis, with cleavage at C329 even increased (Figure 3(a)). In addition, a strong cleavage appeared at C316. Conversely, binding of RepZ mRNA induced main protections in regions C31-C36 and U44-A49 of Inc RNA whereas strong cleavages remained at positions C38-A41 in the hairpin loop. An enhanced cut was observed at C28 (Figure 3(b)). Finally, the formation of the two intermolecular helices (B and B′ by analogy with CopA-CopT) was supported by several RNase V_1 cuts at positions U33-C36 of Inc RNA (Figure 3(c)) and G321-A322 in RepZ mRNA (results not shown).[6] Thus, these results support the formation of the two intermolecular helices (B and B′, Figure 2(d)) connected by single-stranded regions (G37 to A41 in Inc, C324 to C329 in RepZ). These data correlate well with the probing experiment performed on RepZ bound to Inc using several single-strand-specific RNases.[6] Cleavages still occurred at residues U326 to G328 of RepZ bound to Inc, whereas cleavages at positions U318, A319, A322 and A323 became protected. Similar enzymatic cleavage patterns were obtained on the inhibitory RNAI-RNAII complex of plasmid pMU720 (an IncB plasmid, and a close relative of ColIb-P9), where the antisense RNAI carries a loop sequence identical with that of Inc and CopA RNAs (Figure 1).[4] Indeed, the binding of RNAII induced several identical RNase V_1 cleavages at the 5′ side of the hairpin loop of RNAI, and guanines 37 and 42 were accessible to RNase T_1.

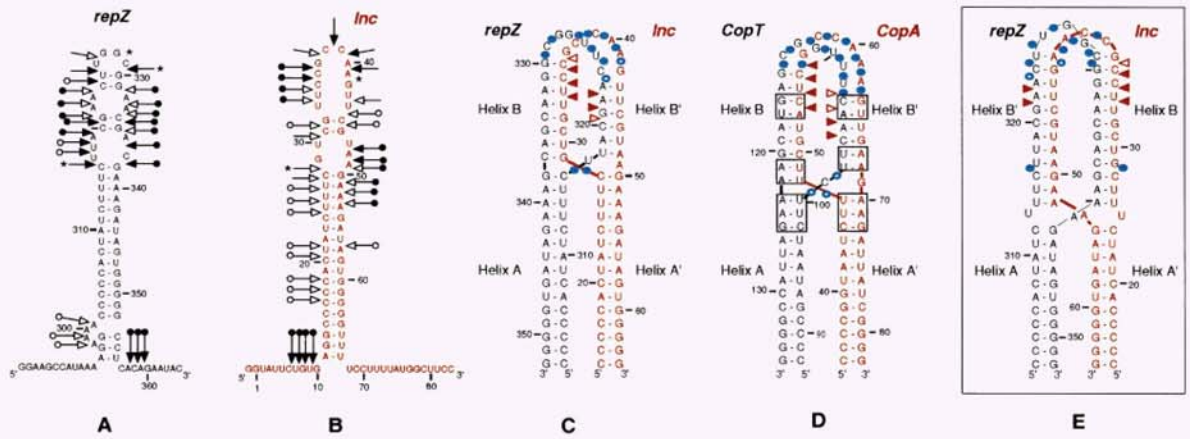

Figure 2. Secondary structure of antisense and target RNA of ColIb-P9, and complexes formed. Structures of the RepZ target (a), the antisense RNA Inc (b), and the complex (c) based on the experimental results in this work. As a comparison, the CopA-CopT model[12] (d), and the Inc-RepZ model as proposed by Asano & Mizobuchi[6] (e) are shown. The secondary structure model for the Inc-RepZ complex (c) was derived from the lead-induced cleavages and RNase V_1 probing and took into account the data from Asano & Mizobuchi.[6] (a) and (b) Strong and moderate Pb^{2+} cleavages are shown by filled and open arrows, respectively. Effect of antisense or target RNA binding: strong and moderate protections are shown by filled and open spheres, respectively, and increased or new cleavages are shown by stars. (c)-(e) strong and moderate Pb^{2+}-induced cleavages in the Inc-RepZ complex are indicated by filled or open blue circles, respectively. New RNase V_1 cleavages are represented by red arrows. (d) The position of the mutations performed in CopA or CopT that disrupt complementarity have been squared on the secondary structure model of CopA-CopT.[11]

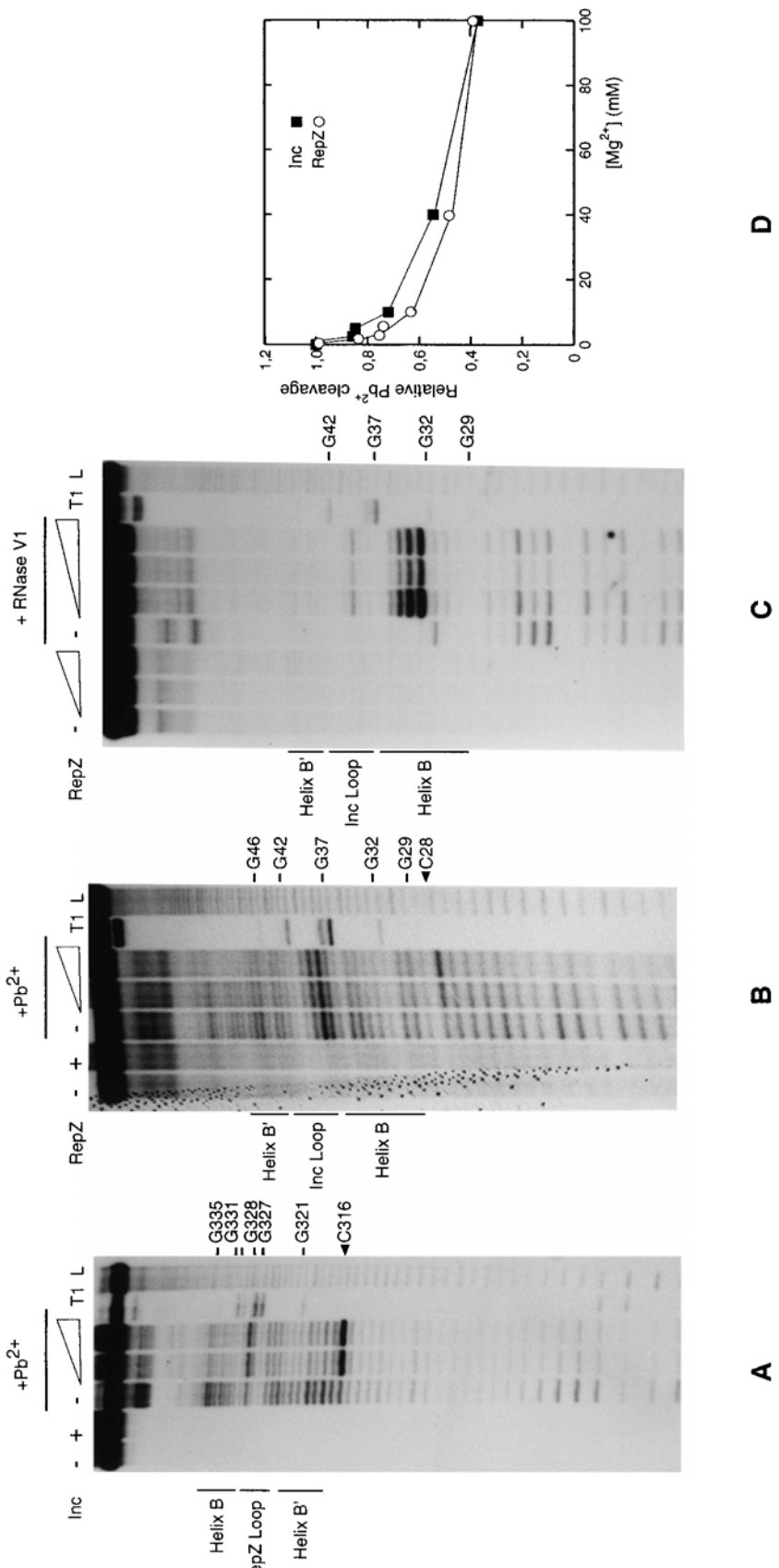

Figure 3. Probing the structure of the Inc-RepZ complex. Pb^{2+}-induced hydrolysis was performed on (a) 5′ end-labeled RepZ alone (–) or in the presence of an excess of Inc RNA (+), and (b) on 5′ end-labeled Inc alone (–) or in the presence of an excess of RepZ RNA (+). Complex formation was performed at 37 °C for ten minutes in TMN buffer. Lanes T1, L, RNases T_1 and alkaline ladders, respectively. (c) RNase V_1 probing on 5′ end-labeled Inc RNA alone (–) or bound to RepZ. Open triangles denoted increasing concentrations of RepZ. (d) Effect of magnesium concentration on Pb^{2+} hydrolysis performed on the Inc-RepZ complex. The intensities of the cleavages at C28 in Inc (square) or at C316 in RepZ (circle), within the complex, are depicted as a function of the magnesium concentration. The intensities of cleavages were quantified using a PhosphorImager (Fuji) and were normalized to the signal intensity of full-length RNA.

Evidence for a divalent metal ion-binding site at the four-way junction

Divalent cations have been shown to stabilize RNA four-way junctions in a precise conformation.[14,15] The 3D model of the CopA-CopT complex indicated a stacked cruciform junction.[12] Such a negatively charged cavity is likely to bind divalent ions with high affinity. The observation of several Pb^{2+}-cleavages at nucleotides forming the junction (Figure 2(d)),[12] which were strongly reduced by the addition of Mg^{2+}, was indicative of competition with divalent cations.[16,17] We have also recently shown that neomycin B, an aminoglycoside that binds at specific divalent metal ion-binding sites in RNAs,[18,19] was able to compete with the Pb^{2+}-induced cleavages at the CopA/CopT junction (data not shown).

Figure 3 shows that two strong Pb^{2+} cleavages were identified at C28 of Inc and C316 of RepZ within the complex. Interestingly, these two cleavages occur at positions that are structurally equivalent to those displayed in the four-way junction model of the CopA-CopT complex (Figure 2(d)). In both complexes, the addition of increasing concentrations of Mg^{2+} resulted in strong reduction of these lead cleavages; 50 % of inhibition was obtained at a sixfold molar excess of Mg^{2+} over Pb^{2+} (Figure 3(d)). Such a competition between Mg^{2+} and Pb^{2+} was also described for M1 RNA of RNase P,[20] where experimental evidence supports the presence of a divalent cation-binding pocket.[21,22] Thus, the strong lead cleavages at C28 and C316 most likely result from a combination of favorable divalent metal ion-binding site and appropriate stereochemistry for phosphodiester cleavage, as described for different RNAs.[16] These two cleavages are therefore not compatible with residues C28 and C316 being in a regular helical conformation (Figure 2(e)). The presence of metal ion-binding sites at equivalent positions in RepZ-Inc and CopA-CopT complexes indicates a similar four-way junction topology. Thus, these Pb^{2+} cleavages can serve as an experimental signature of the position of the junction in these complexes, since the cross-over point in CopA-CopT is independently confirmed by site-directed mutagenesis.[11,12] Thus, we propose that lead-induced cleavage analysis of the functional, stable RNA-RNA complexes may be a useful tool to map such types of four-way junction.

Topological constraints determine the structure and location of the four-way junction

The data presented above permit us to define the boundaries of the two newly formed intermolecular helices B and B′ connected by unpaired residues (G37 to A41 in Inc, C324 to C329 in RepZ). The two single-stranded regions are susceptible to lead-induced cleavages, and the RNase V_1 cleavage pattern supports the two helices B and B′. The resulting secondary structure model of the Inc-RepZ complex is shown in Figure 2(c). The two intermolecular helices (B and B′) are proposed to contain eight and six base-pairs, respectively. The overall form of the resulting model is that of an asymmetrical cruciform structure. The close proximity of the two helical domains at the four-way junction may explain why several RNase V_1 cuts at positions C312 to U314 of RepZ were protected by Inc binding.[6] Again, similar results were also found in CopT bound to CopA.[12] The resulting model is different from that proposed by Asano & Mizobuchi[6] (Figure 2(e)), who suggested two 12 base-pair long intermolecular helices connected by loops of three nucleotides. In CopA-CopT, the number of base-pairs in helices B and B′ was additionally determined by site-directed mutagenesis. Base-pair inversions located above the lower bulged residues in CopA or CopT strongly altered the formation of the two intermolecular helices, and concomitantly decreased binding rate for complex formation *in vitro* and control *in vivo*[11,12] (Figure 2(d)). Wild-type-like behavior was restored by using fully complementary mutant CopA and CopT. Thus base-pairing *per se*, and not the nucleotide sequence in this region, is the critical determinant for the rate of complex formation. Furthermore, inversion of three base-pairs in the intramolecular helices A and A′, immediately below the junction, neither affected the topology of the four-way junction nor the *in vivo* inhibitory efficiency. These data indicate that the corresponding nucleotides are not involved in intermolecular base-pairing between CopA and CopT.[11] Thus, structural and genetic experiments support the CopA-CopT model (Figure 2(d)). Since the experiments presented here indicate that Pb^{2+} cleavages occurred at equivalent positions at the junctions of both the CopA-CopT and the RepZ-Inc complexes, they strongly suggest that in both cases, helix propagation should be arrested at the lower bulged nucleotides of the interacting RNAs to form the four-helix junction (Figure 2(c) and (d)).

Thus, the overall topology of the Inc/RepZ complex is suggested to be similar to that of the CopA/CopT complex, except that the length of helix B is shorter by one base-pair in the Inc-RepZ complex as compared to the CopA-CopT complex (Figure 3(c) and (d)). This is apparently due to the absence of one nucleotide on the 5′ side of the hairpin loop of Inc as compared to CopA (Figure 1). A 3D model of the Inc-RepZ complex was assembled based on experimental data and stereochemical constraints in order to discard improbable folding schemes (Figure 4(a)). As previously discussed,[12] two types of junctions were possible, in principle, since the two helical domains, generated by coaxial stacking, could be constituted by either B-A and B′-A′ (junction 1 in Figure 3(c)), or by B′-A and B-A′ (junction 2, not shown) arrangements. As for CopA-CopT, junction 2 was rejected because topological constraints prevented connection of helices B and B′ *via* the single-stranded regions. The overall form of the resulting Inc-RepZ model is that of

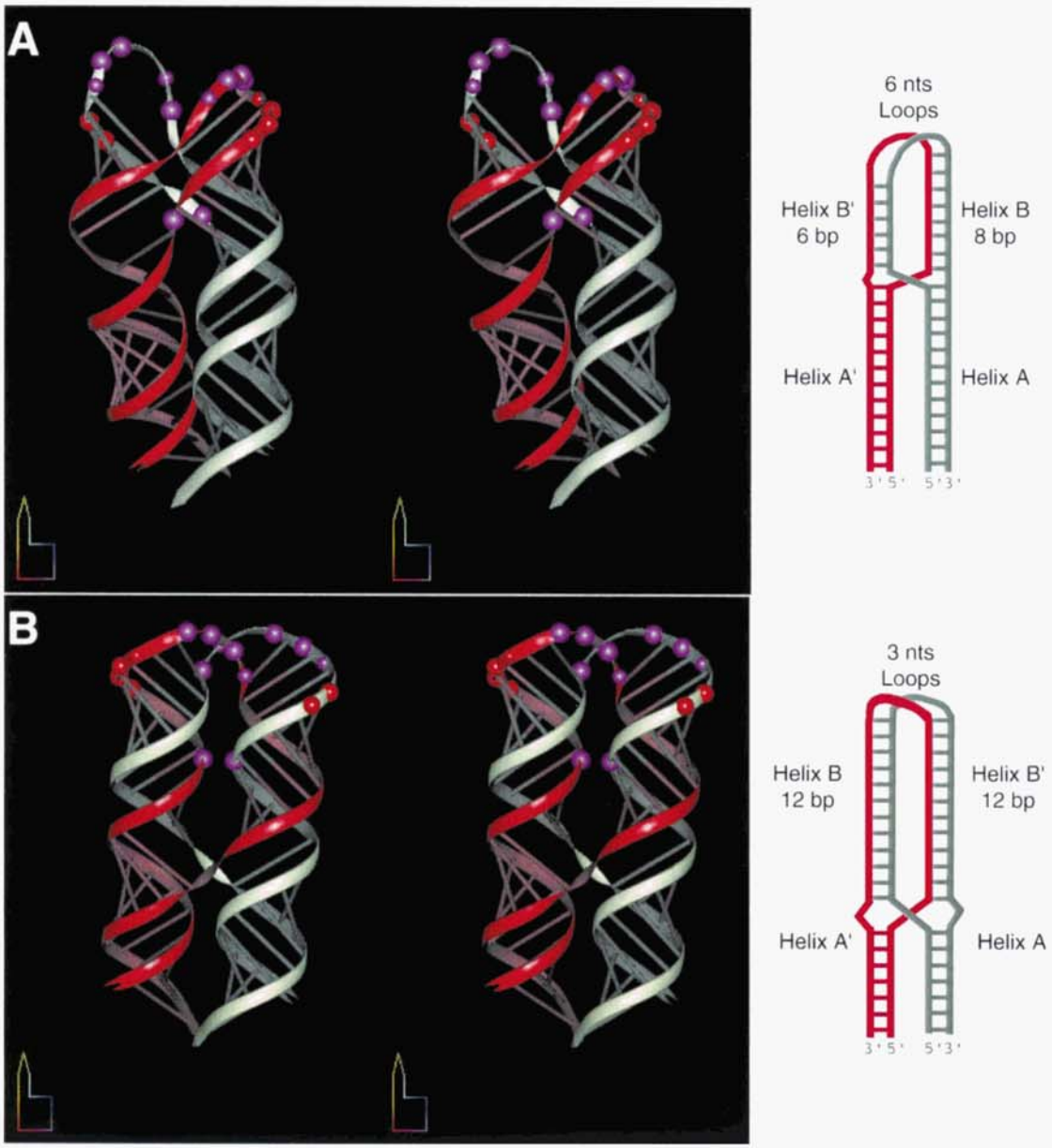

Figure 4. The 3D models of the Inc-RepZ complex. Stereoviews of the 3D models of the Inc-RepZ complex derived from the experimental data (a) and as proposed by Asano & Mizobuchi[6] (b). In the latter model, a knotted structure is formed. Inc RNA [4-54] is in red and RepZ RNA [303-351] in white (numbering according to Asano & Mizobuchi[6]). RNase V_1 and Pb^{2+} cleavages are represented by red and purple dots, respectively. The models were drawn with the program DRAWNA.[37] For clarity, schematic 2D views of the Inc-RepZ complexes are shown (right panel).

an asymmetrical cruciform structure with a side-by-side helical alignment (Figure 4(a)). This topology is identical with that previously deduced for the CopA-CopT complex.[12] Although the six nucleotide-long loops appear to be optimal to connect helices B and B′, significant variation in local geometry for both loops may occur because of the limited number of constraints. In contrast, graphic modeling of the model proposed by Asano & Mizobuchi[6] revealed that the extension of helices B and B′ by four and six base-pairs, respectively, interferes with the connection of the loops to the intermolecular helices (Figure 4(b)). These extensions force the extremities of the two intermolecular helices to be rotated by 132° (4 × 33°) and 198° (6 × 33°), respectively, relative to the four-way junction. Consequently, the two helices cannot be connected without crossing of the two loops, which would imply knot formation (Figure 4(b)). In line with this, we were also unable to connect the two loops in the CopA-CopT complex by extending helices B and B′ even by only one or two

base-pairs (data not shown). Thus, we conclude that both the localization of lead cleavages (Figure 3(c) and (e)) and the geometrical, and structural constraints favor our proposed base-pairing scheme for the Inc-RepZ complex (Figure 4(a)). However, additional mutational analyses, as carried out for CopA-CopT,[11] will be required to confirm and validate the present model and rule out other models.[6] Furthermore, determination of the crystal structure of either CopA-CopT or Inc-RepZ complexes should provide detailed structural information on these inhibitory complexes.

A conserved step-wise binding pathway

We have recently proposed a binding pathway for plasmid R1 that is characterized by a hierarchy of intermediates that leads to a stable and inhibitory CopA-CopT complex.[11] The structural features of CopA and CopT determine the topologically possible and kinetically favored pathway. There are good indications that other plasmids, closely and distantly related to R1, share similar properties with respect to steps in the binding pathway and structures formed. Structural and functional studies on the inhibitory RNA-RNA complexes from plasmids pMU720 (IncB) and ColIb-P9 (IncIα) have indicated similarities to that of plasmid R1. In these three systems, the antisense and target stem-loops contain identical loop sequences, bulged residues in the upper stems, and a single-stranded region adjacent to the major stem-loops (Figure 1). Furthermore, full duplex formation is not required for control. Instead, a stable antisense-target RNA complex is formed that involves inter-strand pairing within the upper stems of antisense and target RNAs (pMU720,[4,23] Col1b-P9[5]). We showed here that the stable Inc-RepZ complex adopts a topology very similar to that present in CopA-CopT. Hence, the structural similarity of both complementary RNAs and the complexes formed argues that the same binding pathway is followed by CopA-CopT and Inc-RepZ and, by analogy, to the other members of this plasmid family.

In all these systems, the first recognition step is a loop-loop interaction that involves a subset of bases in the hairpin loops of both RNAs. This interaction, which is fully reversible and most likely very unstable,[11] depends mainly on the loop structure of both RNAs,[24] which presents the sites for initial recognition in a proper conformation. The sequences of the loops in both RNAs are strictly conserved in all these plasmids (Figure 1). It is noteworthy that the target loops contain a highly conserved UUNG motif, which was proposed to form a U-turn structure[6,25] (Figure 1) that may position the downstream bases in a pre-formed A-helical conformation. The ubiquitous occurrence of this structural motif in antisense/target RNA systems has been noted and was suggested to reflect the requirement for binding specificity and rate.[25] Subsequent helix progression then proceeds into the upper stems, resulting in disruption of the initiating base-pairs and formation of two intermolecular helices. We have previously shown that helix progression proceeds unidirectionally in CopA-CopT, such that helix B is formed first and is a prerequisite for all further steps.[11] Only mutations that altered helix B but not helix B′ in CopA or CopT affected binding and regulation. Also, an RNA fragment complementary to the 3′ side of the CopT stem, but not the 5′ side, competed efficiently with CopA for CopT binding.[11] This conclusion is supported by work on pMU720 (IncB): changes at residues located at the 3′ side of the antisense RNA loop had no significant effect on antisense efficiency *in vivo* or *in vitro*.[23] Thus, the formation of helix B is an early step required to commit the interacting RNAs to stable complex formation. In IncIα[26] and IncB[27] plasmids and their relatives (but not in R1-related plasmids, Figure 1), the antisense RNA also competes with the formation of a pseudoknot structure required for translation. Interestingly, this pseudoknot within the mRNA involves the target hairpin loop and nucleotides involved in helix B formation. Thus, the pseudoknot structure is mutually exclusive with loop-loop interaction and with helix B formation, the two essential steps required for the formation of the stable RNA-RNA complex. Alignment of the closely related target RNAs reveals the presence of bulged residues in all the upper stems, although their positions are not conserved (Figure 1). These bulged residues are essential, since their deletions strongly affected rapid *in vitro* binding and *in vivo* control (R1,[28] pMU720,[23] Col1b-P9[29]). They destabilize the upper stem regions, a feature that is crucial for inter-strand helix propagation. In the CopA-CopT and the Inc-RepZ complex, intermolecular base-pairing proceeds to, and is arrested at, the lower bulged nucleotides. We propose that inter-strand progression stops due to topological stress encountered during formation of this structure, since rotational freedom is constrained by the connecting loops. Molecular modeling indicates that the longest helix (b) can consist of maximally nine base-pairs when connecting loops contain at least six nucleotides. Based on the rules defined for both CopA-CopT and Inc-RepZ systems, the lengths of helices B and B′ and thus the positioning of the four-way junction can be predicted for all other closely related plasmids (Figure 1). The four-way junction induces a side-by-side helical alignment (Figure 4(a)) that favors the formation of a third intermolecular helix involving the 5′ tail of the antisense RNA and the complementary sequence of the mRNA. The resulting structure completely and irreversibly inhibits *rep* translation by occluding ribosome binding (pMU720,[23] R1,[13] Col1b-P9[30]).

Most of the plasmids listed in Figure 1 were classified into different incompatibility groups. Incompatibility is defined as the inability of two co-resident plasmids to be stably inherited in the absence of external selection.[31] This implies that the antisense and target RNAs of these plasmids do cross-react significantly. Interestingly, all the

antisense/targets from different incompatibility plasmid groups carry identical loop sequences (5′-CGCCAA-3′ in the antisense loop, and 5′-UUGGCG-3′ in the target loop), whereas the sequences of the upper part of the stems diverge. The sequences of helix A remain highly conserved, especially in the R1-related plasmids (Figure 1). Praszkier *et al.*[32] showed that IncB and IncZ plasmids were incompatible; in plasmids from these classes, the sequences of the loops and helices B are strictly identical whereas several differences are observed in helix A (Figure 1). These data further support the requirement for helix B progression subsequent to initial loop-loop kissing to provide a stable and inhibitory antisense-target complex. Compatibility therefore reflects the mismatches encountered when heterologous antisense and target RNAs initiate binding, either at critical loop nucleotides or at nucleotides forming helix B. The variations in the upper stem sequences of both target and antisense RNAs of the plasmids for which the incompatibility groups was not clearly identified, strongly suggest that some or all of these replicons may be compatible with each other (see the R1-related plasmids in Figure 1).

In summary, we suggest that many antisense-target systems in plasmids encounter the same topological difficulty: helix formation starting from loops cannot be propagated further than a restricted number of base-pairs.[1,33] Instead of forming full duplexes, these RNAs form unusual four-way junction structures that form rapidly due to unimpeded helix invasion of upper stems. We infer that shared sequence (major loop) and structure motif, as well as similar or identical binding pathways, are found in the plasmid, even though sequences can diverge and the mechanism of inhibition can differ substantially.

Experimental Procedures

DNA templates and RNA synthesis

Plasmid pLG272[34] is a derivative of ColIb-P9 (EMBL accession number AB021078), and was used as a DNA template for polymerase chain reactions (PCR). RepZ DNA fragments for use as transcription templates were synthesized with primers T7-repZ (5′GAA ATT AAT ACG ACT CAC TAT AGG AAG CCA TAA AAG GAA AAC3′) and repZ-rev (5′GGT ATT CTG TGA GGC C3′), and Inc DNA fragments with primers T7-inc (5′GAA ATT AAT ACG ACT CAC TAT AGG TAT TCT GTG AGG CC3′) and inc-rev (5′GGA AGC CAT AAA AGG AAA ACC CCC3′).

RepZ and Inc RNAs were synthesized with phage T7 RNA polymerase using PCR-generated DNA fragments. 5′-End-labeling of dephosphorylated RNA and purification were performed as described.[12] Before use, RNAs were dissolved in RNase-free water and renatured by incubation at 90 °C for two minutes followed by slow cooling at 20 °C in TMN buffer (20 mM Tris-acetate (pH 7.5), 10 mM magnesium acetate, 100 mM sodium acetate).

Lead-induced hydrolysis and RNase V_1 probing

Lead-induced hydrolysis was done on end-labeled RepZ or Inc, free or in complex, as described.[11] Hydrolysis was conducted in 20 mM Hepes-NaOH (pH 7.5), 10 mM magnesium acetate, 100 mM sodium acetate, in the presence of 8 or 16 mM lead(II) acetate for five minutes. For competition experiments, magnesium acetate was added to final concentrations ranging from 1 to 100 mM. RNase V_1 hydrolysis on end-labeled Inc RNA, free or bound to RepZ, was performed at 37 °C for five minutes in 10 μl of TMN, in the presence of 1 μg of carrier tRNA and RNase V_1 (0.1 unit). All reactions were stopped by addition of 50 mM EDTA, followed by precipitation in ethanol. Identification and quantification of the cleavage sites was done according to Kolb *et al.*[12]

Molecular modeling

The 3D model of Inc RNA (nucleotides C4 to G54; numbering according to Asano & Mizobuchi[6]) interacting with RepZ (nucleotides C303 to G351) was built using several algorithms[35] incorporated in the program MANIP.[36] The generated model was subjected to restrained least-squares refinement using the programs NUCLIN and NUCLSQ[35] in order to ensure geometry and stereochemistry with allowed distances between interacting atoms and to avoid steric conflicts. The color views were generated with the program DRAWNA.[37] The model was then tested against the experimental data.

Acknowledgments

We thank Dr Brian M. Wilkins (Leicester) for providing us with plasmid pLG272. This work was supported by grants from Centre National de la Recherche Scientifique (CNRS), and by the Swedish Natural Science Research Council (NFR, G.W.) and Swedish Research Council for Engineering Sciences (TFR, G.W.).

References

1. Wagner, E. G. H. & Simons, R. W. (1994). Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.* **48**, 713-742.
2. Tomizawa, J. (1984). Control of ColE1 plasmid replication: the process of binding of RNAI and the primer transcript. *Cell*, **38**, 861-870.
3. Persson, C., Wagner, E. & Nordström, K. (1988). Control of replication of plasmid R1: kinetics of *in vitro* interaction between the antisense RNA, CopA, and its target, CopT. *EMBO J.* **7**, 3279-3288.
4. Siemering, K. R., Praszkier, J. & Pittard, A. J. (1994). Mechanism of binding of the antisense and target RNAs involved in the regulation of IncB plasmid replication. *J. Bacteriol.* **176**, 2677-2688.
5. Asano, K., Niimi, T., Yokoyama, S. & Mizobuchi, K. (1998). Structural basis for binding of the plasmid ColIb-P9 antisense Inc RNA to its target RNA with the 5′-rUUGGCG-3′ motif in the loop sequence. *J. Biol. Chem.* **273**, 11826-11838.
6. Asano, K. & Mizobuchi, K. (2000). Structural analysis of late intermediate complex formed between plasmid ColIb-P9 Inc RNA and its target RNA. How does a single antisense RNA repress

translation of two genes at different rates? *J. Biol. Chem.* **275**, 1269-1274.

7. Nordström, K., Molin, S. & Light, J. (1984). Control of replication of bacterial plasmids: genetics, molecular biology, and physiology of the plasmid R1 system. *Plasmid,* **12**, 71-90.
8. Blomberg, P., Wagner, E. G. H. & Nordström, K. (1990). Control of replication of plasmid R1: the duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically *in vivo* and *in vitro* by RNase III. *EMBO J.* **9**, 2331-2340.
9. Blomberg, P., Nordström, K. & Wagner, E. G. H. (1992). Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. *EMBO J.* **11**, 2675-2683.
10. Malmgren, C., Engdahl, H. M., Romby, P. & Wagner, E. G. H. (1996). An antisense/target RNA duplex or a strong intramolecular RNA structure 5′ of a translation initiation signal blocks ribosome binding: the case of plasmid R1. *RNA,* **2**, 1022-1032.
11. Kolb, F. A., Engdahl, H. M., Slagter-Jäger, J. G., Ehresmann, B., Ehresmann, C., Westhof, E., Wagner, E. G. H. & Romby, P. (2000). Progression of a loop-loop complex to a four-way junction is crucial for the activity of a regulatory antisense RNA. *EMBO J.* **19**, 5905-5915.
12. Kolb, F. A., Malmgren, C., Westhof, E., Ehresmann, C., Ehresmann, B., Wagner, E. G. & Romby, P. (2000). An unusual structure formed by antisense-target RNA binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment. *RNA,* **6**, 311-324.
13. Malmgren, C., Wagner, E. G. H., Ehresmann, C., Ehresmann, B. & Romby, P. (1997). Antisense RNA control of plasmid R1 replication. The dominant product of the antisense RNA-mRNA binding is not a full RNA duplex. *J. Biol. Chem.* **272**, 12508-12512.
14. Duckett, D. R., Murchie, A. I. & Lilley, D. M. (1995). The global folding of four-way helical junctions in RNA, including that in U1 snRNA. *Cell,* **83**, 1027-1036.
15. Walter, F., Murchie, A. I. H. & Lilley, D. M. J. (1998). Folding of the four-way RNA junction of the hairpin ribozyme. *Biochemistry,* **37**, 17629-17636.
16. Pan, T., Long, D. M. & Uhlenbeck, O. C. (1993). Divalent metal ions in RNA folding and catalysis. In *The RNA World* (Gesteland, R. F. & Atkins, J. F., eds), pp. 271-302, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
17. Streicher, B., Von Ahsen, U. & Schroeder, R. (1993). Lead cleavage sites in the core structure of group I intron RNA. *Nucl. Acids Res.* **21**, 311-317.
18. Hermann, T. & Westhof, E. (1998). Aminoglycoside binding to the hammerhead ribozyme: a general model for the interaction of cationic antibiotics with RNA. *J. Mol. Biol.* **276**, 903-912.
19. Mikkelsen, N. E., Brannvall, M., Virtanen, A. & Kirsebom, L. A. (1999). Inhibition of RNase P RNA cleavage by aminoglycosides. *Proc. Natl Acad. Sci. USA,* **96**, 6155-6160.
20. Ciesiolka, J., Hardt, W. D., Schlegl, J., Erdmann, V. A. & Hartmann, R. K. (1994). Lead-ion-induced cleavage of RNase P RNA. *Eur. J. Biochem.* **219**, 49-56.
21. Kazakov, S. & Altman, S. (1991). Site-specific cleavage by metal ion cofactors and inhibitors of M1 RNA, the catalyitc subunit of RNase P from *Escherichia coli. Proc. Natl Acad. Sci. USA,* **88**, 9193-9197.
22. Kufel, J. & Kirsebom, L. A. (1996). Residues in *Escherichia coli* RNase P RNA important for cleavage site selection and divalent metal ion-binding. *J. Mol. Biol.* **263**, 685-698.
23. Siemering, K. R., Praszkier, J. & Pittard, A. J. (1993). Interaction between the antisense and target RNAs involved in the regulation of IncB plasmid replication. *J. Bacteriol.* **175**, 2895-2906.
24. Hjalt, T. & Wagner, E. G. H. (1992). The effect of loop size in antisense and target RNAs on the efficiency of antisense RNA control. *Nucl. Acids Res.* **20**, 6723-6732.
25. Franch, T. & Gerdes, K. (2000). U-turns and regulatory RNAs. *Curr. Opin. Microbiol.* **3**, 159-164.
26. Asano, K. & Mizobuchi, K. (1998). Copy number control of IncIalpha plasmid ColIb-P9 by competition between pseudoknot formation and antisense RNA binding at a specific RNA site. *EMBO J.* **17**, 5201-5213.
27. Wilson, I. W., Praszkier, J. & Pittard, A. J. (1993). Mutations affecting pseudoknot control of the replication of IncB group plasmids. *J. Bacteriol.* **175**, 6476-6483.
28. Hjalt, T. A. & Wagner, E. G. (1995). Bulged-out nucleotides in an antisense RNA are required for rapid target RNA binding *in vitro* and inhibition *in vivo. Nucl. Acids Res.* **23**, 580-587.
29. Asano, K., Kato, A., Moriwaki, H., Hama, C., Shiba, K. & Mizobuchi, K. (1991). Positive and negative regulations of plasmid CoLIb-P9 *repZ* gene expression at the translational level. *J. Biol. Chem.* **266**, 3774-3781.
30. Asano, K., Hama, C., Shin-ichi, I., Moriwaki, H. & Mizobuchi, K. (1999). The plasmid ColIb-P9 antisense Inc RNA controls expression of the RepZ replication protein and its positive regulator repY with different mechanisms. *J. Biol. Chem.* **274**, 17924-17933.
31. Novick, R. P., Clowes, R. C., Cohen, S. N., Curtiss, R. I., Data, N. & Falkow, S. (1976). Uniform nomenclature for bacterial plasmids: a proposal. *Bacteriol. Rev.* **40**, 168-189.
32. Praszkier, J., Wei, T., Siemering, K. & Pittard, J. (1991). Comparative analysis of the replication regions of IncB, IncK, and IncZ plasmids. *J. Bacteriol.* **173**, 2393-2397.
33. Zeiler, B. N. & Simons, R. W. (1998). Antisense RNA structure and function. In *RNA Structure and Function* (Simons, R. W. & Grunberg-Manago, M., eds), pp. 437-464, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
34. Howland, C. J., Rees, C. E., Barth, P. T. & Wilkins, B. M. (1989). The *ssb* gene of plasmid ColIb-P9. *J. Bacteriol.* **171**, 2466-2473.
35. Westhof, E. (1993). Modeling the three-dimensional structure of ribonucleic acids. *J. Mol. Struct.* **286**, 203-211.
36. Massire, C. & Westhof, E. (1999). MANIP: an interactive tool for the modeling of ribonucleic acids. *J. Mol. Graph.* **16**, 197-205.
37. Massire, C., Gaspin, C. & Westhof, E. (1994). A program for drawing schematic views of nucleic acids. *J. Mol. Graph.* **12**, 201-206.
38. Ryder, T. B., Davidson, D. B., Rosen, J. I., Ohtsubo, E. & Ohtsubo, H. (1982). Analysis of plasmid genome evolution based on nucleotide-sequence comparison of two related plasmids of *Escherichia coli. Gene,* **17**, 299-310.

39. Weber, P. C. & Palchaudhuri, S. (1986). Incompatibility repressor in a RepA-like replicon of the IncFI plasmid ColV2-K94. *J. Bacteriol.* **166**, 1106-1112.
40. Saadi, S., Maas, W. K., Hill, D. F. & Bergquist, P. L. (1987). Nucleotide sequence analysis of RepFIC, a basic replicon present in IncFI plasmids P307 and F, and its relation to the RepA replicon of IncFII plasmids. *J. Bacteriol.* **169**, 1836-1846.
41. Lopez, J., Crespo, P., Rodriguez, J. C., Andres, I. & Ortiz, J. M. (1989). Analysis of IncF plasmids evolution: nucleotide sequence of an IncFIII replication region. *Gene*, **78**, 183-187.
42. Vanootegham, J. C. & Cornelis, G. R. (1990). Structural and functional similarities between the replication region of the *Yersinia* virulence plasmid and the RepFIIA replicons. *J. Bacteriol.* **172**, 3600-3608.
43. Perry, R. D., Straley, S. C., Fetherston, J. D., Rose, D. J., Gregor, J. & Blattner, F. R. (1998). DNA sequencing and analysis of the low-Ca^{2+}-response plasmid pCD1 of *Yersinia pestis* KIM5. *Infect. Immun.* **66**, 4611-4623.
44. Rodriguez-Pena, J. M., Buisan, M., Ibanez, M. & Rotger, R. (1997). Genetic map of the virulence plasmid of *Salmonella enteritidis* and nucleotide sequence of its replicons. *Gene*, **188**, 53-61.
45. Osborn, A. M., da Silva Tatley, F. M., Steyn, L. M., Pickup, R. W. & Saunders, J. R. (2000). Mosaic plasmids and mosaic replicons: evolutionary lessons from the analysis of genetic diversity in IncFII-related replicons. *Microbiology*, **146**, 2267-2275.
46. Tobe, T., Hayashi, T., Han, C. G., Schoolnik, G. K., Ohtsubo, E. & Sasakawa, C. (1999). Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. *Infect. Immun.* **67**, 5455-5462.
47. Burland, V., Shao, Y., Perna, N. T., Plunkett, G. I., Sofia, H. J. & Blattner, F. R. (1998). The complete DNA sequence and analysis of the large virulence plasmid of *E. coli* 0157:H7. *Nucl. Acids Res.* **26**, 4196-4204.
48. Hama, C., Takizawa, T., Moriwaki, H., Urasaki, Y. & Mizobuchi, K. (1990). Organization of the replication control region of plasmid ColIb-P9. *J. Bacteriol.* **172**, 1983-1991.
49. Athanasopoulos, V., Praszkier, J. & Pittard, A. J. (1995). The replication of an IncL/M plasmid is subject to antisense control. *J. Bacteriol.* **177**, 4730-4741.

Edited by M. Belfort

(Received 21 November 2000; received in revised form 12 March 2001; accepted 4 April 2001)

V. Cristallogenèse du complexe CopA-CopT

Le modèle du complexe CopA-CopT présenté dans l'article I est essentiellement basé sur les contraintes imposées par les régions appariées dans chacun des deux ARN. La conformation exacte des nucléotides des boucles de connexion (CopA : nts G58–A62 ; CopT : nts U108–C113) ainsi que des nucléotides non appariés à la jonction (CopA : G69 ; CopT : C101) est difficile à prédire par les approches biochimiques utilisées ou par la modélisation graphique. Il en est de même pour la topologie exacte de la jonction à quatre hélices. Pour ces raisons, et dans le but d'obtenir une structure de résolution atomique, la cristallisation du complexe CopA-CopT a été entreprise.

A. Stratégie expérimentale

1. *Conception des ARN*

Nous avons démontré que malgré leur parfaite complémentarité, CopA et CopT ne formaient pas de duplexe étendu *in vitro* lorsqu'ils étaient co-incubés pendant des temps biologiquement significatifs [326, Article I]. Cependant, les essais de cristallisation sont effectués sur des durées beaucoup plus longues. Par ailleurs, les hautes concentrations en ARN utilisées dans les essais de cristallisation pourraient déplacer l'équilibre du complexe irréversible CopA-CopT vers le duplexe étendu (Article III, Fig. 5). Pour ces deux raisons, nous avons changé la séquence de CopA afin d'empêcher la formation d'un duplexe étendu lors des expériences de cristallogenèse.

L'inversion des 3 pb dans la partie inférieure de l'hélice intramoléculaire de CopA n'affecte ni l'interaction ARN-ARN *in vitro*, ni l'efficacité du contrôle *in vivo* (mutation H3 : article I, Fig. 5,6 et article III, Fig. 1 & tableaux I, II). Nous avons tiré profit de cette observation pour construire un ARN CopA mutant dans lequel toute la partie inférieure de la tige II est inversée (Figure 32). Cette région de CopA n'est donc plus complémentaire à sa séquence correspondante dans CopT, mais identique. Par conséquent, un duplexe étendu ne peut pas être formé entre ces deux ARN. Par ailleurs, nous avons restreint la taille des deux ARN à 61 nts de façon à ce que l'hélice C ne contiennent que 13 pb. La taille limitée des ARN $CopA_{61}$ et $CopT_{61}$ devrait en outre faciliter leur purification.

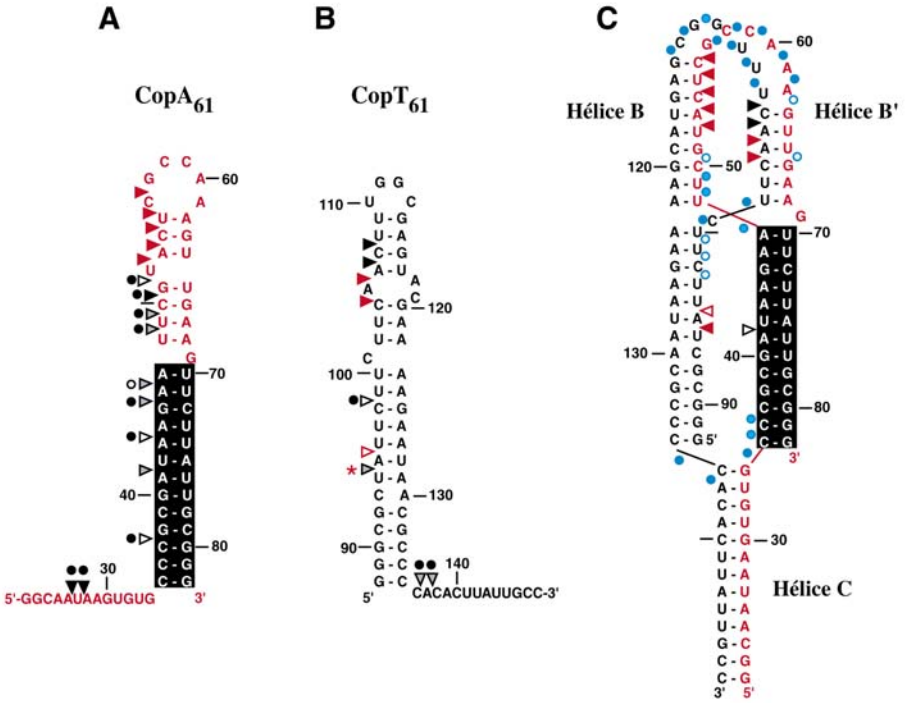

Figure 32 : ***Structures secondaires des ARN CopA$_{61}$ et CopT$_{61}$ utilisés dans les essais de cristallogenèse et du complexe résultant***. Les positions mutées dans CopA$_{61}$ par rapport à l'ARN sauvage sont indiquées par une boîte noire. Les coupures induites par les ions Pb^{2+} (ronds bleus) ou la RNase V1 (triangles) sont indiquées ; rouge : nouvelle coupure ; rond noir sur triangle : position protégée après formation du complexe. Les résidus sont numérotés de la même manière que dans les ARN entiers.

2. *Préparation des ARN CopA$_{61}$ et CopT$_{61}$*

En raison de la taille des deux ARN, leur synthèse chimique n'était pas envisageable. Nous avons par conséquent opté pour leur production *in vitro* par l'ARN polymérase du bactériophage T7 [339]. Les séquences ADN correspondantes aux deux ARN ont été reconstituées par cassettes d'oligonucléotides, flanquées d'une extrémité franche et d'un site de restriction *Pst* I (Figure 33). Celles-ci contiennent également les sites de restriction *Nla* IV et *Nae* I respectivement pour CopA et CopT. Une fois reconstituées par hybridation, les cassettes ont été introduites dans le vecteur pUT7 [340], préalablement linéarisé par les enzymes de restriction *Stu* I et *Pst* I. Le plasmide pUT7 contient la séquence promotrice reconnue par l'ARN polymérase du bactériophage T7, partiellement chevauchante avec la séquence reconnue par *Stu* I (5'-AGG↓CCT-3'), dont les deux G sont les premiers nucléotides de l'ARN transcrit. Après clonage et préparation des plasmides pK7A et pK7T, ceux-ci ont été respectivement linéarisés par les enzymes *Nla* IV et *Nae* I et utilisés comme matrice pour la production des ARN CopA$_{61}$ et CopT$_{61}$. Après transcription, les ARN obtenus sont hétérogènes à leur extrémité 3' [341]. La technique de choix pour purifier les ARN de taille correcte est l'électrophorèse sur gel de polyacrylamide en conditions dénaturantes. Après électroélution du gel, les ARN sont purifiés sur une chromatographie échangeuse d'anions de

type Q-sépharose (Q15 Strong Anion Exchanger, Sartorius™), puis précipités. Les culots sont lavés dans un tampon 5 mM Na cacodylate pH 6.5, 2 mM MgAc et les ARN sont reprécipités. Les ARN sont conservés sous forme de précipités éthanoliques à –20°C.

B. Analyse biochimique du complexe $CopA_{61}$-$CopT_{61}$

1. *Le complexe $CopA_{61}$-$CopT_{61}$ se forme avec une constante de vitesse similaire à celle du complexe sauvage*

Les ARN $CopA_{61}$ et $CopT_{61}$ ont été modifiés dans le but de ne pas former de duplexe étendu. Pour avoir la certitude que les mutations introduites n'affectent pas la reconnaissance ARN-ARN, nous avons mesuré la vitesse de formation (k_{app}) du complexe *in vitro*. Le complexe formé entre les ARN tronqués n'étant pas stable sur gel dénaturant, nous avons suivi sa formation par migration sur gel de polyacrylamide en conditions natives (5 mM MgAc dans le gel et dans le tampon de migration). Cette expérience qui utilise un ARN $CopA_{61}$ marqué radioactivement au ^{32}P à son extrémité 5' et un excès de $CopT_{61}$ froid a été reconduite trois fois (Article III, [63]). La moyenne des mesures réalisées donne une valeur du k_{app} de $2x10^6$ $M^{-1}.s^{-1}$, directement comparable au complexe sauvage ($1,2x10^6$ $M^{-1}.s^{-1}$). Ainsi, l'inversion de la partie inférieure de la tige de CopA n'affecte pas ses propriétés d'interaction avec CopT. Néanmoins, la formation de 13 pb dans l'hélice intermoléculaire C ne semble par suffisante pour stabiliser de façon irréversible le complexe puisqu'il ne peut être séparé par électrophorèse sur gel de polyacrylamide en conditions dénaturantes.

2. *Le complexe $CopA_{61}$-$CopT_{61}$ adopte une topologie identique au complexe sauvage*

Les coupures induites par la ribonucléase V1 et les ions Pb^{2+} dans le complexe sauvage sont une bonne signature de la présence de la jonction à quatre hélices (Articles 1 et IV). Cinq nouvelles coupures à la RNase V1 (nts 52-56) sont observées dans CopA après formation de l'hélice B et deux nouvelles coupures (nts 104-105) sont observées dans CopT après formation de l'hélice B'. Dans le complexe, les boucles des deux ARN (CopA : nts 56-62 ; CopT : nts 107-113) sont quant à elles accessibles aux ions Pb^{2+}. Afin de vérifier si le complexe tronqué adopte la même topologie que le complexe sauvage, nous avons étudié sa structure à l'aide de ces deux sondes. Les autoradiogrammes correspondants sont présentés dans la figure 34. Les clivages observés sont résumés sur les structures secondaires des ARN libres et du complexe (Figure 32).

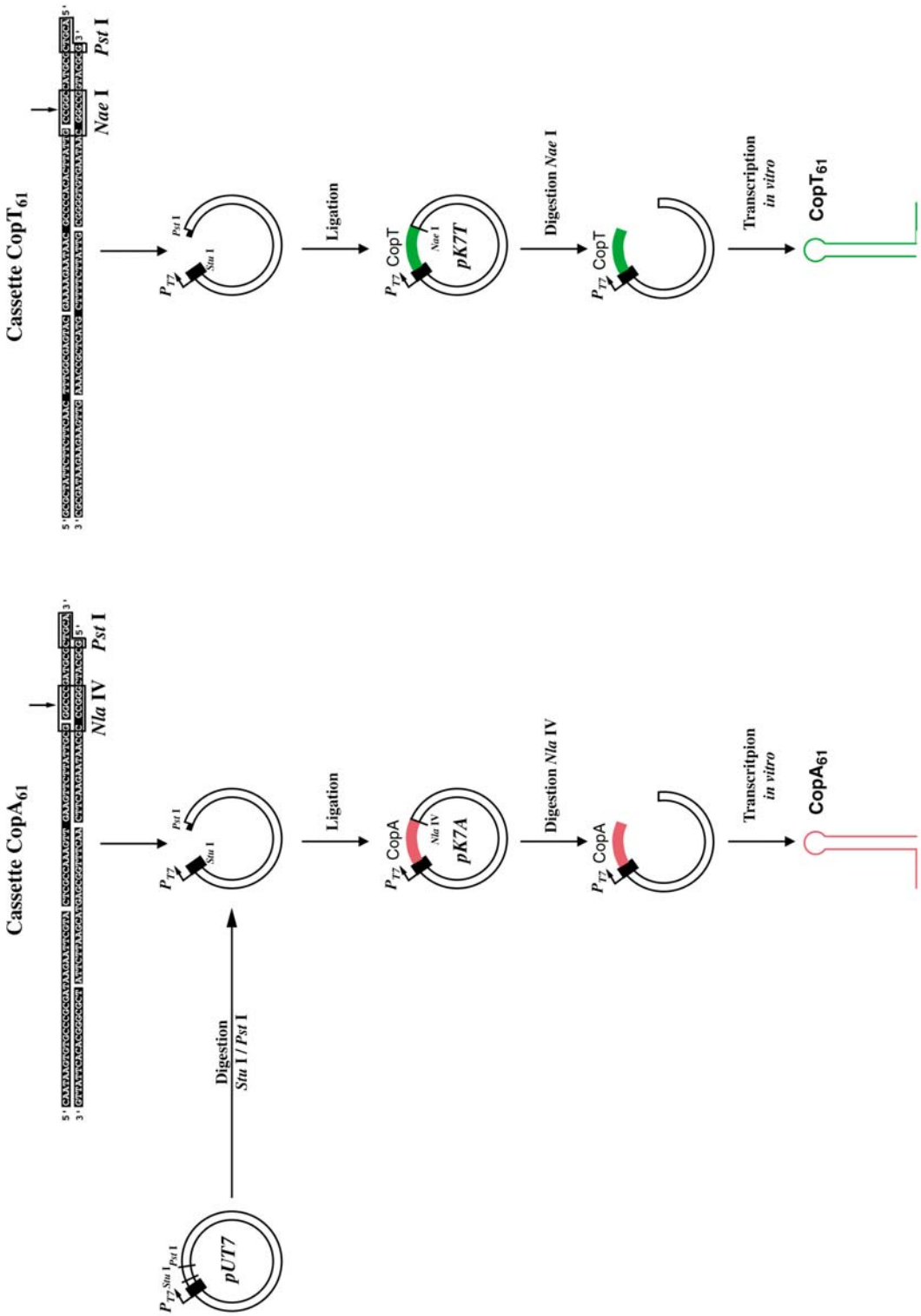

Figure 33 : ***Stratégie de clonage des cassettes codant pour les ARN CopA$_{61}$ (gauche) et CopT$_{61}$ (droite)***. Voir le texte pour les détails.

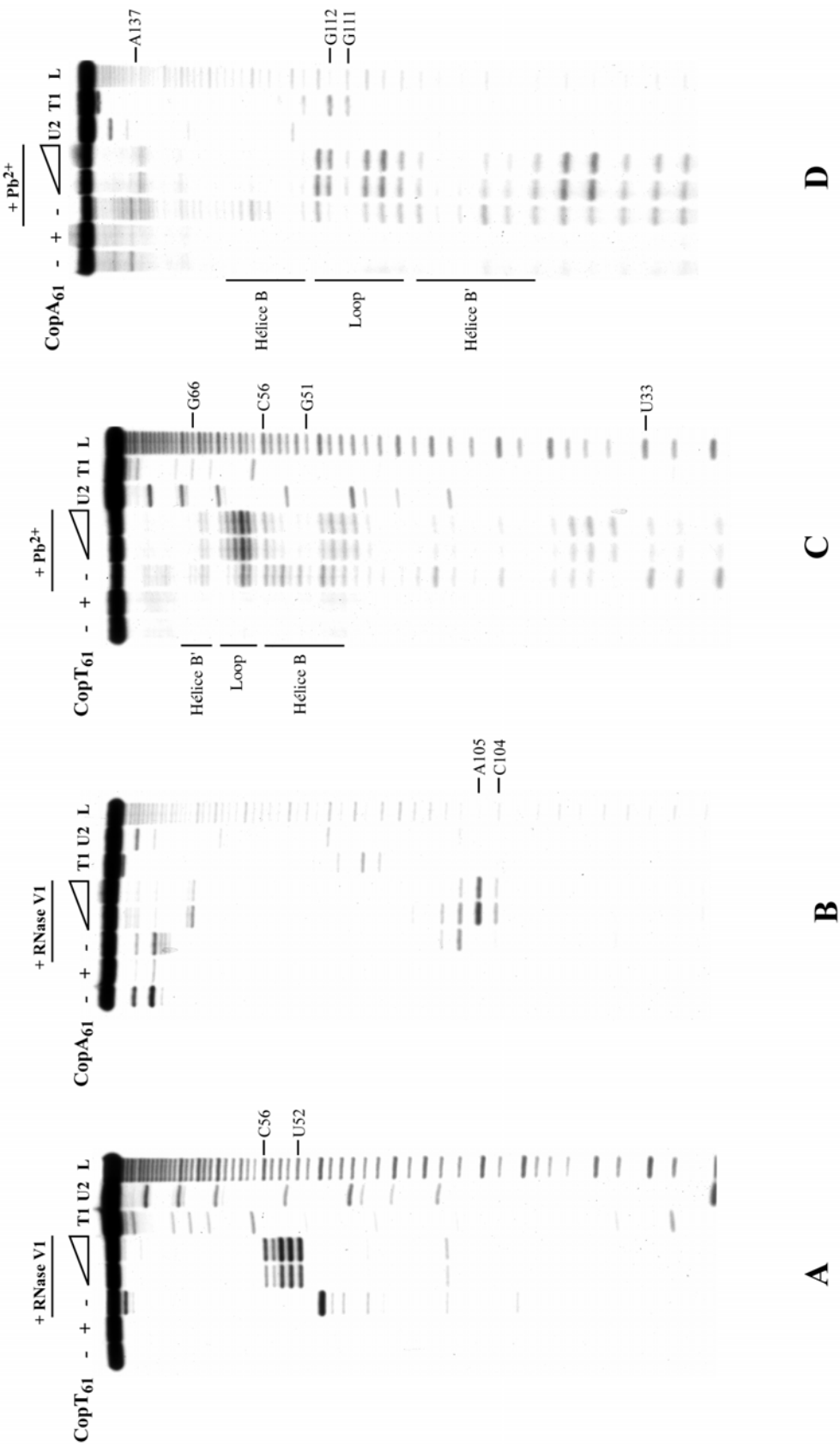

Figure 34 : Expériences de cartographie en solution du complexe $CopA_{61}$-$CopT_{61}$ à la RNase V1 et aux ions Pb^{2+}. Cartographie à la RNase V1 sur (A) un ARN $CopA_{61}$ marqué à son extrémité 5', seul (-) ou en présence d'un excès d'ARN $CopT_{61}$ ou (B) sur un ARN $CopT_{61}$ marqué à son extrémité 5', seul (-) ou en présence d'un excès d'ARN $CopA_{61}$. (C) Cartographie aux ions Pb^{2+} sur un ARN $CopA_{61}$ marqué à son extrémité 5', seul (-) ou en présence d'un excès d'ARN $CopT_{61}$. (D) Cartographie aux ions Pb^{2+} sur un ARN $CopT_{61}$ marqué à son extrémité 5', seul (-) ou en présence d'un excès d'ARN $CopA_{61}$. Les conditions de formation du complexe et de clivages sont décrites dans l'articleIV. Les positions des boucles et celles impliquées dans la formation des hélices B et B' sont indiquées.

Le profil de coupure du complexe $CopA_{61}$-$CopT_{61}$ par la RNase V1 est superposable à celui observé pour le complexe sauvage. Après formation du complexe, des coupures par la RNase V1 apparaissent dans $CopA_{61}$ aux positions 52-56 et aux positions 104-105 dans $CopT_{61}$. De façon concomitante, les coupures par les ions Pb^{2+} aux positions 51-56 dans $CopA_{61}$ et 116-122 dans $CopT_{61}$ sont fortement diminuées (Figure 34 C, D). Ces résultats indiquent la formation des hélices intermoléculaires B et B'. Les boucles de connexion (nts 57-62 dans $CopA_{61}$ et 108-113 dans $CopT_{61}$) sont fortement accessibles aux ions Pb^{2+}. Finalement, même si le complexe $CopA_{61}$-$CopT_{61}$ n'est pas stable sur gel dénaturant, l'hélice C est formée en solution. En effet, les ions Pb^{2+} n'induisent plus de coupures au-delà de la position A137 dans $CopT_{61}$ (Figure 34D) et au niveau des positions 28-33 dans $CopA_{61}$ (Figure 34C). Finalement, des coupures aux ions Pb^{2+} sont présentes à la jonction dans le complexe. Elles résultent probablement de la fixation de cations divalents à la jonction à quatre hélices et sont donc une indication de la formation de cette dernière. Ces résultats indiquent donc que la structure du complexe tronqué dans lequel le pied de la tige de CopA a été inversé est essentiellement identique à celle du complexe CopA-CopT sauvage.

C. Essais de cristallisation du complexe $CopA_{61}$-$CopT_{61}$

1. *Renaturation des ARN et formation du complexe*

Les ARN sont conservés sous forme de précipité éthanolique à –20°C. Avant tout essai de cristallisation, ils sont centrifugés et les culots sont repris dans un volume variable de tampon Na cacodylate 5 mM pH 6.5, MgAc 2 mM, en fonction de la concentration finale souhaitée. Celle-ci est vérifiée par mesure de l'absorbance des solutions d'ARN à 260 nm et les ARN sont renaturés séparément par une incubation de 30 minutes à 37°C. $CopA_{61}$ et $CopT_{61}$ sont alors co-incubés en quantités équimolaires, pendant 30 minutes à 37°C. Nous avons vérifié qu'il ne reste plus d'ARN libre à l'issue de cette incubation (résultat non montré).

2. *Essais de cristallisation à l'aide de la matrice éparse Natrix™ d'Hampton Research*

Dans un premier temps, nous avons réalisé une recherche de conditions de cristallisation à l'aide du kit Natrix™ d'Hampton Research. Les gouttes de cristallisation sont composées de 1,5 µl de solution de complexe ARN-ARN à 10 mg/ml et 1,5 µl de liqueur mère. Les essais ont été effectués à 20°C par la technique des gouttes suspendues. Après 15 h d'incubation, des microcristaux ont été observés dans les conditions 3, 16 et 38 (Figure 35).

Des microcristaux ont également été observés dans la condition 26, ceci après une semaine d'incubation. L'aspect des cristaux ainsi que la composition des solutions sont donnés dans la figure 35. La taille des microcristaux varie entre 10 et 20 µm. Il est important de noter qu'un précipité amorphe est également observé dans ces quatre conditions. Ceci peut indiquer que les concentrations en ARN ou en agent cristallisant utilisées sont trop élevées et/ou que la quantité de sels présente est trop faible.

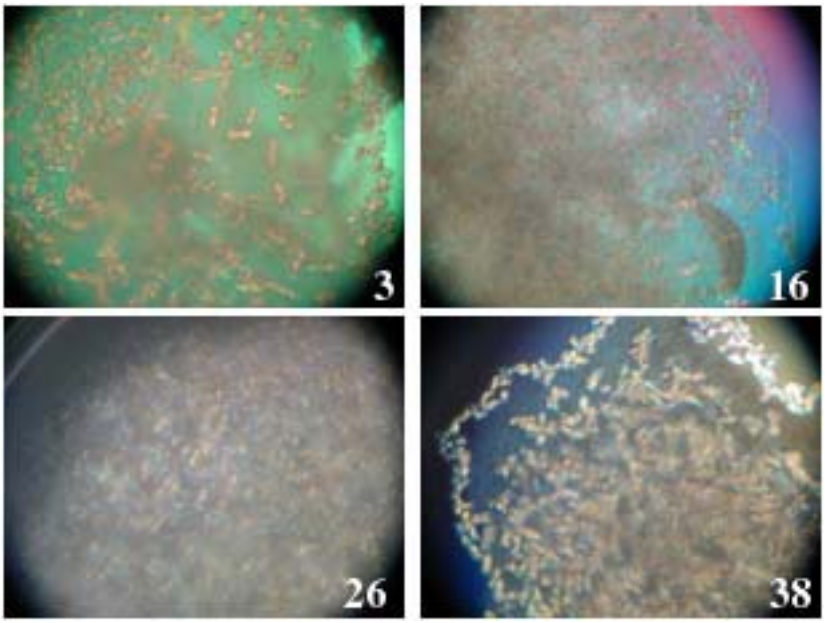

Figure 35 : ***Microcristaux obtenus à l'aide de la matrice clairsemée Natrix^{TM} d'Hampton Research***. (3) : 100 mM MgAc; 50 mM MES pH5.6 ; 20% MPD. (16) : 40 mM MgAc; 50 mM Na cacodylate pH 6.0 ; 30 % MPD. (26) : 100 mM MgAc; 200 mM KCl ; 50 mM Na cacodylate pH 6.5 ; 10% PEG 8000. (38) : 150 mM MgAc; 200 mM NH_4 Ac; 50 mM Hepes-NaOH pH 7.0 ; 5% PEG 4000.

3. *Optimisation des conditions de cristallisation*

Les microcristaux décrits au paragraphe précédent ont tous été obtenus à haute concentration en magnésium (> 40 mM) et en présence de MPD ou de PEG en tant qu'agent cristallisant. Basées sur ces résultats, plusieurs matrices de cristallisation ont été conçues. Nous y avons fait varier la concentration des ARN, le volume des gouttes, la nature et la concentration en agent précipitant et la concentration en ions monovalents. Par manque de place et par soucis de clarté, tous les résultats obtenus ne seront pas décrits. Le tableau 7 recense les matrices dans lesquelles les résultats obtenus ont été les meilleurs. La figure 36 présente des photographies des microcristaux observés. Les résultats obtenus sont reproductibles.

N°	Tampon	*Sels*	Agent Cristallisant	Volume Goutte (µl)	[ARN] (mg/ml)
1	0,05 M Na Cacodylate pH 6,0	0,05 M MgAc	16 % MPD	6	1,67
2	0,05 M Na Cacodylate pH 6,0	0,1 M MgAc	16 % MPD	6	1,67
3	0,05 M Na Cacodylate pH 6,0	0,15 M MgAc	16 % MPD	6	1,67
4	0,05 M Na Cacodylate pH 6,0	0,05 M MgAc	20 % MPD	6	1,67
5	0,05 M Na Cacodylate pH 6,0	0,1 M MgAc	20 % MPD	6	1,67
6	0,05 M Na Cacodylate pH 6,0	0,15 M MgAc	20 % MPD	6	1,67
7	0,05 M Na Cacodylate pH 6,0	0,05 M MgAc	24 % MPD	6	1,67
8	0,05 M Na Cacodylate pH 6,0	0,1 M MgAc	24 % MPD	6	1,67
9	0,05 M Na Cacodylate pH 6,0	0,15 M MgAc	24 % MPD	6	1,67
10	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,05 M MgAc	5 % PEG 4000	6	1,67
11	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,1 M MgAc	5 % PEG 4000	6	1,67
12	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,15 M MgAc	5 % PEG 4000	6	1,67
13	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,05 M MgAc	7,5 % PEG 4000	6	1,67
14	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,1 M MgAc	7,5 % PEG 4000	6	1,67
15	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,15 M MgAc	7,5 % PEG 4000	6	1,67
16	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,05 M MgAc	10 % PEG 4000	6	1,67
17	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,1 M MgAc	10 % PEG 4000	6	1,67
18	0,05 M Na Cacodylate pH 6,5	0,2 M NH_4Ac, 0,15 M MgAc	10 % PEG 4000	6	1,67
19	0,05 M Na Cacodylate pH 6,5	0,05 M MgAc, 0,25 M KCl	7,5 % PEG 8000	6	1,67
20	0,05 M Na Cacodylate pH 6,5	0,05 M MgAc, 0,25 M NH_4Ac	7,5 % PEG 8000	6	1,67
21	0,05 M Na Cacodylate pH 6,5	0,05 M MgAc, 0,25 M LiCl	7,5 % PEG 8000	6	1,67
22	0,05 M Na Cacodylate pH 6,5	0,1 M MgAc	1,8 M Sulfate d'ammonium	6	1,67
23	0,05 M Na Cacodylate pH 6,5	0,1 M MgAc	1,9 M Sulfate d'ammonium	6	1,67
24	0,05 M Na Cacodylate pH 6,5	0,1 M MgAc	2 M Sulfate d'ammonium	6	1,67
25	0,05 M Na Cacodylate pH 6,5	0,1 M MgAc	2,5 M Sulfate d'ammonium	6	1,67
26	0,05 M Hepes Na OH pH 7,0	0,05 M NH_4Ac, 0,15 M MgAc	7 % PEG 4000	6	1,5

Tableau 7 : ***Composition des 26 solutions de cristallisation testées et conditions expérimentales des essais***. La concentration en ARN donnée est celle dans la goutte de cristallisation, avant équilibration de celle-ci. Les conditions dans lesquelles des microcristaux ont été obtenus sont indiquées en rouge. Tous les essais ont été effectués à 20°C.

Plusieurs conditions nous ont permis d'obtenir des microcristaux d'aspects et de tailles proches de ceux obtenus dans les essais précédents. Par ailleurs, une condition (n°26) nous a permis d'obtenir un cristal unique d'environ 150 µm. Par rapport aux résultats obtenus dans les premiers essais effectués à l'aide du kit Natrix^{TM}, ces derniers essais ont permis de résoudre plusieurs problèmes additionnels. La concentration moins élevée en ARN et la manière dont les gouttes sont réalisées nous ont permis d'éliminer totalement la formation d'un précipité amorphe. En effet, les gouttes sont ici composées de 4 µl d'ARN et 2 µl de solution mère contre 1,5 µl + 1,5 µl dans les essais précédents (§ V.C.2). La concentration en agent cristallisant initiale est donc plus faible. Par ailleurs, le volume des gouttes étant plus grand et le volume du réservoir étant plus faible (500 µl contre 900 µl précédemment), l'équilibration des gouttes est plus lente. Cela se traduit par une vitesse d'apparition des

cristaux moins rapide. En moyenne, les microcristaux présentés dans la figure 36 apparaissent entre 1 et 3 jours d'incubation suivant les conditions.

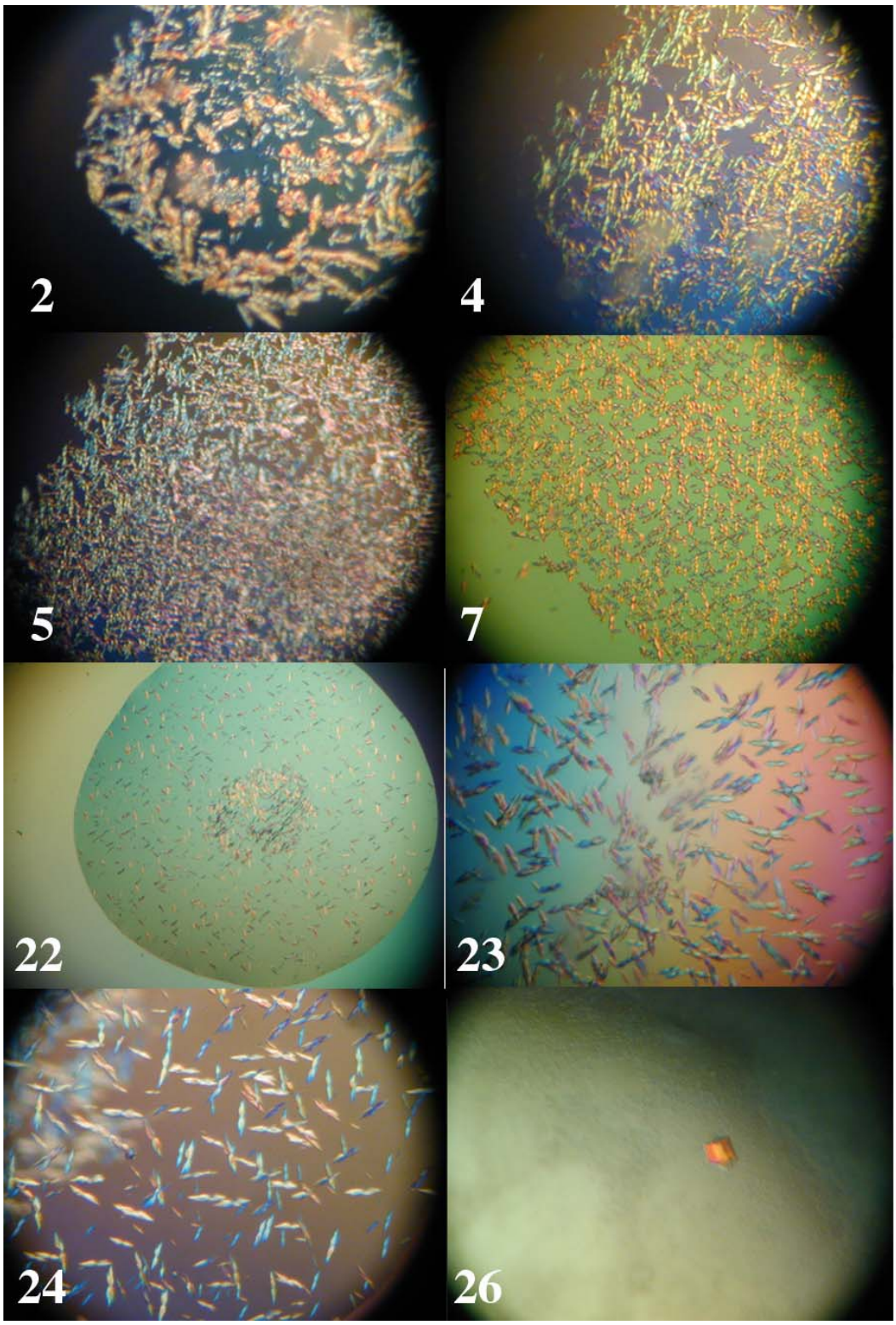

Figure 36 : ***Microcristaux obtenus à l'aide des matrices décrites dans le tableau 7***. Les conditions du tableau dans lesquelles les microcristaux ont été obtenus sont indiquées par les nombres.

4. *Analyse du contenu des cristaux*

Les microcristaux obtenus dans ces derniers essais l'ont été soit en présence de MPD en tant qu'agent cristallisant, soit en présence de sulfate d'ammonium. Nous avons voulu vérifier la présence d'ARN dans les microcristaux de ces deux classes. Après un mois d'incubation, une vingtaine de microcristaux ont été prélevés et lavés trois fois dans la liqueur mère correspondante. Les cristaux ont été dilués dans 10 µl d'une solution de dépôt contenant 30% de glycérol, 0,05% de xylène cyanol, 0,05% de bleu de bromophénol et déposés sur gel de polyacrylamide 8% en conditions natives contenant 5 mM de $MgCl_2$. En parallèle, 0,5 µg d'ARN $CopA_{61}$ et 0,5 µg de complexe formé extemporanément ont été déposés. La figure 37 représente la coloration au Stains-all™ (Sigma-Aldrich) du gel sur lequel les cristaux obtenus dans la condition 4 (50 mM MgAc, 20 % MPD) ont été déposés. Le résultat montre que les cristaux contiennent de l'ARN sous une forme comigrant avec le complexe. Par ailleurs, aucun produit de dégradation ne peut être observé et les solutions de lavage des cristaux ne contiennent pas d'ARN. Le contenu en ARN des cristaux obtenus en présence de sulfate d'ammonium n'a pas pu être testé par cette approche. En effet, la haute concentration en sels présente dans la solution interfère avec la migration électrophorétique.

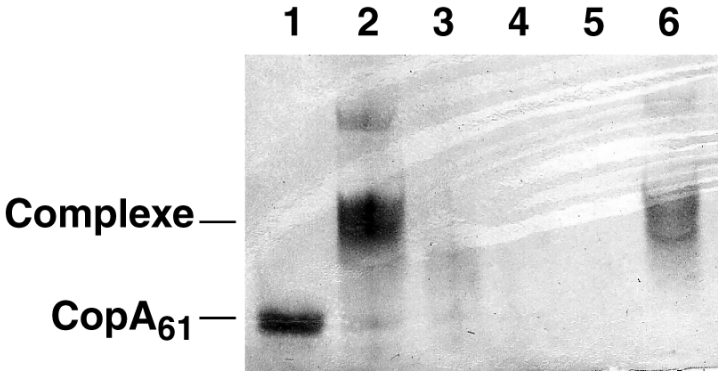

Figure 37 : ***Analyse du contenu des microcristaux obtenus dans la condition 4 par électrophorèse sur gel de polyacrylamide et coloration au Stains-all™***. 1 : 0,5 µg d'ARN $CopA_{61}$; 2 : Complexe $CopA_{61}$-$CopT_{61}$ formé *in vitro* ; 3, 4 et 5 : 6 µl respectivement des premières, secondes et troisième solutions de lavage des cristaux; 6 : microcristaux lavés. Les positions de migration de l'ARN $CopA_{61}$ et du complexe $CopA_{61}$-$CopT_{61}$ sont indiquées.

VI. Conclusions et perspectives

A. Conclusions

Nos résultats, ainsi que les systèmes procaryotiques décrits dans l'introduction de ce manuscrit de thèse, montrent qu'il y a toujours une stricte corrélation entre l'activité biologique des ARN antisens et leur capacité à interagir rapidement avec leur cible. Par ailleurs, dans les systèmes contrôlant la réplication d'un plasmide, l'hybride ARN-ARN formé doit être irréversible pour que le contrôle soit efficace. A travers notre étude du plasmide R1 et d'autres systèmes homologues, nous avons pu mettre en évidence un nouveau mode d'interaction efficace entre deux ARN complémentaires. Celui-ci représente probablement un nouveau concept applicable non seulement à d'autres systèmes antisens naturels mais également au repliement de grands ARN. De plus, un tel mécanisme doit également être pris en compte lors de la conception d'ARN antisens artificiels.

Enfin, les interactions initiales instables sont toujours stabilisées par des interactions additionnelles. Celles-ci impliquent souvent des régions non structurées de l'ARN antisens ou de sa cible. Ces mécanismes de stabilisation varient cependant d'un système à l'autre. Cette diversité indique à quel point les ARN sont des molécules versatiles capables d'adapter leur structure à une fonction commune.

1. *Caractéristiques de l'étape de reconnaissance ARN-ARN*

Dans les cellules, parmi le grand nombre de molécules d'ARN présentes, les ARN antisens ont à reconnaître leur ARN cible rapidement. C'est de la spécificité de cette étape que va dépendre l'efficacité du contrôle. De manière surprenante, elle est particulièrement conservée. Les ARN antisens et/ou leurs cibles contiennent une structure en tige-boucle par laquelle l'interaction est initiée. Selon que les deux partenaires contiennent une telle structure ou non, l'interaction initiale est alors un complexe boucle-boucle (*i.e.* « kissing-complex » comme dans le cas du système CopA/CopT) ou un complexe boucle-région simple brin. La vitesse de formation d'un complexe stable à partir de ces deux types de reconnaissance est optimisée et ne diffère pas de façon significative ($\approx 10^6$ $M^{-1}.s^{-1}$), indiquant des efficacités semblables. Dans tous ces systèmes, les boucles ont une taille de 6 à 7 nts, optimale pour la reconnaissance *in vitro* et le contrôle *in vivo* [325]. La présence d'un faible nombre de nucléotides en conformation simple-brin permet de diminuer la probabilité d'interactions non spécifiques en minimisant le nombre de points de nucléation potentiels. L'observation qu'un complexe stable n'est pas formé plus rapidement à partir de boucles plus larges indique que le mécanisme d'association n'est pas dicté par la stabilité de l'interaction initiale mais par sa

cinétique de propagation vers des interactions additionnelles. En faveur de cette hypothèse, les mutations décrites dans cette étude et affectant l'efficacité de l'interaction CopA-CopT *in vitro* et *in vivo* n'ont qu'un effet ménagé sur la stabilité des complexes formés (résultat non montré).

Un autre intérêt d'utiliser un motif de reconnaissance en tige-boucle est de pouvoir contraindre la structure de la boucle dans une conformation favorable à la propagation ultérieure des appariements. Dans plusieurs systèmes, la boucle de l'antisens ou de sa cible contient un motif conservé présentant un « U-turn » (5'-YUNR-3') également retrouvé dans la boucle de l'anticodon de tous les ARNt ([342], figures 38 & 39). Dans le cas de l'ARN-I du plasmide ColE1 (Figures 21 & 39), ce mimétisme va même plus loin. En effet, les boucles I et II correspondent respectivement aux séquences des boucles des anticodons des $ARNt_1^{Thr}$ et $ARNt_3^{Thr}$ d'*E.coli*. La fonction principale de l'ARNt est le décodage de l'ARNm. On peut donc penser que la structure de la boucle de l'anticodon a évolué de façon à optimiser la reconnaissance du codon. La structure cristallographique de divers ARNt a révélé la présence d'un motif en « U-turn » dans lequel la position N3 de l'uridine 33 conservée interagit avec le phosphate du nucléotide en position 36. Ce motif stabilise l'empilement continu des 3 nts adjacents constituant l'anticodon pré-organisés en conformation de type A (Figure 38). Par conséquent, la formation de l'hélice codon-anticodon ne nécessite pas de réarrangement de la boucle. Ceci a été visualisé dans la structure cristallographique de l'$ARNt^{Asp}$ de levure où l'empilement cristallographique implique des interactions anticodon-anticodon [343]. De même, il a été montré par RMN qu'une boucle de six nucléotides (5'-GUAAAA-3') trouvée dans l'ARN génomique d'HIV-1 mime structuralement une boucle anticodon [344]. Dans ce cas, les trois premières adénines sont empilées de façon continue ce qui doit faciliter leur interaction avec la boucle de l'anticodon de l'$ARNt^{Lys}_3$ lors de l'étape d'initiation de la rétrotranscription du génome viral [345]. Ainsi, dans le système CopA-CopT et dans les systèmes homologues décrits dans l'article IV, le motif 5'-UUGGC-3' extrêmement conservé dans la boucle de l'ARNm adopterait une telle structure afin de permettre une interaction optimale avec la séquence 5'-GCC-3' de la boucle de l'ARN antisens. Des mutations perturbant cette architecture affectent le contrôle dans les systèmes Col1b-P9 [61] et l'association *in vitro* dans les systèmes *hok*/Sok [66], IS10 [242] et ColE1 [289]. L'importance de cette topologie dans les autres cas (FinOP/*traj* et DicF/*ftsZ*) reste à prouver. Néanmoins, la conservation du motif supporte un rôle fonctionnel important.

Notons que la boucle T des ARNt qui contient aussi 7 nts et le motif conservé 5'-TψNRA-3' possède une structure intrinsèque différente [343]. Outre la présence d'un « U-turn » au niveau de la pseudouridine, la structure est contrainte par un appariement de type reverse-Hoogsteen conservé entre les nucléotides T54 et A58. De manière intéressante,

aucune des boucles présentées dans la figure 39 ne contient une adénine en 3' de la séquence 5'-YUNR-3'. Au contraire, cette position est souvent occupée par une pyrimidine (Figure 39, [66]. Les séquences des boucles des ARN antisens ou de leurs cibles auraient donc évolué de façon convergente pour permettre la formation d'une structure de type boucle de l'anticodon et ainsi faciliter les interactions de type boucle-boucle ou boucle-région simple-brin. Dans ce sens, des expériences de sélection évolutive d'aptamères capables d'interagir efficacement avec l'ARNt^{Phe} de levure ont révélé des ARN pouvant former un « kissing complex » avec la boucle de l'anticodon [346]. Le même type d'expériences conduites sur la tige-boucle TAR (contenant un motif de type U-turn) de HIV-1 a mené à des résultats similaires [347].

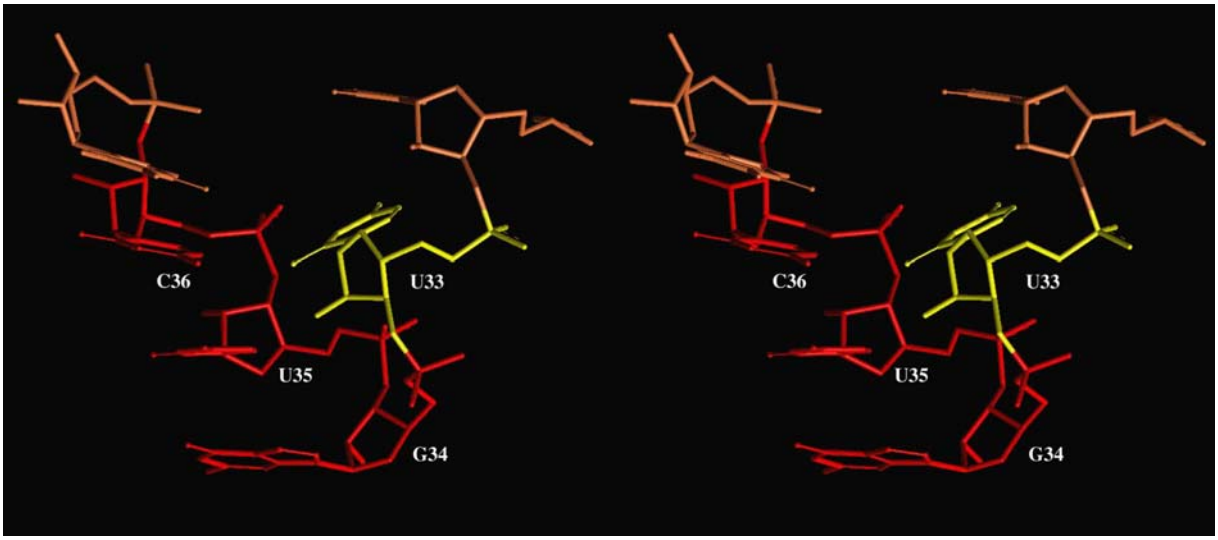

Figure 38 : ***Vue stéréoscopique de la structure en U-turn de la boucle de l'anticodon de l'ARNt^{Asp} de levure (d'après [343])***. L'uridine conservée (U33) est dessinée en jaune. Les nucléotides du triplet de l'anticodon adoptant une conformation de type A sont dessinés en rouge. Les nucléotides de la boucle flanquant le motif en U-turn sont dessinés en bronze.

Malgré les avantages liés à l'implication d'une boucle d'au moins un des ARN dans l'interaction initiale, ce type de contact présente néanmoins un inconvénient majeur. En effet, le nombre restreint de nucléotides impliqués confère un caractère réversible à cette interaction initiale du fait de l'instabilité du complexe formé. Dans de nombreux systèmes, ceci est partiellement contourné par l'implication préférentielle de paires de bases stables de type G≡C dans le complexe initial. Dans le système CopA-CopT, la mutation des paires de bases G≡C en paires de bases de type A=U affecte le contrôle *in vivo* par plus d'un ordre de magnitude [324]. Ainsi, même si la stabilité du complexe boucle-boucle ou boucle-région simple-brin initial n'a pas un rôle prépondérant, elle permet néanmoins de déplacer l'équilibre vers la formation du complexe stable et fonctionnel.

2. *De l'interaction initiale au complexe stable*

Dans le cas des ARN antisens, la formation rapide d'un complexe stable et fonctionnel est essentielle au contrôle. Pour cela, l'interaction boucle-boucle ou boucle-région simple brin initiale doit rapidement être suivie d'une propagation de la première hélice intermoléculaire sans rencontrer de barrières thermodynamiques. Au contraire de la première étape qui est une réaction de second ordre réversible, l'étape de conversion doit être de premier ordre, donc unidirectionnelle afin de conduire à une stabilisation irréversible. Si l'étape initiale est largement conservée entre tous les systèmes, il est fascinant de constater la grande diversité de mécanismes menant aux complexes stables.

Pour les systèmes dans lesquels l'interaction initiale est de type boucle-région simple brin, tout le processus d'appariement peut être considéré comme une étape unique. Le contact initial fait intervenir une extrémité libre d'un des ARN et une boucle du second, et se propage pour former un duplexe étendu dans lequel l'interaction initiale subsiste. Ce mode d'interaction est retrouvé dans les systèmes ARN-IN/ARN-OUT du transposon IS10 [242], *hok*/Sok du plasmide R1 [280] et probablement MicF/*ompF* et DicF/*ftsZ* d'*E. coli* [220,222]. Nous avons montré dans l'article II que la présence de nucléotides non appariés dans les tiges de CopA et de CopT était essentielle à l'étape de conversion et donc à l'efficacité du contrôle. Cette observation a également été faite pour les systèmes IS10 et *hok*/Sok cités ci-dessus [245,276]. Elle souligne l'importance des paramètres thermodynamiques de l'étape de propagation pour l'efficacité du contrôle biologique (Figure 39). En effet, la progression des appariements initiaux est facilitée par des hélices intramoléculaires imparfaites, et donc moins stables.

Au contraire des interactions boucle-région simple brin, la stabilisation d'interactions boucle-boucle ne peut pas être réalisée par une simple propagation des appariements initiaux. En effet, pour des raisons topologiques, la propagation d'un « kissing-complex » simultanément dans les deux directions des boucles n'est pas possible. Par conséquent, la stabilisation doit se faire par des séquences additionnelles distantes du site de contact primaire. La formation d'un complexe stable est alors vue comme un processus à plusieurs étapes. L'étude du système CopA-CopT présentée dans cette thèse a permis de révéler un mécanisme de conversion non seulement inattendu mais également conservé. L'interaction boucle-boucle initiale instable est rapidement convertie de façon unidirectionnelle à travers les hélices intramoléculaires de chacun des deux ARN. Les deux ARN possèdent toute l'information nécessaire à cette étape. Celle-ci nécessite la présence de nucléotides non appariés pour ouvrir les hélices intramoléculaires en cours de conversion. Bien que la mécanistique de cette conversion ne soit pas encore entièrement comprise, nous pouvons

supposer qu'elle se fait par un mécanisme de type « engrenage » où les interactions initiales sont rompues lorsque les appariements suivants sont formés. La propagation de la première hélice intermoléculaire est stoppée en raison de contraintes topologiques et permet la formation de la seconde hélice intermoléculaire. La structure en jonction à quatre hélices ainsi générée est tout à fait atypique puisqu'elle permet une juxtaposition de deux longs segments hélicoïdaux chacun généré par l'empilement d'une hélice intermoléculaire sur une hélice intramoléculaire. Finalement cette topologie permet le rapprochement des régions 5' de CopA et 3' de CopT dont l'interaction stabilise irréversiblement le complexe. La conversion de l'interaction initiale est essentielle au rapprochement de ces régions car les tiges de CopA et de CopT sont longues et les régions 5' de CopA et 3' de CopT sont très éloignées dans le « kissing complex » initial. Nous avons montré dans l'article IV que ce mode de formation du complexe stable est conservé dans le système plasmidique Col1b-P9 et probablement dans tous les systèmes homologues.

Au contraire du système CopA-CopT, dans les systèmes FinOP du plasmide R1 et ARN-I/ARN-II du plasmide ColE1, les ARN ne contiennent pas toute l'information structurale pour garantir une stabilisation efficace de l'interaction initiale (Figure 39). Dans ces deux cas, une protéine facilite cette étape et donc l'efficacité du contrôle biologique. Là encore, le mécanisme par lequel les protéines interviennent peut diverger. Dans le cas du plasmide ColE1, l'association ARN-I-ARN-II est initiée entre une ou deux boucles des deux ARN. Les kissing-complex formés impliquent chacun sept paires de bases et l'hélice intermoléculaire est empilée de façon coaxiale entres les tiges de chacun des ARN [289,290]. Des expériences de RMN montrent néanmoins que les contraintes imposées par une telle topologie induisent un angle d'environ 45 degrés dans l'empilement afin d'accommoder tous les appariements intermoléculaires [291,348]. La protéine homodimérique Rom interagit avec l'interaction boucle-boucle par des contacts symétriques et amplifie l'angle entre les hélices [290,349,350]. Elle stabilise le premier intermédiaire réactionnel plus de 100 fois et déplace donc l'équilibre vers la formation d'un complexe stable. La formation de ce dernier ne se fait pas par une progression des « kissing complex » mais implique une région additionnelle en 5' de l'ARN-I. Dans le cas du système FinOP, la protéine FinO a un double rôle. Elle protège les deux ARN de l'attaque d'endoribonucléases [53,271] et elle stimule la formation d'un complexe stable entre FinP et *traJ* [44]. L'interaction entre FinP et traJ est initiée par les boucles II de chacun des deux ARN. Au contraire de la protéine Rom de ColE1, le mécanisme d'action de FinO n'est pas connu. Néanmoins, l'observation que FinO reconnaisse les tiges-boucles II des deux ARN et qu'elle contient un domaine N-terminal à activité ARN hélicase (M. Glover, Edmonton, Canada, communication personnelle) suggère qu'elle pourrait permettre la conversion des interactions boucle-boucle initiales vers un duplexe étendu en

déstabilisant les tiges des deux ARN. Il est intéressant de noter que les deux systèmes nécessitant l'intervention d'une protéine sont les deux seuls où les structures des tiges initiant l'interaction sont dépourvues de nucléotides non appariés (Figure 39). En vue de nos résultats et des autres systèmes naturels, les ARN antisens peuvent donc être classés en deux catégories selon qu'ils contiennent ou non tous les éléments structuraux nécessaires à la conversion vers un complexe stable. Il est tentant de faire un parallélisme entre les systèmes CopA/CopT et FinOP. L'absence de nucléotides non appariés dans les tiges II des ARN FinP et *traJ* est compensée par l'implication d'une protéine à activité hélicase dans le mécanisme de conversion. Cette observation rejoint la conclusion de l'article II dans lequel nous énoncions la possibilité que les nucléotides non appariés des tiges de CopA et de CopT puissent servir d' « ARN-hélicases intramoléculaires ». Il est également intéressant de noter que la plupart des ARN antisens ou de leurs cibles possédant de longues hélices intramoléculaires présentent des nucléotides non appariés. Ceux-ci protègent généralement les ARN contre la dégradation par la RNase III en plus de favoriser la formation d'un complexe stable.

Nous avons montré dans l'article I que le produit final de l'interaction entre CopA et CopT *in vitro* n'était pas un duplexe étendu. Dans ce cas précis, ce sont probablement des contraintes topologiques qui empêchent la conversion de la jonction à quatre hélices. Cependant, le duplexe étendu n'est pas nécessaire au contrôle *in vivo* puisque le complexe réversible CopI-CopT, uniquement constitué de la jonction à quatre hélices, est suffisant pour assurer le contrôle. Par ailleurs, nous avons récemment montré que le complexe CopA-CopT stable est reconnu de manière efficace par la RNase III *in vitro* et que la formation d'un duplexe étendu n'est pas nécessaire au clivage (Huntzinger E., Kolb F.A. et Romby P., résultats non publiés). La formation d'un duplexe étendu n'est également pas nécessaire au contrôle dans les systèmes homologues à R1 [64,338], ColE1 [351], *hok*/Sok [52], pIP501 [296], Sar/*ant* du bactériophage P22 [251] et probablement d'autres. Ceci est probablement lié au fait que la formation d'un duplexe étendu peut nécessiter plus de 15 minutes [251,326,351], ce qui est trop long pour garantir un contrôle rapide et efficace. Tout comme dans le cas de CopA, la plupart des systèmes antisens naturels auraient évolué de manière à utiliser des intermédiaires réactionnels comme complexes fonctionnels.

La stabilisation d'interactions boucle-boucle initiales n'est pas restreinte aux complexes formés par les ARN antisens avec leur cible. En effet, que ce soit dans le cas de la dimérisation de l'ARN génomique de HIV-1 sous l'action de la protéine Ncp7 [352] ou de celle de l'ARNm *bicoid* de *D. melanogaster* (C. Wagner et C. Brunel, Strasbourg, communication personnelle), une stabilisation ou une conversion des « kissing-complex » initiaux semble être requise pour la fonction des ARN.

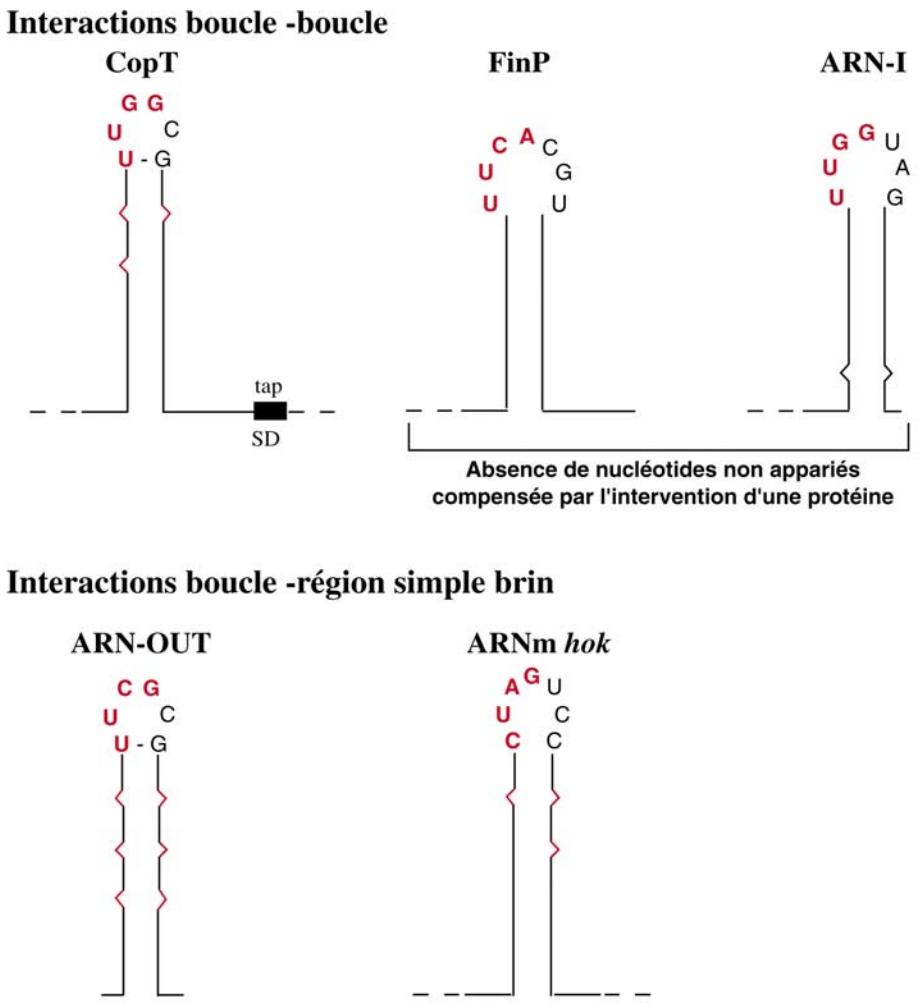

Figure 39 : ***Eléments de séquence et de structure nécessaires à la reconnaissance ARN antisens-ARN cible et à la stabilisation du complexe.*** Sont représentés le motifs CopT de l'ARNm *repA* du plasmide R1, la seconde tige-boucle de l'ARN antisens FinP, la tige-boucle I de l'ARN antisens ARN-I du plasmide ColE1, l'ARN antisens ARN-OUT du transposon IS10 et la tige-boucle interagissant avec l'ARN antisens Sok dans la région 5'-UTR de l'ARNm *hok*. Les nucléotides des motifs de type « U-turn » des boucles apicales des ARN antisens ou cibles sont indiqués en rouge. Les nucléotides non appariés de tiges-boucles, importants pour assurer une conversion rapide de l'interaction initiale vers un complexe stable, sont représentés par des lignes rouges. La séquence Shine et Dalgarno de *tap* est indiquée par une boîte noire. Les structures sont divisées en deux groupes, selon que l'interaction antisens-cible se fasse *via* une interaction boucle-boucle ou *via* une interaction boucle-région simple-brin. Les deux systèmes nécessitant l'intervention d'une protéine pour assurer la stabilisation de l'interaction initiale réversible sont indiqués.

B. Perspectives

1. *Etude de la structure du complexe CopA-CopT formé* in vivo

Nous avons toujours noté une stricte corrélation entre les effets *in vivo* et *in vitro* des mutations introduites dans les ARN CopA et CopT. Par ailleurs, les sites de coupure du complexe à la RNase III sont les mêmes *in vivo* et *in vitro*. Enfin, le complexe stable contenant la jonction à quatre hélices est suffisant pour assurer un contrôle efficace. Ces résultats ne sont cependant pas des preuves directes de la formation de ce complexe dans les cellules. Il est envisageable qu'une protéine de type hélicase puisse favoriser la formation du duplexe étendu *in vivo* pour accélérer la dégradation du complexe par les RNases spécifiques de régions double-brin. Plusieurs sondes de structure des ARN, dont le DMS [353], le kéthoxal [354] et les ions Pb^{2+} (M. Lindell & E. G. H. Wagner, Uppsala, Suède, communication personnelle), peuvent être utilisées *in vivo* parce qu'elles diffusent aisément à l'intérieur des cellules. Cette dernière sonde est particulièrement intéressante puisque nous avons montré qu'elle génère une signature de la jonction à quatre hélices. Son utilisation devrait permettre de discriminer définitivement entre les deux topologies envisageables pour le complexe CopA-CopT fonctionnel formé *in vivo*.

2. *Recherche d'autres intermédiaires réactionnels dans le processus d'appariement de CopA à CopT*

Grâce aux approches utilisées dans ce travail nous avons caractérisé un nouvel intermédiaire réactionnel dans le mécanisme de formation du complexe CopA-CopT stable. Afin de définir d'autres étapes du processus nous envisageons d'utiliser la cartographie en solution en temps résolu. Notre laboratoire vient d'acquérir un appareil de cinétique rapide (KinTech Corporation) qui devrait nous permettre d'utiliser les sondes enzymatiques ou chimiques sur des temps de réactions de l'ordre de la milliseconde. Ainsi, nous espérons pouvoir identifier des intermédiaires réactionnels entre CopA et CopT impossibles à détecter avec les méthodes traditionnelles. Nous esperons ainsi que cette approche, en conjonction avec la modélisation moléculaire, nous permettra de définir au niveau mécanistique comment est réalisée la transition de l'interaction boucle-boucle vers la jonction à quatre hélices.

3. *Cristallogenèse du complexe CopA-CopT*

L'analyse cristallographique du complexe CopA-CopT permettra d'obtenir une structure de résolution atomique de la jonction à quatre hélices. Elle définira précisément la topologie des régions des deux ARN pour lesquelles peu de contraintes structurales existent. Ceci concerne notamment les boucles de connexion et les nucléotides de la jonction.

Les résultats de cristallogenèse présentés dans la partie V de ce chapitre et obtenus jusqu'à présent sont encourageants. Le processus de nucléation va néanmoins devoir être ralenti. En effet, la quantité de cristaux obtenue dans une goutte est encore trop importante et leur croissance est trop rapide. Il en résulte qu'ils sont majoritairement polycristallins. Plusieurs autres paramètres tels que l'effet de la température ou l'addition de polyamines vont devoir être testés afin d'améliorer la qualité des cristaux. Par ailleurs, des essais de cristallisation par ensemencement, en utilisant les microcristaux déjà obtenus, pourraient éventuellement permettre d'obtenir des cristaux plus larges. A l'heure actuelle, le seul cristal suffisamment grand (~150 µm, figure 36: condition 26) pour que ses propriétés de diffraction des rayons X puissent être testées a été perdu lors d'une tentative de cryoprotection. L'obtention d'un tel cristal dans ces conditions n'a pas pu être reproduite.

Jusqu'à présent, la manière dont le complexe CopI-CopT (*i.e.* la jonction à quatre hélices) inhibe la fixation du ribosome demeure obscure. Récemment, une structure cristallographique de la sous-unité ribosomique 30S de *Thermus thermophilus* a été obtenue. Il serait donc intéressant de co-cristalliser le complexe ternaire ARNm/ARNt_i^{fmet}/sous-unité 30S afin de positionner la structure CopT par rapport à la sous-unité 30S.

2ème Chapitre

Etude structurale et fonctionnelle de l'ARN-III de *Staphylococcus aureus*

I. Introduction

A. La virulence des Staphylocoques

Les staphylocoques sont des pathogènes sévères qui causent une grande variété d'infections telles que des abcès cutanés, des endocardites, des pneumonies, des chocs toxiques et des intoxications alimentaires. Comme chez toutes les bactéries Gram-positives, la virulence de *S. aureus* est multifactorielle (pour une revue, [355]). Le tableau 8 répertorie les facteurs de virulence de *S. aureus*. Elle repose sur un grand nombre de toxines et d'enzymes sécrétées (α-toxine, hémolysines, leucocidine, protéases, lipase, nucléase et staphylokinase) et sur des protéines de surface (protéine A, coagulase et récepteurs de la fibronectine). *In vitro*, leur expression est contrôlée selon un schéma temporel précis représenté dans la figure 40.

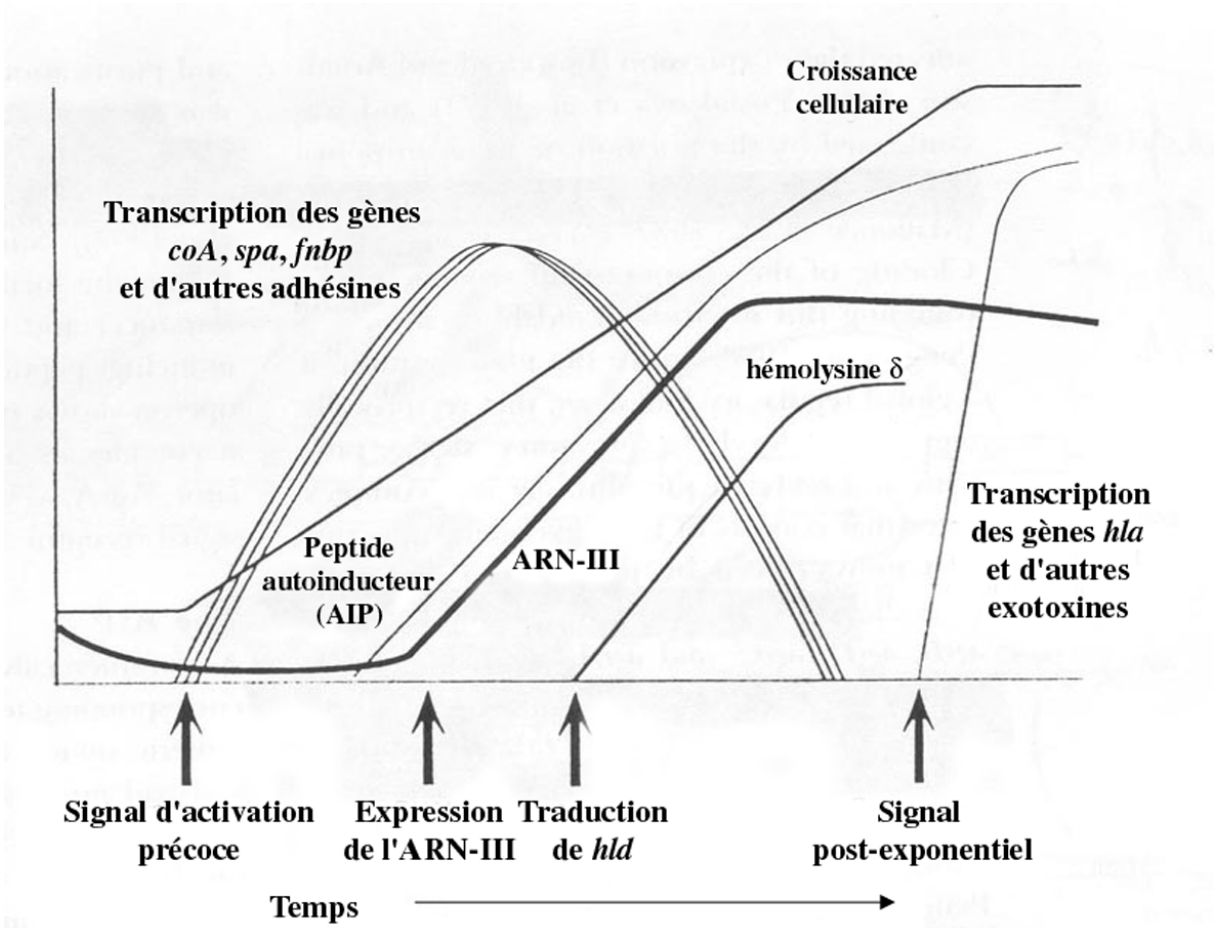

Figure 40 : ***Expression temporelle* in vitro *des différentes toxines de* Staphylococcus aureus *et du riborégulateur ARN-III en cours de croissance cellulaire, d'après [355]***. En début de phase exponentielle, les protéines de surface sont exprimées probablement en réponse à un signal temporel non déterminé. Durant la phase exponentielle, le peptide autoinducteur s'accumule et induit l'expression de l'ARN-III. L'expression des protéines de surface est réprimée et celle des facteurs excrétés est induite. L'hémolysine δ n'est synthétisée qu'une heure après le début de transcription de l'ARN-III.

Pour établir une infection productive *in vivo*, les différentes toxines doivent être synthétisées au moment approprié et en quantité adéquate. L'expression des facteurs commence en début de phase exponentielle (cas de la protéine A, de la coagulase, des protéines fixant la fibronectine et probablement également d'autres protéines de surface). Ces mêmes facteurs de surface sont néanmoins réprimés au niveau transcriptionnel en fin de phase exponentielle de croissance. Lors de la phase post-exponentielle et suite à un signal indéterminé, la transcription des exotoxines est activée. Celle-ci continue sur un court intervalle de temps et est réprimée en cours de phase stationnaire.

Plusieurs loci sont responsables conjointement de la régulation temporelle de l'expression des gènes de virulence de *S. aureus*. Ces gènes constituent un réseau complexe dans lequel tous les composants ne sont probablement pas encore caractérisés. Le locus *agr* (accessory gene regulator) est le régulateur principal et a été identifié par une insertion du transposon Tn551 atténuant sévèrement la virulence de *S. aureus* [356]. Depuis, le locus *sar* (staphylococcal accessory regulator) a également été identifié [357]. Il est impliqué dans l'activation transcriptionnelle du locus *agr* en début de croissance exponentielle. Parallèlement, le locus *sar* intervient directement dans le contrôle transcriptionnel de l'expression de la protéine A, indépendamment d'*agr* [358]. L'expression de *sar* est elle-même sous le contrôle de *sigB* [359]. Enfin les loci *sae* (*staphylococcus aureus* exoprotein expression) et *xpr* contrôlent l'expression des facteurs de virulence de manière *agr*-indépendante pour le premier [360] et par un mode d'action indéterminé pour le second [361].

Facteur de Virulence	Rôle dans la pathogénie	Régulation *agr*	*sar*	*sae*
Superantigènes				
Entérotoxines :				
A	Intoxication alimentaire, TSS			
B	Intoxication alimentaire, TSS	+	+	
C, D, E	Intoxication alimentaire, TSS	+		
Toxines de chocs toxiques :				
TSST-1	TSS	+	+	
TSST-0	Aucune (?)	+		
Exfoliatines (ETA, ETB)	Syndrome de la peau ébouillantée	+		
Cytotoxines				
Hémolysine α	Hémolyse, nécrose	+	+	+
Hémolysine β	Hémolyse, nécrose	+	+	+
Hémolysine δ	Hémolyse légère	+	+	
Hémolysine γ	Hémolyse, nécrose	+		
Leucocidine	Leucolyse, infections suppuratives, pneumopathies communautaires	+		
Enzymes				
Protéases	Diffusion, nutrition	+	-	
Nucléases	Diffusion, nutrition	+		+
Lipases	Diffusion, nutrition	+	-	
Hyaluromidase	Diffusion, nutrition	+		
Estérase	Inactivation des acides gras toxiques	+		
Protéines de surface				
Protéine A	Antiphagocytose	-	-	+
Coagulase	Antiphagocytose	-		+
FAF	Adhésion, antiphagocytose	-	+	+
FBP	Adhésion, internalisation	-	-	+
CAP	Adhésion		-	

Tableau 8 : ***Rôle des facteurs de virulence de* S. aureus *dans la pathogénie et leur régulation positive (+) ou négative (-) par les loci* agr, sar et sae**. TSS : Syndrome de choc toxique ; FAF : Facteur d'affinité pour le fibrinogène ; FBP : Protéine de liaison au fibrinogène ; CAP : Protéine d'adhésion au collagène. D'après [355].

B. Le locus *agr*

L'organisation génétique du locus *agr* est schématisée dans la figure 41. Il contient deux opérons divergents transcrits à partir des promoteurs P2 et P3. Cette région chromosomique est suffisante pour réprimer l'expression des protéines de surface et activer celle des exotoxines [362]. P2 dirige la synthèse de l'ARN-II, long de 3 kb et P3 dirige la synthèse de l'ARN-III, long de 0.5 kb.

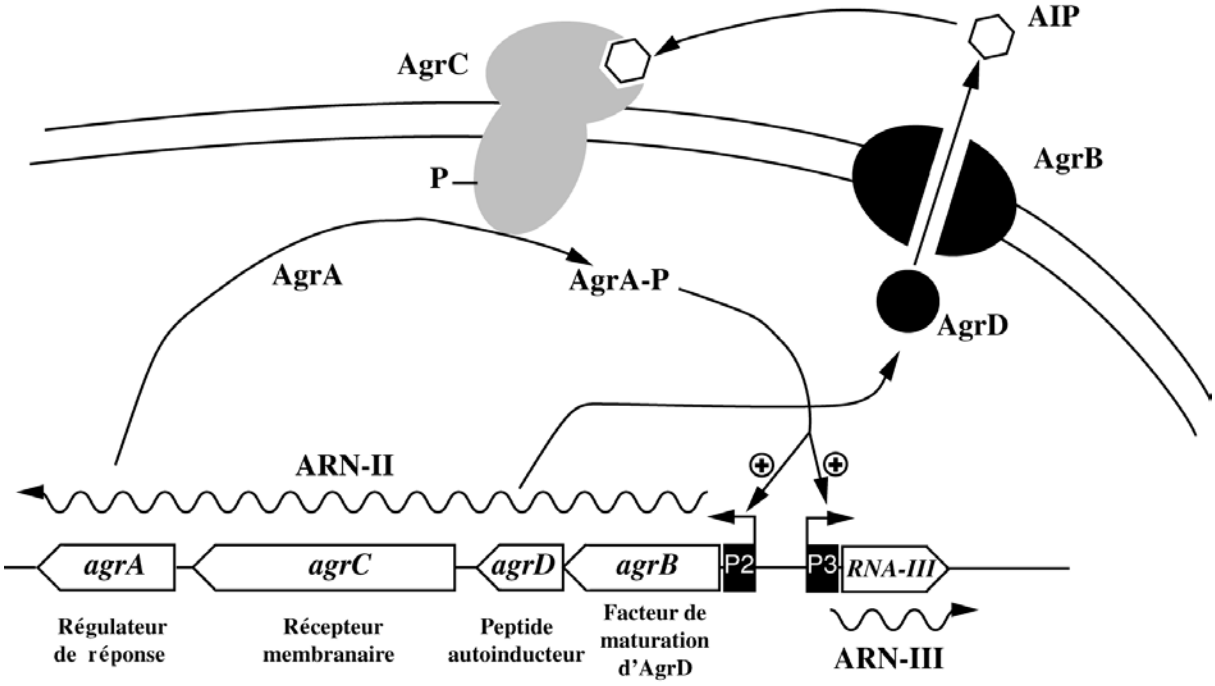

Figure 41 : ***Le locus* agr *et son système d'autoinduction***. Le transcrit polycistronique ARN-II et le riborégulateur ARN-III, produits respectivement à partir des promoteurs P2 et P3, sont indiqués par des flèches ondulées. La fonction des protéines codées par les gènes *agrA, B, C et D* sont indiqués sous l'opéron P2. La maturation d'AgrD en AIP par la protéine transmembranaire AgrB est représentée. L'AIP se fixe à la région extracellulaire d'AgrC et cause son autophosphorylation. L'histidine phosphorylée du domaine C-terminal d'AgrC est schématisée. Les formes non activée (AgrA) et activée (AgrA-P) d'AgrA sont indiquées. L'activation de la transcription de P2 et de P3 par AgrA-P est représentée par des (+) en fin de flèches.

1. *L'opéron P2*

L'opéron P2 code pour quatre gènes (*agrD*, *agrC*, *agrB* et *agrA*) nécessaires à l'activation du système *agr* [363]. La fonction de ces quatre gènes est d'autoactiver la transcription à partir du promoteur P2 et d'activer celle à partir du promoteur P3, en fonction de la densité cellulaire [355].

La protéine AgrD est un précurseur peptidique de 46 kDa. Sa maturation et sa sécrétion sont effectuées de manière concertée par la protéine transmembranaire AgrB. AgrB a une masse de 26 kDa et ne peut maturer qu'un précurseur AgrD homologue (*i.e.* de la même souche bactérienne) [364]. La forme maturée d'AgrD est un court peptide de 7 à 9 acides

aminés suivant la souche de Staphylocoque contenant un cycle thiolactone essentiel à son activité [365]. Un tel peptide naturel ou synthétique est capable d'activer l'expression du promoteur P2 *in vivo* dans un modèle d'infection murine [365]. Ce peptide, capable d'activer l'expression du locus à partir duquel il est exprimé, a été appelé AIP (pour AutoInducing Peptide). L'opéron P2 code donc pour un système assurant son auto-activation. Celle-ci est fonction de la densité cellulaire puisqu'elle est directement proportionnelle à la concentration d'AIP dans le milieu extracellulaire.

AgrC est une protéine transmembranaire de 46 kDa [366]. Sa région N-terminale est extracellulaire et sa région C-terminale cytoplasmique contient une histidine conservée caractéristique des récepteurs de signaux bactériens (Figure 41). AgrC est phosphorylée *in vitro* et *in vivo* sur son histidine et en réponse à l'AIP [366]. L'AIP se fixe sur la région N-terminale d'AgrC et stimule la cascade d'activation de P2 et de P3 lorsque les deux protéines sont issues d'une même souche de Staphylocoque. Cependant pour un AIP et un récepteur AgrC provenant de deux souches différentes, l'interaction inhibe la cascade et par conséquent la virulence bactérienne [364]. L'AIP joue alors le rôle d'un inhibiteur compétitif pour la fixation à AgrC et des subtilités conformationnelles des cycles thiolactones sont probablement responsables de la spécificité d'interaction entre l'AIP et son récepteur homologue [365]. Cette propriété pourrait permettre à une souche colonisant un hôte d'empêcher une surinfection par une souche différente et constitue une voix intéressante pour concevoir de nouveaux agents anti-staphylocoques.

AgrA est une protéine cytoplasmique de 34 kDa avec des caractéristiques communes aux régulateurs de réponse bactériens [367]. AgrC et AgrA constituent ensemble une cascade de transduction de signal extracellulaire. AgrA est activée par phosphorylation d'un acide aspartique conservé par AgrC~P. AgrA~P peut alors activer la transcription à partir des deux promoteurs P2 et P3 [362]. L'interaction d'AgrA avec la région intergénique P2-P3 n'a cependant jamais été montrée [368]. Ceci suggère soit que seule la forme phosphorylée d'AgrA pourrait reconnaître cette région, soit que l'intervention d'autres facteurs (comme par exemple SarA) pourrait faciliter cette interaction [369].

Ensemble, les quatre gènes codés par l'opéron P2 constituent donc un système de communication au sein d'une population bactérienne. Ils permettent d'activer l'expression de P2 et de P3 en réponse à une densité cellulaire élevée.

2. *L'opéron P3*

L'opéron P3 code pour un seul ARN, l'ARN-III. Celui-ci contient une phase ouverte de lecture (*hld*) codant pour l'hémolysine δ. Novick et collaborateurs ont montré que l'expression de l'ARN-III dans une souche de *S. aureus* délétée du locus *agr* restaurait le contrôle de l'expression des toxines [370]. L'ARN-III est donc l'effecteur global du locus *agr*. L'inactivation de *hld* n'affecte cependant pas les propriétés régulatrices de l'ARN-III. Celui-ci est donc un riborégulateur et *hld* a probablement été acquis de façon fortuite par l'ARN-III au cours de l'évolution. Ainsi, l'ARN-III *de S. lugdunensis* a les mêmes propriétés régulatrices que son homologue de *S. aureus* [371] mais ne contient pas la phase codante *hld* [372].

C. L'ARN-III : un riborégulateur à effet pléiotropique

En plus de coder pour *hld*, l'ARN-III intervient comme régulateur de l'expression des gènes de virulence chez *S. aureus*. Ce contrôle s'effectue au niveau transcriptionnel et traductionnel.

1. *L'ARN-III, régulateur transcriptionnel*

Grâce à l'étude de fusions transcriptionnelles entre les promoteurs de différents gènes cibles et le gène de la β-galactosidase, Novick et collaborateurs ont montré que l'ARN-III régule leur expression principalement au niveau transcriptionnel [370]. Les mécanismes par lesquels cette régulation est effectuée demeurent cependant inconnus. Plusieurs hypothèses peuvent néanmoins être envisagées. Une interaction directe entre l'ARN-III et l'ADN au niveau des séquences promotrices est peu probable. En effet, les promoteurs des différents gènes cibles ne présentent pas de caractéristiques communes qui pourraient constituer un site de reconnaissance par l'ARN-III. Une autre possibilité serait que l'ARN-III régule la traduction de facteurs de transcription impliqués dans l'expression des gènes de virulence. Enfin, l'ARN-III pourrait moduler l'activité de divers facteurs de transcription en interagissant directement avec eux. En faveur de cette hypothèse, l'ARN-III migre sous la forme de complexes de différents coefficients de sédimentation sur gradient de sucrose [370].

2. *L'ARN-III, régulateur traductionnel*

En plus de son intervention en tant que régulateur de la transcription d'un grand nombre de gènes, l'ARN-III régule la traduction d'au moins une toxine. Dans une souche mutante *agr*, la transcription de *hla* a lieu mais l'hémolysine α est indétectable dans le cytoplasme des cellules. Soit la protéine est rapidement dégradée, soit sa traduction est inhibée. L'identification d'une région complémentaire à la partie 5'-UTR de l'ARNm *hla* dans la région 5' de l'ARN-III a conduit à l'hypothèse que l'ARN-III pourrait interagir avec *hla* et activer sa traduction [370]. En l'absence d'ARN-III, l'ARNm *hla* n'est pas traduit [50,370]. En effet, la région de fixation du ribosome de *hla* est séquestrée dans une structure secondaire stable (Figure 42) [50]. L'ARN-III est complémentaire sur 75% à la région anti-SD. L'appariement ARN-III-*hla* a été caractérisé *in vitro* et *in vivo* [50]. Dans le complexe, la région Shine et Dalgarno devient accessible, confirmant que l'ARN-III intervient en tant qu'ARN antisens activateur de la traduction de *hla*. Ce mécanisme est particulier puisqu'un seul autre cas d'ARN antisens activateur a été décrit [167].

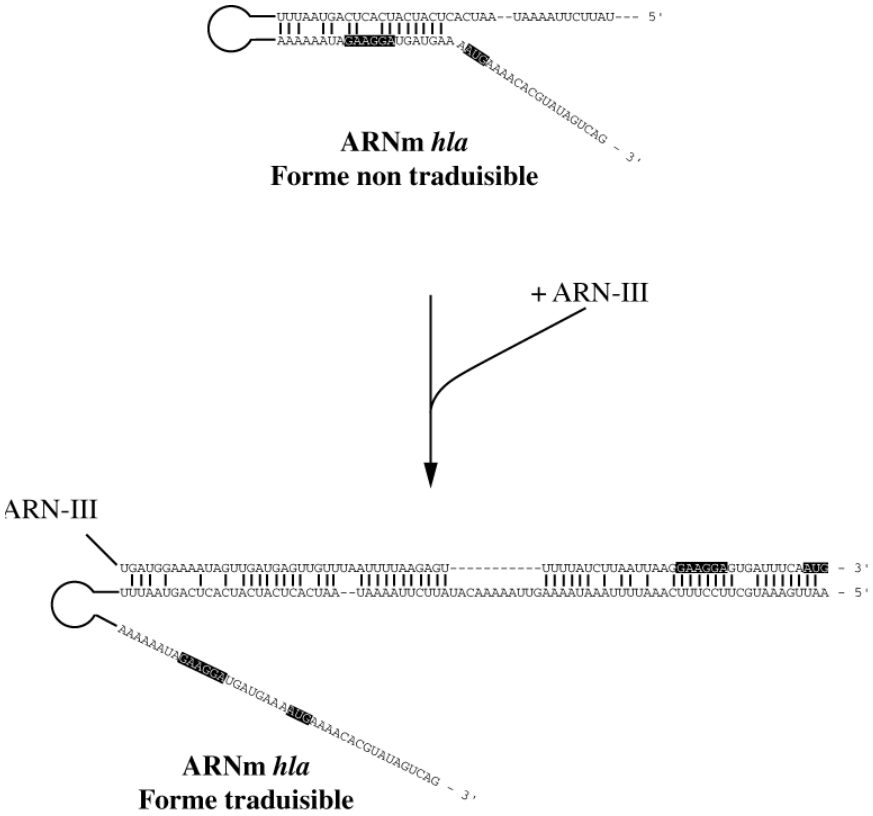

Figure 42 : ***Activation traductionnelle de*** **hla** ***par l'ARN-III***. Un modèle de repliement secondaire de l'ARNm *hla* libre non traduisible est représenté en haut. Le complexe putatif formé par l'interaction de *hla* avec l'ARN-III est représenté en bas. Les séquences Shine et Dalgarno et les codons AUG initiateurs des deux ARN sont indiqués par des boîtes noires.

L'ARN-III régule positivement à la fois la transcription et la traduction de *hla*. La raison de ce double contrôle est encore inconnue. Néanmoins, l'ARNm *hla* a une demi-vie relativement longue ($t_{1/2} \approx$ 10–15 minutes) par rapport aux ARNm des autres gènes de toxines [50]. Ce double contrôle pourrait ainsi permettre de diminuer le taux de synthèse de l'hémolysine α plus efficacement lors de l'entrée des cellules en phase stationnaire. Ceci ne serait pas permis par une simple répression de la transcription puisque l'ARNm subsisterait encore dans les cellules après l'arrêt de la transcription.

Dans tous les systèmes procaryotiques décrits dans l'introduction de ce mémoire, des structures particulières des ARN antisens sont responsables de l'efficacité des interactions avec leurs ARN cibles. Dans le cas de l'ARN-III, celles-ci n'ont pas encore été caractérisées. Dans le modèle de repliement bidimensionnel que nous proposons (voir article V), la région de l'ARN-III complémentaire à la région 5-UTR de *hla* contient plusieurs structures en tige-boucle. Celles-ci sont de bons candidats et leur implication dans le mécanisme de reconnaissance entre les deux ARN devra être testée. De manière alternative, un facteur protéique encore indéterminé pourrait faciliter l'interaction entre les deux ARN.

Il est envisageable que l'ARN-III puisse réguler la traduction d'autres gènes. Dans ce sens, il semblerait que l'ARN-III forme un complexe ARN-ARN avec l'ARNm du gène *tsst-1*, codant pour la toxin shock syndrom toxin-1 (Morfeldt *et al.*, résultats non publiés). Le mode de contrôle de l'expression de *tsst-1* doit néanmoins être différent de celui de *hla*. En effet, dans ce cas, l'ARNm est dépourvu de structure inhibitrice de la traduction. Par ailleurs, l'ARN-III est dépourvu de longues régions de complémentarité avec *tsst-1*. La réalité fonctionnelle de cette interaction devra être validée.

3. *L'ARN-III, riborégulateur dynamique*

Comme cela est schématisé dans la figure 40, il existe un délai d'une heure entre le moment où l'ARN-III est transcrit et celui où *hld* est traduit. Les fonctions activatrices de l'ARN-III sur les gènes des exoprotéines ne surviennent encore qu'une heure plus tard. Ce délai est maintenu dans une souche *agr*- dans laquelle l'ARN-III est exprimé sous la dépendance d'un promoteur inductible [373]. Il existe donc un signal temporel, indépendant d'*agr*, qui va permettre de moduler les différentes fonctions de l'ARN-III en cours de croissance cellulaire [374]. Cette modulation se fait probablement par des changements conformationnels de l'ARN-III et/ou par l'intervention de facteurs en *trans* [373].

L'ARN-III joue un rôle pléiotropique au sein des cellules et interagit pour cela avec de nombreuses macromolécules parmi lesquelles les ribosomes et l'ARNm *hla*. Le site d'initiation de la traduction de *hld* et la séquence complémentaire à la région 5'-UTR de *hla*

sont chevauchantes. Les deux types d'interaction sont donc mutuellement exclusifs. Il est envisageable que l'ARN-III alterne entre deux conformations lui permettant d'être spécifiquement reconnu par l'un de ces deux partenaires.

II. Article V : Etude structurale et fonctionnelle de l'ARN-III

Agissant sur différentes cibles par des mécanismes distincts, l'ARN-III est un ARN multirégulateur semblable à ceux décrits dans le § IV de l'introduction de ce mémoire. Dans les cas des ARN DsrA et OxyS, les différentes propriétés régulatrices reposent sur des domaines fonctionnels distincts [168,172]. Par conséquent il est fort probable que les différents domaines fonctionnels de l'ARN-III puissent également être génétiquement et structurellement séparés. Des délétions conséquentes ont permis à différents groupes de localiser sommairement les régions importantes pour différents contrôles [370,371]. Par ailleurs, le haut degré de conservation des 50 premiers et des 150 derniers nucléotides entre des ARN-III de différentes souches de Staphylocoques [375] suggère que ces régions sont importantes pour les fonctions régulatrices.

Dans cette étude, nous avons voulu déterminer la structure de l'ARN-III par des approches expérimentales, tant *in vitro* qu'*in vivo*. Dans ce but, nous avons cartographié l'ARN-III en solution à l'aide des différentes sondes de structure décrites dans le § I de la première partie des *Résultats et Discussion*. Nos résultats ont été appuyés par une simulation de repliement de l'ARN-III basée sur un algorithme génétique [376]. Celui-ci simule le repliement d'une molécule d'ARN en cours de transcription et favorise donc les interactions à courte plutôt qu'à longue distance.

Les résultats obtenus *in vitro* nous ont permis de proposer la présence de 14 structures en tige-boucle dans l'ARN-III de *S. aureus*. Leur existence est supportée par la réactivité des boucles vis-à-vis de sondes spécifiques de régions simple-brin et par la présence de clivages par la RNase V1 dans les hélices. Certaines d'entre elles sont cependant métastables puisqu'on peut noter les présences simultanées de coupures à la RNase V1 et de modifications par des sondes spécifiques de régions non structurées. Dans ce modèle, le repliement de l'ARN-III est compact, grâce à la formation de trois interactions à longue distance. Dans ce cas également, celles-ci n'ont pas toutes des stabilités semblables et leur formation nécessite la présence de cations divalents. Comme cela a été mentionné ci-dessus, l'ARN-III peut probablement adopter différentes conformations pour remplir ses différentes fonctions. Nous avons donc chercher à savoir si le modèle de repliement que nous proposons est la forme traduite. Dans ce but, nous avons testé la formation d'un complexe ternaire entre des sous-unités ribosomiques 30S purifiées, l'ARNt_i^{fmet} et l'ARN-III *in vitro*. Nos résultats montrent

que le modèle de structure proposé est la forme reconnue par le ribosome et pourrait correspondre à la forme traduite.

Les résultats obtenus *in vivo* par cartographie au DMS (N3C, N1A) confirment le modèle de repliement établi *in vitro*. Nous avons cependant caractérisé deux différences. En effet, *in vivo*, le site de fixation du ribosome est protégé contre la méthylation par le DMS. Ces résultats suggèrent que lors de l'étape de la phase de croissance à laquelle la cartographie a été effectuée (en fin de phase exponentielle), l'ARN-III est traduit. La deuxième différence observée est l'augmentation de réactivité des nucléotides appariés avec la région complémentaire à l'ARNm *hla in vitro*. En fin de phase exponentielle, l'ARN-III s'apparie à *hla* en détruisant des appariements intramoléculaires au profit d'appariements intermoléculaires. La séquence ainsi libérée deviendrait accessible à la modification chimique. Ces deux types d'interactions (avec le ribosome ou avec l'ARNm *hla*) sont mutuellement exclusives pour une molécule d'ARN-III donnée. Les expériences réalisées *in vivo* révèlent donc un mélange de population de molécules d'ARN-III, chacune intéragissant avec un type de partenaire différent.

Le but de cette étude n'était pas uniquement de déterminer la structure de l'ARN-III, mais de s'en servir comme base de travail pour l'identification de domaines fonctionnels indépendants. Les 150 derniers nucléotides sont hautement conservés et sont importants pour la répression de la transcription du gène de la protéine A [370,371,375]. Nous avons synthétisé *in vitro* un fragment correspondant aux deux dernières tiges-boucles de l'ARN-III. Cette région adopte la même topologie que dans l'ARN entier, confirmant qu'elle constitue un domaine structural indépendant. Nous avons également établi que ce domaine était suffisant pour assurer le contrôle de l'expression de *spa*. De manière intéressante, la tige-boucle 14, constituant le terminateur de transcription, permet un contrôle partiel. Ceci laisse suggérer que le terminateur de transcription participe au rôle régulateur de ce domaine, par exemple en créant un site de fixation bipartite à un facteur protéique non identifié.

RNA (2000), *6*:668–679. Cambridge University Press. Printed in the USA.
Copyright © 2000 RNA Society.

Probing the structure of RNAIII, the *Staphylococcus aureus agr* regulatory RNA, and identification of the RNA domain involved in repression of protein A expression

YVONNE BENITO,[1] FABRICE A. KOLB,[2] PASCALE ROMBY,[2] GERARD LINA,[1] JEROME ETIENNE,[1] and FRANÇOIS VANDENESCH[1]

[1]EA1655, Faculté de Médecine Laennec, 69372 Lyon Cedex 08, France
[2]UPR 9002 du Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 67084 Strasbourg Cedex, France

ABSTRACT

RNAIII, a 514-nt RNA molecule, regulates the expression of many *Staphylococcus aureus* genes encoding exoproteins and cell-wall-associated proteins. We have studied the structure of RNAIII in solution, using a combination of chemical and enzymatic probes. A model of the secondary structure was derived from experimental data with the help of computer simulation of RNA folding. The model contains 14 hairpin structures connected by unpaired nucleotides. The data also point to three helices formed by distant nucleotides that close off structural domains. This model was generally compatible with the results of in vivo probing experiments with dimethylsulfate in late exponential-phase cultures. Toe-printing experiments revealed that the ribosome binding site of *hld*, which is encoded by RNAIII, was accessible to the *Escherichia coli* 30S ribosomal subunit, suggesting that the in vitro structure represented a translatable form of RNAIII. We also found that, within the 3′ end of RNAIII, the conserved hairpin 13 and the terminator form an intrinsic structural domain that exerts specific regulatory activity on protein A gene expression.

Keywords: *agr* locus; chemical and enzymatic probing; in vivo probing; regulation; RNAIII

INTRODUCTION

In *Staphylococcus aureus*, expression of several virulence factors is regulated by the global regulon *agr*. The *agr* system is composed of two divergent operons designated P2 and P3. The P2 operon combines a density-sensing cassette (*agrD* and *B*) and a two-component sensory transduction system (*agrA* and *C*), both of which are required for autocatalytic activation of the promoter P2 and for transcription from the divergent promoter P3 (Ji et al., 1995; Lina et al., 1998). The effector of the *S. aureus agr* system is a 514-nt transcript of the P3 operon, designated RNAIII, which contains the 26-amino-acid delta-hemolysin gene (*hld*). RNAIII stimulates the expression of postexponentially expressed extracellular toxins and enzymes such as alpha and beta-hemolysins, enzymes (lipases, proteases and nucleases) and toxins (toxic shock syndrome toxin and enterotoxins), and represses the expression of exponential-phase surface proteins such as protein A and coagulase. RNAIII appears to be the specific regulator, as cloned RNAIII, defective in delta-hemolysin production, and expressed in an RNAIII-deletion mutant, restored both the positive and negative effects on the expression of exoproteins and surface proteins (Janzon & Arvidson, 1990) and in an *agr*-null strain (Kornblum et al., 1990; Novick et al., 1993). RNAIII acts primarily on target gene transcription initiation and, in the case of alpha-hemolysin, independently stimulates translation by a mechanism that appears to involve direct interaction between the 5′ end of RNAIII and a stretch of approximately 80 bp within the ribosome binding site of the *hla* transcript (Morfeldt et al., 1995). The mechanism by which RNAIII controls the transcription of its target genes is unknown. One possibility is that RNAIII acts via *trans*-acting factors, as RNAIII is potentially capable of forming stable secondary structures that may create protein binding sites (Novick et al., 1993). The importance of the 5′ and 3′ regions of RNAIII for its regulatory function has been tested by deletion

Reprint requests to: Francois Vandenesch, EA1655, Faculté de Médecine Laennec, Rue Guillaume Paradin, 69372 Lyon Cedex 08, France; e-mail: denesch@univ-lyon1.fr.

analysis and by studying the ability of RNAIII homologs from non-*aureus* staphylococcal species to complement *agr*-deficient strains of *S. aureus* (Benito et al., 1998; Tegmark et al., 1998). The 3′ end of RNAIII may be important to repress protein A expression, and may therefore be an intrinsic domain (Novick et al., 1993; Benito et al., 1998). As mentioned above, besides its regulatory function, RNAIII is also an mRNA that encodes the delta-hemolysin gene. Translation of RNAIII into delta-hemolysin was delayed by 1 h following transcription and this delay was abolished by deletion of the 3′ half of RNAIII, suggesting that a conformational change of the RNA is required for translation (Balaban & Novick, 1995). Secondary structure predictions indicate that the 5′ and 3′ ends of RNAIII may pair, thereby sequestering the ribosomal binding site of *hld* (Novick et al., 1993). Taken together, these data indicate that the structure of RNAIII plays an essential role not only in *hld* translation but also in regulatory functions.

Here, we probed the secondary structure of *S. aureus* RNAIII in vitro, using chemical and enzymatic probes. The results were compared with those of in vivo probing of RNAIII produced during the late exponential phase. The proposed model of RNAIII secondary structure is characterized by 14 hairpin structures and three long-range interactions that bring the 3′ and 5′ ends of RNAIII into close proximity. A number of hairpin structures appear to be conserved within the *Staphylococcus* species. One, located in the 3′ end of *S. aureus* RNAIII, efficiently repressed protein A expression in vivo. The proposed model for the secondary structure of RNAIII is discussed in the light of its regulatory functions.

RESULTS

Structural probing of *Staphylococcus aureus* RNAIII

We investigated the conformation of RNAIII (nt 1–514) with various chemical and enzymatic probes (Ehresmann et al., 1987) (Fig. 1). We chose an identical pH, temperature, ionic strength, and divalent ion concentration for the different probes to minimize the effects of these parameters on conformational changes. The four bases were tested at one of their Watson–Crick pairing positions by using DMS (N1A > N3C) and CMCT (N3U > N1G). Single-stranded regions were probed with RNases T2 and T1, and helical regions were detected with the double-stranded specific RNase V1. Information on the 3′ and 5′ ends of RNAIII was obtained using enzymes, DMS (for N3C), and nickel complex (which modifies N7G). Nickel complex is sensitive to stacking of base rings; for example, the N7 position of a guanine within a helix is not reactive unless the deep groove is widened (Chen et al., 1993). The RNA molecules were renatured in the presence of magnesium ions prior to modification. We also verified that the RNAs folded into a homogenous population by using gel electrophoresis in nondenaturing conditions. A secondary structure model for RNAIII was derived from the experimental data with the help of a computer simulation of RNA folding (Gultyaev et al., 1995). This procedure folds the molecule by generating consecutive transitions between favorable intermediate structures, that is, by simulating folding pathways rather than by searching for the global minimum state (Gultyaev et al., 1995; Gerdes et al., 1997). The secondary structure model is characterized by 14 hairpin structures and three helices (A–C) which involve distant nucleotides (Fig. 2). The model can be divided into three mains domains: the 5′ end domain (nt 1–31), a central domain delimitated by helix B (nt 32–382) and the 3′ end domain (nt 383–514).

The 5′ end domain contains a well-characterized stem-loop motif. This hairpin structure 1 bears a GNRA loop that is known to adopt a particular structure closed by a sheared G-A base pair (Heus & Pardi, 1991; Pley et al., 1994). The poor accessibility of G9 and A12 towards chemicals and enzymes supports the existence of such a G-A base pair (Fig. 1A,E). Two long-range interactions involving residues A24-G30/U452-458 (helix C) and U32-A42/U371-G382 (helix B) can be proposed. Both helices are susceptible to RNase V1 hydrolysis and several nucleotides are weakly or not reactive at one of their Watson–Crick positions. However helix C presents a weak stability as evidenced by the concomitant presence of several single-strand-specific RNase cleavages (Fig. 2). Thus, helix C brings in close proximity the 3′ and 5′ domains whereas helix B delimitates the central domain.

In this large central domain, the reactivity patterns of nucleotides in hairpin 2–11 were characteristic of hairpin structures (Fig. 2), as most of the nucleotides located in the loops were reactive at their Watson–Crick positions. This was accompanied by single-strand-specific RNase cleavages (Figs. 1 and 2). Furthermore, the stems were all cleaved by RNase V1, and most of the nucleotides were weakly or not reactive towards chemicals. This is particularly well-illustrated for the stem-loop motifs 3, 4, 5, and 7–11 (Fig. 2). Helix 2, which is A-U rich, forms a metastable helical structure, which may account for the concomitant presence of RNase V1 and RNase T2 cleavages and for the reactivity of several nucleotides at Watson–Crick positions (Fig. 2). The long hairpin structure 6, characterized by a stem rich in A-U base pairs and interrupted by an internal loop, required magnesium for stabilization: most of the residues within the helix became strongly reactive in the absence of magnesium (Fig. 1B). The central domain also contained a branched structure constituted by hairpins 7–9 and closed by helix A (Fig. 2). This helix involving the two distant regions G207–U216 and A310–U319 is supported by the presence of sev-

eral RNase V1 cleavages in both strands, and by the nonreactivity of several residues towards chemicals (Figs. 1 and 2). The single-stranded conformation of the interhelical regions is supported mainly by the reactivity of nucleotides at their Watson–Crick position (Figs. 1 and 2). This was notably the case for nucleotides A42–U47, U65–A68, A105–A109, C138–C140, A157–U163, A218–A220, U258–A260, U321–U327, and U350–C353.

The 3′ end domain is constituted by three hairpins, 12–14, and by a long single-stranded region involving residues 459–483 (Fig. 2). The stem-loop structure 13 is clearly identified, but its folding required the presence of magnesium, because many nucleotides within the helix became reactive under semidenaturing conditions (Fig. 2D). The stem, rich in A-U base pairs, is interrupted by two bulged nucleotides and closed by a 13-nt long loop. Finally, hairpin 14, corresponding to the terminator, is rich in G-C base pairs and is one of the most stable stem-loops of RNAIII. Its existence is supported by the presence of RNAse V1 cuts on both sides of the helix, and by the fact that none of the cytosines at position N3 and the guanines at position N7 (located in the helix) were reactive towards chemicals (Fig. 1C). In addition, cytosines 495–496 and 498–500 at their N3 positions were highly reactive with DMS, and several RNase T2 cleavages occurred in the loop (Fig. 1C).

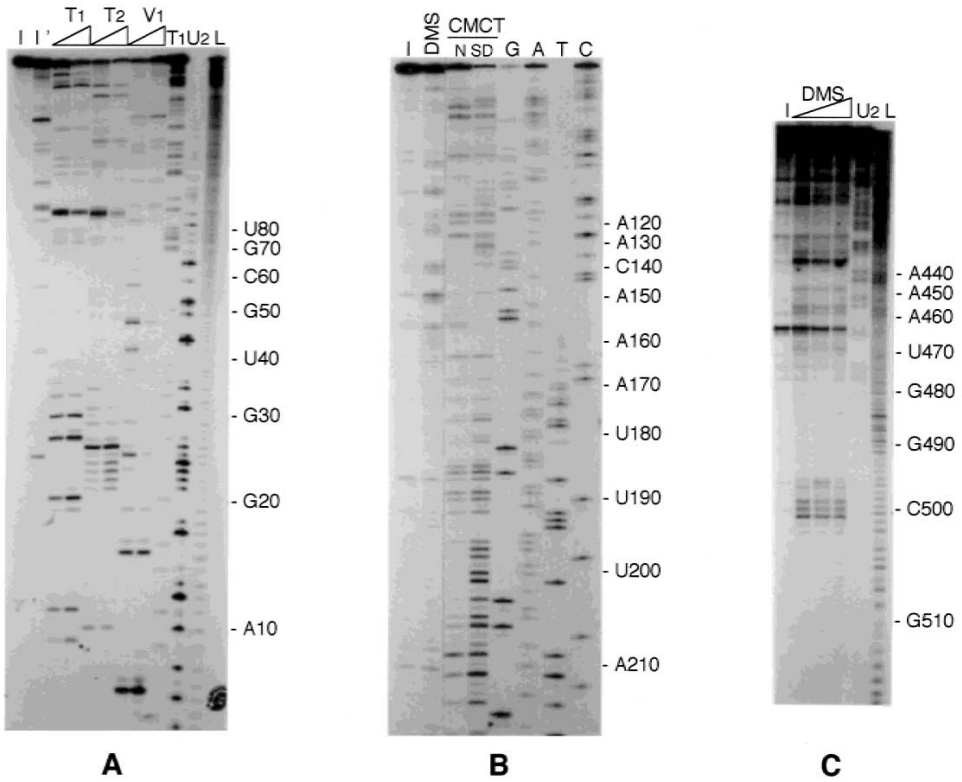

FIGURE 1. Probing the structure of RNAIII in vitro and in vivo. Selected autoradiograms showing enzymatic (**A**) and chemical (**B**,**C**) probing performed at 37 °C on in vitro transcribed RNAIII. **A**: RNase T1 (T1), RNase T2 (T2) and RNase V1 (V1) hydrolysis on 5′-labeled RNAIII, short migration. Lanes I, I′: incubation controls in buffer used for RNase T1, and RNase V1 hydrolysis, respectively; lanes T1: 0.05 and 0.15 U of RNase T1; lanes T2: 0.02 and 0.04 U of RNase T2; lanes V1: 0.1 and 0.2 U of RNase V1; lanes T1 and U2: RNases T1 and U2 in denaturing conditions, respectively; lane L: alkaline ladder. **B**: DMS and CMCT modifications, long migration. Lane I: incubation control in buffer used for DMS; lane DMS: native conditions, 1 μL of DMS diluted 1:16; lanes CMCT: native conditions (N) and semidenaturing conditions (SD), 2 μL of CMCT (100 mg/mL); lanes G, A, T, and C correspond to sequencing ladders. **C**: DMS modification of cytosines at N3 on 3′-end-labeled RNAIII, short migration. Lane I: incubation control; lanes DMS: native conditions, 1, 2, and 4 μL of DMS diluted 1:8; lanes U2 and L as in **A**. **D** and **E**: selected autoradiograms showing DMS and CMCT modifications on in vitro transcribed RNAIII (as in **B**) and a comparison with DMS modification on RNAIII in vivo, short migration. The differences in reactivities are denoted by arrows. Lanes I: incubation controls in buffer used for DMS; lanes DMS in vivo: 25 μL of DMS in 20 mL of staphylococcal culture at late exponential growth phase; lanes DMS in vitro: native conditions (N), semi-denaturing conditions (SD), 1 μL of DMS diluted 1:16; lanes CMCT: native conditions (N), semi-denaturing conditions (SD), 1 μL of CMCT (100 mg/mL); lanes G, A, T, and C correspond to sequencing ladders. (*Figure continues on facing page.*)

Probing the structure of the 3′ end of RNAIII

Because it had been suggested that the 5′ and 3′ ends of RNAIII would pair, and because specific regulatory activity was attributed to the 3′ end (Novick et al., 1993; Benito et al., 1998), the conformation of this region (nt 394–490) was probed separately with chemical and enzymatic probes (Fig. 3). Interestingly, the enzymatic cleavage and chemical reactivity patterns of nt 409–483 in the 3′ domain were identical to those in the entire RNAIII molecule and are thus consistent with the existence of stem-loop structure 13 (Figs. 2 and 3). Indeed, RNase V1 cuts occurred at positions 410–418, 421–422, 439, 443–444, and 447–450, whereas the main RNase T2 cleavage sites were located in the hairpin loop at positions 423–430 and 434–437 (Fig. 3). Furthermore, the nucleotides in the external loop were all highly reactive with chemical probes at their Watson–Crick position (Fig. 3). The presence of several RNase V1 cuts in the upper part of the stem (A419–U424/A438–U443) and the reactivity of several uridines pointed to the formation of a metastable helix.

The main reactivity changes observed between the entire RNAIII molecule and the 3′ end domain alone were located at nt U452–U466. Indeed, in the 3′ domain alone, all these nucleotides were reactive at one of their Watson–Crick positions and were cleaved by RNase T2 (Fig. 3A), indicating that this region is single stranded.

In vivo DMS probing of RNAIII

The structure of RNAIII was studied in vivo by using DMS, which alkylates cytosine at N3 and adenine at N1 (Mayford & Weisblum, 1989). In a typical experiment, *S. aureus* strain RN6390 was grown to the late exponential phase and treated with DMS. Sites of RNAIII modification were detected by the primer extension reaction on RNA extracts and were compared to modifications observed in vitro. As shown in Figure 1D,E, reactivity with RNAIII in vivo and in vitro was essentially identical, as only nucleotides in the hairpin loop and interhelical regions were highly sensitive to DMS. Differences between the in vitro and in vivo mapping data mainly concerned the reactivity of the adenine residues involved in the proposed long-range interactions, particularly helix B (U371–G382/U32–A42). These adenines were more sensitive to DMS alkylation at their N1 position in vivo than in vitro (Fig. 1D). Another sig-

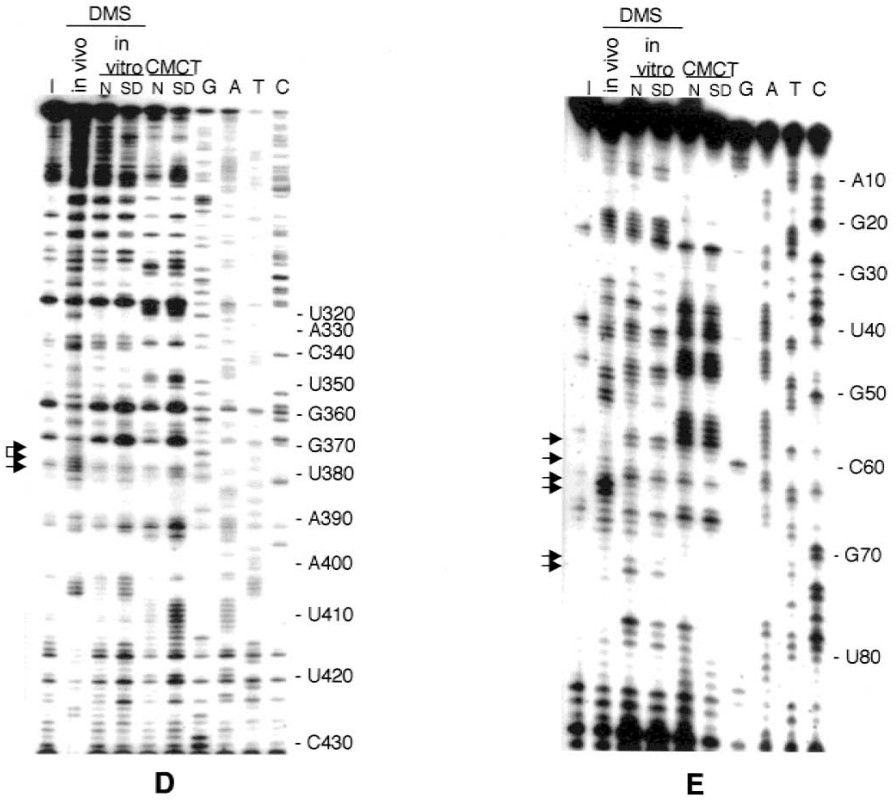

FIGURE 1. (*continued.*)

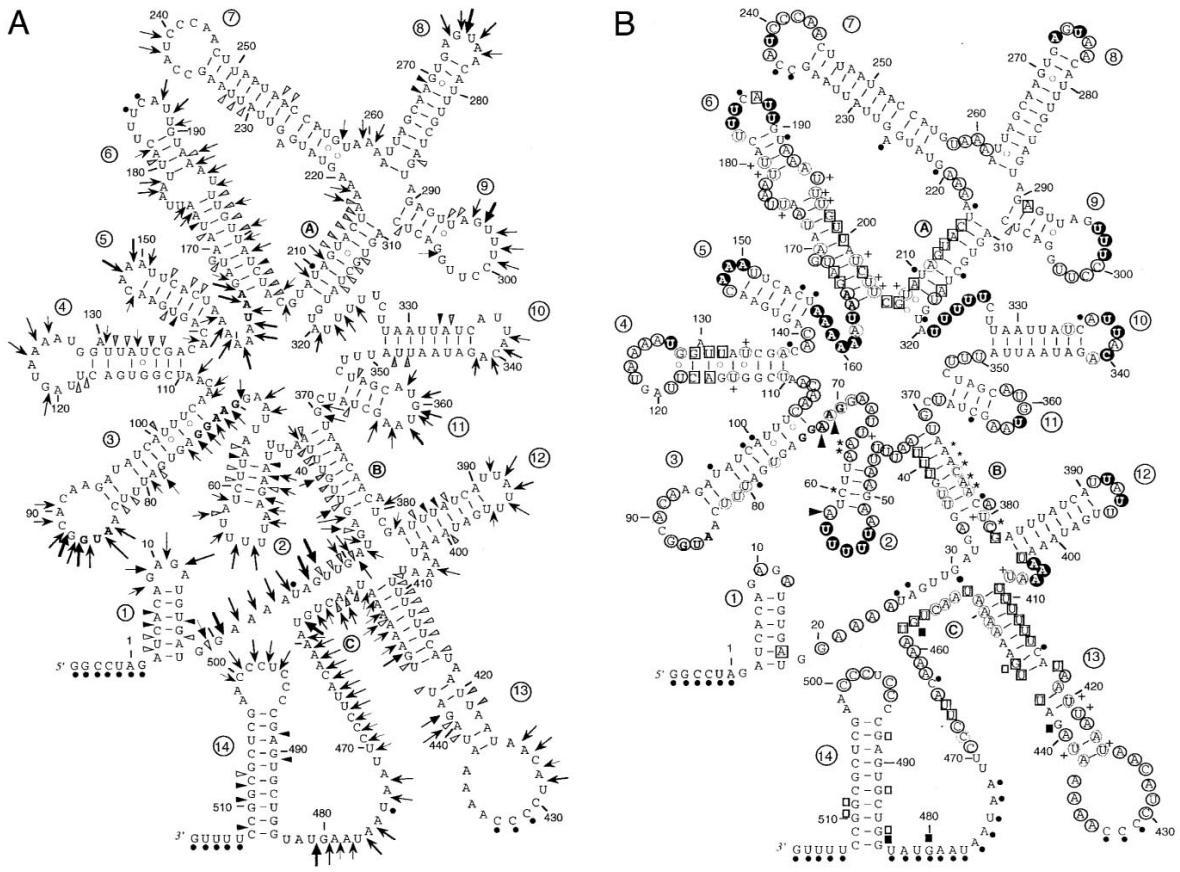

FIGURE 2. Secondary structure model of RNAIII showing the enzymatic cleavages (**A**) and reactivity of nucleotides to chemicals probes (**B**). The hairpin structures are numbered 1 to 14; long distance interactions are designated A, B and C; the Shine–Dalgarno sequence and the start and stop codons of *hld* are in bold face. **A**: RNase T1 cleavages are denoted by (→), RNase T2 by (→), thick, medium, or thin arrows for strong, moderate or low reactivity, respectively; RNase V1 cleavages are denoted by black, gray, or empty triangles for strong, moderate, or low reactivity, respectively. **B:** Reactivity of Watson–Crick positions to DMS (N1A, N3C) and CMCT (N3U, N1G) are represented as follows: reactive in native conditions (darkened black or grey circles for strong, moderate or low reactivity, respectively); reactive only in semidenaturing conditions (squares); more reactive in semidenaturing conditions (cross); nonreactive in both native and semidenaturing conditions (no symbol). Dots indicate not determined due to unspecific cleavages or pauses of RT in the incubation control. Guanines reactive at position N7 to NiCr are denoted by black squares, and empty squares denote unreactive guanines. Information about guanine accessibility at N7 position is only provided for region 410–512. The differences between the reactivity of adenine at N1 and cytosines at N3 to DMS in vitro and in vivo are indicated as follows: increased reactivity in vivo by an asterisk and protection in vivo by black triangles. These data were obtained from four independent experiments.

nificant difference between in vivo and in vitro mapping results concerned several adenine residues located at the ribosome initiation site. Indeed, adenines 58, 71, and 72 were clearly nonreactive in vivo, whereas adenines 58 and 57 became more reactive at their N1 positions in vivo (Figs. 1E, 2).

Ribosomes bind to *S. aureus* RNAIII

Although RNAIII is transcribed in the mid-exponential phase of growth, it is translated into delta-hemolysin only 1 h later (Balaban & Novick, 1995). This delay is not dependent on a specific timing signal during the growth cycle, and is abolished by deletion of the 3′ end of RNAIII (Balaban & Novick, 1995). We therefore probed the accessibility of the *hld* ribosome binding site in vitro by using a toe-printing assay (Hartz et al., 1988). In this primer extension inhibition assay, formation of the ternary complex composed of the 30S subunit, RNAIII, and the initiator $tRNA^{Met}$ blocks the elongation of a cDNA primer by reverse transcriptase (Hartz et al., 1988). Indeed, the $30S/tRNA_f^{Met}$ complex incubated with the refolded RNAIII produced a strong

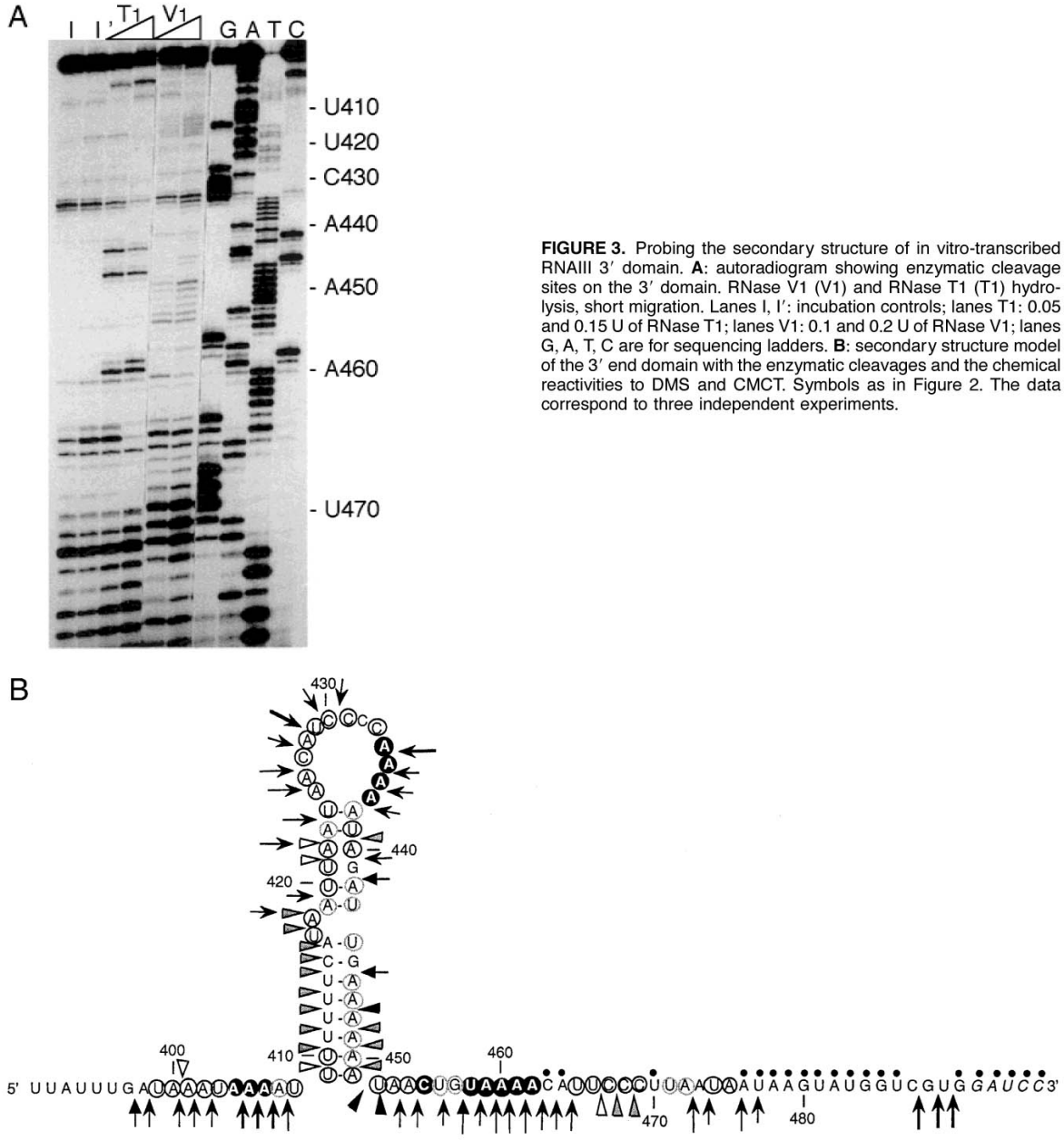

FIGURE 3. Probing the secondary structure of in vitro-transcribed RNAIII 3′ domain. **A**: autoradiogram showing enzymatic cleavage sites on the 3′ domain. RNase V1 (V1) and RNase T1 (T1) hydrolysis, short migration. Lanes I, I′: incubation controls; lanes T1: 0.05 and 0.15 U of RNase T1; lanes V1: 0.1 and 0.2 U of RNase V1; lanes G, A, T, C are for sequencing ladders. **B**: secondary structure model of the 3′ end domain with the enzymatic cleavages and the chemical reactivities to DMS and CMCT. Symbols as in Figure 2. The data correspond to three independent experiments.

toe-print at A100, 15 nt downstream of the initiation codon (Fig. 4), as previously shown (Hartz et al., 1988). The yield of the toe-print was identical when RNAIII was denatured prior to the experiment (Fig. 4). Given the possibility that the primer used for reverse transcription perturbed the structure of RNAIII, formation of the ternary 30S/RNAIII/tRNA^{Met} complex was also studied in filter binding assays (not shown). Again the data showed no major differences between native and denatured RNAIII. Altogether, these results indicate that ribosomes bind efficiently to *hld* RBS and that the in vitro structure represents a translatable form of RNAIII.

The 3′ domain of RNAIII represses protein A gene expression

We have shown here that the 3′ end of RNAIII is an independent structural domain. Furthermore, we and others have previously shown that the 3′ end of the

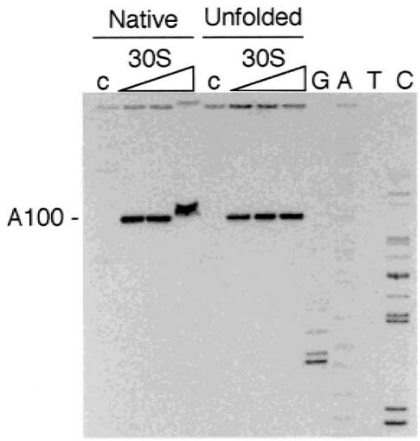

FIGURE 4. Accessibility of the ribosome binding site in RNAIII in vitro. The experiments were conducted in parallel on native and denatured (unfolded) RNAIII molecules. The position of the toeprint (A100) is indicated. Lane c: incubation control of free RNA in the presence of 30S subunit; lanes 30S: experiment in the presence of increasing concentrations of 30S subunits (125, 250 and 625 nM); lanes G, A, T and C are for sequencing ladders.

RNAIII is important for repressing protein A gene (*spa*) expression, and may thus be a functional domain (Novick et al., 1993; Benito et al., 1998). To further investigate this question, we subcloned selected fragments of RNAIII under the control of the RNAIII promoter on the staphylococcal plasmid pE194 and introduced the resulting constructs in *S. aureus* strain WA400 (*agr*$^-$, lacking RNAIII). Down regulation of *spa* expression was achieved normally in WA400/pE194 expressing either the RNAIII fragment comprising residues 391–514 (Fig. 5, lane 4) or residues 403–455 and 484–514 (both RNA fragments contain hairpin 13 and the terminator) (Fig. 5, lane 5). Interestingly, WA400/pE194 expressing a fragment of RNAIII that contains only the transcription terminator (nt 484–514) was still able to repress *spa* expression, albeit to a lesser extent (Fig. 5, lane 3).

DISCUSSION

The secondary structure model of *S. aureus* RNAIII

S. aureus RNAIII regulates the expression of several genes encoding exoproteins and cell-wall-associated proteins (Novick et al., 1993). RNAIII also encodes delta-hemolysin, but inactivation of the open reading frame does not affect the regulatory function of RNAIII (Janzon & Arvidson, 1990; Novick et al., 1993). There is also evidence that RNAIII may undergo conformational changes that activate either its translation or its regulatory functions (Novick et al., 1993; Balaban & Novick, 1995). A computer-simulated model of the secondary structure of RNAIII has been proposed (Novick et al., 1993). This model was based on minimum energy solutions and was characterized by extensive long-range interactions between the 5′ and 3′ domains (Novick et al., 1993). This prompted us to analyze the conformation of RNAIII in vitro by using several chemical and enzymatic probes, and in vivo by using DMS. These experimental data formed the basis for a secondary structure model derived with the help of computer simulation of RNAIII folding (Gultyaev et al., 1995). The present secondary structure of RNAIII is characterized by a number of hairpin structures (Fig. 2), most of which had been predicted (Novick et al., 1993; Tegmark et al., 1998). We also bring evidence that most of these stem-loop structures occurred in vivo. The alignment of RNAIII from *S. aureus* with that of non-*aureus* staphylococcal species (*S. epidermidis*, *S. warneri*, *S. simulans*, and *S. lugdunensis*) reveals that the stem-loop motifs 1, 7, 13, and 14 are predicted to be highly conserved by computer folding, and the hairpins 6, 8, 9, 11, and 12 must be present in all RNAIII but their sequences diverged (Vandenesch et al., 1993; Tegmark et al., 1998). Interestingly, several base pair compensatory changes are found in helices 1, 7, 13, and 14 among *Staphylococcus* species (Fig. 6). Thus the high conservation of hairpins 1, 7, 13, and 14 suggests a common function. For instance, with the exception of *S. epidermidis* RNAIII, which carries a 5-nt-long loop (Fig. 6), hairpin 1 contains a conserved GAGA loop motif. It is well known that stable stem-loop structures located at the 5′ end of mRNA may increase its stability (for a review, see Bechhofer, 1993). It is therefore tempting to propose a 5′ stabilizer function for this conserved hairpin motif in

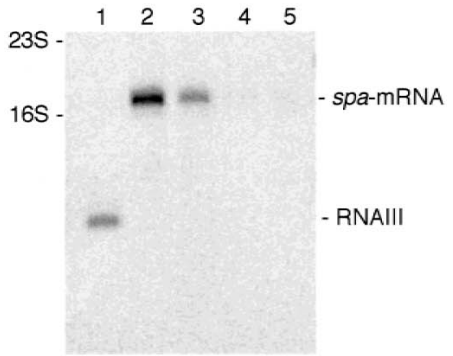

FIGURE 5. Effect of truncated RNAIII on protein A gene transcription. Northern blot analysis of RNAIII and protein A mRNA expression. RNAs from postexponential phase cultures were hybridized with probes corresponding to RNAIII and protein A gene (*spa*). Lane 1: RN6390 (*agr*$^+$); lane 2: WA400 (*agr* mutant); lane 3: WA400/pLUG310 (terminator transcription alone, nt 484–514); lane 4: WA400/pLUG300 (3′ end RNAIII, nt 391–514); lane 5: WA400/pLUG315 (hairpin structure 13 transcription terminator, nt 403–455::484–514). RNAIII transcripts are not seen in lanes 4 and 5 because of their very small sizes. Positions of migration of 16S rRNA (1541 nt) and 23S rRNA (2904 nt) indicated.

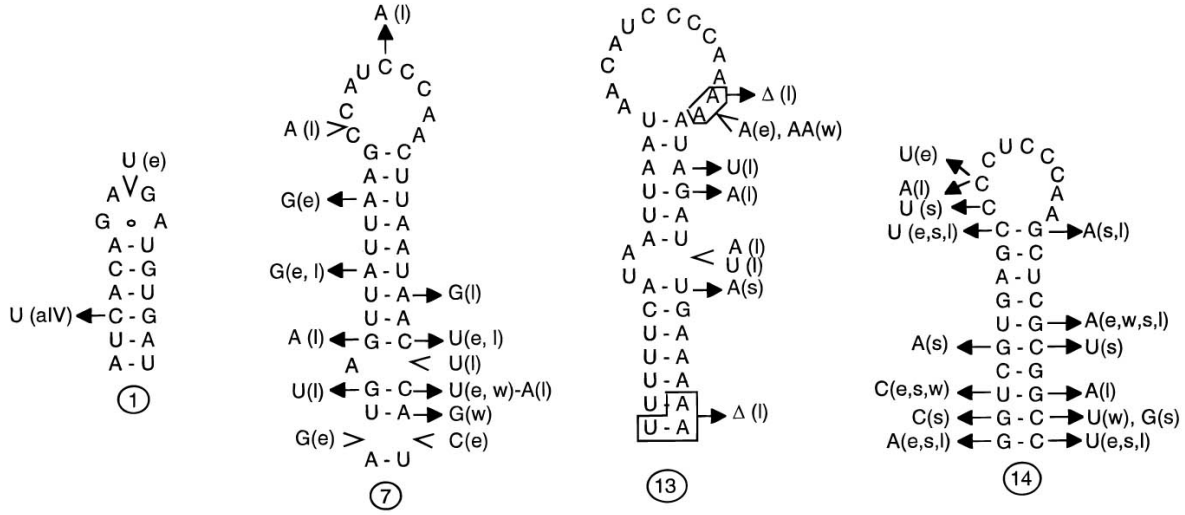

FIGURE 6. Sequence of several conserved hairpin structures of RNAIII from *S. aureus* (types I–IV), and comparison with four different species, *S. epidermidis* (e), *S. warneri* (w), *S. simulans* (s), and *S. lugdunensis* (l). Nucleotide insertion is shown by an open arrow, and Δ is for deletion. The hairpin structures have been assigned according to the secondary structure model presented in Figure 2.

RNAIII. The central domain (nt 32–382) is closed by helix B (Fig. 2). Analogous long-range interaction may also arise in *S. epidermidis*, *S. simulans*, and *S. warneri* as suggested by computer folding, despite the fact that the sequence is not conserved (data not shown). The central domain encloses the open reading frame encoding for delta-hemolysin that is present in all RNAIII, except *S. lugdunensis*, with a high degree of sequence identity (Vandenesch et al., 1993; Tegmark et al., 1998). The central domain also contains a branched structure constituted by three hairpins, 7–9, closed by a long-range interaction (helix A). The existence of helix A, which was clearly demonstrated in vitro, was also supported by the in vivo probing. In this branched structure, hairpin 7 is of particular interest. Several base pair compensatory changes are found among *Staphylococcus* species, and the loop contains a conserved sequence (UCCCAA) that resembles that of the loops of hairpin 13 and the terminator (Fig. 6). This high conservation indicates that hairpin 7 is a potential recognition site for *trans*-acting factor(s).

The central domain may undergo specific conformational changes

The 5′ side of the central domain of RNAIII contains the regulatory region (nt 19–87) involved in the activation of alpha-toxin translation (Novick et al., 1993; Morfeldt et al., 1995). Binding of RNAIII to the leader region of alpha-toxin mRNA induced a conformational change of the *hla* mRNA, rendering the ribosomal initiation site available for translation (Morfeldt et al., 1995). In the proposed model of RNAIII, the two long-range interactions B and C hide part of the *hla*-mRNA binding sequence, indicating that conformational changes of RNAIII are required for efficient binding. Interestingly, the greater accessibility of adenines 372–378 in vivo suggests that helix B is destabilized, likely as the result of *hla* mRNA binding to the 5′ region of RNAIII, explaining the apparent discrepancy with the in vitro data. Different pairing pathways that result in the formation of stable antisense RNA and mRNA complexes have been described (for a review, see Zeiler & Simons, 1998). In most cases, binding initiates between two loops, or between a loop and a single-stranded RNA segment involving a restricted number of nucleotides, and more stable complexes are subsequently formed by pairing of distal RNA segments. In region 42–72 of RNAIII, several nucleotides show high accessibility to single-strand-specific probes, indicating that they are potential sites for initial pairings. A fragment of RNAIII containing nt 1–380, which potentially forms helix B, is indeed competent for binding to *hla* mRNA in vitro (Morfeldt et al., 1995).

The central domain also contains the ribosomal initiation site required for the translation of delta-hemolysin, which is the only known translation product of RNAIII. Thus, the two functions of RNAIII (translation of *hld* and translational regulation of *hla*) are mutually exclusive. Structural probing of RNAIII in vitro revealed that the RBS sequence and the AUG initiation codon of *hld* are both accessible to single-strand-specific probes. The toe-printing and filter binding assays showed that the ribosome efficiently binds both to native RNAIII and to the denatured molecule at the *hld* initiation site (Fig. 4).

Footprinting experiments indicated that the initiating ribosome protects ~40–50 nt (Murakawa & Nierlich, 1989; Huttenhofer & Noller, 1994). As the 3′ side occurred at position +15 from the adenine of the initiation codon (as determined by the toe-printing approach), we assume that the ribosome may cover part of stem-loop structure 2 from the 5′ side. Many groups have investigated the effects of RNA structure on translation initiation (e.g., for a review, see de Smit & van Duin, 1994). However, chemical probing indicated that stem-loop structure 2 is metastable, and the ribosome is thus able to bind to the unfolded form. It was also shown that a 5-bp hairpin located 3 nt upstream of the SD sequence of the T4 *rIIB* gene can be increased to 9 bp without having an inhibitory effect (Shinedling et al., 1987). Therefore, our data indicate that the in vitro structure of RNAIII corresponds to its translatable form. One intriguing finding is that translation of *hld* in vivo is delayed by 1 h following transcription (Balaban & Novick, 1995). It was postulated that a conformational change of RNAIII to a translatable form preceded its regulation of target genes (Balaban & Novick, 1995). This apparently conflicts with the fact that the in vitro structure is efficiently recognized by the ribosome. It is possible that, owing to transcription–translation coupling, some transient inhibitory structures occurring in vivo sequester the ribosome binding site or influence the kinetic folding pathway of RNAIII in vivo. Interestingly, the overall pattern of nucleotide reactivity to chemical probes was very similar in vivo and in vitro. One major difference was located within the RBS: in particular, the two adenines (71 and 72) of the Shine–Dalgarno sequence became protected in vivo (Fig. 2), a phenomenon which may correlate with occupancy by the ribosome of its initiation *hld* binding site. Indeed, in the initiation complex, base-specific protection mainly occurs in the RBS and at the initiation codon, reflecting the formation of stable base pairings with the 3′ end of 16S rRNA and the anticodon sequence of the $tRNA_f^{Met}$, respectively (Huttenhofer & Noller, 1994). However, in vivo probing of RNAIII took place in the late exponential phase of growth when RNAIII is also involved in transcriptional regulation of target genes. It is likely that RNAIII undergoes specific conformational changes that are necessary for defined functions. Thus, several populations of RNAIII may coexist in vivo but are difficult to distinguish by chemical mapping.

The 3′ end domain is sufficient for the control of protein A expression

We also identified an intrinsic structural domain in the 3′ end of RNAIII, with hairpin 13 at its center. Stem-loop 13 has been predicted to be partially base paired with the region encompassing nt 130–98 (Novick et al., 1993). However, structural mapping of the full-length RNAIII and of the RNAIII domain from nt 391–490 showed an almost identical nucleotide reactivity pattern, a finding consistent with the existence of a stem-loop motif, as further supported by the in vivo probing (Fig. 3). The stem of hairpin 13 is interrupted by two bulged residues (U417 and A418). Despite the apparently lower stability of the upper stem, the probing data point to the existence of a 13-nt-long loop (Fig. 2). Moreover, this 3′ domain exerted specific regulatory activity, as the last 124 nt cloned downstream of the P3 promoter in an RNAIII-deficient strain were sufficient to repress *spa* expression (Fig. 6). It is remarkable that only hairpin 13 (nt 403–455) and the terminator hairpin (nt 484–514) were required for efficient *spa* transcription control, as deletion of the conserved residues U456 to U489 did not affect regulation (Fig. 2). It has previously been shown that RNAIII from *S. lugdunensis* does not repress *spa* transcription, while a chimeric construct consisting of the 5′ end of *S. lugdunensis* RNAIII and the 3′ end of *S. aureus* RNAIII repressed *spa* expression more efficiently in a *S. aureus* background (Benito et al., 1998). Several sequence differences were observed between *S. lugdunensis* and *S. aureus* RNAIII in the hairpin loop of the terminator and stem-loop 13 (Fig. 6). Conversely, other coagulase-negative staphylococci in which the last 150 nt of RNAIII (including stem-loop 13, Fig. 6) are more strongly conserved were able to repress *spa* expression (Tegmark et al., 1998). These data point to a shared function of the 3′ end domain of several RNAIII species in *spa* regulation. We also found that the last 31 nt of RNAIII (containing the transcription terminator) partially repressed *spa* expression, suggesting that the terminator also participates in the regulatory mechanism. Whether the 3′ domain regulates transcription directly or indirectly is unknown. One plausible explanation is that hairpin 13 and the terminator of RNAIII act via a ribonucleoprotein complex. A possible role of RNA in modulating the activity of DNA binding proteins has been reported (Retallack & Friedman, 1995; Sledjeski & Gottesman, 1995). It is also noteworthy that both hairpin loops contain a conserved CCCAA sequence that may provide a bipartite protein binding site (Fig. 6).

In summary, RNAIII has the unique property of acting as a messenger RNA and having pleiotropic regulatory functions. The secondary structure model of RNAIII proposed here should help to clarify functional domains. We demonstrated here that the 3′ domain of RNAIII contains an independent structure and has a regulatory function in protein A synthesis.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Escherichia coli TG1 and *S. aureus* RN4220, a nitrosoguanidine-induced mutant capable of accepting *E. coli* DNA (Kreiswirth et al., 1983), were used for plasmid amplification

and genetic manipulations. *S. aureus* RN6390 derives from 8325-4 and is our standard agr^+ strain. *S. aureus* WA400 is a derivative of 8325-4 in which the P2 operon is functional but the P3 operon is deleted and replaced by the chloramphenicol transacetylase gene (*cat86*) (Janzon & Arvidson, 1990) (see Table 1). Staphylococci were grown either on BM agar plates (1% peptone, 0.5% yeast extract, 0.1% glucose, 0.5% NaCl, 0.1% K_2HPO_4), or in brain-heart infusion (BHI), with erythromycin (5 μg mL^{-1}) when appropriate.

Molecular cloning and plasmid construction

Total DNA and plasmid DNA were prepared with standard methods (Sambrook et al., 1989). Transformation of *E. coli* TG1 was performed with cells rendered competent by treatment with $CaCl_2$, and *S. aureus* RN4220 and WA400 were transformed by electroporation (Bio-Rad gene pulser).

RNAIII was expressed in *Staphylococcus* by using plasmid pE194 (Horinouchi & Weisblum, 1982), which had been modified by adding an *Eco*RV restriction site next to the *Xba*I site forming pLUG274. PCR was performed from the RN6390 *agr* sequence (GenBank accession number X52543). To clone the 3′ domain of RNAIII under control of the P3 promoter, the promoter sequence was amplified using primers agr-sa1819/Stu1569 (Table 2), and the 3′ end of RNAIII was amplified using primers agr-sa757/Stu1180. The PCR products were digested by *Stu*I and ligated together before being reamplified with the external primers agr-sa1819 and agr-sa757. The resulting PCR product was digested with *Hin*dII and *Xba*I (site present on each primer) and ligated to the modified pE194 vector digested by *Eco*RV and *Xba*I. The ligation product was electroporated to RN4220 and the resulting plasmid (pLUG300) was introduced into strain WA400. To express intramolecular sequences of RNAIII under the control of the P3 promoter, we constructed, using a similar method, plasmid pLUG310 that contained the P3 promoter region linked to the transcriptional terminator (amplified using primers Bam1081 and agr-sa757); these two regions were separated by a short cloning sequence that contained the restriction sites *Stu*I and *BamH*I. To express hairpin structure 13 under the control of the P3 promoter, the corresponding sequence (nt 1168–1116) was amplified using primers Stu1168 and Bam1116, digested by *Stu*I and *BamH*I, and ligated to pLUG310 digested by the same enzymes.

For in vitro transcription, the RNAIII coding sequence (corresponding to nt 1–514 of RNAIII) was amplified using primers Stu1572 and Bam1055. The PCR product was digested by *Stu*I and *Bam*HI and ligated to the pUT7 vector (Serganov et al., 1997) cleaved with the same restriction enzymes. The resulting plasmid was introduced by transformation in *E. coli* TG1. Similarly, the 3′ end of RNAIII (nt 391–490) lacking the transcription terminator was amplified by using primers Stu1180 and Bam1079 and was then cloned onto pUT7 vector in *E. coli* TG1.

Northern blots

Electrophoresis of total RNA through 1% agarose gel containing 2.2 M formaldehyde, vacuum transfer to nylon membrane, hybridizations with specific digoxigenin-labeled RNA probes and luminescent detection were carried out as described (Benito et al., 1998). PCR primers for the production of *agr* RNAIII and protein A-specific gene probes were selected from GenBank accession number X52543 and A04518 respectively.

Preparation of RNA

RNAIII and the 3′ domain of RNAIII were transcribed in vitro with T7 RNA polymerase from *Bam*HI linearized pLUG322 and pLUG281, respectively. Transcription of RNAIII yielded a run-off product of 522 nt including five additional nucleotides at the 5′ end and three at the 3′ end, which all came from the vector. The 3′ domain corresponded to nt 391–490 of RNAIII.

TABLE 1. Main characteristics of strains and plasmids.

Strains or plasmids	Relevant characteristics	Reference or source
S. aureus strains		
8325-4	NCTC8325 cured of three prophages	Novick, 1963
RN4220	restriction-mutant of 8325-4	Kreiswirth et al., 1983
RN6390	derivative of 8325-4, *agr* positive	Peng et al., 1988
WA400	8325-4 Δ RNAIII region::*cat*86	Janzon & Arvidson, 1990
LUG450	WA400/pLUG300	this work
LUG460	WA400/pLUG310	this work
LUG467	WA400/pLUG315	this work
E. coli plasmids		
pUT7	pUC119::T7 promoter	Serganov et al., 1997
PLUG322	pUT7::RNAIII nt 1–514	this work
pLUG281	pUT7::RNAIII nt 391–490	this work
Staphylococcal plasmids		
pE194	3.728 kb *S. aureus* plasmid, inducible MLS resistance (*erm*)	Horinouchi & Weisblum, 1982
pLUG300	pE194::P3 promoter link to 3′-end RNAIII (nt 391–514)	this work
PLUG310	pE194::P3 promoter-TT (nt 484–514)	this work
pLUG315	pLUG310::hairpin 13 (nt 403–455)	this work

TABLE 2. Oligonucleotides.

Primer sense (+/−)	Position on *agr* sequence[a]	Position on RNAIII sequence[b]	5′-to-3′ sequence[c,d]	Restriction site
agr-sa757 (+)	757–771	—	AAC**GTT**AACGCGAAAATATAC	*Hin*dII
agr-sa1819 (−)	1819–1805	—	TTTT**CT**AGATACGTGGCAAAC	*Xba*I
Stu 1569	1569–1588	—	CTG**AGGC**CTAGTTATATTAAAACATGC	*Stu*I
Bam1081 (−)	1087–1068	484–503	**ATA**GGA**TCC**GAGCCCCTCCCAAG	*Bam*HI
Stu1168 (−)	1168–1144	403–427	TGA**GGCC**TAAAATTTTTTTCATAATTAATAAC	*Stu*I
Bam1116 (+)	1116–1133	455–438	GT**GGAT**CCGTTATTTTTTCAATCTAT	*Bam*HI
Stu1572 (−)	1572–1560	1–11	TA**GGC**CTAGATCACAGAG	*Stu*I
Stu1180 (−)	1180–1161	391–410	T**AGGCC**TTATTTGATAAATAAAATTT	*Stu*I
Bam1079 (+)	1081–1094	490–477	AGGG**ATC**CACGACCATACTTA	*Bam*HI
Bam1055 (+)	1055–1065	505–516	TCA**GGATCC**AAAAGGCCGCG	*Bam*HI
agr-sa1329 (+)	1329–1350	242–221	GGGATGGCTTAATAACTCATAC	—
agr-sa1101 (+)	1101–1120	470–451	AGGGAATGTTTTACAGTTAT	—

[a]According to GenBank sequence X52543.
[b]According to the direction of transcription (nt 1 of RNAIII (A) corresponds to nt 1570 (T) of the *agr* sequence).
[c]Restriction sites underlined.
[d]Additional nucleotides in bold.

In vitro transcribed RNAs were purified by 8% polyacrylamide-8 M urea gel electrophoresis. After elution in 0.5 M ammonium acetate/1 mM EDTA buffer, the RNAs were precipitated twice with ethanol.

To prepare RNAs with free 5′-OH groups, the in vitro transcription step using T7 RNA polymerase was performed in the presence of 4 mM ApG, and 1 mM NTPs according to Pitulle et al. (1992). 5′-end labeling of RNA or DNA oligonucleotides was performed with T4 polynucleotide kinase and [γ-^{32}P]ATP (Sambrook et al., 1989). 3′-end labeling of RNA was performed with T4 RNA ligase and [^{32}P]-pCp (England & Uhlenbeck, 1978). Labeled RNAs or DNA were purified by 8% polyacrylamide/8 M urea gel electrophoresis, eluted and precipitated twice with ethanol. Before use, the different RNAs were renatured by incubation at 90 °C for 2 min in RNase-free water, followed by slow cooling at 20 °C in the buffer used for enzymatic hydrolysis or chemical modifications (see below).

Enzymatic probing

Enzymatic hydrolysis was performed on either 50,000 cpm of end-labeled RNAIII or 2 pmol of cold RNA in 10 μL of TMN buffer (20 mM Tris-acetate, pH 7.5, 10 mM Mg-acetate, 100 mM sodium-acetate), in the presence of 1 μg of carrier tRNA at 20 °C for 5 min with one of the following enzymes: RNase T1 (0.005 and 0.01 U), RNase V1 (0.1 and 0.05 U) or RNase T2 (0.05 U).

Chemical probing

Chemical modifications were performed on 2 pmol of RNAIII or of the 3′ domain at 20 °C in 20 μL of reaction buffer containing 2 μg of carrier tRNA. Alkylation of the C(N3) and A(N1) positions was done with 1 μL DMS (diluted 1/8 and 1/16 in ethanol) for 2 min in TMN buffer. Modification of U(N3) and G(N1) were done with 5 μL of CMCT (100 mg/mL) for 15 min in a buffer containing 50 mM Na-borate, pH 8, 5 mM MgAc, 150 mM KOAc. DMS and CMCT reactions were also carried out in semi-denaturing conditions in the same buffers containing 1 mM EDTA instead of magnesium and KOAc.

When using end-labeled RNA, cleavage at modified C residues was induced by hydrazine subsequently followed by aniline treatment (Peattie & Gilbert, 1980). Modification of G(N7) by NiCR was performed at 20 °C for 30 min according to Chen et al. (1993) in a volume of 20 μL containing 25 mM phosphate-KOH pH 7.0, 10 mM $MgCl_2$, 100 mM NaCl, 2 μg carrier tRNA, in the presence of 3 μM NiCr and 200 μM $KHSO_5$. Cleavage at modified G residues was induced by aniline treatment (Peattie & Gilbert, 1980).

Identification of cleavage and modification sites

End-labeled RNA fragments were sized on 12% polyacrylamide/8 M slab gels. Cleavage positions were identified by running RNase T1, RNase U2 and alkaline ladders of the RNA in parallel (Donis-Keller et al., 1977). The cleavage or modification sites of unlabeled RNAs were detected by primer extension. The modified RNAs were hybridized with 4 pmol of 5′-end-labeled DNA primers agr-sa1101 (complementary to nt 470–451) or agr-sa1329 (complementary to nt 242–221). Details for hybridization conditions, primer extension and analysis of the data have been described previously (Brunel et al., 1991).

In vivo DMS modification of RNA

Late log-phase cultures of RN6390 (OD_{540} = 1.5) were treated with 25 and 50 μL of DMS for 5 min at 37 °C according to Mayford and Weisblum (1989). Reactions were stopped by adding 10 mL of cold 10 mM Tris-HCl, pH 8.0, 5 mM EDTA buffer. The cells were then pelleted, treated with lysostaphine (50 μg/mL) and 1% SDS. Total RNA was extracted with an equal volume (500 μL) of phenol/chloroform saturated in 0.1 M Na-acetate pH 4, precipitated with ethanol and resuspended in RNase-free water to a final concentration of 2 mg/mL. Primer extension was carried out with 5 μg of RNA as described above.

Toe-printing experiments

The formation of a simplified translational initiation complex and the extension inhibition conditions were strictly identical to those described by Moine et al. (1990). 30S subunits were prepared from tight couples according to a procedure adapted from Makhno et al. (1988). Relative toe-printing (toe-print band over 5′ ends + toe-print) was calculated by scanning the gel with a Bio-imager Analyzer BAS 2000 (Fuji).

ACKNOWLEDGMENTS

We thank G. Wagner, S. Arvidson, R.P. Novick, H. Moine, C. Ehresmann, and B. Ehresmann for stimulating discussions. This work was supported in part by grants from Centre National de la Recherche Scientifique (CNRS).

Received December 10, 1999; returned for revision February 3, 2000; revised manuscript received February 28, 2000

REFERENCES

Balaban N, Novick RP. 1995. Translation of RNAIII, the *Staphylococcus aureus agr* regulatory RNA molecule, can be activated by a 3′-end deletion. *FEMS Microbiol Lett 133*:155–161.

Bechhofer D. 1993. 5′ mRNA stabilizers. In: Belasco JG, Brawerman G, eds. *Control of messenger RNA stability*. London: Academic Press. pp 31–52.

Benito Y, Lina G, Greenland T, Etienne J, Vandenesch F. 1998. *Trans*-complementation of a *Staphylococcus aureus agr* mutant by *Staphylococcus lugdunensis agr* RNAIII. *J Bacteriol 180*:5780–5783.

Brunel C, Romby P, Westhof E, Ehresmann C, Ehresmann B. 1991. Three-dimensional model of *Escherichia coli* ribosomal 5 S RNA as deduced from structure probing in solution and computer modeling. *J Mol Biol 221*:293–308.

Chen X, Woodson SA, Burrows CJ, Rokita SE. 1993. A highly sensitive probe for guanine N7 in folded structures of RNA: Application to tRNA(Phe) and Tetrahymena group I intron. *Biochemistry 32*:7610–7616.

de Smit MH, van Duin J. 1994. Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data. *J Mol Biol 244*:144–150.

Donis-Keller H, Maxam AM, Gilbert W. 1977. Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res 4*:2527–2538.

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res 15*:53–71.

England TE, Uhlenbeck OC. 1978. 3′-terminal labelling of RNA with T4 RNA ligase. *Nature 275*:560–561.

Gerdes K, Gultyaev AP, Franch T, Pedersen K, Mikkelsen ND. 1997. Antisense RNA-regulated programmed cell death. *Annu Rev Genet 31*:1–31.

Gultyaev AP, van Batenburg FH, Pleij CW. 1995. The computer simulation of RNA folding pathways using a genetic algorithm. *J Mol Biol 250*:37–51.

Hartz D, McPheeters DS, Traut R, Gold L. 1988. Extension inhibition analysis of translation initiation complexes. *Methods Enzymol 164*:419–425.

Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science 253*:191–194.

Horinouchi S, Weisblum B. 1982. Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamine, and streptogramin type B antibiotics. *J Bacteriol 150*:804–814.

Huttenhofer A, Noller HF. 1994. Footprinting mRNA-ribosome complexes with chemical probes. *EMBO J 13*:3892–3901.

Janzon L, Arvidson S. 1990. The role of the delta lysin gene in the regulation of virulence genes by the accessory gene regulator (*agr*) in *Staphylococcus aureus*. *EMBO J 9*:1391–1399.

Ji G, Beavis RC, Novick RP. 1995. Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. *Proc Natl Acad Sci USA 92*:12055–12059.

Kornblum J, Kreiswirth B, Projan SJ, Ross H, Novick R. 1990. *Agr*: a polycistronic locus regulating exoprotein synthesis in *Staphylococcus aureus*. In: Novick RP, ed. *Molecular biology of the staphylococci*. New York: VCH Publishers. pp 373–401.

Kreiswirth B, Lofdahl S, Betley M, O'Reilly M, Schlievert P, Bergdoll M, Novick RP. 1983. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature 305*:709–712.

Lina G, Jarraud S, Ji G, Greenland T, Pedraza A, Etienne J, Novick RP, Vandenesch F. 1998. Transmembrane topology and histidine protein kinase activity of AgrC, the *agr* signal receptor in *Staphylococcus aureus*. *Mol Microbiol 28*:655–662.

Makhno VI, Peshin NN, Semenkov Iu P, Kirillov SV. 1988. A modified method of isolation of "tight" 70S ribosomes from *Escherichia coli* highly active at different stages of the elongation cycle. *Mol Biol* (Mosk) *22*:670–679.

Mayford M, Weisblum B. 1989. Conformational alterations in the *erm*C transcript in vivo during induction. *EMBO J 8*:4307–4314.

Moine H, Ehresmann B, Romby P, Ebel JP, Grunberg-Manago M, Springer M, Ehresmann C. 1990. The translational regulation of threonyl-tRNA synthetase. Functional relationship between the enzyme, the cognate tRNA and the ribosome. *Biochim Biophys Acta 1050*:343–350.

Morfeldt E, Taylor D, von Gabain A, Arvidson S. 1995. Activation of alpha-toxin translation in *Staphylococcus aureus* by the *trans*-encoded antisense RNA, RNAIII. *EMBO J 14*:4569–4577.

Murakawa GJ, Nierlich DP. 1989. Mapping the *lac*Z ribosome binding site by RNA footprinting. *Biochemistry 28*:8067–8072.

Novick RP. 1963. Analysis by transduction of mutations affecting penicillinase formation in *Staphylococcus aureus*. *J Gen Microbiol 33*:121–136.

Novick RP, Ross HF, Projan SJ, Kornblum J, Kreiswirth B, Moghazeh S. 1993. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J 12*:3967–3975.

Peattie DA, Gilbert W. 1980. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci USA 77*:4679–4682.

Peng HC, Novick RP, Kreiswirth B, Kornblum J, Schlievert PM. 1988. Cloning, characterization, and sequencing of an accessory gene regulator *agr* in *Staphylococcus aureus*. *J Bacteriol 170*:4365–4372.

Pitulle C, Kleineidam RG, Sproat B, Krupp G. 1992. Initiator oligonucleotides for the combination of chemical and enzymatic RNA synthesis. *Gene 112*:101–105.

Pley HW, Flaherty KM, McKay DB. 1994. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature 372*:111–113.

Retallack DM, Friedman DI. 1995. A role for a small stable RNA in modulating the activity of DNA-binding proteins. *Cell 83*:227–235.

Sambrook J, Fritsch E, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Serganov A, Rak A, Garber M, Reinbolt J, Ehresmann B, Ehresmann C, Grunberg-Manago M, Portier C. 1997. Ribosomal protein S15 from *Thermus thermophilus*: Cloning, sequencing, overexpression of the gene and RNA-binding properties of the protein. *Eur J Biochem 246*:291–300.

Shinedling S, Gayle M, Pribnow D, Gold L. 1987. Mutations affecting translation of the bacteriophage T4 rIIB gene cloned in *Escherichia coli*. *Mol Gen Genet 207*:224–232.

Sledjeski D, Gottesman S. 1995. A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli*. *Proc Natl Acad Sci USA 92*:2003–2007.

Tegmark K, Morfeldt E, Arvidson S. 1998. Regulation of *agr*-dependent virulence genes in *Staphylococcus aureus* by RNAIII from coagulase-negative staphylococci. *J Bacteriol 180*:3181–3186.

Vandenesch F, Projan S, Kreiswirth B, Etienne J, Novick R. 1993. *Agr*-related sequences in *Staphylococcus lugdunensis*. *FEMS Microbiol Lett 111*:115–122.

Zeiler BN, Simons RW. 1998. Antisense RNA structure and function. In: Simons RW, Grunberg-Manago M, eds. *RNA structure and function*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 437–464.

III. Recherche de facteurs se fixant à l'ARN-III par pontages aux UV

Dans l'étude précédente, nous avons montré que le domaine 3' de l'ARN-III, contenant les deux dernières structures en tige-boucle, était suffisant pour contrôler l'expression de la protéine A. De plus, le terminateur de transcription est impliqué dans ce contrôle (Article V, [377]). Celui-ci se faisant probablement *via* un facteur protéique, nous avons entrepris de caractériser la ou les protéine(s) reconnaissant ce sous-domaine de l'ARN-III. Dans ce but, nous avons synthétisé cette région par transcription *in vitro* (Figure 43A : ARN LUG361) et réalisé des pontages non dirigés aux UV à l'aide d'extraits bactériens de différentes souches de *S. aureus*. Par ailleurs, Benito et collaborateurs avaient montré que le domaine 3' de l'ARN-III de *S. lugdunensis* ne permettait pas de restaurer ce contrôle dans une souche *agr*- de *S. aureus* [371]. Ceci suggère que ce dernier fragment ne contient pas tous les déterminants de séquence pour être reconnu par le(s) facteur(s) de *S. aureus* nécessaire(s) au contrôle de l'expression de la protéine A. Parmi les différences de séquence entre les deux ARN, une position varie dans la boucle 14. Nous avons donc également synthétisé une ARN chimérique contenant la tige-boucle 13 de l'ARN-III de *S.aureus* et la tige-boucle 14 (*i.e.* le terminateur) de *S. lugdunensis* (Figure 43B : ARN LUG362). Nous l'avons soumis au même type d'expériences de pontages avec l'espoir d'identifier un facteur reconnaissant le domaine 3' de l'ARN-III de *S. aureus* mais pas l'ARN-III chimérique.

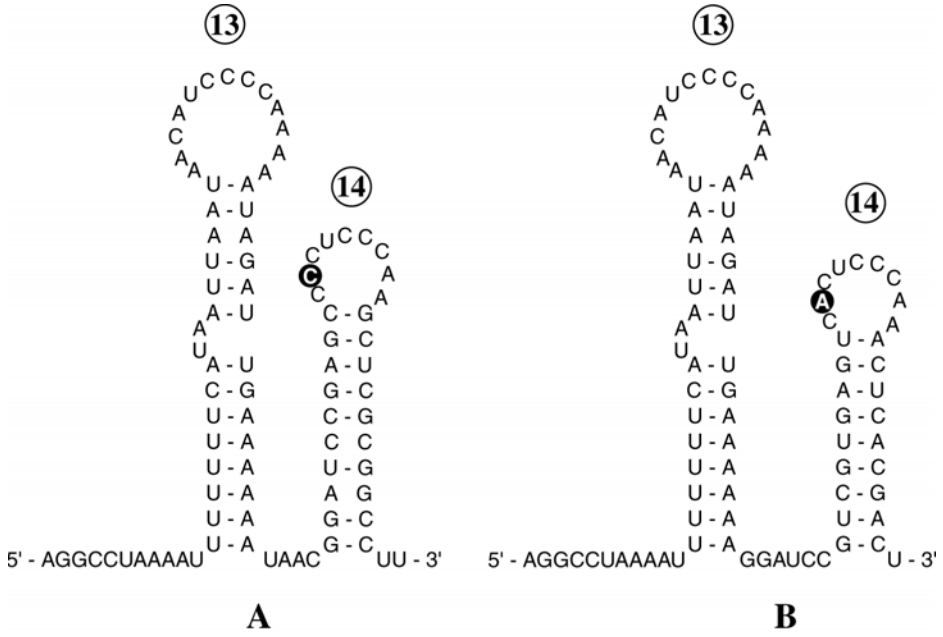

Figure 43 : ***Modèles de structures secondaires des ARN LUG361 (A) et LUG362 (B)***. Les tiges-boucles 13 et 14 sont indiquées. Les divergences de séquence dans la boucle 14 sont schématisées en caractères blancs sur fonds noirs.

Après marquage radioactif uniforme au [α-^{32}P]-ATP, les ARN ont été renaturés 15 min à 20°C dans un tampon 150 mM Na Hepes pH 7,5 ; 150 mM KAc ; 1,5 mM MgAc ; 5 mM β-mercaptoéthanol ; glycérol 2%. Ils ont alors été incubés 15 min sur la glace dans ce même tampon en présence d'extraits bactériens provenant d'une souche *agr*+ (RN6390) ou *agr*- (RN6911) de *S. aureus* cultivées jusqu'à une densité optique de 0,5 ou de 1,5. Toutes les incubations ont été effectuées en présence de 2,5 µg d'ARNt total afin de diminuer les reconnaissances aspécifiques des ARN par des protéines. Les milieux ont alors été irradiés à 254 nm, 50 min à 120 mJ sur la glace, puis traités 1 h à 37°C par 1 unité de RNase A. Les échantillons ont finalement été déposés sur gel de polyacrylamide-SDS 10%, migrés à 250 V et les complexes ARN-protéine pontés ont été visualisées par autoradiographie (Figure 44B). Des contrôles non irradiés ont été réalisés en parallèle (Figure 44A).

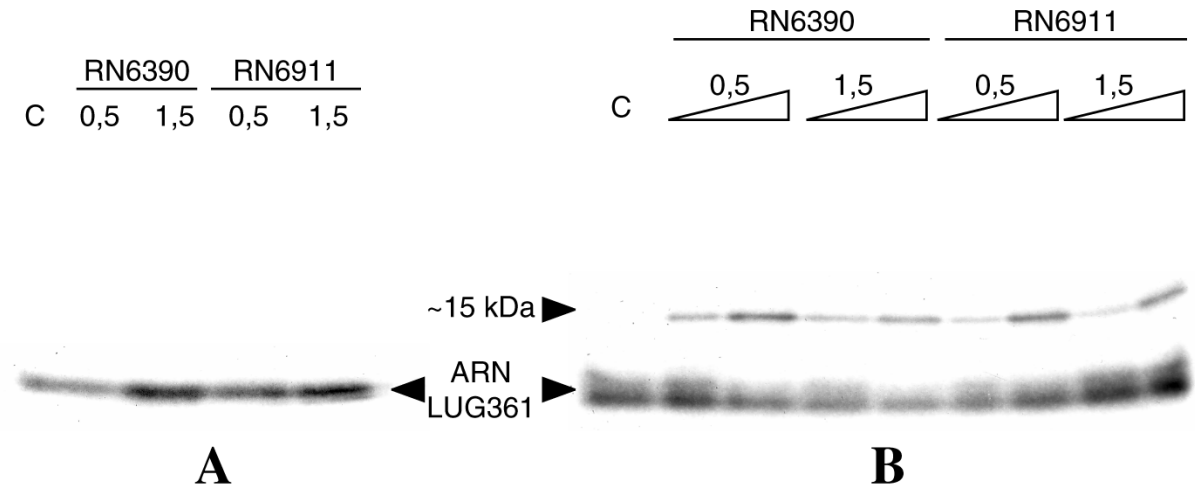

Figure 44 : ***Expériences de pontage aux UV sur l'ARN LUG361***. Autoradiographies des gels de polyacrylamide sur lesquels les échantillons non soumis à l'irradiation aux UV (A) ou soumis aux UV (B) ont été déposés. Les essais dans lesquels l'ARN a été incubé avec des extraits réalisés à une DO de 0,5 ou de 1,5 sont indiqués par 0,5 et 1,5, respectivement. Les souches utilisées sont indiquées au-dessus des gels. Les triangles représentent une incubation avec 1 µl ou 5 µl d'extrait bactérien. Les positions de l'ARN LUG361 libre et du complexe ponté (15 kDa) sont indiquées. Pistes « C » : contrôles d'incubations sans extrait protéique.

Sans irradiation, aucun retard de migration sur gel du domaine 3' de l'ARN-III n'est observé (Figure 44A), et aucune bande ne comigre au niveau du complexe ponté lorsque les ARN sont irradiés en absence d'extrait bactérien (Figure 44B, ligne C). En revanche, lorsque les échantillons sont traités aux UV, un signal reproductible correspondant à une protéine de ~15kDa est détectable (Figure 44B). Celui-ci est observé avec les extraits bruts des souches RN6390 ou RN6911, suggérant que la protéine n'est pas codée par le locus *agr*, ou que son expression n'est pas sous son contrôle. Enfin, l'expression de ce facteur ne semble pas être régulée en cours de croissance cellulaire. Ceci est indiqué par la présence d'un signal d'intensité comparable pour les essais effectués avec des extraits réalisés à DO 0,5 et 1,5. De

manière inattendue, le pontage est également maintenu lorsque l'ARN-III chimérique contenant le terminateur de transcription de *S. lugdunensis* est utilisé (résultat non montré). Ce résultat indique soit que la protéine révélée par ces expériences de pontage n'est pas impliquée dans le contrôle de l'expression de la protéine A, soit qu'elle est impliquée mais que les divergences de séquence dans la tige-boucle 14 n'affectent pas sa fixation. Finalement, des expériences de compétition indiquent que les pontages identifiés ici sont spécifiques de l'ARN-III puisque l'ARNt est incapable de déplacer le complexe ponté (résultat non montré).

Dans le but d'identifier le facteur reconnaissant l'ARN-III, la bande correspondant au complexe ponté dans la figure 44 a été découpée. Après transfert sur membrane de nitrocellulose et élution, la protéine pontée a été séquencée par son extrémité N-terminale suivant la méthode d'Edman [378]. Malheureusement aucun signal significatif n'a pu être obtenu en raison de la superposition de plusieurs peptides. La résolution du gel utilisé pour séparer le complexe ponté n'était probablement pas suffisante et celui-ci comigrait probablement avec d'autres protéines. Une solution à ce problème est de fractionner les extraits bruts par différents procédés de purification et de tester la capacité de chaque fraction à être pontée à l'ARN LUG361. Le séquençage devrait alors permettre de caractériser le peptide reconnaissant spécifiquement le domaine 3' de l'ARN-III. De manière alternative, le facteur ponté à l'ARN-III pourrait être un ARN. Ceci pourra être vérifié en traitant les extraits bactériens à la protéinase K, préalablement à l'irradiation aux UV.

IV. Conclusions et Perspectives

A. Les différents domaines fonctionnels de l'ARN-III

La grande taille de l'ARN-III et la grande diversité de ses cibles laissaient suggérer qu'il possède plusieurs domaines fonctionnels distincts. Les régions impliquées dans l'expression de l'hémolysine δ et dans l'activation traductionnelle de *hla* était déjà connues lorsque nous avons entamé ce travail (Figure 45). Cependant, l'implication d'autres régions dans les processus de contrôle n'était que spéculative. La fonction des ARN reposant essentiellement sur leur structure, nous avons déterminé le repliement de l'ARN-III *in vitro* et *in vivo*. Nous avons ainsi pu établir que l'ARN-III était constitué de 14 structures en tige-boucle et de plusieurs interactions à longue distance (Figure 45). Par ailleurs, nous avons montré que cette conformation est la forme de l'ARN-III reconnue par les ribosomes. Cette hypothèse pourra être confirmée par des essais de traduction *in vitro*.

Le but principal de cette étude était de proposer un modèle de structure de l'ARN-III et de tester l'implication des différents motifs dans le contrôle de l'expression de gène cibles. Nous avons ainsi déterminé que le domaine 3', constitué des deux dernières structures en tige-boucle, était suffisant pour assurer la répression de *spa* (Figure 45). Ce résultat rejoint ceux obtenus indirectement par des approches génétiques en utilisant des ARN-III chimériques [371,375]. Le même type d'analyse pourra être effectué avec d'autres domaines isolés. Parmi eux, les régions présentant un haut degré de conservation de séquence et/ou de structure entre différentes espèces de Staphylocoques sont des candidats de choix pour les prochaines études. Ainsi, l'implication des tiges-boucles 1, 6, 7, 8, 11 et 12 dans des fonctions précises (stabilité de l'ARN-III, contrôles transcriptionnels,...) devra être testée. Des résultats préliminaires semblent indiquer que la délétion du sous-domaine 7-8-9, défini par l'interaction à longue distance A, affecte le contrôle transcriptionnel du gène *ssp* codant pour la sérine-protéase (Benito Y. & Vandenesch F., Lyon, communication personnelle). Une autre approche complémentaire consistera à exprimer dans une souche *agr-* des ARN-III délétés d'un domaine particulier. Finalement, les rôles des trois interactions à longue distance sur la fonction de l'ARN-III devront également être testés par des essais de complémentation d'une souche *agr-* à l'aide d'ARN mutants.

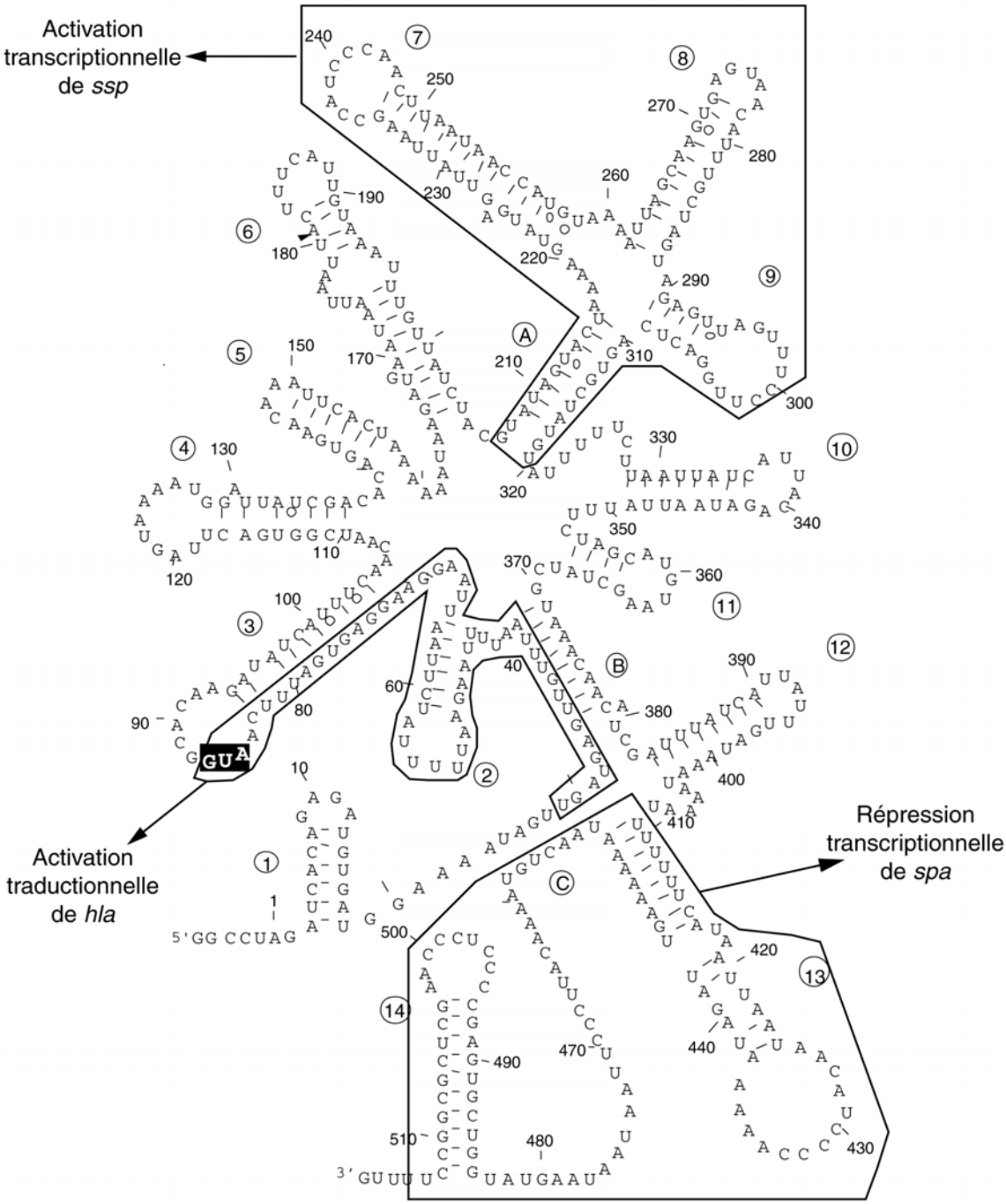

Figure 45 : ***Modèle de structure secondaire de l'ARN-III de*** **S. aureus** ***et fonctions de différents domaines structuraux***. Les motifs en tige-boucle sont numérotés de 1 à 14 et les trois interactions à longue distance sont indiquées de A à C. Le codon d'initiation de la traduction du gène *hld* codant pour l'hémolysine δ est indiqué en caractères blancs sur fond noir. Les domaines impliqués dans le contrôle de l'expression des gènes *hla, ssp et spa*, codant respectivement pour l'hémolysine α, la sérine-protéase et la protéine A sont encadrés.

B. Facteurs interagissant avec l'ARN-III

L'ARN-III agit comme riborégulateur vraisemblablement par différents mécanismes décrits dans l'introduction de ce manuscrit. Il active la traduction de l'ARNm *hla* par un

mécanisme de type antisens et module la transcription de plusieurs gènes par un mécanisme inconnu. Il est fort probable que ce contrôle s'effectue en modulant l'activité de facteurs protéique comme cela a été décrit pour les ARN CsrB, VAI et OxyS [123,126,172]. Dans le cadre de l'étude du domaine 3' de l'ARN-III, nous avons entrepris la caractérisation des facteurs protéiques se fixant à cette région. Des expériences de pontages aux UV nous ont permis d'identifier une protéine de ~15 kDa reconnaissant spécifiquement un ARN contenant les tiges-boucles 13 et 14 de l'ARN-III. Après nous être assuré qu'il ne s'agit pas d'un ARN, le facteur devra être séquencé et la détermination de son identité devrait être facilitée grâce à la séquence complète du génome de *S. aureus* disponible depuis peu [379]. L'étape suivante sera le clonage du gène correspondant et l'analyse de son implication dans les différents contrôles exercés par l'ARN-III, notamment sur l'expression de *spa*. Parallèlement aux expériences de pontages, différentes approches dont le triple-hybride et le système de co-purification par insertion d'une étiquette nucléotidique (aptamères reconnus pas la streptavidine, [380]) sont ou seront développées. Elles devraient permettre de caractériser différents facteurs protéiques ou ARN reconnaissant l'ARN-III. Des recherches dans le génome de *S. aureus* devrait aussi permettre d'identifier de nouvelles cibles ARN présentant une complémentarité avec l'ARN-III. Finalement, à plus long terme, les interactions qui seront établies entre l'ARN-III et les différents facteurs pourront servir de cibles à l'élaboration de nouveaux antibiotiques spécifiques des Staphylocoques.

C. Vers de nouvelles cibles de l'ARN-III

Le tableau 8 recense les gènes de virulence dont l'expression est sous le contrôle du locus *agr*. Ils ont été caractérisés parce que l'inactivation du locus *agr* se traduisait par l'absence de certains symptômes dans des modèles d'infections. Cependant, si l'ARN-III contrôle l'expression de ces gènes en interagissant avec une protéine régulatrice à effets pléiotropiques (de la même manière que l'ARN OxyS sur la protéine Hfq), ses cibles potentielles ne se limitent sûrement pas aux gènes de virulence. Dans le but d'identifier de nouveaux gènes dont l'expression est contrôlée par l'ARN-III, nous avons entrepris un projet de type protéome sur des souches *agr*+ et *agr*- de *S. aureus*. La figure 46 montre une expérience préliminaire qui devra être renouvelée plusieurs fois pour améliorer la séparation des protéines. L'analyse de ces résultats préliminaires montre d'ores et déjà plusieurs différences entre la souche sauvage et la souche mutante (Figure 46). Les protéines pour lesquelles l'expression varie seront identifiées par spectrométrie de masse ou par séquençage à partir de l'extrémité N-terminale. Cette approche sera utilisée sur une souche *agr*- complémentée par des ARN-III sauvages ou mutants afin de révéler les gènes dont

l'expression est directement contrôlée par l'ARN-III. On peut ainsi espérer mettre en évidence de nouvelles cascades d'événements sous le contrôle de l'ARN-III.

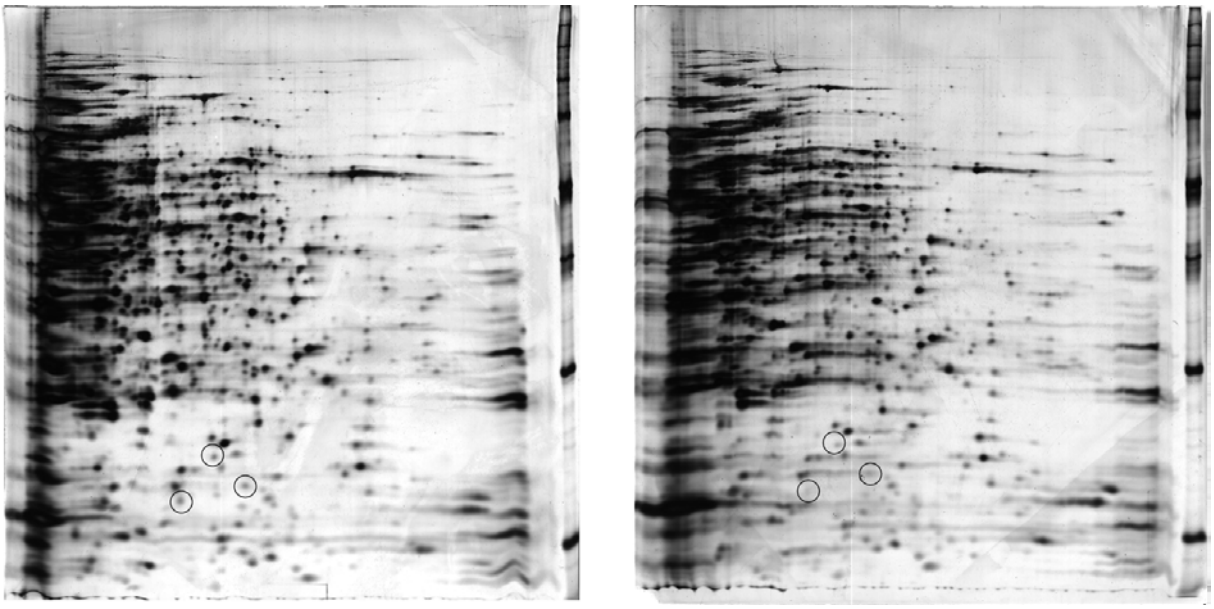

Figure 46 : ***Analyse du profil d'expression protéique d'une souche de* S. aureus *sauvage (gauche) et* agr- *(droite) par migration bidimensionnelle d'extraits protéiques sur gel de polyacrylamide***. Des exemples de protéines dont le profil d'expression change entre les deux souches sont indiquées par des cercles. La gamme de pH utilisée (5 à 8 de gauche à droite) n'est pas idéale, comme l'indique la grande concentration de protéines dans la partie gauche des gels.

CONCLUSION GENERALE

Grâce à l'étude de deux riborégulateurs bactériens, nous avons pu montrer à quel point les mécanismes d'action de cette classe d'ARN pouvaient être variés et surprenants. La réplication du plasmide R1 est contrôlée par l'ARN antisens CopA au niveau post-transcriptionnel. Malgré une parfaite complémentarité avec sa cible, le complexe ARN-ARN stable et fonctionnel n'est pas un duplexe étendu et adopte une topologie originale de jonction à quatre hélices. Le mode de formation de cette structure est également particulier à la fois *in vitro* et *in vivo*, avec une progression asymétrique d'une interaction boucle-boucle initiale. Loin d'être un cas particulier, les résultats obtenus pour le système R1 sont applicables à d'autres systèmes homologues. Dans le cas de l'ARN-III, nous avons déterminé un modèle de repliement intégrant des données obtenues à la fois *in vitro* et *in vivo* et défini un domaine structural et fonctionnel indépendant. Dans les deux cas, la fonction régulatrice de l'ARN repose sur des motifs de structure particuliers.

Les ARN non codants (dont les riborégulateurs) décrits dans l'introduction de ce manuscrit et caractérisés à l'heure actuelle ne constituent probablement qu'une infime proportion de ceux présents dans les cellules, tant procaryotiques qu'eucaryotiques. Ceux-ci ont été découverts de façon fortuite dans la plupart des cas. La disponibilité des séquences génomiques de plusieurs organismes a permis de caractériser de nombreuses protéines mais pas les gènes non traduits. Ainsi, les petits ARN non codants sont passés inaperçus dans ces programmes de recherche de phases ouvertes de lecture. A vrai dire, peu de groupes avaient conscience de l'énorme variété de fonction dont l'ARN peut faire preuve. Ce n'est que récemment, devant l'intérêt grandissant de la communauté scientifique envers les propriétés régulatrices des ARN, que différentes études ont eu pour but l'identification de petits ARN non codants que ce soit dans la levure *S. cerevisiae* [381] ou chez les mammifères [382]. Dans ce dernier cas, 201 nouveaux petits ARN non codants (ARNsnm) ont été caractérisés dans les cellules du cerveau de souris par purification sur gel des ARN de taille comprise entre 50 et 500 nts. L'expression de plus de la moitié a été testée par hybridation de type northern. Parmi les 201 ARNsnm, 113 correspondent à des snoRNA dont certains sont exprimés de façon tissu-spécifique. Finalement, 57 ARNsnm ne font pas partie de cette dernière classe et ne correspondent pas à des produits de dégradation d'ARNm. Ils représentent par conséquent de bons candidats pour de nouveaux riborégulateurs. Récemment, deux groupes différents ont identifié de nouveaux petits ARN non codants chez *E. coli* [383,384]. Ces deux groupes ont caractérisé respectivement 14 et 17 nouveaux petits ARN, certains étant retrouvés dans les

deux études. Par ailleurs, ceux-ci ont montré que l'expression de certains de ces ARN était induite lors de conditions particulières de croissance (choc thermique, croissance en milieu minimal et entrée des cellules en phase stationnaire de croissance). Cette observation, et celle que 8 parmi les 17 ARN sont reconnus par la protéine chaperone Hfq, les relie aux riborégulateurs DsrA et OxyS dont les modes d'action ont été décrits dans l'introduction de ce manuscrit. Que ce soit chez la levure, les mammifères ou les bactéries, l'étape suivante et nécessaire sera de déterminer la fonction de ces nouveaux ARN *in vivo*.

La comparaison des ARN caractérisés à ce jour en combinaison avec le grand nombre de séquences génomiques maintenant disponibles devrait permettre d'utiliser des outils informatiques pour la recherche de nouveaux riborégulateurs. Plusieurs ARN régulateurs bactériens sont courts et contiennent un terminateur de transcription ρ-indépendant à leur extrémité 3'. Ainsi, une approche pouvant s'avérer fructueuse est la recherche de séquence promotrices et de terminaison séparées par une centaine de nucléotides. Une autre caractéristique redondante d'un grand nombre d'antisens est la présence d'une structure en tige-boucle imparfaitement appariée, coiffée par une boucle contenant une séquence très conservée de type 5'-YUNR-3'. La recherche d'un tel motif dans les génomes disponibles devrait également permettre la caractérisation de nouveaux ARN antisens. Des approches plus globales de type transcriptome et protéome devraient permettre de déterminer quels sont les gènes dont l'expression est sous leur contrôle.

Ainsi, si un monde « Tout ARN » a dû exister il y a plus de 3,5 milliards d'années, les années à venir seront inévitablement tournées vers l'étude de l'ARN et de ses multiples fonctions.

REFERENCES

1. Crick FH: **The origin of the genetic code**. *J Mol Biol* 1968, **38**:367-379.
2. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S: **The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme**. *Cell* 1983, **35**:849-857.
3. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR: **Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena**. *Cell* 1982, **31**:147-157.
4. Tanner NK: **Ribozymes: the characteristics and properties of catalytic RNAs**. *FEMS Microbiol Rev* 1999, **23**:257-275.
5. Gilbert W: **The RNA world**. *Nature* 1986, **319**:618.
6. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA: **The structural basis of ribosome activity in peptide bond synthesis**. *Science* 2000, **289**:920-930.
7. Bartel DP, Szostak JW: **Isolation of new ribozymes from a large pool of random sequences**. *Science* 1993, **261**:1411-1418.
8. Lorsch JR, Szostak JW: **In vitro evolution of new ribozymes with polynucleotide kinase activity**. *Nature* 1994, **371**:31-36.
9. Tarasow TM, Tarasow SL, Eaton BE: **RNA-catalysed carbon-carbon bond formation**. *Nature* 1997, **389**:54-57.
10. Wilson C, Szostak JW: **In vitro evolution of a self-alkylating ribozyme**. *Nature* 1995, **374**:777-782.
11. Zhang B, Cech TR: **Peptide bond formation by in vitro selected ribozymes**. *Nature* 1997, **390**:96-100.
12. Hindley J: **Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting**. *J Mol Biol* 1967, **30**:125-136.
13. Eddy SR: **Noncoding RNA genes**. *Curr Opin Genet Dev* 1999, **9**:695-699.
14. Wassarman KM, Zhang A, Storz G: **Small RNAs in Escherichia coli**. *Trends Microbiol* 1999, **7**:37-45.
15. Filipowicz W, Pelczar P, Pogacic V, Dragon F: **Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification**. *Acta Biochim Pol* 1999, **46**:377-389.
16. Peluso P, Herschlag D, Nock S, Freymann DM, Johnson AE, Walter P: **Role of 4.5S RNA in assembly of the bacterial signal recognition particle with its receptor**. *Science* 2000, **288**:1640-1643.
17. Karzai AW, Roche ED, Sauer RT: **The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue**. *Nat Struct Biol* 2000, **7**:449-455.
18. Chen C, Sheng S, Shao Z, Guo P: **A dimer as a building block in assembling RNA. A hexamer that gears bacterial virus phi29 DNA-translocating machinery**. *J Biol Chem* 2000, **275**:17510-17516.
19. Wassarman KM, Storz G: **6S RNA regulates E. coli RNA polymerase activity**. *Cell* 2000, **101**:613-623.
20. Hatfull GF, Joyce CM: **Deletion of the spf (spot 42 RNA) gene of Escherichia coli**. *J Bacteriol* 1986, **166**:746-750.
21. Chen X, Quinn AM, Wolin SL: **Ro ribonucleoproteins contribute to the resistance of Deinococcus radiodurans to ultraviolet irradiation**. *Genes Dev* 2000, **14**:777-782.
22. Quentin Y: **Emergence of master sequences in families of retroposons derived from 7sl RNA**. *Genetica* 1994, **93**:203-215.
23. Reed R: **Mechanisms of fidelity in pre-mRNA splicing**. *Curr Opin Cell Biol* 2000, **12**:340-345.
24. Weinstein LB, Steitz JA: **Guided tours: from precursor snoRNA to functional snoRNP**. *Curr Opin Cell Biol* 1999, **11**:378-384.
25. Simpson L, Thiemann OH, Savill NJ, Alfonzo JD, Maslov DA: **Evolution of RNA editing in trypanosome mitochondria**. *Proc Natl Acad Sci U S A* 2000, **97**:6986-6993.
26. Green CD, Long KS, Shi H, Wolin SL: **Binding of the 60-kDa Ro autoantigen to Y RNAs: evidence for recognition in the major groove of a conserved helix**. *RNA* 1998, **4**:750-765.
27. Fabini G, Raijmakers R, Hayer S, Fouraux MA, Pruijn GJ, Steiner G: **The heterogeneous nuclear ribonucleoproteins I and K interact with a subset of the ro ribonucleoprotein-associated Y RNAs in vitro and in vivo**. *J Biol Chem* 2001, **276**:20711-20718.
28. Blackburn EH: **The end of the (DNA) line**. *Nat Struct Biol* 2000, **7**:847-850.
29. Ohno M, Mattaj IW: **Meiosis: MeiRNA hits the spot**. *Curr Biol* 1999, **9**:R66-69.
30. McKechnie SW, Halford MM, McColl G, Hoffmann AA: **Both allelic variation and expression of nuclear and cytoplasmic transcripts of Hsr-omega are closely associated with thermal phenotype in**

Drosophila. *Proc Natl Acad Sci U S A* 1998, **95**:2423-2428.

31. Henry RA, Tews B, Li X, Scott MJ: **Recruitment of the MSL dosage compensation complex to an autosomally integrated roX chromatin entry site correlates with an increased expression of an adjacent reporter gene in male Drosophila**. *J Biol Chem* 2001, **11**:11.
32. Kelley RL, Kuroda MI: **Noncoding RNA genes in dosage compensation and imprinting**. *Cell* 2000, **103**:9-12.
33. Milligan L, Antoine E, Bisbal C, Weber M, Brunel C, Forne T, Cathala G: **H19 gene expression is up-regulated exclusively by stabilization of the RNA during muscle cell differentiation.[In Process Citation]**. *Oncogene* 2000, **19**:5810-5816.
34. Spiegelman WG, Reichardt LF, Yaniv M, Heinemann SF, Kaiser AD, Eisen H: **Bidirectional transcription and the regulation of Phage lambda repressor synthesis**. *Proc Natl Acad Sci U S A* 1972, **69**:3156-3160.
35. Itoh T, Tomizawa J: **Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H**. *Proc Natl Acad Sci U S A* 1980, **77**:2450-2454.
36. Lacatena RM, Cesareni G: **Base pairing of RNA I with its complementary sequence in the primer precursor inhibits ColE1 replication**. *Nature* 1981, **294**:623-626.
37. Wagner EG, Simons RW: **Antisense RNA control in bacteria, phages, and plasmids**. *Annu Rev Microbiol* 1994, **48**:713-742.
38. Stolt P, Zillig W: **Antisense RNA mediates transcriptional processing in an archaebacterium, indicating a novel kind of RNAse activity**. *Mol Microbiol* 1993, **7**:875-882.
39. Hildebrandt M, Nellen W: **Differential antisense transcription from the Dictyostelium EB4 gene locus: implications on antisense-mediated regulation of mRNA stability**. *Cell* 1992, **69**:197-204.
40. Slack F, Ruvkun G: **Temporal pattern formation by heterochronic genes**. *Annu Rev Genet* 1997, **31**:611-634.
41. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G: **The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans**. *Nature* 2000, **403**:901-906.
42. Blomberg P, Nordstrom K, Wagner EG: **Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation**. *Embo J* 1992, **11**:2675-2683.
43. Altuvia S, Zhang A, Argaman L, Tiwari A, Storz G: **The Escherichia coli OxyS regulatory RNA represses fhlA translation by blocking ribosome binding**. *Embo J* 1998, **17**:6069-6075.
44. van Biesen T, Frost LS: **The FinO protein of IncF plasmids binds FinP antisense RNA and its target, traJ mRNA, and promotes duplex formation**. *Mol Microbiol* 1994, **14**:427-436.
45. Andersen J, Delihas N: **micF RNA binds to the 5' end of ompF mRNA and to a protein from Escherichia coli**. *Biochemistry* 1990, **29**:9249-9256.
46. Wu TH, Liao SM, McClure WR, Susskind MM: **Control of gene expression in bacteriophage P22 by a small antisense RNA. II. Characterization of mutants defective in repression**. *Genes Dev* 1987, **1**:204-212.
47. Asano K, Mizobuchi K: **An RNA pseudoknot as the molecular switch for translation of the repZ gene encoding the replication initiator of IncIalpha plasmid ColIb-P9**. *J Biol Chem* 1998, **273**:11815-11825.
48. Athanasopoulos V, Praszkier J, Pittard AJ: **Analysis of elements involved in pseudoknot-dependent expression and regulation of the repA gene of an IncL/M plasmid**. *J Bacteriol* 1999, **181**:1811-1819.
49. Lease RA, Belfort M: **A trans-acting RNA as a control switch in Escherichia coli: DsrA modulates function by forming alternative structures**. *Proc Natl Acad Sci U S A* 2000, **97**:9919-9924.
50. Morfeldt E, Taylor D, von Gabain A, Arvidson S: **Activation of alpha-toxin translation in Staphylococcus aureus by the trans-encoded antisense RNA, RNAIII**. *Embo J* 1995, **14**:4569-4577.
51. Blomberg P, Wagner EG, Nordstrom K: **Control of replication of plasmid R1: the duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically in vivo and in vitro by RNAse III**. *Embo J* 1990, **9**:2331-2340.
52. Franch T, Thisted T, Gerdes K: **Ribonuclease III processing of coaxially stacked RNA helices**. *J Biol Chem* 1999, **274**:26572-26578.
53. Jerome LJ, van Biesen T, Frost LS: **Degradation of FinP antisense RNA from F-like plasmids: the RNA-binding protein, FinO, protects FinP from**

ribonuclease E. *J Mol Biol* 1999, **285**:1457-1473.

54. Krinke L, Wulff DL: **RNAse III-dependent hydrolysis of lambda cII-O gene mRNA mediated by lambda OOP antisense RNA**. *Genes Dev* 1990, **4**:2223-2233.
55. Kumar M, Carmichael GG: **Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes**. *Microbiol Mol Biol Rev* 1998, **62**:1415-1434.
56. Okamoto K, Freundlich M: **Mechanism for the autogenous control of the crp operon: transcriptional inhibition by a divergent RNA transcript**. *Proc Natl Acad Sci U S A* 1986, **83**:5000-5004.
57. Novick RP, Iordanescu S, Projan SJ, Kornblum J, Edelman I: **pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator**. *Cell* 1989, **59**:395-404.
58. Tomizawa J: **Control of ColE1 plasmid replication: binding of RNA I to RNA II and inhibition of primer formation**. *Cell* 1986, **47**:89-97.
59. Kaberdin VR, Chao YH, Lin-Chao S: **RNAse E cleaves at multiple sites in bubble regions of RNA I stem loops yielding products that dissociate differentially from the enzyme**. *J Biol Chem* 1996, **271**:13103-13109.
60. Soderbom F, Wagner EG: **Studies on the degradation pathway of CopA, the antisense RNA that controls replication of plasmid R1**. *Microbiology* 1998, **144**:1907-1917.
61. Asano K, Niimi T, Yokoyama S, Mizobuchi K: **Structural basis for binding of the plasmid ColIb-P9 antisense Inc RNA to its target RNA with the 5'-rUUGGCG-3' motif in the loop sequence**. *J Biol Chem* 1998, **273**:11826-11838.
62. Hjalt TA, Wagner EG: **Bulged-out nucleotides in an antisense RNA are required for rapid target RNA binding in vitro and inhibition in vivo**. *Nucleic Acids Res* 1995, **23**:580-587.
63. Kolb FA, Engdahl HM, Slagter-Jager JG, Ehresmann B, Ehresmann C, Westhof E, Wagner EG, Romby P: **Progression of a loop-loop complex to a four-way junction is crucial for the activity of a regulatory antisense RNA**. *Embo J* 2000, **19**:5905-5915.
64. Siemering KR, Praszkier J, Pittard AJ: **Mechanism of binding of the antisense and target RNAs involved in the regulation of IncB plasmid replication**. *J Bacteriol* 1994, **176**:2677-2688.
65. Zeiler BN, Simons RW: **Antisense RNA Structure and Function**. In *RNA Structure and Function*. Edited by Press CSHL; 1998:437-464.,
66. Franch T, Petersen M, Wagner EG, Jacobsen JP, Gerdes K: **Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure**. *J Mol Biol* 1999, **294**:1115-1125.
67. Grosjean H, Marquet R, Romby P: **Modulatory role of modified nucleotides in RNA loop-loop interaction**. In *Modification and editing of RNA*. Edited by:; 1988:113-133.,
68. Bennett CF: **Antisense oligonucleotides: is the glass half full or half empty?** *Biochem Pharmacol* 1998, **55**:9-19.
69. Agrawal S, Kandimalla ER: **Antisense therapeutics: is it as simple as complementary base recognition?** *Mol Med Today* 2000, **6**:72-81.
70. Nellen W, Sczakiel G: **In vitro and in vivo action of antisense RNA**. *Mol Biotechnol* 1996, **6**:7-15.
71. Agrawal S, Zhao Q: **Antisense therapeutics**. *Curr Opin Chem Biol* 1998, **2**:519-528.
72. Chakrabarti MC, Schwarz FP: **Thermal stability of PNA/DNA and DNA/DNA duplexes by differential scanning calorimetry**. *Nucleic Acids Res* 1999, **27**:4801-4806.
73. Rittner K, Burmester C, Sczakiel G: **In vitro selection of fast-hybridizing and effective antisense RNAs directed against the human immunodeficiency virus type 1**. *Nucleic Acids Res* 1993, **21**:1381-1387.
74. Patzel V, Sczakiel G: **Theoretical design of antisense RNA structures substantially improves annealing kinetics and efficacy in human cells [see comments]**. *Nat Biotechnol* 1998, **16**:64-68.
75. Patzel V, Steidl U, Kronenwett R, Haas R, Sczakiel G: **A theoretical approach to select effective antisense oligodeoxyribonucleotides at high statistical probability**. *Nucleic Acids Res* 1999, **27**:4328-4334.
76. Tu GC, Cao QN, Zhou F, Israel Y: **Tetranucleotide GGGA motif in primary RNA transcripts. Novel target site for antisense design**. *J Biol Chem* 1998, **273**:25125-25131.
77. Stocks MR, Rabbitts TH: **Masked antisense: a molecular configuration for discriminating similar RNA targets**. *EMBO Reports* 2000, **1**:59-64.
78. Famulok M, Jenne A: **Oligonucleotide libraries--variatio delectat**. *Curr Opin Chem Biol* 1998, **2**:320-327.

79. Famulok M, Mayer G: **Aptamers as tools in molecular biology and immunology**. *Curr Top Microbiol Immunol* 1999, **243**:123-136.
80. Toulme JJ, Tinevez RL, Brossalina E: **Targeting RNA structures by antisense oligonucleotides**. *Biochimie* 1996, **78**:663-673.
81. Guo S, Kemphues KJ: **par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed**. *Cell* 1995, **81**:611-620.
82. Fire A, Albertson D, Harrison SW, Moerman DG: **Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle**. *Development* 1991, **113**:503-514.
83. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: **Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans**. *Nature* 1998, **391**:806-811.
84. Kennerdell JR, Carthew RW: **Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway**. *Cell* 1998, **95**:1017-1026.
85. Ngo H, Tschudi C, Gull K, Ullu E: **Double-stranded RNA induces mRNA degradation in Trypanosoma brucei**. *Proc Natl Acad Sci U S A* 1998, **95**:14687-14692.
86. Sanchez Alvarado A, Newmark PA: **Double-stranded RNA specifically disrupts gene expression during planarian regeneration**. *Proc Natl Acad Sci U S A* 1999, **96**:5049-5054.
87. Lohmann JU, Endl I, Bosch TC: **Silencing of developmental genes in Hydra**. *Dev Biol* 1999, **214**:211-214.
88. Gura T: **A silence that speaks volumes**. *Nature* 2000, **404**:804-808.
89. Vaucheret H, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Mourrain P, Palauqui JC, Vernhettes S: **Transgene-induced gene silencing in plants**. *Plant J* 1998, **16**:651-659.
90. Waterhouse PM, Graham MW, Wang MB: **Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA**. *Proc Natl Acad Sci U S A* 1998, **95**:13959-13964.
91. Baulcombe DC: **Fast forward genetics based on virus-induced gene silencing**. *Curr Opin Plant Biol* 1999, **2**:109-113.
92. Cogoni C, Irelan JT, Schumacher M, Schmidhauser TJ, Selker EU, Macino G: **Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation**. *Embo J* 1996, **15**:3153-3163.
93. Pal-Bhadra M, Bhadra U, Birchler JA: **Cosuppression in Drosophila: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent [see comments]**. *Cell* 1997, **90**:479-490.
94. Bahramian MB, Zarbl H: **Transcriptional and posttranscriptional silencing of rodent alpha1(I) collagen by a homologous transcriptionally self-silenced transgene**. *Mol Cell Biol* 1999, **19**:274-283.
95. Angell SM, Baulcombe DC: **Consistent gene silencing in transgenic plants expressing a replicating potato virus X RNA**. *Embo J* 1997, **16**:3675-3684.
96. Tabara H, Grishok A, Mello CC: **RNAi in C. elegans: soaking in the genome sequence**. *Science* 1998, **282**:430-431.
97. Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJ: **Transcriptional silencing and promoter methylation triggered by double- stranded RNA**. *Embo J* 2000, **19**:5194-5201.
98. Matzke M, Matzke AJ, Kooter JM: **RNA: guiding gene silencing**. *Science* 2001, **293**:1080-1083.
99. Montgomery MK, Xu S, Fire A: **RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans**. *Proc Natl Acad Sci U S A* 1998, **95**:15502-15507.
100. Tuschl T, Zamore PD, Lehmann R, Bartel DP, Sharp PA: **Targeted mRNA degradation by double-stranded RNA in vitro**. *Genes Dev* 1999, **13**:3191-3197.
101. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T: **Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells**. *Nature* 2001, **411**:494-498.
102. Caplen NJ, Parrish S, Imani F, Fire A, Morgan RA: **Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems**. *Proc Natl Acad Sci U S A* 2001, **98**:9742-9747.
103. Vance V, Vaucheret H: **RNA silencing in plants--defense and counterdefense**. *Science* 2001, **292**:2277-2280.
104. Hammond SM, Caudy AA, Hannon GJ: **Post-transcriptional gene silencing by double-stranded RNA**. *Nat Rev Genet* 2001, **2**:110-119.
105. Hammond SM, Bernstein E, Beach D, Hannon GJ: **An RNA-directed nuclease mediates**

post-transcriptional gene silencing in Drosophila cells. *Nature* 2000, **404**:293-296.

106. Zamore PD, Tuschl T, Sharp PA, Bartel DP: **RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals [In Process Citation]**. *Cell* 2000, **101**:25-33.
107. Bernstein E, Caudy AA, Hammond SM, Hannon GJ: **Role for a bidentate ribonuclease in the initiation step of RNA interference**. *Nature* 2001, **409**:363-366.
108. Smardon A, Spoerke JM, Stacey SC, Klein ME, Mackin N, Maine EM: **EGO-1 is related to RNA-directed RNA polymerase and functions in germ- line development and RNA interference in C. elegans [published erratum appears in Curr Biol 2000 May 18;10(10):R393-4]**. *Curr Biol* 2000, **10**:169-178.
109. Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Jouette D, Lacombe AM, Nikic S, Picault N, Remoue K, Sanial M, Vo TA, Vaucheret H: **Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance**. *Cell* 2000, **101**:533-542.
110. Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC: **An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus**. *Cell* 2000, **101**:543-553.
111. Cogoni C, Macino G: **Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase**. *Nature* 1999, **399**:166-169.
112. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ: **Argonaute2, a link between genetic and biochemical analyses of RNAi**. *Science* 2001, **293**:1146-1150.
113. Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH: **Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNAseD**. *Cell* 1999, **99**:133-141.
114. Grishok A, Tabara H, Mello CC: **Genetic requirements for inheritance of RNAi in C. elegans**. *Science* 2000, **287**:2494-2497.
115. Wu-Scharf D, Jeong B, Zhang C, Cerutti H: **Transgene and transposon silencing in chlamydomonas reinhardtii by a DEAH-Box RNA helicase**. *Science* 2000, **290**:1159-1163.
116. Cogoni C, Macino G: **Posttranscriptional gene silencing in Neurospora by a RecQ DNA helicase**. *Science* 1999, **286**:2342-2344.
117. Dalmay T, Horsefield R, Braunstein TH, Baulcombe DC: **SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in Arabidopsis**. *Embo J* 2001, **20**:2069-2078.
118. Fagard M, Boutet S, Morel JB, Bellini C, Vaucheret H: **AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals [In Process Citation]**. *Proc Natl Acad Sci U S A* 2000, **97**:11650-11654.
119. Baker MW, Macagno ER: **RNAi of the receptor tyrosine phosphatase HmLAR2 in a single cell of an intact leech embryo leads to growth-cone collapse [In Process Citation]**. *Curr Biol* 2000, **10**:1071-1074.
120. Lam G, Thummel CS: **Inducible expression of double-stranded RNA directs specific genetic interference in Drosophila [In Process Citation]**. *Curr Biol* 2000, **10**:957-963.
121. Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC: **Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control C. elegans Developmental Timing**. *Cell* 2001, **106**:23-34.
122. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD: **A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA**. *Science* 2001, **12**:12.
123. Liu MY, Gui G, Wei B, Preston JF, 3rd, Oakford L, Yuksel U, Giedroc DP, Romeo T: **The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli**. *J Biol Chem* 1997, **272**:17502-17510.
124. Liu Y, Cui Y, Mukherjee A, Chatterjee AK: **Characterization of a novel RNA regulator of Erwinia carotovora ssp. carotovora that controls production of extracellular enzymes and secondary metabolites**. *Mol Microbiol* 1998, **29**:219-234.
125. Aarons S, Abbas A, Adams C, Fenton A, O'Gara F: **A regulatory RNA (PrrB RNA) modulates expression of secondary metabolite genes in Pseudomonas fluorescens F113**. *J Bacteriol* 2000, **182**:3913-3919.

126. Mathews MB, Shenk T: **Adenovirus virus-associated RNA and translation control**. *J Virol* 1991, **65**:5657-5662.
127. Clemens MJ: **The small RNAs of Epstein-Barr virus**. *Mol Biol Rep* 1993, **17**:81-92.
128. Romeo T, Gong M: **Genetic and physical mapping of the regulatory gene csrA on the Escherichia coli K-12 chromosome**. *J Bacteriol* 1993, **175**:5740-5741.
129. Sabnis NA, Yang H, Romeo T: **Pleiotropic regulation of central carbohydrate metabolism in Escherichia coli via the gene csrA**. *J Biol Chem* 1995, **270**:29096-29104.
130. Yang H, Liu MY, Romeo T: **Coordinate genetic regulation of glycogen catabolism and biosynthesis in Escherichia coli via the CsrA gene product**. *J Bacteriol* 1996, **178**:1012-1017.
131. Macnab RM: **Flagella and motility**. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edn 2nd. Edited by: Washington, DC: American Society for Microbiology Press; 1996:123-145. [Neidhardt FC, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B (Series Editor):, vol 1.]
132. Romeo T, Gong M, Liu MY, Brun-ZinkeRNAgel AM: **Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties**. *J Bacteriol* 1993, **175**:4744-4755.
133. Liu MY, Yang H, Romeo T: **The product of the pleiotropic Escherichia coli gene csrA modulates glycogen biosynthesis via effects on mRNA stability**. *J Bacteriol* 1995, **177**:2663-2672.
134. Romeo T: **Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB**. *Mol Microbiol* 1998, **29**:1321-1330.
135. Chatterjee A, Cui Y, Liu Y, Dumenyo CK, Chatterjee AK: **Inactivation of rsmA leads to overproduction of extracellular pectinases, cellulases, and proteases in Erwinia carotovora subsp. carotovora in the absence of the starvation/cell density-sensing signal, N-(3-oxohexanoyl)-L-homoserine lactone**. *Appl Environ Microbiol* 1995, **61**:1959-1967.
136. Ma W, Cui Y, Liu Y, Dumenyo CK, Mukherjee A, Chatterjee AK: **Molecular characterization of global regulatory RNA species that control pathogenicity factors in Erwinia amylovora and Erwinia herbicola pv. gypsophilae**. *J Bacteriol* 2001, **183**:1870-1880.
137. Mukherjee A, Cui Y, Liu Y, Dumenyo CK, Chatterjee AK: **Global regulation in Erwinia species by Erwinia carotovora rsmA, a homologue of Escherichia coli csrA: repression of secondary metabolites, pathogenicity and hypersensitive reaction**. *Microbiology* 1996, **142**:427-434.
138. Cui Y, Chatterjee A, Liu Y, Dumenyo CK, Chatterjee AK: **Identification of a global repressor gene, rsmA, of Erwinia carotovora subsp. carotovora that controls extracellular enzymes, N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting Erwinia spp**. *J Bacteriol* 1995, **177**:5108-5115.
139. Murata H, Chatterjee A, Liu Y, Chatterjee AK: **Regulation of the production of extracellular pectinase, cellulase, and protease in the soft rot bacterium Erwinia carotovora subsp. carotovora: evidence that aepH of E. carotovora subsp. carotovora 71 activates gene expression in E. carotovora subsp. carotovora, E. carotovora subsp. atroseptica, and Escherichia coli**. *Appl Environ Microbiol* 1994, **60**:3150-3159.
140. Altier C, Suyemoto M, Ruiz AI, Burnham KD, Maurer R: **Characterization of two novel regulatory genes affecting Salmonella invasion gene expression**. *Mol Microbiol* 2000, **35**:635-646.
141. Blumer C, Heeb S, Pessi G, Haas D: **Global GacA-steered control of cyanide and exoprotease production in Pseudomonas fluorescens involves specific ribosome binding sites**. *Proc Natl Acad Sci U S A* 1999, **96**:14073-14078.
142. Sen GC, Ransohoff RM: **Interferon-induced antiviral actions and their regulation**. *Adv Virus Res* 1993, **42**:57-102.
143. Cuddihy AR, Li S, Tam NW, Wong AH, Taya Y, Abraham N, Bell JC, Koromilas AE: **Double-stranded-RNA-activated protein kinase PKR enhances transcriptional activation by tumor suppressor p53**. *Mol Cell Biol* 1999, **19**:2475-2484.
144. Cuddihy AR, Wong AH, Tam NW, Li S, Koromilas AE: **The double-stranded RNA activated protein kinase PKR physically associates with the tumor suppressor p53 protein and phosphorylates human p53 on serine 392 in vitro**. *Oncogene* 1999, **18**:2690-2702.
145. Donze O, Jagus R, Koromilas AE, Hershey JW, Sonenberg N: **Abrogation of translation initiation factor eIF-2**

phosphorylation causes malignant transformation of NIH 3T3 cells. *Embo J* 1995, **14**:3828-3834.

146. Robertson HD, Mathews MB: **The regulation of the protein kinase PKR by RNA**. *Biochimie* 1996, **78**:909-914.
147. Gale M, Jr., Katze MG: **Molecular mechanisms of interferon resistance mediated by viral- directed inhibition of PKR, the interferon-induced protein kinase**. *Pharmacol Ther* 1998, **78**:29-46.
148. Manche L, Green SR, Schmedt C, Mathews MB: **Interactions between double-stranded RNA regulators and the protein kinase DAI**. *Mol Cell Biol* 1992, **12**:5238-5248.
149. Ryter JM, Schultz SC: **Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA**. *Embo J* 1998, **17**:7505-7513.
150. Sharp TV, Schwemmle M, Jeffrey I, Laing K, Mellor H, Proud CG, Hilse K, Clemens MJ: **Comparative analysis of the regulation of the interferon-inducible protein kinase PKR by Epstein-Barr virus RNAs EBER-1 and EBER-2 and adenovirus VAI RNA**. *Nucleic Acids Res* 1993, **21**:4483-4490.
151. Rahman A, Malhotra P, Dhar R, Kewalramani T, Thimmapaya B: **Effect of single-base substitutions in the central domain of virus- associated RNA I on its function**. *J Virol* 1995, **69**:4299-4307.
152. Ma Y, Mathews MB: **Structure, function, and evolution of adenovirus-associated RNA: a phylogenetic approach**. *J Virol* 1996, **70**:5083-5099.
153. Bhat RA, Thimmappaya B: **Construction and analysis of additional adenovirus substitution mutants confirm the complementation of VAI RNA function by two small RNAs encoded by Epstein-Barr virus**. *J Virol* 1985, **56**:750-756.
154. Carpick BW, Graziano V, Schneider D, Maitra RK, Lee X, Williams BRG: **Characterization of the solution complex between the interferon-induced, double-stranded RNA-activated protein kinase and HIV-I trans- activating region RNA**. *J Biol Chem* 1997, **272**:9510-9516.
155. Nogueira T, Springer M: **Post-transcriptional control by global regulators of gene expression in bacteria**. *Curr Opin Microbiol* 2000, **3**:154-158.
156. Hengge-Aronis R: **Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in Escherichia coli**. *Mol Microbiol* 1996, **21**:887-893.
157. Lange R, Hengge-Aronis R: **The cellular concentration of the sigma S subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability**. *Genes Dev* 1994, **8**:1600-1612.
158. Muffler A, Traulsen DD, Lange R, Hengge-Aronis R: **Posttranscriptional osmotic regulation of the sigma(s) subunit of RNA polymerase in Escherichia coli**. *J Bacteriol* 1996, **178**:1607-1613.
159. Muffler A, Fischer D, Hengge-Aronis R: **The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for rpoS translation in Escherichia coli**. *Genes Dev* 1996, **10**:1143-1151.
160. Brown L, Elliott T: **Mutations that increase expression of the rpoS gene and decrease its dependence on hfq function in Salmonella typhimurium**. *J Bacteriol* 1997, **179**:656-662.
161. Barth M, Marschall C, Muffler A, Fischer D, Hengge-Aronis R: **Role for the histone-like protein H-NS in growth phase-dependent and osmotic regulation of sigma S and many sigma S-dependent genes in Escherichia coli**. *J Bacteriol* 1995, **177**:3455-3464.
162. Yamashino T, Ueguchi C, Mizuno T: **Quantitative control of the stationary phase-specific sigma factor, sigma S, in Escherichia coli: involvement of the nucleoid protein H-NS**. *Embo J* 1995, **14**:594-602.
163. Kajitani M, Ishihama A: **Identification and sequence determination of the host factor gene for bacteriophage Q beta**. *Nucleic Acids Res* 1991, **19**:1063-1066.
164. Sledjeski D, Gottesman S: **A small RNA acts as an antisilencer of the H-NS-silenced rcsA gene of Escherichia coli**. *Proc Natl Acad Sci U S A* 1995, **92**:2003-2007.
165. Lease RA, Cusick ME, Belfort M: **Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci**. *Proc Natl Acad Sci U S A* 1998, **95**:12456-12461.
166. Sledjeski DD, Gupta A, Gottesman S: **The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli**. *Embo J* 1996, **15**:3993-4000.
167. Majdalani N, Cunning C, Sledjeski D, Elliott T, Gottesman S: **DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of**

transcription. *Proc Natl Acad Sci U S A* 1998, **95**:12462-12467.

168. Lease RA, Belfort M: **Riboregulation by DsrA RNA: trans-actions for global economy**. *Mol Microbiol* 2000, **38**:667-672.
169. Sledjeski DD, Whitman C, Zhang A: **Hfq is necessary for regulation by the untranslated RNA DsrA**. *J Bacteriol* 2001, **183**:1997-2005.
170. Majdalani N, Chen S, Murrow J, St John K, Gottesman S: **Regulation of RpoS by a novel small RNA: the characterization of RprA**. *Mol Microbiol* 2001, **39**:1382-1394.
171. Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G: **A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator**. *Cell* 1997, **90**:43-53.
172. Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, Storz G: **The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein**. *Embo J* 1998, **17**:6061-6068.
173. Senear AW, Steitz JA: **Site-specific interaction of Qbeta host factor and ribosomal protein S1 with Qbeta and R17 bacteriophage RNAs**. *J Biol Chem* 1976, **251**:1902-1912.
174. Williams RM, Rimsky S: **Molecular aspects of the E. coli nucleoid protein, H-NS: a central controller of gene regulatory networks**. *FEMS Microbiol Lett* 1997, **156**:175-185.
175. De Rijcke M, Seneca S, Punyammalee B, Glansdorff N, Crabeel M: **Characterization of the DNA target site for the yeast ARGR regulatory complex, a sequence able to mediate repression or induction by arginine**. *Mol Cell Biol* 1992, **12**:68-81.
176. Baccigalupi L, Marasco R, Ricca E, De Felice M, Sacco M: **Control of ilvIH transcription during amino acid downshift in stringent and relaxed strains of Escherichia coli**. *FEMS Microbiol Lett* 1995, **131**:95-98.
177. Levinthal M, Lejeune P, Danchin A: **The H-NS protein modulates the activation of the ilvIH operon of Escherichia coli K12 by Lrp, the leucine regulatory protein**. *Mol Gen Genet* 1994, **242**:736-743.
178. Oh H, Park Y, Park C: **A mutated PtsG, the glucose transporter, allows uptake of D-ribose**. *J Biol Chem* 1999, **274**:14006-14011.
179. Argaman L, Altuvia S: **fhlA repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex**. *J Mol Biol* 2000, **300**:1101-1112.
180. Wiggerich HG, Klauke B, Koplin R, Priefer UB, Puhler A: **Unusual structure of the tonB-exb DNA region of Xanthomonas campestris pv. campestris: tonB, exbB, and exbD1 are essential for ferric iron uptake, but exbD2 is not**. *J Bacteriol* 1997, **179**:7103-7110.
181. Ha I, Wightman B, Ruvkun G: **A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation**. *Genes Dev* 1996, **10**:3041-3050.
182. Moss EG, Lee RC, Ambros V: **The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA**. *Cell* 1997, **88**:637-646.
183. Kimelman D, Kirschner MW: **An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes**. *Cell* 1989, **59**:687-696.
184. Murphy PR, Knee RS: **Identification and characterization of an antisense RNA transcript (gfg) from the human basic fibroblast growth factor gene**. *Mol Endocrinol* 1994, **8**:852-859.
185. Skeiky YA, Iatrou K: **Developmental regulation of covalent modification of double-stranded RNA during silkmoth oogenesis**. *J Mol Biol* 1991, **218**:517-527.
186. Noguchi M, Miyamoto S, Silverman TA, Safer B: **Characterization of an antisense Inr element in the eIF-2 alpha gene**. *J Biol Chem* 1994, **269**:29161-29167.
187. Lagunoff M, Randall G, Roizman B: **Phenotypic properties of herpes simplex virus 1 containing a derepressed open reading frame P gene**. *J Virol* 1996, **70**:1810-1817.
188. Farrell CM, Lukens LN: **Naturally occurring antisense transcripts are present in chick embryo chondrocytes simultaneously with the down-regulation of the alpha 1 (I) collagen gene**. *J Biol Chem* 1995, **270**:3400-3408.
189. Munroe SH, Lazar MA: **Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA**. *J Biol Chem* 1991, **266**:22083-22086.
190. Okano H, Aruga J, Nakagawa T, Shiota C, Mikoshiba K: **Myelin basic protein gene and the function of antisense RNA in its repression in myelin-deficient mutant mouse**. *J Neurochem* 1991, **56**:560-567.
191. Khochbin S, Lawrence JJ: **An antisense RNA involved in p53 mRNA maturation in**

murine erythroleukemia cells induced to differentiate. *Embo J* 1989, **8**:4107-4114.

192. Prang N, Wolf H, Schwarzmann F: **Latency of Epstein-Barr virus is stabilized by antisense-mediated control of the viral immediate-early gene BZLF-1**. *J Med Virol* 1999, **59**:512-519.
193. Javier RT, Stevens JG, Dissette VB, Wagner EK: **A herpes simplex virus transcript abundant in latently infected neurons is dispensable for establishment of the latent state**. *Virology* 1988, **166**:254-257.
194. Michael NL, Vahey MT, d'Arcy L, Ehrenberg PK, Mosca JD, Rappaport J, Redfield RR: **Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat**. *J Virol* 1994, **68**:979-987.
195. Spencer CA, Gietz RD, Hodgetts RB: **Overlapping transcription units in the dopa decarboxylase region of Drosophila**. *Nature* 1986, **322**:279-281.
196. Lankenau S, Corces VG, Lankenau DH: **The Drosophila micropia retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase**. *Mol Cell Biol* 1994, **14**:1764-1775.
197. Bedford M, Arman E, Orr-Urtreger A, Lonai P: **Analysis of the Hoxd-3 gene: structure and localization of its sense and natural antisense transcripts**. *DNA Cell Biol* 1995, **14**:295-304.
198. Nepveu A, Marcu KB: **Intragenic pausing and anti-sense transcription within the murine c-myc locus**. *Embo J* 1986, **5**:2859-2865.
199. Chang Y, Spicer DB, Sonenshein GE: **Effects of IL-3 on promoter usage, attenuation and antisense transcription of the c-myc oncogene in the IL-3-dependent Ba/F3 early pre-B cell line**. *Oncogene* 1991, **6**:1979-1982.
200. Miyajima N, Horiuchi R, Shibuya Y, Fukushige S, Matsubara K, Toyoshima K, Yamamoto T: **Two erbA homologs encoding proteins with different T3 binding capacities are transcribed from opposite DNA strands of the same genetic locus**. *Cell* 1989, **57**:31-39.
201. Armstrong BC, Krystal GW: **Isolation and characterization of complementary DNA for N-cym, a gene encoded by the DNA strand opposite to N-myc**. *Cell Growth Differ* 1992, **3**:385-390.
202. Lerner A, D'Adamio L, Diener AC, Clayton LK, Reinherz EL: **CD3 zeta/eta/theta locus is colinear with and transcribed antisense to the gene encoding the transcription factor Oct-1**. *J Immunol* 1993, **151**:3152-3162.
203. Laabi Y, Gras MP, Brouet JC, Berger R, Larsen CJ, Tsapis A: **The BCMA gene, preferentially expressed during B lymphoid maturation, is bidirectionally transcribed**. *Nucleic Acids Res* 1994, **22**:1147-1154.
204. Eccles MR, Grubb G, Ogawa O, Szeto J, Reeve AE: **Cloning of novel Wilms tumor gene (WT1) cDNAs; evidence for antisense transcription of WT1**. *Oncogene* 1994, **9**:2059-2063.
205. Borsu L, Presse F, Nahon JL: **The AROM gene, spliced mRNAs encoding new DNA/RNA-binding proteins are transcribed from the opposite strand of the melanin-concentrating hormone gene in mammals**. *J Biol Chem* 2000, **275**:40576-40587.
206. Black AR, Dolnick BJ: **Expression of rTS correlates with altered growth regulation of thymidylate synthase**. *Cancer Res* 1996, **56**:700-705.
207. Fu XD, Maniatis T: **Isolation of a complementary DNA that encodes the mammalian splicing factor SC35**. *Science* 1992, **256**:535-538.
208. Adelman JP, Bond CT, Douglass J, Herbert E: **Two mammalian genes transcribed from opposite strands of the same DNA locus**. *Science* 1987, **235**:1514-1517.
209. Bevilacqua A, Ceriani MC, Schiavone N, Nicolin A: **BCL-2 regulation targeting the AU-rich domain**. *Nucleosides Nucleotides Nucleic Acids* 2000, **19**:1273-1279.
210. Chalfie M, Horvitz HR, Sulston JE: **Mutations that lead to reiterations in the cell lineages of C. elegans**. *Cell* 1981, **24**:59-69.
211. Lee RC, Feinbaum RL, Ambros V: **The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14**. *Cell* 1993, **75**:843-854.
212. Wightman B, Ha I, Ruvkun G: **Posttranscriptional regulation of the heterochronic gene lin-14 by lin- 4 mediates temporal pattern formation in C. elegans**. *Cell* 1993, **75**:855-862.
213. Rougvie AE, Ambros V: **The heterochronic gene lin-29 encodes a zinc finger protein that controls a terminal differentiation event in Caenorhabditis elegans**. *Development* 1995, **121**:2491-2500.
214. Slack FJ, Basson M, Liu Z, Ambros V, Horvitz HR, Ruvkun G: **The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory**

RNA and the LIN-29 transcription factor. *Mol Cell* 2000, **5**:659-669.

215. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Muller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G: **Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA**. *Nature* 2000, **408**:86-89.
216. Hildebrandt M, Humbel BM, Nellen W: **The Dictyostelium discoideum EB4 gene product and a truncated mutant form of the protein are localized in prespore vesicles but absent from mature spores**. *Dev Biol* 1991, **144**:212-214.
217. Saccomanno L, Bass BL: **A minor fraction of basic fibroblast growth factor mRNA is deaminated in Xenopus stage VI and matured oocytes**. *RNA* 1999, **5**:39-48.
218. Scadden AD, Smith CW: **A ribonuclease specific for inosine-containing RNA: a potential role in antiviral defence?** *Embo J* 1997, **16**:2140-2149.
219. Kumar M, Carmichael GG: **Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts**. *Proc Natl Acad Sci U S A* 1997, **94**:3542-3547.
220. Schmidt M, Zheng P, Delihas N: **Secondary structures of Escherichia coli antisense micF RNA, the 5'-end of the target ompF mRNA, and the RNA/RNA duplex**. *Biochemistry* 1995, **34**:3621-3631.
221. Andersen J, Forst SA, Zhao K, Inouye M, Delihas N: **The function of micF RNA. micF RNA is a major factor in the thermal regulation of OmpF protein in Escherichia coli**. *J Biol Chem* 1989, **264**:17961-17970.
222. Tetart F, Bouche JP: **Regulation of the expression of the cell-cycle gene ftsZ by DicF antisense RNA. Division does not require a fixed number of FtsZ molecules**. *Mol Microbiol* 1992, **6**:615-620.
223. Waldbeser LS, Tolmasky ME, Actis LA, Crosa JH: **Mechanisms for negative regulation by iron of the fatA outer membrane protein gene expression in Vibrio anguillarum 775**. *J Biol Chem* 1993, **268**:10433-10439.
224. Waldbeser LS, Chen Q, Crosa JH: **Antisense RNA regulation of the fatB iron transport protein gene in Vibrio anguillarum**. *Mol Microbiol* 1995, **17**:747-756.
225. Poulsen LK, Refn A, Molin S, Andersson P: **The gef gene from Escherichia coli is regulated at the level of translation**. *Mol Microbiol* 1991, **5**:1639-1648.
226. Grundy FJ, Henkin TM: **Conservation of a transcription antitermination mechanism in aminoacyl- tRNA synthetase and amino acid biosynthesis genes in gram-positive bacteria**. *J Mol Biol* 1994, **235**:798-804.
227. Janssen PJ, Jones DT, Woods DR: **Studies on Clostridium acetobutylicum glnA promoters and antisense RNA**. *Mol Microbiol* 1990, **4**:1575-1583.
228. Chen Q, Crosa JH: **Antisense RNA, fur, iron, and the regulation of iron transport genes in Vibrio anguillarum**. *J Biol Chem* 1996, **271**:18885-18891.
229. Chai S, Welch TJ, Crosa JH: **Characterization of the interaction between Fur and the iron transport promoter of the virulence plasmid in Vibrio anguillarum**. *J Biol Chem* 1998, **273**:33841-33847.
230. Mizuno T, Chou MY, Inouye M: **A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA)**. *Proc Natl Acad Sci U S A* 1984, **81**:1966-1970.
231. Chou JH, Greenberg JT, Demple B: **Posttranscriptional repression of Escherichia coli OmpF protein in response to redox stress: positive control of the micF antisense RNA by the soxRS locus**. *J Bacteriol* 1993, **175**:1026-1031.
232. Esterling L, Delihas N: **The regulatory RNA gene micF is present in several species of gram- negative bacteria and is phylogenetically conserved**. *Mol Microbiol* 1994, **12**:639-646.
233. Suzuki T, Ueguchi C, Mizuno T: **H-NS regulates OmpF expression through micF antisense RNA in Escherichia coli**. *J Bacteriol* 1996, **178**:3650-3653.
234. Tétart F, Albigot R, Conter A, Mulder E, Bouché JP: **Involvement of FtsZ in coupling of nucleoid separation with septation**. *Mol Microbiol* 1992, **6**:621-627.
235. Bouché F, Bouché JP: **Genetic evidence that DicF, a second division inhibitor encoded by the Escherichia coli dicB operon, is probably RNA**. *Mol Microbiol* 1989, **3**:991-994.
236. Cam K, Cuzange A, Bouche JP: **Sigma S-dependent overexpression of ftsZ in an Escherichia coli K-12 rpoB mutant that is resistant to the division inhibitors DicB and DicF RNA**. *Mol Gen Genet* 1995, **248**:190-194.
237. Faubladier M, Cam K, Bouché JP: **Escherichia coli cell division inhibitor**

DicF-RNA of the dicB operon. Evidence for its generation in vivo by transcription termination and by RNAse III and RNAse E-dependent processing. *J Mol Biol* 1990, **212**:461-471.

238. Faubladier M, Bouche JP: **Division inhibition gene dicF of Escherichia coli reveals a widespread group of prophage sequences in bacterial genomes**. *J Bacteriol* 1994, **176**:1150-1156.
239. Case CC, Roels SM, Gonzalez JE, Simons EL, Simons RW: **Analysis of the promoters and transcripts involved in IS10 anti-sense RNA control**. *Gene* 1988, **72**:219-236.
240. Simons RW, Kleckner N: **Translational control of IS10 transposition**. *Cell* 1983, **34**:683-691.
241. Roberts D, Hoopes BC, McClure WR, Kleckner N: **IS10 transposition is regulated by DNA adenine methylation**. *Cell* 1985, **43**:117-130.
242. Kittle JD, Simons RW, Lee J, Kleckner N: **Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5' end of the target RNA and a loop in the anti-sense RNA**. *J Mol Biol* 1989, **210**:561-572.
243. Ma C, Simons RW: **The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site**. *Embo J* 1990, **9**:1267-1274.
244. Case CC, Simons EL, Simons RW: **The IS10 transposase mRNA is destabilized during antisense RNA control**. *Embo J* 1990, **9**:1259-1266.
245. Jain C: **IS10 antisense control in vivo is affected by mutations throughout the region of complementarity between the interacting RNAs**. *J Mol Biol* 1995, **246**:585-594.
246. Case CC, Roels SM, Jensen PD, Lee J, Kleckner N, Simons RW: **The unusual stability of the IS10 anti-sense RNA is critical for its function and is determined by the structure of its stem-domain**. *Embo J* 1989, **8**:4297-4305.
247. Biere AL, Citron M, Schuster H: **Transcriptional control via translational repression by c4 antisense RNA of bacteriophages P1 and P7**. *Genes Dev* 1992, **6**:2409-2416.
248. Heinrich J, Riedel HD, Ruckert B, Lurz R, Schuster H: **The lytic replicon of bacteriophage P1 is controlled by an antisense RNA**. *Nucleic Acids Res* 1995, **23**:1468-1474.
249. Sabbattini P, Forti F, Ghisotti D, Deho G: **Control of transcription termination by an RNA factor in bacteriophage P4 immunity: identification of the target sites**. *J Bacteriol* 1995, **177**:1425-1434.
250. Ranade K, Poteete AR: **A switch in translation mediated by an antisense RNA**. *Genes Dev* 1993, **7**:1498-1507.
251. Schaefer KL, McClure WR: **Antisense RNA control of gene expression in bacteriophage P22. II. Kinetic mechanism and cation specificity of the pairing reaction**. *RNA* 1997, **3**:157-174.
252. Citron M, Schuster H: **The c4 repressors of bacteriophages P1 and P7 are antisense RNAs**. *Cell* 1990, **62**:591-598.
253. Citron M, Schuster H: **The c4 repressor of bacteriophage P1 is a processed 77 base antisense RNA**. *Nucleic Acids Res* 1992, **20**:3085-3090.
254. Stougaard P, Molin S, Nordstrom K: **RNAs involved in copy-number control and incompatibility of plasmid R1**. *Proc Natl Acad Sci U S A* 1981, **78**:6008-6012.
255. Shiba K, Mizobuchi K: **Posttranscriptional control of plasmid ColIb-P9 repZ gene expression by a small RNA**. *J Bacteriol* 1990, **172**:1992-1997.
256. Tomizawa J, Itoh T, Selzer G, Som T: **Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA**. *Proc Natl Acad Sci U S A* 1981, **78**:1421-1425.
257. Brantl S, Birch-Hirschfeld E, Behnke D: **RepR protein expression on plasmid pIP501 is controlled by an antisense RNA-mediated transcription attenuation mechanism**. *J Bacteriol* 1993, **175**:4052-4061.
258. Le Chatelier E, Ehrlich SD, Janniere L: **Countertranscript-driven attenuation system of the pAM beta 1 repE gene**. *Mol Microbiol* 1996, **20**:1099-1112.
259. Sugiyama T, Itoh T: **Control of ColE2 DNA replication: in vitro binding of the antisense RNA to the Rep mRNA**. *Nucleic Acids Res* 1993, **21**:5972-5977.
260. del Solar G, Espinosa M: **The copy number of plasmid pLS1 is regulated by two trans-acting plasmid products: the antisense RNA II and the repressor protein, RepA**. *Mol Microbiol* 1992, **6**:83-94.
261. Kim K, Meyer RJ: **Copy-number of broad host-range plasmid R1162 is regulated by a small RNA**. *Nucleic Acids Res* 1986, **14**:8027-8046.
262. Gerdes K, Gultyaev AP, Franch T, Pedersen K, Mikkelsen ND: **Antisense RNA-regulated programmed cell death**. *Annu Rev Genet* 1997, **31**:1-31.
263. Koraimann G, Koraimann C, Koronakis V, Schlager S, Hogenauer G: **Repression and**

derepression of conjugation of plasmid R1 by wild-type and mutated finP antisense RNA. *Mol Microbiol* 1991, **5**:77-87.

264. Woese CR, Kandler O, Wheelis ML: **Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya**. *Proc Natl Acad Sci U S A* 1990, **87**:4576-4579.
265. Paranchych W, Frost LS: **The physiology and biochemistry of pili**. *Adv Microb Physiol* 1988, **29**:53-114.
266. Willetts N: **The transcriptional control of fertility in F-like plasmids**. *J Mol Biol* 1977, **112**:141-148.
267. van Biesen T, Frost LS: **Differential levels of fertility inhibition among F-like plasmids are related to the cellular concentration of finO mRNA**. *Mol Microbiol* 1992, **6**:771-780.
268. Koraimann G, Teferle K, Markolin G, Woger W, Hogenauer G: **The FinOP repressor system of plasmid R1: analysis of the antisense RNA control of traJ expression and conjugative DNA transfer**. *Mol Microbiol* 1996, **21**:811-821.
269. van Biesen T, Soderbom F, Wagner EG, Frost LS: **Structural and functional analyses of the FinP antisense RNA regulatory system of the F conjugative plasmid**. *Mol Microbiol* 1993, **10**:35-43.
270. Jerome LJ, Frost LS: **In vitro analysis of the interaction between the FinO protein and FinP antisense RNA of F-like conjugative plasmids**. *J Biol Chem* 1999, **274**:10356-10362.
271. Lee SH, Frost LS, Paranchych W: **FinOP repression of the F plasmid involves extension of the half-life of FinP antisense RNA by FinO**. *Mol Gen Genet* 1992, **235**:131-139.
272. Ghetu AF, Gubbins MJ, Oikawa K, Kay CM, Frost LS, Glover JN: **The FinO repressor of bacterial conjugation contains two RNA binding regions**. *Biochemistry* 1999, **38**:14036-14044.
273. Hiraga S: **Chromosome and plasmid partition in Escherichia coli**. *Annu Rev Biochem* 1992, **61**:283-306.
274. Gerdes K, Bech FW, Jorgensen ST, Lobner-Olesen A, Rasmussen PB, Atlung T, Boe L, Karlstrom O, Molin S, von Meyenburg K: **Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon**. *Embo J* 1986, **5**:2023-2029.
275. Gultyaev AP, Franch T, Gerdes K: **Programmed cell death by hok/sok of plasmid R1: coupled nucleotide covariations reveal a phylogenetically conserved folding pathway in the hok family of mRNAs**. *J Mol Biol* 1997, **273**:26-37.
276. Franch T, Gultyaev AP, Gerdes K: **Programmed cell death by hok/sok of plasmid R1: processing at the hok mRNA 3'-end triggers structural rearrangements that allow translation and antisense RNA binding**. *J Mol Biol* 1997, **273**:38-51.
277. Thisted T, Nielsen AK, Gerdes K: **Mechanism of post-segregational killing: translation of Hok, SrnB and Pnd mRNAs of plasmids R1, F and R483 is activated by 3'-end processing**. *Embo J* 1994, **13**:1950-1959.
278. Dam Mikkelsen N, Gerdes K: **Sok antisense RNA from plasmid R1 is functionally inactivated by RNAse E and polyadenylated by poly(A) polymerase I**. *Mol Microbiol* 1997, **26**:311-320.
279. Thisted T, Gerdes K: **Mechanism of post-segregational killing by the hok/sok system of plasmid R1. Sok antisense RNA regulates hok gene expression indirectly through the overlapping mok gene**. *J Mol Biol* 1992, **223**:41-54.
280. Thisted T, Sorensen NS, Wagner EG, Gerdes K: **Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5'-end single-stranded leader and competes with the 3'-end of Hok mRNA for binding to the mok translational initiation region**. *Embo J* 1994, **13**:1960-1968.
281. Weaver KE, Walz KD, Heine MS: **Isolation of a derivative of Escherichia coli-Enterococcus faecalis shuttle vector pAM401 temperature sensitive for maintenance in E. faecalis and its use in evaluating the mechanism of pAD1 par-dependent plasmid stabilization**. *Plasmid* 1998, **40**:225-232.
282. Pritchard RH, Barth PT, Collins J: *Symp. Soc. Gen. Microbiol.* 1969, **19**:263-297.
283. Morita M, Oka A: **The structure of a transcriptional unit on colicin E1 plasmid**. *Eur J Biochem* 1979, **97**:435-443.
284. Gultyaev AP, van Batenburg FH, Pleij CW: **The influence of a metastable structure in plasmid primer RNA on antisense RNA binding kinetics**. *Nucleic Acids Res* 1995, **23**:3718-3725.
285. Tomizawa J, Itoh T: **Plasmid ColE1 incompatibility determined by**

interaction of RNA I with primer transcript. *Proc Natl Acad Sci U S A* 1981, **78**:6096-6100.

286. Lin-Chao S, Cohen SN: **The rate of processing and degradation of antisense RNAI regulates the replication of ColE1-type plasmids in vivo**. *Cell* 1991, **65**:1233-1242.
287. Xu F, Cohen SN: **RNA degradation in Escherichia coli regulated by 3' adenylation and 5' phosphorylation**. *Nature* 1995, **374**:180-183.
288. Tomizawa J: **Control of ColE1 plasmid replication. Intermediates in the binding of RNA I and RNA II**. *J Mol Biol* 1990, **212**:683-694.
289. Eguchi Y, Tomizawa J: **Complexes formed by complementary RNA stem-loops. Their formations, structures and interaction with ColE1 Rom protein**. *J Mol Biol* 1991, **220**:831-842.
290. Eguchi Y, Tomizawa J: **Complex formed by complementary RNA stem-loops and its stabilization by a protein: function of CoIE1 Rom protein**. *Cell* 1990, **60**:199-209.
291. Lee AJ, Crothers DM: **The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence**. *Structure* 1998, **6**:993-1005.
292. Tomizawa J: **Control of ColE1 plasmid replication. Interaction of Rom protein with an unstable complex formed by RNA I and RNA II**. *J Mol Biol* 1990, **212**:695-708.
293. Kumar CC, Novick RP: **Plasmid pT181 replication is regulated by two countertranscripts**. *Proc Natl Acad Sci U S A* 1985, **82**:638-642.
294. Brantl S, Wagner EG: **Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181**. *Mol Microbiol* 2000, **35**:1469-1482.
295. Ermolaeva MD, Khalak HG, White O, Smith HO, Salzberg SL: **Prediction of transcription terminators in bacterial genomes**. *J Mol Biol* 2000, **301**:27-33.
296. Brantl S, Wagner EG: **Antisense RNA-mediated transcriptional attenuation occurs faster than stable antisense/target RNA pairing: an in vitro study of plasmid pIP501**. *Embo J* 1994, **13**:3599-3607.
297. Brantl S, Wagner EG: **An unusually long-lived antisense RNA in plasmid copy number control: in vivo RNAs encoded by the streptococcal plasmid pIP501**. *J Mol Biol* 1996, **255**:275-288.
298. Brantl S: **The copR gene product of plasmid pIP501 acts as a transcriptional repressor at the essential repR promoter**. *Mol Microbiol* 1994, **14**:473-483.
299. Brantl S, Wagner EG: **Dual function of the copR gene product of plasmid pIP501**. *J Bacteriol* 1997, **179**:7016-7024.
300. Henkin TM: **tRNA-directed transcription antitermination**. *Mol Microbiol* 1994, **13**:381-387.
301. Condon C, Putzer H, Grunberg-Manago M: **Processing of the leader mRNA plays a major role in the induction of thrS expression following threonine starvation in Bacillus subtilis**. *Proc Natl Acad Sci U S A* 1996, **93**:6992-6997.
302. Osborn AM, da Silva Tatley FM, Steyn LM, Pickup RW, Saunders JR: **Mosaic plasmids and mosaic replicons: evolutionary lessons from the analysis of genetic diversity in IncFII-related replicons**. *Microbiology* 2000, **146**:2267-2275.
303. Asano K, Hama C, Inoue S, Moriwaki H, Mizobuchi K: **The plasmid ColIb-P9 antisense Inc RNA controls expression of the RepZ replication protein and its positive regulator repY with different mechanisms**. *J Biol Chem* 1999, **274**:17924-17933.
304. Asano K, Moriwaki H, Mizobuchi K: **An induced mRNA secondary structure enhances repZ translation in plasmid ColIb-P9**. *J Biol Chem* 1991, **266**:24549-24556.
305. Praszkier J, Wilson IW, Pittard AJ: **Mutations affecting translational coupling between the rep genes of an IncB miniplasmid**. *J Bacteriol* 1992, **174**:2376-2383.
306. Morris DR, Geballe AP: **Upstream open reading frames as regulators of mRNA translation**. *Mol Cell Biol* 2000, **20**:8635-8642.
307. Asano K, Kato A, Moriwaki H, Hama C, Shiba K, Mizobuchi K: **Positive and negative regulations of plasmid CoLIb-P9 repZ gene expression at the translational level**. *J Biol Chem* 1991, **266**:3774-3781.
308. Wilson IW, Praszkier J, Pittard AJ: **Mutations affecting pseudoknot control of the replication of B group plasmids**. *J Bacteriol* 1993, **175**:6476-6483.
309. Praszkier J, Bird P, Nikoletti S, Pittard J: **Role of countertranscript RNA in the copy number control system of an IncB miniplasmid**. *J Bacteriol* 1989, **171**:5056-5064.
310. Asano K, Mizobuchi K: **Structural analysis of late intermediate complex formed between plasmid ColIb-P9 Inc RNA and**

its target RNA. How does a single antisense RNA repress translation of two genes at different rates? *J Biol Chem* 2000, **275**:1269-1274.

311. Siemering KR, Praszkier J, Pittard AJ: **Interaction between the antisense and target RNAs involved in the regulation of IncB plasmid replication**. *J Bacteriol* 1993, **175**:2895-2906.
312. Meynell E, Datta N: **Mutant drug resistant factors of high transmissibility**. *Nature* 1967, **214**:885-887.
313. Nordstrom K, Molin S, Light J: **Control of replication of bacterial plasmids: genetics, molecular biology, and physiology of the plasmid R1 system**. *Plasmid* 1984, **12**:71-90.
314. Masai H, Arai K: **RepA and DnaA proteins are required for initiation of R1 plasmid replication in vitro and interact with the oriR sequence**. *Proc Natl Acad Sci U S A* 1987, **84**:4781-4785.
315. Riise E, Stougaard P, Bindslev B, Nordstrom K, Molin S: **Molecular cloning and functional characterization of a copy number control gene (copB) of plasmid R1**. *J Bacteriol* 1982, **151**:1136-1145.
316. Riise E, Molin S: **Purification and characterization of the CopB replication control protein, and precise mapping of its target site in the R1 plasmid**. *Plasmid* 1986, **15**:163-171.
317. Berzal-Herranz A, Wagner EG, Diaz-Orejas R: **Control of replication of plasmid R1: the intergenic region between copA and repA modulates the level of expression of repA**. *Mol Microbiol* 1991, **5**:97-108.
318. Ohman M, Wagner EG: **Regulation of replication of plasmid R1: an analysis of the intergenic region between copA and repA**. *Mol Gen Genet* 1991, **230**:321-328.
319. Blomberg P, Engdahl HM, Malmgren C, Romby P, Wagner EG: **Replication control of plasmid R1: disruption of an inhibitory RNA structure that sequesters the repA ribosome-binding site permits tap- independent RepA synthesis**. *Mol Microbiol* 1994, **12**:49-60.
320. Malmgren C, Engdahl HM, Romby P, Wagner EG: **An antisense/target RNA duplex or a strong intramolecular RNA structure 5' of a translation initiation signal blocks ribosome binding: the case of plasmid R1**. *RNA* 1996, **2**:1022-1032.
321. Persson C, Wagner EG, Nordstrom K: **Control of replication of plasmid R1: kinetics of in vitro interaction between the antisense RNA, CopA, and its target, CopT**. *Embo J* 1988, **7**:3279-3288.
322. Wagner H, Nordstrom K: **Structural analysis of an RNA molecule involved in replication control of plasmid R1**. *Nucleic Acids Res* 1986, **14**:2523-2538.
323. Persson C, Wagner EG, Nordstrom K: **Control of replication of plasmid R1: formation of an initial transient complex is rate-limiting for antisense RNA--target RNA pairing**. *Embo J* 1990, **9**:3777-3785.
324. Givskov M, Molin S: **Copy mutants of plasmid R1: effects of base pair substitutions in the copA gene on the replication control system**. *Mol Gen Genet* 1984, **194**:286-292.
325. Hjalt T, Wagner EG: **The effect of loop size in antisense and target RNAs on the efficiency of antisense RNA control**. *Nucleic Acids Res* 1992, **20**:6723-6732.
326. Malmgren C, Wagner EGH, Ehresmann C, Ehresmann B, Romby P: **Antisense RNA control of plasmid R1 replication. The dominant product of the antisense RNA-mRNA binding is not a full RNA duplex**. *J Biol Chem* 1997, **272**:12508-12512.
327. Wagner EG, Blomberg P, Nordstrom K: **Replication control in plasmid R1: duplex formation between the antisense RNA, CopA, and its target, CopT, is not required for inhibition of RepA synthesis**. *Embo J* 1992, **11**:1195-1203.
328. Li H, Nicholson AW: **Defining the enzyme binding domain of a ribonuclease III processing signal. Ethylation interference and hydroxyl radical footprinting using catalytically inactive RNAse III mutants**. *Embo J* 1996, **15**:1421-1433.
329. Massire C, Westhof E: **MANIP: an interactive tool for the modelling of ribonucleic acids**. *J Mol Graph & Modelling* 1998, **16**:197-205.
330. Peattie DA, Gilbert W: **Chemical probes for higher-order structure in RNA**. *Proc Natl Acad Sci U S A* 1980, **77**:4679-4682.
331. Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B: **Probing the structure of RNAs in solution**. *Nucleic Acids Res* 1987, **15**:9109-9128.
332. Moine H: **Probing RNA structure and function in solution**. In *RNA structure and function*. Edited by:; 1997:77-115.,
333. Latham JA, Cech TR: **Defining the inside and outside of a catalytic RNA molecule**. *Science* 1989, **245**:276-282.
334. Persson C, Wagner EG, Nordstrom K: **Control of replication of plasmid R1: structures and sequences of the antisense RNA, CopA, required for its**

binding to the target RNA, CopT. *Embo J* 1990, **9**:3767-3775.

335. Kolb FA, Malmgren C, Westhof E, Ehresmann C, Ehresmann B, Wagner EG, Romby P: **An unusual structure formed by antisense-target RNA binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment**. *RNA* 2000, **6**:311-324.
336. Kolb FA, Westhof E, Ehresmann C, Ehresmann B, Wagner EG, Romby P: **Bulged residues promote the progression of a loop-loop interaction to a stable and inhibitory antisense-target RNA complex**. *in press* 2001.
337. Wilson IW, Siemering KR, Praszkier J, Pittard AJ: **Importance of structural differences between complementary RNA molecules to control of replication of an IncB plasmid**. *J Bacteriol* 1997, **179**:742-753.
338. Asano K, Mizobuchi K: **Copy number control of IncIalpha plasmid ColIb-P9 by competition between pseudoknot formation and antisense RNA binding at a specific RNA site**. *Embo J* 1998, **17**:5201-5213.
339. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning, a laboratory manual*. Edited by Press CSHL. Cold Spring Harbor, N. Y.; 1989.
340. Serganov A, Rak A, Garber M, Reinbolt J, Ehresmann B, Ehresmann C, Grunberg-Manago M, Portier C: **Ribosomal protein S15 from Thermus thermophilus--cloning, sequencing, overexpression of the gene and RNA-binding properties of the protein**. *Eur J Biochem* 1997, **246**:291-300.
341. Kao C, Rudisser S, Zheng M: **A simple and efficient method to transcribe RNAs with reduced 3' heterogeneity**. *Methods* 2001, **23**:201-205.
342. Franch T, Gerdes K: **U-turns and regulatory RNAs**. *Curr Opin Microbiol* 2000, **3**:159-164.
343. Westhof E, Dumas P, Moras D: **Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals**. *Acta Crystallogr A* 1988, **44**:112-123.
344. Puglisi EV, Puglisi JD: **HIV-1 A-rich RNA loop mimics the tRNA anticodon structure**. *Nat Struct Biol* 1998, **5**:1033-1036.
345. Isel C, Marquet R, Keith G, Ehresmann C, Ehresmann B: **Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription**. *J Biol Chem* 1993, **268**:25269-25272.
346. Scarabino D, Crisari A, Lorenzini S, Williams K, Tocchini-Valentini GP: **tRNA prefers to kiss**. *Embo J* 1999, **18**:4571-4578.
347. Duconge F, Toulme JJ: **In vitro selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1**. *RNA* 1999, **5**:1605-1614.
348. Marino JP, Gregorian RS, Jr., Csankovszki G, Crothers DM: **Bent helix formation between RNA hairpins with complementary loops**. *Science* 1995, **268**:1448-1454.
349. Chang KY, Tinoco I, Jr.: **The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement**. *J Mol Biol* 1997, **269**:52-66.
350. Predki PF, Nayak LM, Gottlieb MB, Regan L: **Dissecting RNA-protein interactions: RNA-RNA recognition by Rop**. *Cell* 1995, **80**:41-50.
351. Tomizawa J: **Control of ColE1 plasmid replication: the process of binding of RNA I to the primer transcript**. *Cell* 1984, **38**:861-870.
352. Muriaux D, De Rocquigny H, Roques BP, Paoletti J: **NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop- loop complex into a stable dimer**. *J Biol Chem* 1996, **271**:33686-33692.
353. Balzer M, Wagner R: **Mutations in the leader region of ribosomal RNA operons cause defective 30 S ribosomes as revealed by in vivo structural probing**. *J Mol Biol* 1998, **276**:547-557.
354. Balzer M, Wagner R: **A chemical modification method for the structural analysis of RNA and RNA-protein complexes within living cells**. *Anal Biochem* 1998, **256**:240-242.
355. Novick RP: **Regulation of pathogenicity in Staphylococcus aureus by a peptide-based density-sensing system**. In *Cell-cell signaling in bacteria*. Edited by:; 1999:129-146.,
356. Recsei P, Kreiswirth B, O'Reilly M, Schlievert P, Gruss A, Novick RP: **Regulation of exoprotein gene expression in Staphylococcus aureus by agar**. *Mol Gen Genet* 1986, **202**:58-61.
357. Heinrichs JH, Bayer MG, Cheung AL: **Characterization of the sar locus and its interaction with agr in Staphylococcus aureus**. *J Bacteriol* 1996, **178**:418-423.
358. Cheung AL, Eberhardt K, Heinrichs JH: **Regulation of protein A synthesis by the sar and agr loci of Staphylococcus aureus**. *Infect Immun* 1997, **65**:2243-2249.

359. Cheung AL, Chien YT, Bayer AS: **Hyperproduction of alpha-hemolysin in a sigB mutant is associated with elevated SarA expression in Staphylococcus aureus**. *Infect Immun* 1999, **67**:1331-1337.
360. Giraudo AT, Cheung AL, Nagel R: **The sae locus of Staphylococcus aureus controls exoprotein synthesis at the transcriptional level**. *Arch Microbiol* 1997, **168**:53-58.
361. Smeltzer MS, Hart ME, Iandolo JJ: **Phenotypic characterization of xpr, a global regulator of extracellular virulence factors in Staphylococcus aureus**. *Infect Immun* 1993, **61**:919-925.
362. Morfeldt E, Janzon L, Arvidson S, Lofdahl S: **Cloning of a chromosomal locus (exp) which regulates the expression of several exoprotein genes in Staphylococcus aureus**. *Mol Gen Genet* 1988, **211**:435-440.
363. Novick RP, Projan SJ, Kornblum J, Ross HF, Ji G, Kreiswirth B, Vandenesch F, Moghazeh S: **The agr P2 operon: an autocatalytic sensory transduction system in Staphylococcus aureus**. *Mol Gen Genet* 1995, **248**:446-458.
364. Ji G, Beavis R, Novick RP: **Bacterial interference caused by autoinducing peptide variants**. *Science* 1997, **276**:2027-2030.
365. Mayville P, Ji G, Beavis R, Yang H, Goger M, Novick RP, Muir TW: **Structure-activity analysis of synthetic autoinducing thiolactone peptides from Staphylococcus aureus responsible for virulence**. *Proc Natl Acad Sci U S A* 1999, **96**:1218-1223.
366. Lina G, Jarraud S, Ji G, Greenland T, Pedraza A, Etienne J, Novick RP, Vandenesch F: **Transmembrane topology and histidine protein kinase activity of AgrC, the agr signal receptor in Staphylococcus aureus**. *Mol Microbiol* 1998, **28**:655-662.
367. Peng HL, Novick RP, Kreiswirth B, Kornblum J, Schlievert P: **Cloning, characterization, and sequencing of an accessory gene regulator (agr) in Staphylococcus aureus**. *J Bacteriol* 1988, **170**:4365-4372.
368. Morfeldt E, Tegmark K, Arvidson S: **Transcriptional control of the agr-dependent virulence gene regulator, RNAIII, in Staphylococcus aureus**. *Mol Microbiol* 1996, **21**:1227-1237.
369. Cheung AL, Projan SJ: **Cloning and sequencing of sarA of Staphylococcus aureus, a gene required for the expression of agr**. *J Bacteriol* 1994, **176**:4168-4172.
370. Novick RP, Ross HF, Projan SJ, Kornblum J, Kreiswirth B, Moghazeh S: **Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule**. *Embo J* 1993, **12**:3967-3975.
371. Benito Y, Lina G, Greenland T, Etienne J, Vandenesch F: **trans-complementation of a Staphylococcus aureus agr mutant by Staphylococcus lugdunensis agr RNAiii**. *J Bacteriol* 1998, **180**:5780-5783.
372. Vandenesch F, Projan SJ, Kreiswirth B, Etienne J, Novick RP: **Agr-related sequences in Staphylococcus lugdunensis**. *FEMS Microbiol Lett* 1993, **111**:115-122.
373. Balaban N, Novick RP: **Translation of RNAIII, the Staphylococcus aureus agr regulatory RNA molecule, can be activated by a 3'-end deletion**. *FEMS Microbiol Lett* 1995, **133**:155-161.
374. Vandenesch F, Kornblum J, Novick RP: **A temporal signal, independent of agr, is required for hla but not spa transcription in Staphylococcus aureus**. *J Bacteriol* 1991, **173**:6313-6320.
375. Tegmark K, Morfeldt E, Arvidson S: **Regulation of agr-dependent virulence genes in Staphylococcus aureus by RNAIII from coagulase-negative staphylococci**. *J Bacteriol* 1998, **180**:3181-3186.
376. Gultyaev AP, van Batenburg FH, Pleij CW: **The computer simulation of RNA folding pathways using a genetic algorithm**. *J Mol Biol* 1995, **250**:37-51.
377. Benito Y, Kolb FA, Romby P, Lina G, Etienne J, Vandenesch F: **Probing the structure of RNAIII, the Staphylococcus aureus agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression**. *RNA* 2000, **6**:668-679.
378. Edman P: **Sequence determination**. *Mol Biol Biochem Biophys* 1970, **8**:211-255.
379. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K: **Whole genome sequencing of meticillin-resistant Staphylococcus aureus**. *Lancet* 2001, **357**:1225-1240.
380. Srisawat C, Engelke DR: **Streptavidin aptamers: affinity tags for the study of**

RNAs and ribonucleoproteins. *RNA* 2001, **7**:632-641.

381. Olivas WM, Muhlrad D, Parker R: **Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs**. *Nucleic Acids Res* 1997, **25**:4619-4625.
382. Huttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie JP, Brosius J: **RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse**. *Embo J* 2001, **20**:2943-2953.
383. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S: **Novel small RNA-encoding genes in the intergenic regions of Escherichia coli**. *Curr Biol* 2001, **11**:941-950.
384. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S: **Identification of novel small RNAs using comparative genomics and microarrays**. *Genes Dev* 2001,**15**:1637-1651.