



HAL
open science

Identification, caractérisation et études structurales d'ARN non-codants

Benoît Masquida

► **To cite this version:**

Benoît Masquida. Identification, caractérisation et études structurales d'ARN non-codants. Sciences du Vivant [q-bio]. Université Louis Pasteur - Strasbourg I, 2007. tel-00271873

HAL Id: tel-00271873

<https://theses.hal.science/tel-00271873>

Submitted on 10 Apr 2008

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.

**IDENTIFICATION, CARACTERISATION ET ETUDES STRUCTURALES
D'ARN NON-CODANTS**

Habilitation à diriger des recherches
présentée par Benoît Masquida

*dans l'équipe : "Modélisation et Simulation des Acides Nucléiques"
de l'UPR 9002 dirigée par Eric WESTHOF
Institut de Biologie Moléculaire et Cellulaire du CNRS
15 rue René Descartes
67084 Strasbourg CEDEX
FRANCE*

Sommaire

1	INTRODUCTION.....	12
2	RESUME DES TRAVAUX ANTERIEURS (1995-2000).....	15
2.1	TRAVAIL DE THESE	15
2.1.1	Modélisation de fragments d'ARN ribosomiques	16
2.1.2	Modélisation du ribozyme en épingle à cheveux	19
2.1.3	Résolution à 0,97 Å de la structure cristallographique d'un nonamère d'ARN contenant des paires GoU 20	
2.2	TRAVAUX DE STAGE POST-DOCTORAL.....	23
2.2.1	Reconstruction du complexe ARN/Protéine de la RNase P procaryote.....	23
2.2.2	Caractérisation d'un motif universel de reconnaissance ARN/ARN.....	24
2.3	TRAVAUX EN TEMPS QUE CHARGE DE RECHERCHES	26
2.4	RESUME DES TRAVAUX DE MODELISATION MOLECULAIRE DES INTRONS DE GROUPE I ET DE LA RNASE P PROCARYOTE ET DE DEUX OPERATEURS D'ARNt SYNTHETASE.....	28
2.4.1	Modélisation des introns de groupe I.....	28
2.4.2	Modélisation de l'holoenzyme de la RNase P procaryote	31
2.4.3	Modélisation des opérateurs de deux ARNt-synthétases (Thréonyl et Aspartyl).....	32
2.5	CONTRIBUTION CRISTALLOGRAPHIQUE A L'ETUDE DU SYSTEME SITE-A/AMINOGLYCOSIDES.....	34
2.6	MODELISATION ET CARTOGRAPHIE DE LA REGION DE L'ELEMENT CONSTITUTIF DE TRANSPORT (CTE) DE L'ARN DU VIRUS DE MASON-PFIZER	36
3	PROJETS	39
3.1	STRUCTURE ET EVOLUTION DE L'INTRON DIR956-1 DE <i>DIDYMIUM IRIDIS</i>	41
3.2	IDENTIFICATION ET CARACTERISATION DE NOUVEAUX ARN CHEZ <i>S.AUREUS</i> ET <i>L.MONOCYTOGENES</i>	46
3.2.1	L'interférence à ARN chez les eucaryotes.....	46
3.2.2	Les petits ARN régulateurs chez <i>E. coli</i>	47
3.2.3	La régulation de la virulence chez les Gram ⁺ pathogènes	48
3.2.4	Les interrupteurs à ARN.....	49
3.2.5	Identifier de nouveaux ARN chez des bactéries pathogènes	51
4	BIBLIOGRAPHIE.....	54
5	Articles publiés dans des revues à comité de lecture	

Benoît Masquida

Date de naissance : 01/06/1970, Strasbourg, France

Etat civil : Marié, 3 enfants

Nationalité : Française

Service militaire effectué (08/94-05/95)

Institut de Biologie Moléculaire et Cellulaire - CNRS

15, rue René Descartes

67084 - Strasbourg

03 88 41 70 45

03 88 60 22 18 (Fax)

B.Masquida@ibmc.u-strasbg.fr

Adresse personnelle

19, rue Charles Appell

67000 Strasbourg

03 88 60 28 61

06 70 23 79 50

b.masquida@gmail.com

Diplômes universitaires

- 1995-1999 :** **Doctorat ès sciences** de l'université Louis Pasteur, Strasbourg I
Mention : Biologie Structurale, obtenue le 18 juin 1999 mention très honorable avec les félicitations du jury
- 1993-1994 :** **DEA National de Cristallographie et RMN Biologiques** de l'université Louis Pasteur (Strasbourg I), Université Joseph Fourier (Grenoble I), Université d'Orsay (Paris-sud). Obtention du **Magistère de Chimie-Biologie** de l'Université Louis Pasteur
- 1992-1993 :** **Maîtrise de Biochimie** de l'université Louis Pasteur, Strasbourg I
Deuxième année de **Magistère Chimie-Biologie** de l'Université Louis Pasteur
- 1991-1992 :** **Licence de Biochimie** de l'université Louis Pasteur, Strasbourg I
Première année de **Magistère Chimie-Biologie** de l'Université Louis Pasteur
- 1989-1991 :** **DEUG Bio-physico-chimie** de l'université Louis Pasteur, Strasbourg I
- 1988-1989 :** **Baccalauréat série D** Mention Assez Bien

Activité en matière de recherche

- Depuis Oct. 2005 :** **Chargé de Recherches 1^e classe au CNRS**
Institut de Biologie Moléculaire et Cellulaire, CNRS, Strasbourg, France
Domaines de Recherche : Identification par des approches génomiques d'ARN non-codants régulateurs chez les bactéries, Modélisation et cristallographie d'ARN non-codants
Directeur scientifique : Professeur Eric Westhof, ULP de Strasbourg, CNRS UPR 9002
- Oct. 2001 – Oct.2005 :** **Chargé de Recherches 2^e classe au CNRS**
Institut de Biologie Moléculaire et Cellulaire, CNRS, Strasbourg, France
Domaines de Recherche : Modélisation et cristallographie d'ARN non-codants
Directeur scientifique : Professeur Eric Westhof, ULP de Strasbourg, CNRS UPR 9002
- Oct. 2000 – Sept. 2001 :** **Stage Postdoctoral**
Institut de Biologie Moléculaire et Cellulaire, CNRS, Strasbourg, France
Domaines de Recherche : Détermination de la structure tridimensionnelle de la ribonucléase P eubactérienne et de l'IRES de HCV par cristallographie et modélisation moléculaire
Directeur scientifique : Professeur Eric Westhof, ULP de Strasbourg, CNRS UPR 9002

Financement : European Framework Program 5

Oct. 1999 - Sept. 2000 : Stage Postdoctoral

Yale University, Howard Hughes Medical Institute, Department of Molecular Biophysics & Biochemistry, New Haven, Connecticut, USA

Domaines de Recherche : Détermination de la structure tridimensionnelle de la ribonucléase P eubactérienne par cristallographie, application de la modélisation d'ARN a des problèmes de repliement

Directeur scientifique : Professeur Jennifer Doudna, Yale University, New-Haven, CT, USA

Financement : Howard Hughes Research Fellow

Oct. 1995 - Juin 1999 : Doctorat d'université ; Mention Biologie Structurale

Institut de Biologie Moléculaire et Cellulaire, CNRS, Strasbourg, France

Domaines de Recherche : Investigation théorique et expérimentale de la structure des ARN : “ Etude de petits motifs d'ARN par modélisation et cristallographie ”

Directeur de thèse : Professeur Eric Westhof, ULP de Strasbourg, CNRS UPR 9002

Mention : Très bien avec les félicitations du jury

Financement : Human Frontiers Science Program (Oct. 1998 – Juin 1999)

Bourse pour Docteur-Ingénieur CNRS/Rhône-Poulenc Rorer (Oct. 1995 – Sept. 1998)

Distinctions :

Prix de thèse ADRERUS 1999

L'ADRERUS (Association pour le développement des Relations entre l'Economie et la Recherche auprès des Universités de Strasbourg et de l'Université de Haute-Alsace) décerne chaque année des prix de thèse destinés à valoriser les travaux les plus remarquables et à faire connaître leurs auteurs. Le conseil général soutient cette action.

Publications avec comité de lecture:

1. M. Ryckelynck, **B. Masquida**, R. Giegé, M. Frugier, « An intricate RNA structure with two tRNA-derived motifs directs complex formation between yeast aspartyl-tRNA synthetase and its mRNA », *J. Mol. Biol.*, 354, 614-29 (2005)
2. B. François, R.J. Russel, J.B. Murray, F. Aboul-ela, **B. Masquida**, Q. Vicens, E. Westhof, « Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides : role of the number of rings and positive charges in the specific binding leading to miscoding », *Nucleic Ac. Res.*, 33, 5677-90 (2005)
3. O. Chaloin, J.-C. Peter, J.-P. Briand, **B. Masquida**, C. Desgranges, S. Muller, J. Hoebeke, « The N-terminus of HIV-1 Tat protein is essential for Tat-TAR RNA interaction », *Cell. Mol. Life Sci.*, 62, 355-61 (2005)
4. P. Rangan, **B. Masquida**, E. Westhof, S. A. Woodson, « Architecture and folding mechanism of the *Azoarcus* group I pre-tRNA », *J. Mol. Biol.*, 339, 41-51 (2004)
5. P. Rangan, **B. Masquida**, E. Westhof, S. A. Woodson, « Metal ion-dependent folding pathway of the *Azoarcus* group I ribozyme », *Proc. Natl. Acad. Sci. USA*, 100, 1574-1579 (2003)
6. J. Caillet, T. Nogueira, **B. Masquida**, M. Graffe, F. Eyermann, A.C. Dock-Bregeon, R. Sankaranarayanan, E. Westhof, D. Moras, B. Ehresmann, C. Ehresmann, P. Romby, M. Springer, « The modular structure of *Escherichia coli* threonyl-tRNA synthetase as an enzyme and as a regulator », *Mol. Microbiol.*, 47, 961-974 (2003)
7. Hsin-Yue Tsai, **Benoît Masquida**, Roopa Biswas, Eric Westhof & Venkat Gopalan « Molecular modelage of the three-dimensional structure of the bacterial RNase P », *J. Mol. Biol.* 325, 661-675 (2003)
8. Christina Waldsich, **Benoît Masquida**, Eric Westhof & Renée Schroeder, « Monitoring intermediate folding states of *td* group I intron *in vivo* », *EMBO J.*, 21, 5281-5291 (2002)
9. Pascal Auffinger, **Benoît Masquida** & Eric Westhof, “ Structural and dynamical characterization of nucleic acids water”, Springer Verlag’s Lecture Notes Series in computational methods for macromolecules : challenges and applications, 61-70, (2002)
10. Elizabeth A. Doherty, Robert T. Batey, **Benoît Masquida** & Jennifer A. Doudna, “ A universal mode of helix RNA packing ”, *Nature Struct. Biol.*, 8, 339-343 (2001)
11. **Benoît Masquida** & Eric Westhof, “ On the wobble GoU and related pairs ”, *RNA*, 6, 9-15 (2000)
12. **Benoît Masquida**, Claude Sauter & Eric Westhof, “ A sulfate pocket formed by three GoU pairs in the 0.97 Å crystal structure of a nonameric RNA ”, *RNA*, 5, 1384-1395 (1999)
13. **Benoît Masquida**, David J. Earnshaw, Sabine Müller, Snorri Th. Sigurdson, Fritz Eckstein, Eric Westhof, & Michael J. Gait, “ Inter-domain cross-linking and molecular modelling of the hairpin ribozyme ”, *J. Mol. Biol.*, 274, 197-212 (1997).

14. **Benoît Masquida**, Brice Felden, & Eric Westhof, “ Context dependent RNA-RNA recognition in a three-dimensional model of the 16S rRNA core ”, *BioMed. Chem.*, **5**, 1021-1035 (1997).
15. Sacha Serganov, **Benoît Masquida**, Eric Westhof, C. Cachia, Claude Portier, Maria Garber, Bernard Ehresmann, & Chantal Ehresmann, “ The 16S rRNA binding site of *Thermus thermophilus* ribosomal protein S15: comparison with *Escherichia coli* S15, minimum site and structure ”, *RNA*, **2**, 1124-1138 (1996).
16. Luttman, C.; Morize, Isabelle; Laoui, A.; **Masquida, B.**; James-Surcouf, E., « Specific Amino-Acid-Like Database Generation: Application to Bradykinin Antagonist Design », *Chemical Design Automation News*. **10** (4): 40 (1995)

Articles de revue :

1. **Benoît Masquida**, Boris François, Andreas Werner & Eric Westhof, « RNA crystallogenesi », in *High-throughput and conventional macromolecular crystallography* (M. Sanderson ; J. Skelly eds.), Oxford University Press, *sous presse*.
2. **Benoît Masquida**, & Eric Westhof, « A modular and hierarchical approach for all-atom RNA modelling », in *The RNA world, 3rd edition* (T. Cech ed.), pp. 659-681 , CSHLP (2006).
3. **Benoît Masquida**, & Eric Westhof, « Modelling tertiary structure of RNA », in *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*, pp. , Wiley-VCH Verlag GmbH & Co. KgaA (2005).
4. **Benoît Masquida**, & Eric Westhof, « Modelling the architecture of structured RNAs within a modular and hierarchical framework », in *Wiley Handbook of Crystallography* (Hartmann, R., Schön, A. & Westhof, E., eds.), pp.536-545, Wiley-VCH Verlag GmbH & Co. KgaA, (2005).
5. Boris François, Aurélie Lescoute-Philipps, Andreas Werner & **Benoît Masquida**, « Preparation and handling of RNA crystals », in *Wiley Handbook of Crystallography* (Hartmann, R., Schön, A. & Westhof, E. eds.), pp.438-452, Wiley-VCH Verlag GmbH & Co. KgaA, (2005).
6. **Benoît Masquida**, & Eric Westhof, “ Crystallographic structures of RNA oligoribonucleotides and ribozymes ”, in *Oxford Handbook of Nucleic Acid Structure* (Steve Neidle, ed.) pp 533-565, Oxford University Press (Oxford), (1999).
7. Eric Westhof, **Benoît Masquida**, & Luc Jaeger, “ RNA tectonics and molecular modeling of RNA”, in *Molecular modeling of nucleic acids* (Neocles B. Leontis and John SantaLucia, Jr, ed) pp. 346-358, ACS symposium series 682 American Chemical Society (Washington, DC), (1997).
8. Eric Westhof, **Benoît Masquida**, & Luc Jaeger, “ RNA tectonics : towards RNA design ”, *Folding & Design*, **1**, r78-r88 (1996).

Participations à des colloques et congrès nationaux et internationaux :

- 2006.** Henrik Nielsen, **Benoît Masquida**, Christer Einvik, Eric Westhof & Steinar Johansen « Structure and evolution of the GIR1 branching ribozyme »
RNA meeting 2006 (contribution orale d'Henrik Nielsen), Seattle (USA), 20-25 juin 2006
- Bertrand Beckert, Steinar Johansen, Eric Westhof & **Benoît Masquida**
 « Modélisation moléculaire d'un système à ribozymes siamois codant pour l'endonucléase I-DirI chez *Didymium iridis* »
 6^e rencontre SifrARN (contribution orale de Bertrand Beckert), Rennes (France), 3-6 juillet 2006
- 2005.** **Benoît Masquida** « L'architecture de l'ARN à portée de main : la modélisation moléculaire »
Andreas Werner, **Benoît Masquida** & Eric Westhof « Tertiary interactions in RNase P RNA investigated by UV melting »
RNA meeting 2005 (poster) – Banff (Canada), 24-29 Mai 2005
- 2003.** **Benoît Masquida** « Un essai sur la RNase P »
Séminaire interne de l'IBMC (conférence) – Strasbourg (France), 22 Juin 2003
- Benoît Masquida** & Eric Westhof « Comparison between RNA models and crystal structures »
Gordon Conference Nucleic Acids (poster) – Newport (RI, USA), 1 – 6 Juin 2003.
RNA meeting 2003 (poster) – Vienne (Autriche), 1 – 6 Juillet 2003
- 2002.** **Benoît Masquida**, Hsin-Yue Tsai, Roopa Biswas, Venkat Gopalan, Eric Westhof
 « Three-dimensional models for the two types of RNase P holoenzymes »
- C. Waldsich, **Benoît Masquida**, E. Westhof, R. Schröder « Monitoring intermediate folding states of the *td* group I intron in vivo »
Ribozymes and RNA catalysis (poster) – Dundee (Ecosse), 23 – 27 Août 2002.
- 2001.** Hsin-Yue Tsai, Roopa Biswas, **Benoît Masquida**, Eric Westhof, Venkat Gopalan
 « RNA-protein interactions in *Escherichia coli* ribonuclease P »
RNA meeting 2001 (poster) – Banff (Alberta, Canada), 29 Mai au 3 Juin 2001.
- 1999.** **Benoît Masquida**, Claude Sauter, & Eric Westhof, “ Crystal structure of a nonameric RNA at 0.97 Å of resolution ”
RNA meeting 99 (poster) – Edimbourg, Ecosse, 23 au 27 Juin 1999.
- 1998.** **Benoît Masquida**, Claude Sauter, & Eric Westhof, “ Crystal structure of an RNA nonamer at 1.12 Å resolution ”
12th REGIO meeting of crystallography, (Présentation orale) - Strasbourg, 30 Septembre - 2 Octobre 1998.
- 1997.** **Benoît Masquida**, David J. Earnshaw, Sabine Müller, Snorri Th. Sigurdson, Fritz Eckstein, Eric Westhof, & Michael J. Gait, “ Modelling and crosslinking of the hairpin ribozyme ”

- SIFRA^{RN}*, (poster) - Strasbourg (France), 28-30 Avril, 1998
Spetsai summer school (poster), Spetses (Greece), 1-14 Septembre, 1997
Journées du Groupe Graphique Moléculaire, (poster) - Dourdan (France), 21-23 Mai, 1997
- 1995.** Claude Luttmann, Isabelle Morize, Abdelaziz Laoui, **Benoît Masquida**, Evelyne James-Surcouf, « Specific Amino-Acid-Like Database Generation: Application to Bradykinin Antagonist Design »
Abstracts of the Molecular Graphics and Modelling Society Conference, University of Leeds, UK, 10-12 April, 1995
- 1994.** **Benoît Masquida**, Brice Felden, & Eric Westhof, “ 3D modelling of the 16S RNA core of *Escherichia coli* ”
13^e Colloque de la société française de Biophysique, (poster) - Annecy, 8-12 Septembre, 1994.
- 1994.** **Benoît Masquida**, Isabelle Morize, Abdelaziz Laoui, & Evelyne James-Surcouf, “ Génération d'une base spécifique de structures 3D mimes d'amino acides : application a la conception d'antagonistes de la bradykinine ”
Science Day Rhône Poulenc Rorer. (poster) - Collegeville USA, février 1994.

Invitations à donner des conférences :

- 2006.** *Ph. D. course* : « Molecular Biology of RNA », Cours intitulé : RNA-protein interactions : Plasticity and dynamics – Copenhague (Danemark), 26 janvier
- 2005.** *Séminaire de l'IRISA, bioinfo Ouest et symbiose* « La structure des ARN à portée de main : la modélisation moléculaire » – Rennes (France), 7 Avril
- 1999.** Yale University, MB&B department seminars « Crystallographic study of small RNA motifs », Présentation orale, (USA), 18 Février

Participation à l'obtention de contrats de recherche :

1. Lundbeck foundation (Lundbeckfonden, Danemark) : natural science research (2007)

« Base structurales de la formation du lasso catalysée par le ribozyme GIR1 : Applications à la protection de l'extrémité 5' des ARNm »

2. ANR programme blanc : Secteur Biologie et Santé (2006)

« Etude structurale et fonctionnelle de l'IRES cellulaire dans l'ARNm de l'histone H4 »

Rapporteur dans des revues à comité de lecture :

- Acta Crystallographica section D
- Angewandte Chemie
- Biochemical Journal
- ChemBioChem
- Nucleic Acids Research
- RNA journal

Membre de sociétés savantes :

Société Française de Biophysique

Contributions à l'encadrement d'étudiants

DEPUIS SEPTEMBRE 2006.

- ENCADREMENT D'UN ETUDIANT EN THESE : BERTRAND BECKERT (90%)
2005-2006 :
- ENCADREMENT D'UN ETUDIANT EN DEUXIEME ANNEE DE MASTER GENOMIQUE STRUCTURALE ET BIOINFORMATIQUE : BERTRAND BECKERT (100%)
2004-2005.
- ENCADREMENT DE DEUX ETUDIANTS DE L'IUP TECHNIQUES AVANCEES DES SCIENCES DU VIVANT (STAGE DE 2 MOIS) : SOLENE BOICHOT ET SIMON FREY (100%)
2003-2004.
- JURY DE THESE : ALAIN XAYAPHOUMINE (DOCTORAT DE L'UNIVERSITE LOUIS PASTEUR) « DYNAMIQUE DE REPLIEMENT DE L'ARN : SIMULATIONS ET EXPERIENCES » (18 JUIN 2004), DIDIER CHATENAY (DIRECTEUR DE THESE), HERVE ISAMBERT (DIRECTEUR DE THESE), THOMAS SIMONSON (RAPPORTEUR EXTERNE), JEAN-LOUIS VIOVY (RAPPORTEUR EXTERNE), JORGE BASCHNAGEL (RAPPORTEUR INTERNE), BENOIT MASQUIDA (EXAMINATEUR).
2002-2006.
- ENCADREMENT PARTIEL DE TROIS ETUDIANTS DE THESE : ANDREAS WERNER (50%), BORIS FRANÇOIS (20%) ET AURELIE LESCOUTE-PHILIPPS (20%)
2001-2002.
- ENCADREMENT D'UN ETUDIANT EN DEA DE BIOLOGIE MOLECULAIRE ET CELLULAIRE : AURELIE LESCOUTE-PHILIPPS (100%)
- ENCADREMENT D'UN ETUDIANT EN THESE : BORIS FRANÇOIS (20%)
1998-1999.
- ENCADREMENT D'UN ETUDIANT EN THESE : QUENTIN VICENS (50%)
1997-1998.
- ENCADREMENT D'UN ETUDIANT EN DEA DE CRISTALLOGRAPHIE ET RMN BIOLOGIQUES: QUENTIN VICENS (100%)
1993-1994.
- ENCADREMENT D'UN ETUDIANT EN DEUXIEME ANNEE DE MAGISTERE CHIMIE-BIOLOGIE : VALERIE LEHNERT (50%)

Activité d'enseignement

2006-2007 :

- **TRAVAUX DIRIGES DE BIOINFORMATIQUE EN L3 SCIENCES DU VIVANT, PARCOURS BIOLOGIE ET INFORMATIQUE « STRUCTURE DES PROTEINES, DES ACIDES NUCLEIQUES ET DE LEURS COMPLEXES » (40 H)**
- **TRAVAUX DIRIGES DANS LE CADRE DE L'UE « LES ANTIBIOTIQUES ET LA SYNTHÈSE PROTÉIQUE » (12 H)**
- **ANIMATION D'UN STAND A LA FETE DE LA SCIENCE 2006**

2005-2006 :

- **TRAVAUX DIRIGES DE BIOINFORMATIQUE EN L3 SCIENCES DU VIVANT, PARCOURS BIOLOGIE ET INFORMATIQUE « STRUCTURE DES PROTEINES, DES ACIDES NUCLEIQUES ET DE LEURS COMPLEXES » (12 H)**
- **INTERVENANT DANS L'UE CHOISIR/VALORISER EN L3**
- **ANIMATION D'UN STAND A LA FETE DE LA SCIENCE 2005**

1 Introduction

Notre vision des ARN a largement évolué ces dernières années. De simple intermédiaire moléculaire entre le génome et les protéines, cette molécule s'est révélée comme l'acteur central de nombreux processus biologiques.

Historiquement, l'importance de la structure des ARN a tout d'abord été révélée dans les mécanismes de la traduction des ARN messagers en protéines par les ribosomes. De nombreux ARN sont impliqués dans ces phénomènes : les ARN messagers bien sûr mais également les ARN de transfert et les ARN ribosomiques qui portent d'ailleurs l'activité peptidyl-transférase. Des ARN jouent également des rôles centraux en amont de la traduction. La ribonucléase P permet la maturation 5' des ARN de transfert qui peuvent alors subir d'autres maturations puis être amino-acylés pour entrer dans la traduction. Les introns de groupe I et de groupe II, en s'auto-excisant des messagers, aboutissent à la formation de transcrits matures qui peuvent alors être décodés par le ribosome. De plus, ces introns jouent un rôle dans l'évolution en tant qu'éléments mobiles encodant souvent des enzymes de type nucléase ou encore polymérase. Ces ARN actifs ou ribozymes (pour ribo-enzymes) sont donc doués d'activité catalytique. La découverte de cette propriété des ARN a valu à leurs auteurs Sydney Altman et Thomas Cech le prix Nobel de chimie 1989.

Dans les années 1990, des évidences, génétiques cette fois, démontraient que la traduction de certaines protéines bactérienne semblait régulée par le produit de leur gène. La caractérisation de ce mécanisme sur le plan fonctionnel dans le cas de gènes de certaines ARNt-synthétases (thréonyl par exemple) a montré que la structure de l'extrémité 5' du messenger sequestrait le codon d'initiation et la séquence Shine-Dalgarno lors de l'interaction avec la protéine. Ce type de régulation était également observé dans le cas de la protéine ribosomique S15 dont le messenger se replie pour former une structure en pseudonœud qui piège le ribosome rendant celui-ci incapable de procéder à la traduction. Ces expériences démontraient que des structures pouvaient se former dans des ARN a priori voués uniquement à stocker de l'information génétique.

Voici maintenant une quinzaine d'années que l'équipe animée par le professeur Eric Westhof s'intéresse aux relations structure/fonction de ces ribozymes et plus généralement aux ARN régulateurs structurés qui reconnaissent des ligands spécifiques tels

que des protéines ou des antibiotiques. Des méthodes comme la modélisation moléculaire, la radiocristallographie et la dynamique moléculaire ont été spécifiquement développées pour répondre aux questions que pose la structure des ARN. C'est dans ce domaine que s'inscrit la majeure partie de mon travail. Nous tentons donc d'élucider les structures tridimensionnelles d'ARN pour en déduire leur architecture et les mécanismes fonctionnels qui en découlent. Pour avoir accès aux architectures d'ARN, nous construisons d'une part des modèles 3D par modélisation moléculaire en nous basant sur des analyses comparatives de séquences et sur des données de cartographie en solution. D'autre part, afin d'avoir accès à des détails moléculaires plus fins comme la géométrie des interactions entre nucléotides, nous étudions certains systèmes par radiocristallographie.

Nos efforts ont principalement porté sur plusieurs modèles architecturaux d'ARN non-codants tels des introns de groupe I, des ribonucleases P bactériennes, de nombreux ribozymes identifiés dans des viroïdes de plantes, ou encore des fragments non-traduits (5'-UTR) d'ARN messagers impliqués dans la régulation de leur propre expression ainsi que des fragments d'ARN ribosomiques. Les modèles construits permettent de rationaliser les données expérimentales accumulées. En effet, le modèle 3D d'ARN est plus puissant que la représentation sous forme de structure secondaire en ceci que cette dernière est une pure représentation de l'esprit très éloignée de la réalité. Le modèle lui, en intégrant de « réels » aspects structuraux, donne l'opportunité de voir la molécule sous une forme proche de ce à quoi elle ressemble dans son environnement. De plus la modélisation moléculaire permet de construire des modèles d'ARN plus rapidement que la cristallographie ne permet d'en résoudre les structures, les ARN restant des molécules difficiles à cristalliser. La modélisation moléculaire est donc une méthode de choix pour travailler en synergie en guidant les expérimentalistes dans le choix des expériences à réaliser. Au laboratoire, nous caractérisons également du point de vue biochimique certains ARN que nous étudions par radiocristallographie. Plusieurs structures cristallographiques ont pu être publiées ces dernières années et nous poursuivons actuellement nos efforts de ce sens.

Plus récemment, il a été mis en évidence chez les eucaryotes et plus spécifiquement chez les plantes et chez le nématode *C. elegans* un mécanisme de régulation de l'expression génétique dans lequel un court ARN se fixant en *trans* au messager déclenche la dégradation de ce dernier par la protéine Dicer. Ces ARN clivés sont ensuite pris en charge par un complexe protéique (RISC) qui détruit alors systématiquement tous les messagers

identiques qui subsistent dans la cellule. Ce système, l'interférence à ARN, permet donc de réduire au silence un gène donné avant même qu'il ait pu être traduit. Leurs découvreurs, Andrew Fire et Craig Mello, se sont vus décerner le prix Nobel 2006. Chez les procaryotes, un système apparenté basé sur la dégradation par des ribonucléases de duplexes d'ARN formés entre une région de messenger et un petit ARN non-codant régulateur a été mis en évidence. Mais la régulation génétique chez les bactéries livre encore d'autres surprises comme les interrupteurs à ARN (riboswitchs). Ces riboswitchs contenus dans les extrémités 5' non-traduites sont capables de changer de structure tridimensionnelle suite à la reconnaissance d'une petite molécule effectrice (typiquement une métabolite) ou en fonction de paramètres physico-chimique comme le pH ou la température. Ce changement conformationnel a pour conséquence de réguler positivement ou négativement l'expression des gènes dans lesquels ils se situent.

Ces découvertes récentes qui bouleversent notre vision de l'ARN doivent être intégrées à nos travaux de recherche. D'une part, quel que soit le mode d'action de ces ARN non-codants ou les mécanismes qu'ils régissent, ils doivent nécessairement interagir *in vivo* avec des partenaires moléculaires pour remplir leur fonction et donc adopter des structures précises que nous souhaitons étudier. D'autre part, il semble que seule une petite fraction de ces ARN régulateurs a été découverte à ce jour. Actuellement, il est estimé qu'environ 5% du génome de certaines bactéries pourraient contenir des ARN régulateurs. Les bactéries contrôlent également l'expression de certains gènes par des interrupteurs à ARN (riboswitch) qui représenteraient potentiellement de 2% du génome. C'est donc une proportion importante des génomes qui code exclusivement pour des ARN. En supposant que ces ARN font entre 20 et 200 nucléotides, les génomes bactériens de 2 millions de bases en moyenne contiendrait donc plusieurs centaines de ces ARN. Nous souhaitons donc également apporter notre contribution à la découverte et à la caractérisation de nouveaux ARN.

Par ailleurs, nous nous intéressons également à identifier et à caractériser de nouveaux ARN chez divers pathogènes comme *Staphylococcus aureus* et *Listeria monocytogenes* respectivement dans le cadre de collaborations avec l'équipe de Pascale Romby (IBMC-CNRS UPR 9002) et avec l'équipe de Pascale Cossart de l'institut Pasteur. Notre travail consiste à identifier des ARN qui pourraient être impliqués dans la régulation de la réponse de ces organismes aux variations physico-chimiques de leur environnement avec un intérêt particulier pour la régulation de leurs gènes de virulence.

2 Résumé des travaux antérieurs (1995-2000)

2.1 Travail de thèse

Mon travail de thèse a largement contribué à la compréhension de la logique d'assemblage des structures d'ARN à travers l'identification de modules élémentaires de construction. Diverses méthodes théoriques et expérimentales d'investigation de la structure des ARN sont mises en œuvre au laboratoire. L'analyse comparative de séquences permet de déduire la structure secondaire des ARN et certaines interactions tertiaires. La structure secondaire dévoilée par l'analyse de séquences est ensuite affinée à l'aide de méthodes biochimiques de cartographie en solution. La modélisation moléculaire permet ensuite de proposer un modèle d'organisation relative des divers éléments de structure secondaire au sein de l'édifice tridimensionnel que constitue la structure native de la molécule. La modélisation des ARN se pratique au niveau atomique, c'est-à-dire qu'elle prend intrinsèquement en compte les contraintes stéréochimiques. Dans certains cas, l'analyse de séquences permet d'associer un module structural à un motif dont la structure est connue par cristallographie ou RMN. Ce nombre de cas va en augmentant puisque la connaissance des structures des ARN ribosomiques a augmenté le volume de données structurales sur les ARN d'un facteur dix. Pour construire le modèle, les modules structuraux sont « décalqués » à partir des structures cristallographiques, puis assemblés interactivement à l'écran graphique. Dans les cas où sont identifiés des motifs dont les structures ne sont pas connues, le modèle repose sur la qualité des données en solution. Toutefois cette situation est de plus en plus rare aujourd'hui car l'attention grandissante des laboratoires du monde entier pour l'ARN fait que le nombre de structures cristallographiques augmentent significativement depuis quelques années. L'adéquation entre les données et le modèle permet de tester la validité du choix des différents motifs structuraux et de suggérer de nouvelles expériences lorsque des ambiguïtés sont révélés. Le modèle sera alors réexaminé à la lumière de ces nouveaux résultats. Toutefois, la structure fine des ARN résulte également de la mise en place d'interactions non-Watson-Crick. Dans ces cas-là, la modélisation tridimensionnelle des ARN trouve ses limites même si elle a permis d'identifier et de répertorier certains motifs récurrents de structure tertiaire dont les ARN sont composés. Ainsi, la modélisation d'ARN de fonctions variées permet de supposer que la diversité des architectures rencontrées est due au brassage d'un nombre fini de motifs autonomes. La modélisation d'ARN peut donc être appréhendée à la

manière d'un jeu de construction présentant des modules structuraux essentiels qui permettent d'échafauder des édifices fonctionnels.

Comment avoir accès à la structure fine de ces modules ? A ce stade, les interactions tertiaires peuvent être explorées par mutagenèse dirigée ; l'ARN peut également être cartographié par des sondes chimiques et enzymatiques ce qui permet d'émettre des hypothèses sur la géométrie des motifs constituant la structure tridimensionnelle entre lesquelles la modélisation permettra de trancher (1). Ces méthodes chimiques et biochimiques, couplées à l'analyse comparative de séquences et à l'étude des structures cristallographiques, permettent de conclure que les modules structuraux trouvés de façon récurrente dans les ARN adoptent des structures tridimensionnelles similaires (2) parce que leur formation repose sur principes physico-chimiques intrinsèques. Leur observation par des méthodes directes (RMN, diffraction des rayons X) constitue donc la clé de détermination des règles phylogénétiques auxquelles ils obéissent. Il est alors possible de comprendre comment les contraintes structurales d'un motif donné sélectionnent des séquences particulières qui rendent ensuite ces motifs identifiables par analyse comparative de séquences (3,4). Les modules concernés pourront ainsi être transplantés dans les divers ARN dans lesquels ils auront été identifiés. L'étude de la structure des ARN, ainsi que des motifs de base qui les composent peut donc être envisagée suivant une approche duale regroupant la modélisation des ARN, et leur étude par cristallographie.

2.1.1 Modélisation de fragments d'ARN ribosomiques

2.1.1.1 Le cœur de l'ARN 16S d'*Escherichia coli*

Nous nous sommes tout d'abord intéressés à la modélisation du cœur de l'ARN ribosomique 16S d'*Escherichia coli* (5). Nous avons choisi de modéliser la région du cœur de l'ARN 16S car elle comprend trois pseudonœuds qui contraignent fortement la structure. La modélisation était donc guidée par des contraintes géométriques fortes. Pour la première fois, un modèle atomique englobant des contraintes stéréochimiques précises a été proposé pour cette région de l'ARN ribosomique 16S. Il est d'autant plus intéressant de parler de ce modèle de structure aujourd'hui que la structure cristallographique de la sous-unité 30S est connue (6). La topologie proposée dans notre modèle du cœur de l'ARN 16S est correcte, c'est-à-dire que les passages de brins se font toujours dans les sillons profonds ou peu-profonds comme observé dans la structure cristallographique de la sous-unité 30S (Figure1). En conséquence, les différents domaines du modèle se situent bien à leur place au sein de la sous-unité 30S. La

majorité des nucléotides pour lesquelles des interactions avaient été proposées se produisent, néanmoins souvent selon des géométries différentes. Certaines interactions ne se produisent pas comme celles dans la région où le modèle s'éloigne le plus de la structure cristallographique. La région du pied de l'hélice H27 a en effet subi un remaniement de structure secondaire qui n'était pas encore identifié au moment où notre travail de modélisation a été entrepris. Néanmoins le déplacement quadratique moyen normalisé (*nrmsd*, (7)) sur les atomes du squelette ribose-phosphate englobant le premier pseudonoeud et les hélices H3, H19 et H27 est de 11,8 Å alors que ce paramètre n'excède pas 8,0 Å pour les autres régions du modèle. Certains domaines pris à part ont une structure très proche de la structure cristallographique correspondante. De même que dans la structure cristallographique, les trois pseudonoeuds se trouvent très proches les uns des autres dans notre modèle. Leur proximité spatiale soulève de nombreuses questions relatives à leur fonction ainsi qu'à de possibles interactions. Les rôles de ces pseudonoeuds sont en partie élucidés (8) aujourd'hui puisque les structures obtenues par cristallographie (9) et microscopie électronique (10) correspondant à différents états du ribosome au cours de la traduction ont été résolus.

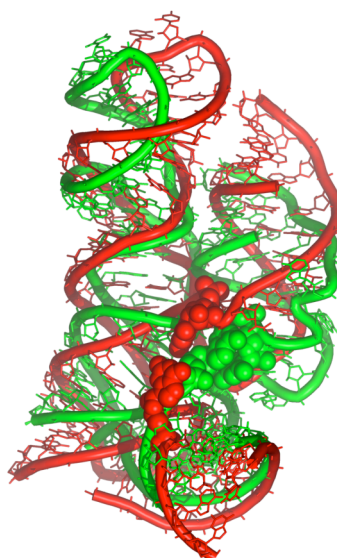


Figure 1 : Vue de la superposition entre les hélices H3, H4, H5 et H15 du modèle du cœur de l'ARN 16S d'*E. coli* (rouge) avec la région correspondante de la structure cristallographique de la sous-unité 30S (vert). Le déplacement quadratique moyen normalisé sur 100 résidus (*nrmsd*) sur les atomes du squelette sucre-phosphate est de 3,7 Å. La prise en compte d'un pontage covalent (nucléotides représentés en mode CPK) a permis d'élucider l'architecture de cette région de l'ARN 16S.

2.1.1.2 Le site de fixation ribosomique de S15 de *Thermus thermophilus*

Nous avons ensuite modélisé l'ARN fixant la protéine ribosomique S15 chez *Thermus thermophilus* (11). L'ARN reconnu par S15 se compose de la jonction triple entre trois épingles à cheveux de l'ARN 16S (H20, H21 et H22) reliées entre-elles par leur base. Deux régions simple-brins connectent H22 à H20 et H21. H20 et H21 sont directement connectées (Figure 2). Les données biochimiques indiquent que le site de fixation de la protéine S15 est constitué par la jonction triple qui contribue principalement à l'affinité, et par une boucle interne asymétrique de H22 qui contribue dans une moindre mesure à la fixation de la protéine. L'analyse de séquences nous a permis de corriger la structure secondaire de la boucle interne. Nous avons modélisé le site de fixation de la protéine sur H22 avec un excellent accord entre le modèle et les données en solution. Concernant la modélisation de la jonction triple, la structure secondaire de l'ARN contenait une erreur que l'analyse comparative de séquences ne pouvait pas déceler en raison de l'absence de covariations. La structure cristallographique a révélé le rôle précis de ces nucléotides qui consiste à interagir avec d'autres résidus de la jonction triple. En conséquence, la jonction entre les 3 hélices H20, H21 et H22 a été modélisée de façon incorrecte ; H20 a été empilée sur H22 alors que cette dernière est en fait empilée sur H21.

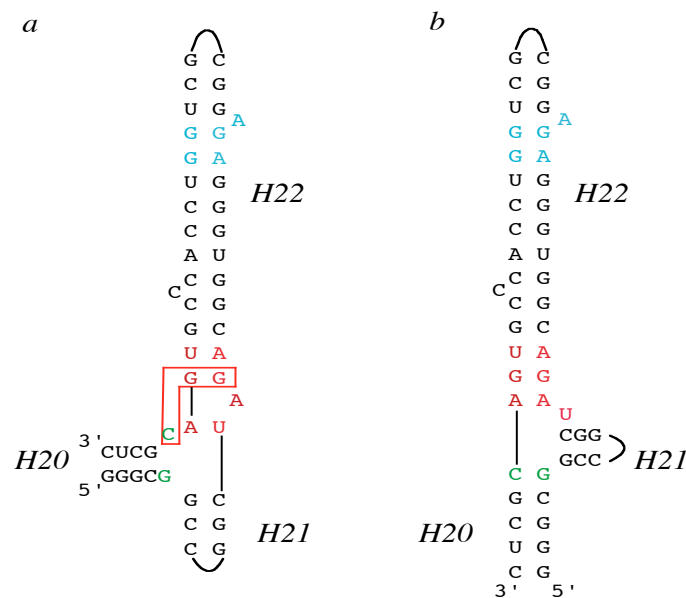


Figure 2 : Comparaison entre les structures secondaires observées dans la structure cristallographique (a) et le modèle (b) du site de fixation ribosomique de la protéine S15. La nomenclature des hélices est indiquée. La boucle interne asymétrique de H22 est représentée en bleu. Les appariements prenant place dans la jonction triple sont en rouge. Les nucléotides que la structure cristallographique a révélé non-appariés sont en vert. (a), Les hélices H21 et H22 sont empilées et H20 forme un angle avec celles-ci. Un triplet de bases se forme entre la paire GoG et le résidu C (vert) non-apparié. (b) L'appariement des bases vertes, en diminuant la flexibilité de la jonction triple force l'empilement des hélices H20 et H22, positionnant H21 perpendiculairement à l'empilement.

2.1.2 Modélisation du ribozyme en épingle à cheveux

Nous avons également construit un modèle de structure du ribozyme en épingle à cheveux (12) dans son état pré-catalytique, c'est-à-dire avant la formation du complexe permettant la réaction d'hydrolyse. Ce ribozyme se compose de deux domaines principaux A et B chacun contenant une boucle interne formant une majorité d'appariements non-canoniques (Figure 3). L'interaction de ces deux boucles est pré requise pour que la catalyse se produise. La façon dont ces boucles interagissent est aujourd'hui connue grâce à la structure cristallographique (13). Cette dernière montre que l'approche des domaines A et B est suivie par la reconnaissance de la G substrat au niveau de laquelle le clivage intervient.

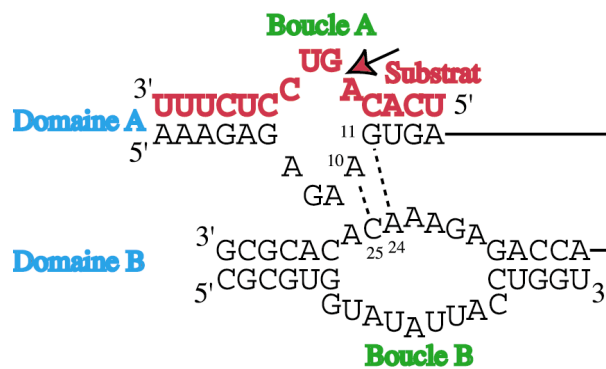


Figure 3 : Structure secondaire du ribozyme en épingle à cheveux. Les nucléotides formant le cœur catalytique sont en noir et le brin substrat en rouge. L'emplacement de la coupure est matérialisé par une flèche noire à fond rouge. Quatre résidus constituent chaque brin de la boucle A alors que la boucle B se compose de sept et neuf résidus. Notre modèle décrit une interaction possible entre les boucles A et B sous forme d'une glissière à ribose entre les résidus A10-C25 et G11-A24 (lignes pointillées).

Les seules informations structurales directes dont nous disposions reposaient sur la structure de la boucle A obtenue par RMN (14,15). L'homologie de séquences entre la boucle B et la boucle de l' α sarcine (16) nous a permis de proposer un schéma d'appariements pour trois paires de base de la boucle B. Nous avons alors puisé dans les bases de données cristallographiques les motifs structuraux dont les séquences s'apparentaient le plus aux régions à modéliser *de novo* du ribozyme. La structure du domaine P4-P6 des introns de groupe I (17) nous a été très utile pour proposer un mode d'interaction entre les deux boucles sous forme de glissière à riboses dont l'existence était soutenue expérimentalement (18). La structure cristallographique (13) montre que la glissière à riboses se forme de la même manière. L'angle formé entre les deux domaines est donc correct (Figure 4). En revanche, les appariements des 2 boucles non canoniques ne sont prédits correctement que dans 5 cas sur 9. Le fait que G+1 s'apparie à un résidu de la boucle B est le caractère, par essence imprévisible,

qui nous a empêché de réaliser un schéma d'appariements totalement correct. Le fait que l'architecture globale et une bonne partie des appariements non canoniques est été correctement prédits constitue toutefois un résultat important.



Figure 4 : Vue de la superposition entre le modèle (rouge) et la structure cristallographique (vert) du ribozyme en épingle à cheveux. Le nucléotide 5' est représenté en mode CPK. Le *nrmsd* (voir § 2.1.1.1) entre les atomes des riboses est de 6,9 Å.

2.1.3 Résolution à 0,97 Å de la structure cristallographique d'un nonamère d'ARN contenant des paires GoU

Nous avons ensuite identifié parmi ces modèles et ceux du laboratoire, des motifs de structure secondaire dont l'étude par cristallographie a été entreprise. Après avoir pallié les problèmes techniques liés à l'obtention de grandes quantités d'ARN à des taux de pureté élevés (revus dans (19)), nous avons synthétisé différents oligoribonucléotides englobant différents motifs que nous avons tenté de cristalliser. Parmi ceux-ci, nous avons pu résoudre la structure d'un oligonucléotide autoccomplémentaire de séquence $(r(\text{CGUAGUCG})\text{dC})_2$ à 0,97 Å de résolution. Ce duplex de forme A renferme deux paires de bases GoU dans des environnement ioniques différents (20). Cette résolution représentait la plus élevée jamais atteinte pour un oligoribonucléotide jusqu'à la publication d'une structure de quartets de G (21) à 0,66 Å. Nous avons donc pu observer de manière très précise la géométrie de l'hélice et des appariements qui la composent (Figure 5). Ce travail permettra notamment d'améliorer les paramètres des acides nucléiques dans les programmes d'affinement cristallographiques. De plus, la première sphère d'hydratation a pu être observée quasiment dans son ensemble,

permettant ainsi de caractériser les différents sites d'hydratation de l'ARN. Dans cette structure, une des deux paires GoU de l'unité asymétrique se trouve au contact d'un ion sulfate dans le sillon peu profond. Les atomes d'oxygène de l'ion sulfate sont reconnus par le groupement N2 de la guanine. Bien que cette observation résulte en partie d'un effet d'empilement cristallin, la similitude chimique du sulfate et du phosphate permet de proposer que le groupement N2 des guanines peut jouer un rôle dans la reconnaissance d'un brin d'ARN par l'intermédiaire d'un groupement phosphate du squelette. Ce cas de figure se présente souvent dans les pseudonœuds et un exemple a permis d'étayer notre conclusion (22). De manière plus générale, la reconnaissance, par le N2 d'une guanine formant une paire GoU, d'un partenaire chargé négativement est également observé entre substrat des introns de groupe I par l'intermédiaire du N3 (δ) d'une adénine faisant partie de la jonction J4/5 (23) élargissant encore la portée de notre observation.

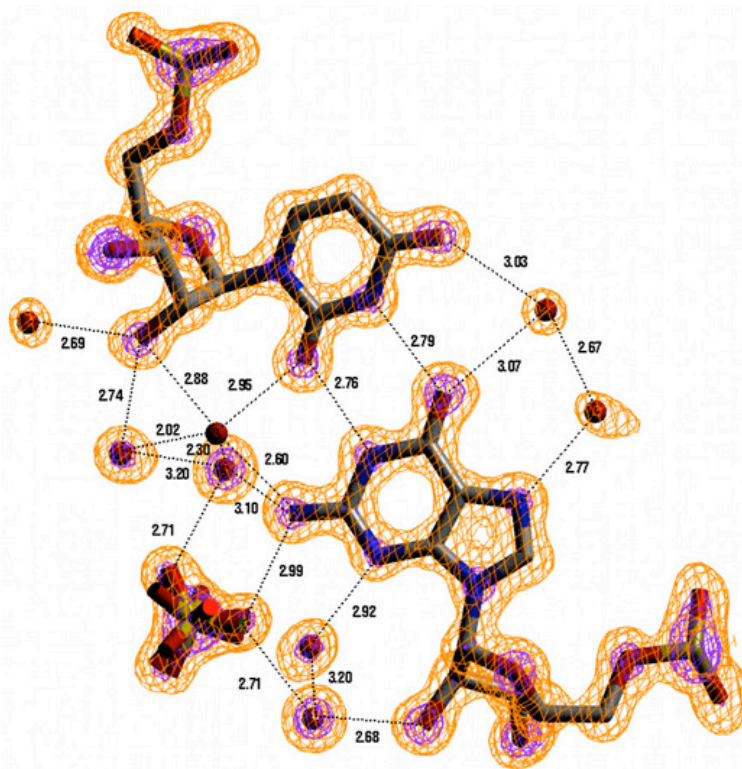


Figure 5 : Représentation de la densité électronique autour de la paire GoU en interaction avec l'ion sulfate dans AR0019. La résolution est 0,97 Å. Le contour de la carte orange est à $2,5 \sigma$, le contour de la carte pourpre est à 4σ .

L'obtention de ces résultats nous a permis de faire le point sur les différents types de paires GoU et sur les paires de bases qui leur sont isostériques (24). Les paires GoU ne sont pas tout à fait isostériques aux paires Watson-Crick. Nous avons donc étudié comment l'incorporation au sein d'une hélice d'une paire GoU ou d'un tandem de paires GoU (ou bien

d'appariements isostériques telles les paires A⁺oC ou UoU) se répercutait sur les paramètres de l'hélice. De façon caractéristique, en 5' de la U d'une paire GoU, la torsion de l'hélice diminue systématiquement alors qu'elle augmente en 3'. Si deux hélices sont superposées au niveau d'un plateau de bases, l'un étant une paire GoU, on constate que le déplacement quadratique moyen sur les atomes du squelette ribose-phosphate devient significatif (rmsd > 2 Å) cinq paires de bases en amont et en aval. Dans l'hypothèse où un ARN est reconnu entre autres au niveau d'une paire GoU, on comprend que la mutation de cette paire ou son renversement se traduira par un déplacement significatif du squelette qui diminuera la stabilité du complexe.

2.2 Travaux de stage post-doctoral

2.2.1 Reconstruction du complexe ARN/Protéine de la RNase P procaryote

Mon stage post-doctoral, (Octobre 1999 - Septembre 2000) dans le laboratoire du professeur Jennifer Doudna à l'université Yale aux Etats-Unis, a été dédié à l'étude de la ribonucléase P (RNase P) procaryote. Cette enzyme est responsable de la maturation de l'extrémité 5' des ARN de transfert (ARNt) par coupure de la séquence additionnelle présente sur les transcrits de pré-ARNt. Celle-ci constitue une cible thérapeutique potentielle dans la mesure où les différences avec l'enzyme eucaryote permettent d'envisager d'inhiber la fonction de la première sans interférer avec la seconde. J'ai travaillé sur les RNase P de *Thermotoga maritima* et de *Staphylococcus aureus* en vue d'une étude structurale par cristallographie. La ribonucléase P se compose chez les procaryotes d'une sous-unité ribonucléique qui catalyse la réaction d'hydrolyse et d'une sous-unité protéique (25). Un travail préliminaire à la cristallisation consistait à réaliser des expériences de reconstruction du complexe entre les deux sous-unités de façon à générer une population homogène d'enzyme tant du point de vue chimique que conformationnel et ce pour initier des essais de cristallogenèse. Il a fallu pour cela cloner, surproduire et purifier l'ARN et la protéine de *Thermotoga maritima*, les composants de l'enzyme de *Staphylococcus aureus* m'étant fournis par la firme Smith Kline Beecham. Il a ensuite fallu procéder à la renaturation de l'ARN et à la reconstruction du complexe. Ces complexes ont été partiellement caractérisés par des tests de mobilité sur gel et de rétention sur filtre de nitrocellulose. Les résultats de ces expériences mettent en évidence une propension importante du complexe à former des agrégats qui précipitent rapidement en solution. Divers protocoles de reconstruction du complexe ont donc été testés afin d'identifier des conditions chimiques neutralisant l'agrégation. La renaturation thermique de l'ARN suivie d'une incubation avec la protéine se montre peu efficace par rapport à la dénaturation chimique du complexe. Cette méthode consiste à dialyser lentement une solution contenant l'ARN, la protéine et de l'urée 7 M contre un tampon de dialyse contenant les concentrations finales de solutés dans lesquelles le complexe sera testé. Cette méthode a permis d'éviter l'agrégation des composés à des concentrations de l'ordre du μM . De façon intéressante, l'incubation avec la protéine de l'ARN renaturé par dialyse ne permet pas d'éviter l'agrégation. A noter que ces concentrations devraient encore être multipliées par un facteur cent pour espérer entreprendre des tests de cristallisation.

La reconstruction du complexe ARN/protéine pose donc des problèmes importants. Or l'ARN de la RNase P est catalytique chez les procaryotes, signifiant qu'il est capable d'adopter sans l'aide de la protéine une structure native. Il semble donc raisonnable de résoudre également la structure de l'ARN seul. Pour cela, j'ai déterminé les conditions optimales de renaturation de l'ARN et j'ai tenté de cristalliser l'ARN de la RNase P de *Thermotoga maritima*. J'ai également tenté de cristalliser deux domaines d'ARN qui ont été montrés indépendants du point de vue structural (26), le domaine catalytique et le domaine de fixation du substrat. Ces expériences ont permis d'obtenir des quasi-cristaux du domaine catalytique qui n'ont pas pu être améliorés pendant le temps de mon stage. Depuis lors plusieurs structures cristallographiques de l'ARN de la RNase P ont été obtenues dans d'autres équipes (27-29).

2.2.2 Caractérisation d'un motif universel de reconnaissance ARN/ARN

Lors de mon stage post-doctoral, j'ai également participé à un autre projet qui consistait à déterminer les contributions énergétiques de mutations introduites dans un motif d'interaction tertiaire dans le domaine P4P6 de l'intron de *Tetrahymena thermophila*. Ce motif est constitué par l'interaction de deux A en simple brin dans le sillon peu profond de deux paires G=C. J'ai, d'une part, proposé des modèles de structures des différents mutants afin de vérifier comment les mutations se répercutaient sur la structure. D'autre part, le motif d'interaction étudié ayant été trouvé de nombreuses fois dans la structure cristallographique de la sous unité 50S de *Haloarcula marismortui* (30), j'ai procédé à une analyse de séquences qui a permis de vérifier l'accord entre phylogénie et données thermodynamiques (31). L'analyse des motifs trouvés dans l'ARN 23S montre que les liaisons hydrogène ne sont pas toujours optimisées, c'est-à-dire qu'elles ne se forment pas toutes simultanément. Au contraire l'enfouissement des surfaces accessibles confère en majorité sa stabilité au motif. En effet, alors qu'une interaction triple aboutit généralement à enfouir 150 \AA^3 , l'interaction triple observée dans l'intron de *Tetrahymena thermophila* impliquant A184 permet d'enfouir 195 \AA^3 (Figure 7) Il a fallu tout d'abord rechercher systématiquement ces motifs dans la structure cristallographique de la sous-unité 50S pour procéder ensuite à une analyse de séquence dirigée. Ceci parce que les liaisons hydrogène observées dans ce motif font intervenir les groupements 2' hydroxyles des riboses d'une part et les groupements O2/N3 des Y/R qui sont des éléments communs à tous les nucléotides. Ce motif ne génère pas de signal détectable dans les séquences ce qui explique que son existence n'a pu être révélée que par

des structures cristallographiques d'ARN adoptant une architecture complexe. Les données ont alors été moyennées afin d'augmenter le signal tout en diminuant le bruit de fond et d'extraire les valeurs suivantes pour A184 qui implique la plus grande surface enfouie, GC 69% ; CG 6% ; UA 3% ; AU 3% (Figure 7). Nous expliquons donc d'une part, comment un motif ubiquitaire d'interaction entre sillon peu-profond et adénosines rend compte de la forte proportion de A trouvée dans les régions simple-brins. D'autre part, nous expliquons aussi pourquoi ce motif n'est pas identifiable par analyse de séquences.

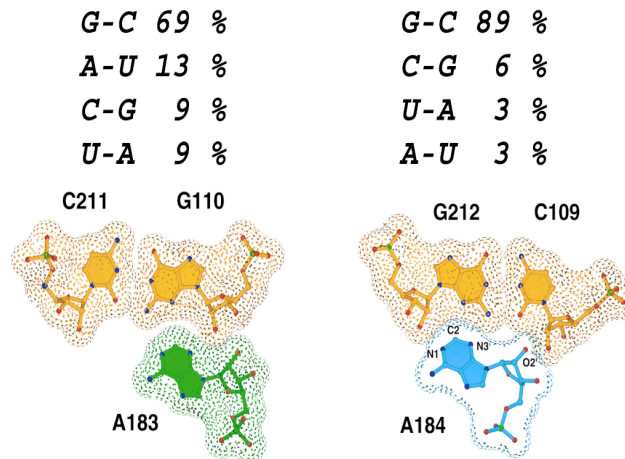


Figure 7 : Représentation des interactions triples de l'intron de groupe I de *Tetrahymena* étudiées par Doherty et collaborateurs. Les paires de bases G=C adjacentes sont représentées avec l'adénine qui interagit spécifiquement avec chacune d'elle, respectivement A183 (vert) et A184 (bleu). Les surfaces de Van der Waals sont représentées pour illustrer la complémentarité de ces interactions. L'analyse de séquences montre qu'alors que la base en interaction dans le sillon peu profond est une A dans 95 % des cas, les préférences phylogénétiques pour les paires G=C sont particulièrement marquées dans ce type d'interaction.

2.3 Travaux en temps que chargé de recherches

De retour des USA, j'ai réalisé un stage post-doctoral d'une année (octobre 2000-septembre 2001) dans mon laboratoire de thèse où j'ai initié une partie des travaux décrits ci-après. J'y ai ensuite été recruté en temps que chargé de recherche stagiaire en octobre 2001, puis titularisé en 2002. Mon activité de recherche se décompose en deux parties. D'une part l'étude théorique de l'architecture des ARN par modélisation moléculaire et d'autre part des approches expérimentales qui visent à mieux comprendre comment les ARN se replient et se structurent. Avec Bertrand Beckert (étudiant en thèse dans notre laboratoire), nous avons également pour projet de contribuer à identifier de nouveaux ARN fonctionnels dans les génomes de certains procaryotes.

J'évoquerai tout d'abord diverses études de modélisation moléculaire réalisées depuis mon retour des USA dans le cadre de cinq collaborations (voire § 2.4). Typiquement, ces collaborations sont l'initiative de laboratoires qui souhaitent voir rationaliser leurs données biochimiques de manière à orienter la suite de leurs recherches. La construction d'un modèle permet aux biochimistes de représenter leurs résultats sur une forme de l'ARN étudié proche de sa structure en solution. En effet, les travaux du laboratoire montrent que la modélisation permet de déduire correctement l'architecture et la topologie des ARN structurés même si localement des hypothèses simplificatrices doivent être faites quand une région souffre d'un déficit de caractérisation. Ces études formalisent les problèmes de repliement et d'architecture, et nous permettent donc de poser les bases des règles qui les gouvernent. La formalisation de ces règles est développée par Eric Westhof en partenariat avec le professeur Neocles Leontis de l'université de Bowling Green (Ohio, USA). Ainsi les modèles de deux introns de groupe I ont été construits en collaboration avec les équipes de Renée Schroeder de l'Université de Vienne et de Sarah Woodson de l'université de Baltimore, Maryland, USA. Nous avons poursuivi cette dernière collaboration pour tenter de mieux comprendre le repliement de l'intron dans son contexte biologique, c'est-à-dire inséré au milieu de la boucle anticodon de l'ARN de transfert isoleucine de la cyanobactérie *Azoarcus*. Des modèles d'holoenzymes des RNase P d'*Escherichia coli* et de *Bacillus subtilis* complexés à leur substrat ont été proposés en collaboration avec l'équipe de Venkat Gopalan de l'université de Columbus, Ohio, USA. Enfin deux modèles d'opérateurs d'ARN messenger de synthétases ont

été réalisés en collaboration avec l'équipe de Pascale Romby et de Magali Frugier au sein de l'UPR 9002. Ces travaux ont été publiés entre fin 2002 et 2005.

Une autre facette de mon activité concerne la cristallographie d'ARN (voire §2.1.3 et §2.5) et les expériences biophysiques et biochimiques visant à caractériser sa structure et leur repliement (voire §2.6). Lors de ma thèse, les ARN synthétiques proposés dans le commerce étaient chers et peu fiables. J'avais donc initié la synthèse d'ARN basée sur la chimie des phosphoramidites au laboratoire afin de donner accès aux chercheurs de l'IBMC à des séquences courtes d'ARN ne présentant pas d'hétérogénéité en 5' ou en 3' ce qui n'est en général pas le cas avec les transcrits obtenus avec la T7-ARN-polymérase (32,33). La possibilité de synthétiser un grand choix de molécules au laboratoire m'a permis de tester de nombreuses séquences et de résoudre la structure d'un ARN nonamérique (20). De retour de mon séjour post-doctoral aux USA, j'ai poursuivi des recherches basées sur la cristallisation de courtes séquences d'ARN mais sans plus les synthétiser moi-même puisque l'entreprise Dharmacon commençait à proposer des synthèses très fiables basées sur une nouvelle chimie (chimie 5'-silyl-2'-acetoxy-ethyl-orthoester : 2'-ACE™). J'ai réussi à cristalliser plusieurs de ces séquences. Deux d'entre elles m'ont permis de collecter des jeux de données au synchrotron. Pour l'un des ARN correspondant à la boucle interne du domaine III de l'IRES du VHC (34), les données étaient correctes sur une majeure partie de l'espace réciproque, mais une zone d'environ 8° n'était pas enregistrable ce qui a empêché l'obtention de jeux complets. Le second ARN consistait en une double hélice intégrant un nucléotide dérivé du noyau acridine (35). Il semble que l'ARN cristallisé ne correspondait pas à la totalité de l'ARN utilisé pour confectionner les gouttes d'essais de cristallisation et aucun modèle de remplacement moléculaire ne s'est révélé satisfaisant. J'ai également contribué à affiner des structures de complexes formés entre le site A de l'ARN 16S bactérien et plusieurs antibiotiques (36).

L'activité cristallographique est complétée par une expertise biophysique et biochimique m'amenant à manipuler les techniques de fonte UV, d'électrophorèse native et dénaturantes d'ARN et les techniques de cartographie en solution. Mon expertise en modélisation moléculaire renforce ma capacité à exploiter les résultats obtenus par ces techniques. J'ai donc développé une collaboration avec Olivier Chaloin (UPR9021, ICT) pour étudier la structure du complexe formé entre l'ARN TAR et la protéine Tat chez HIV. Tat peut être synthétisé en phase solide en deux morceaux qui sont ensuite reliés (37). Ce

laboratoire est un des seuls à utiliser cette technique ce qui confère un avantage dans ce projet en nous permettant de disposer de quantité de protéines conciliables avec des expériences de cristallisation. En premier lieu, nous caractérisons le complexe en solution par résonance plasmonique de surface (RPS) de manière à concevoir des constructions qui nous amèneront à avoir un complexe stable. Par ailleurs, j'ai été contacté par une équipe de l'EMBL (Michael Sattler, groupe de RMN) pour obtenir des informations structurales sur un ARN viral présentant la faculté de reconnaître une protéine de l'hôte impliquée dans l'export nucléocytoplasmique, par modélisation et cartographie en solution. Cette collaboration a constitué un axe majeur de mon travail de ces derniers mois.

2.4 Résumé des travaux de modélisation moléculaire des introns de groupe I et de la RNase P procaryote et de deux opérateurs d'ARNt synthétase

2.4.1 Modélisation des introns de groupe I

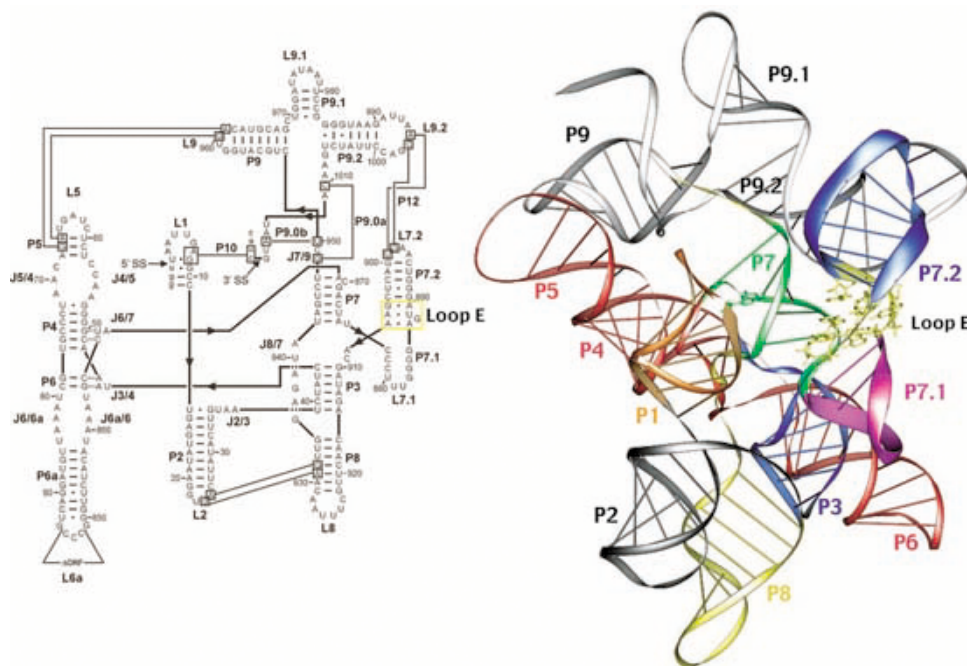


Figure 8 : Structures secondaire (gauche) et tridimensionnelle (droite) de l'intron *td* du bactériophage T4. Les hélices sont numérotés de P1 à P10, les boucles de L1 à L9.2 et les jonctions de J2/3 à J8/9. Les résidus formant le motif boucle E sont en jaune sur le modèle 3D et encadré en jaune sur le diagramme de structure secondaire. On distingue l'interaction entre P3 et J6/6a en bleu et rouge.

Dans le cadre de deux collaborations, les structures des introns de groupe I de *td* du bactériophage T4 (Renée Schroeder, Vienne, Autriche) et de *Azoarcus* (Sarah Woodson, Baltimore, MD, USA) ont été modélisés. Il est intéressant de noter qu'au moment où ces modèles ont été construits, la seule structure cristallographique d'intron disponible (38) ne contenait qu'une partie de l'ARN et ne permettait pas de concilier les données connues publiées. Le premier modèle a consisté à affiner le modèle de (39) à la lumière de nouvelles

données qui ont abouti à identifier un motif boucle E dans l'extension P7.1/2. L'analyse des modes de reconnaissance d'hélices par des boucles E observés dans les structures cristallographiques des sous-unités ribosomiques (30,40) et du ribozyme en épingle à cheveux (13) ont permis d'élucider la conformation que l'extension P7.1/2 adopte pour s'amarrer à l'hélice P7 (Figure 8). L'interaction P12 a contribué de manière importante au choix de l'orientation de l'extension P7.1/2. Par ailleurs, l'exploitation de la structure de la sous-unité ribosomique 50S a permis de mettre en évidence un motif d'interaction non-spécifique entre le sillon peu-profond de l'hélice P3 et la jonction J6/6a. L'analyse des séquences et des modèles publiés a permis de vérifier que la majorité des introns semble posséder ce motif d'interaction qui est décrit dans cette étude pour la première fois (41). Cette interaction se retrouve donc dans tous les introns suivant des variantes structurales qui vont parfaitement dans le sens d'interactions aspécifiques dépendantes du contexte structural dans lequel elles se produisent, comme dans l'interaction entre le renflement riche en A et l'hélice P4 de l'intron de *Tetrahymena thermophila* (31). Seul l'angle entre les deux hélices (45°) en contact semble être caractéristique de l'interaction. La structure cristallographique de l'intron du bactériophage *Twort* (42) montre que notre analyse a mené à des résultats corrects.

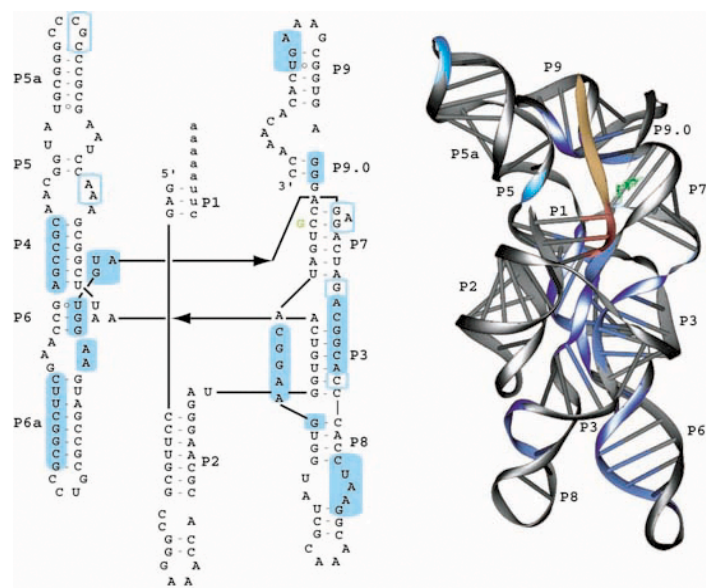


Figure 9 : Structures secondaire et tridimensionnelle de l'intron de *Azorarcus*. Les éléments de structure secondaire sont numérotés comme à la figure 1. Les résidus protégés des attaques des radicaux hydroxyles sont indiqués en bleu. Les formes géométriques évidées représentent les protections faibles et sont matérialisées en cyan sur le modèle 3D et les formes pleines correspondent aux protections fortes et sont en bleu foncé sur le modèle 3D.

Le second modèle a été réalisé *ab initio* en s'appuyant sur les modèles d'introns existant (43,44) et sur les données de cartographie au Fe-EDTA de nos collaborateurs. Les résultats apportés par ces expériences différencient les résidus accessibles au solvant de ceux

qui en sont protégés. Le modèle (45) permet de proposer une architecture qui rend compte de ces différences en vérifiant que l'accessibilités théoriques des résidus est corrélées à l'intensité des coupures chimiques observées (Figure 9).

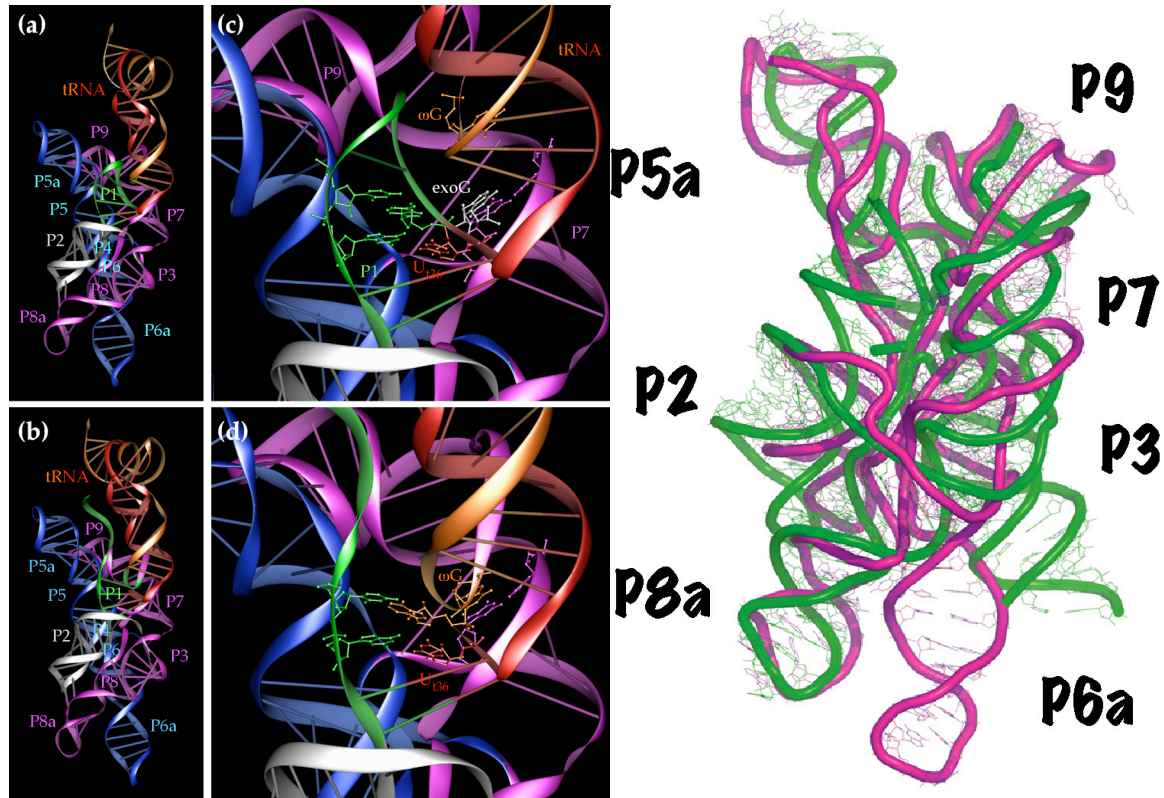


Figure 10 : (panneau gauche) Les différents étapes catalytiques de l'intron présent dans le pré-ARNt^{Ile} d'*Azoarcus*. Modèle de l'hydrolyse de l'exon 5' lors de l'étape 1 (a) et de l'hydrolyse de l'exon 3' lors de l'étape 2 (b). Dans l'étape 1, P1 (en vert) forme une épingle à cheveux. Le cofacteur exoG aborde le grand sillon de P7 et son groupement 3' hydroxyl attaque la liaison scissile entre la base U36 de l'ARNt (rouge) et A1 de l'intron (c). Lors de la 2^e étape (d), l'exoG a été remplacé par la dernière base de l'intron ωG et U36 attaque la liaison entre ωG et A37 ce qui aboutira à la ligature des deux exons. (panneau droit) Superposition du modèle (magenta) de l'étape 2 de l'intron d'*Azoarcus* avec la structure cristallographique (vert). L'ARN de transfert a été omis pour clarifier la vue.

Ce travail a été poursuivi au cours de l'année 2003 par la modélisation de l'intron inséré dans la boucle de l'ARN de transfert isoleucine (46), le pré-ARNt^{Ile}. Des données de cartographie ont été accumulées sur le pré-ARNt^{Ile}. Elles ont permis de déduire des modèles de structure pour chaque étape catalytique. Ils rendent parfaitement compte de la plasticité du ribozyme qui doit d'abord hydrolyser l'exon 5' puis l'exon 3' et enfin ligaturer les exons entre eux. Les nouvelles contraintes imposées à P1 ont permis de mieux comprendre pourquoi le ribozyme se replie plus lentement dans son contexte biologique que dans le cadre de la réaction d'hydrolyse étudiée dans les travaux précédents (45). La réponse tient principalement

dans le fait que l'hélice de l'anticodon (et globalement l'ARN de transfert) conserve sa structure dans le ribozyme natif et permet à P1 de se former plus facilement (Figure 10). Nous avons conclu que le repliement de l'intron et de l'ARNt était couplé via la réorganisation de la structure secondaire. Il semblerait que ceci puisse être lié au besoin de régulation du mécanisme d'épissage dans la cellule. Depuis, la structure cristallographique de l'intron présent dans le pré-ARNt^{lle} d'*Azoarcus* a été publiée (47). Il en ressort que les contraintes imposées par les expériences de cartographie en solution ont permis de déduire correctement l'architecture de l'intron. En effet, l'écart quadratique moyen normalisé calculé sur les atomes du squelette ribose-phosphate de P2 à P9 est inférieur à 4 Å.

2.4.2 Modélisation de l'holoenzyme de la RNase P procaryote

Les modèles représentatifs des deux sous-types d'holoenzymes bactériens (*Escherichia coli* et *Bacillus subtilis*) ont été construits sur la base des modèles correspondant des sous-unités ribonucléiques de la RNase P (48) en collaboration avec l'équipe de Venkat Gopalan de l'université de Columbus, OH, USA qui apportait des données obtenues en solution. Les expériences réalisées par nos collaborateurs consistent à greffer chimiquement à un résidu cystéine de la protéine de la RNase P (C5 chez *E. coli*) un groupement Fe-EDTA qui induit l'hydrolyse de l'ARN dans son environnement proche par réaction radicalaire (49). Plusieurs protéines recombinantes contenant une cystéine à diverses positions sont ainsi construites afin de cartographier l'environnement de diverses zones de la protéine. Ce type d'expérience offre une fiabilité statistique importante par rapport aux méthodes de pontages ou d'empreintes chimiques (50). Ainsi l'environnement proche de trois zones distinctes de la protéine a été analysé, délimitant un plan qui a permis l'ancrage de la protéine à la sous-unité ARN dont le modèle a dû subir des modifications pour accommoder la sous-unité protéique. Ces modifications ont consisté à générer une poche dans laquelle la sous-unité protéique - dix fois plus petite que l'ARN - pouvait se loger. Pour cela l'architecture du bloc contenant les hélices P2, P3, P5, P6, P15, P16 et P18 a dû être totalement remanié. De manière frappante, les adénines 248 et 249 restent au contact du pré-ARNt rendant très plausible leur implication au niveau de la réaction d'hydrolyse de la séquence 5'-leader du substrat. Les éléments de structure secondaires conservés de l'ARN et la protéine dans les modèles des holoenzymes d'*Escherichia coli* et de *Bacillus subtilis* présentent des positions similaires (51). Cette

collaboration est toujours en cours puisque nous modélisons actuellement la RNase P archaéale.

2.4.3 Modélisation des opérateurs de deux ARNt-synthétases (Thréonyl et Aspartyl)

L'opérateur de la thréonyl-ARNt-synthétase (ThrRS) est une région de l'extrémité 5' non-traduite de l'ARNm de la ThrRS qui permet le contrôle de la traduction du gène. Cette région de l'ARNm se replie en formant deux tiges-boucles (domaines 2 et 4) qui miment la région anticodon de l'ARNt et sont reliées par une région simple-brin de 23 nucléotides (52).

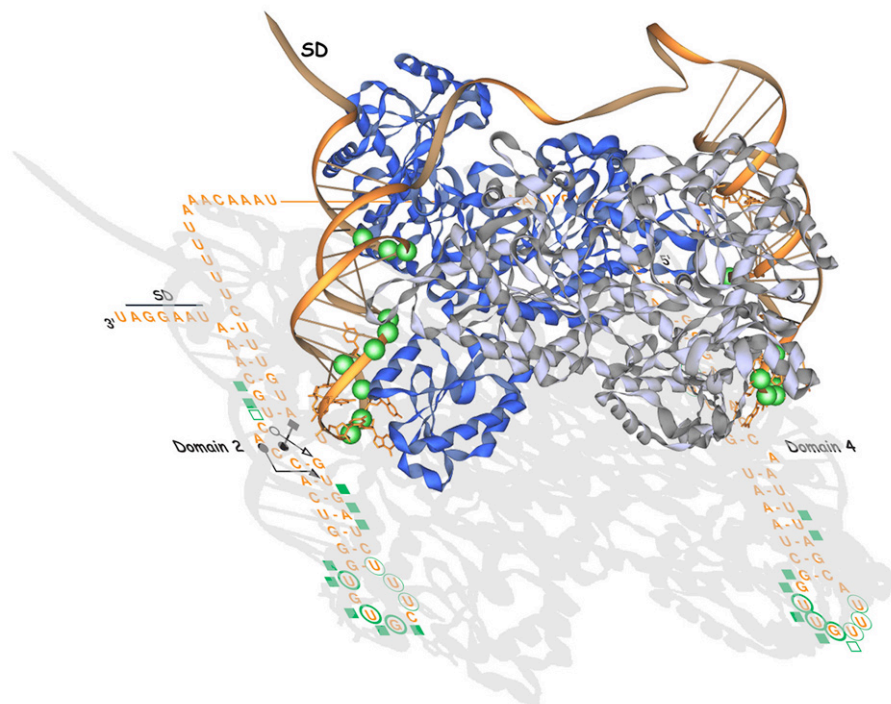


Figure 11 : Modèle du complexe formé par l'interaction de la région opératrice de l'ARNm de la thréonyl-ARNt synthétase avec l'enzyme dimérique. Le modèle structural est superposé à la structure secondaire de l'opérateur (orange). Les deux sous-unités protéiques sont dessinées en bleu et gris. Les deux boucles apicales miment le structure de la boucle anticodon de l'ARNt. Toutefois la boucle du domaine 2 est exactement identique à l'anticodon de l'ARNt (anticodon CGU) alors que la boucle du domaine 4 est légèrement différente sans doute pour permettre aux ARNt de pouvoir déplacer avec aisance l'opérateur.

Ces boucles ont la faculté de se lier à l'enzyme qu'elle code, contrôlant ainsi négativement sa propre expression. Nous avons établi un modèle de structure tridimensionnel du complexe formé entre le dimère cristallographique de la synthétase et la région 5' de l'ARNm en nous appuyant sur les structures cristallographiques des complexes ARNt/synthétase (53). Dans le temps de notre étude, la structure cristallographique du complexe entre la ThrRS et le domaine 2 de l'opérateur a été résolu (54) montrant une

excellente corrélation avec le modèle. Cette étude a permis de rationaliser les données abondantes sur ce complexe en reliant notamment les différences structurales aux différences d'affinités pour la protéine des deux domaines (55). Le modèle permet également de mieux comprendre comment l'ARNm reconnaît à l'aide de deux domaines *cis* en épingle à cheveux différents les deux sites de reconnaissance de l'ARNt (Figure 11). Nous avons donc pu formuler des hypothèses structurales pour expliquer l'orchestration de la compétition entre l'ARNt et l'ARNm pour lier l'enzyme. Le modèle montre également que la formation du complexe entre le dimère de synthétase et l'opérateur du messenger masque la séquence Shine-Dalgarno au ribosome permettant de proposer un mécanisme d'inhibition de la traduction basé sur la compétition entre le ribosome et la synthétase pour une même région de l'ARNm.

L'opérateur de l'aspartyl-ARNt-synthétase fonctionne de manière assez similaire chez la levure (56) bien que présentant une structure plus complexe dans laquelle l'extrémité 3' de l'opérateur interagit avec l'extrémité 5'. Nous avons élaboré un modèle de structure tridimensionnelle en nous basant sur les résultats obtenus en solution par l'équipe de Magali Frugier (IBMC-CNRS UPR 9002).

2.5 Contribution cristallographique à l'étude du système Site-A/aminoglycosides

Le site A de l'ARN ribosomique 16S bactérien reconnaît l' aminoacyl-ARNt en formant une hélice de trois paires de bases entre l'anticodon et le codon correspondant de l'ARNm. Le site A a été observé pour la première fois par cristallographie au sein de la sous-unité 30S de *Thermus thermophilus* (40). Un faisceau d'expériences de microscopie électronique(10,57), de cristallographie (58), de spectroscopie FRET (59) et de cinétique (60) converge pour voir le site A ribosomique comme un interrupteur moléculaire dont le rôle est de rapidement reconnaître si un ARNt correspond au codon présenté à l'instant t . Pour ce faire, deux adénines en bulge de l'hélice 44 (A1492 et A1493 ; Figure 11) forment un motif en A-mineur (31,61) dans le sillon peu-profond de l'hélice codon-anticodon si celle-ci est formée de paires de bases Watson-Crick. Ces deux adénines restent séquestrées dans l'hélice dans le cas contraire (62). Ceci semble déclencher ou non l'hydrolyse du GTP dans un mécanisme communément appelé proofreading.

Le site A constitue donc une cible logique pour des antibiotiques naturels tels que les aminoglycosides produits par certains champignons (Figure 12). Un antibiotique tel que la paromomycine se fixe au niveau du site A dans la cavité laissée libre par le basculement des deux adénines en bulge en direction de l'hélice codon-anticodon. Cette cavité est bordée d'une adénine qui reste empilée dans l'hélice et qui reconnaît le cycle I de l'antibiotique qui se trouve alors intercalé entre les plateaux de bases adjacents. Les trois autres cycles interagissent dans le sillon profond en direction de la paire de base U-U caractéristique de tous les sites A connus. Notre laboratoire a acquis une bonne expertise dans la résolution de structures cristallographiques de complexes entre un court ARN mimant le site A (63) et des aminoglycosides. Dans ce projet, ma contribution a consisté à faire bénéficier les étudiants Quentin Vicens et Boris François de mon expertise expérimentale et théorique. Ces dernières années, une quinzaine de structures cristallographiques du site A en présence de divers antibiotiques ont été résolues dans l'équipe (36,64-70).

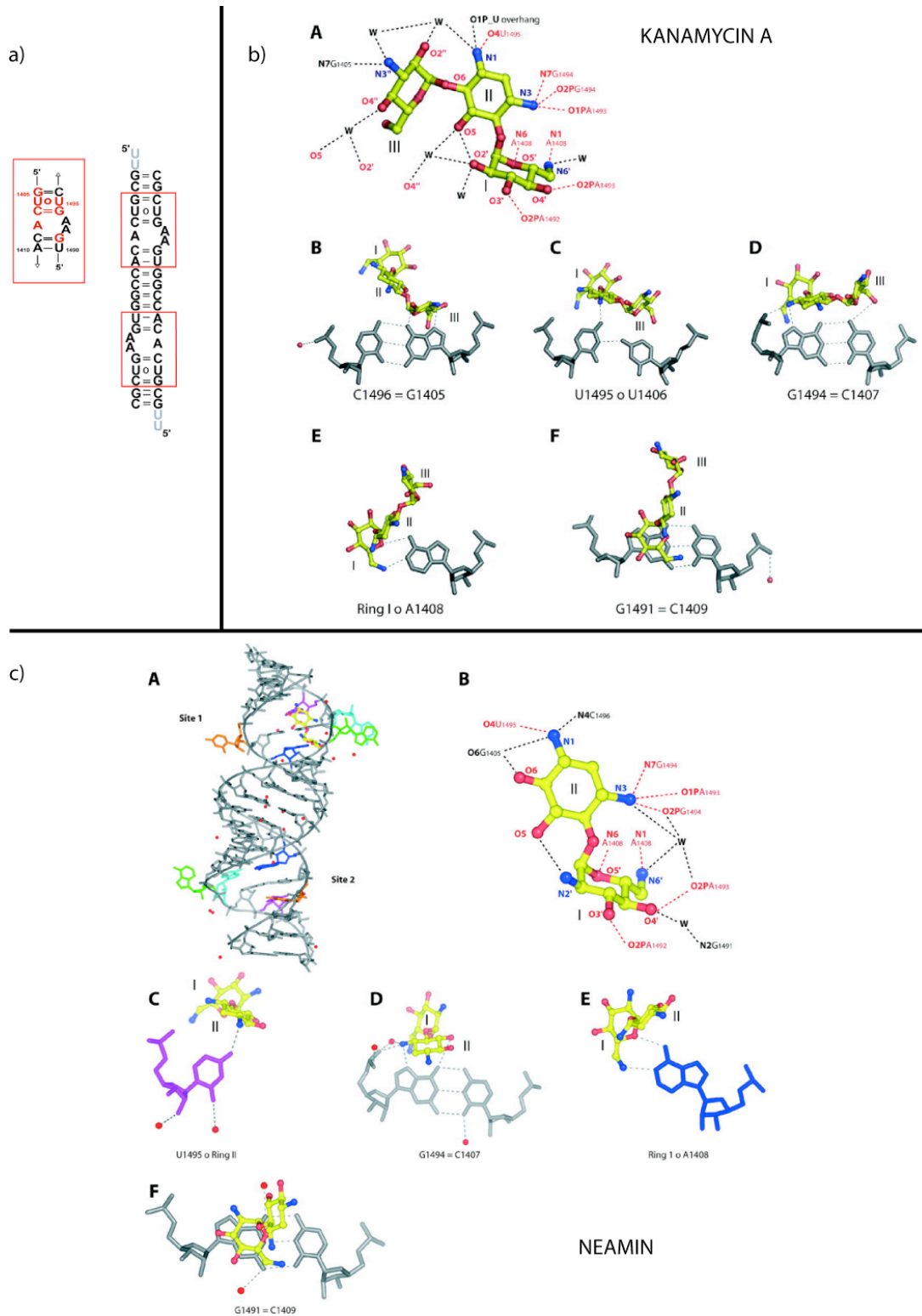


Figure 12 : (a) Le site A extrait du contexte de l'ARN 16S et un des ARN utilisé pour cristalliser le site A en présence d'antibiotiques. Cet ARN autocomplémentaire permet de construire deux sites A dans le même duplex (64). Deux structures d'ARN complexés à la kanamycine A (b) et à la néamine (c) pour lesquelles j'ai procédé à l'affinement cristallographique (36). (bA) vue globale des interactions entre l'antibiotique et l'ARN, suivie de représentations par plateau de bases successifs en commençant du côté distal de A1408. (c) idem à (b) mais (cA) représente une vue globale du duplex tel qu'il apparaît dans la structure cristallographique en contact avec une seule molécule de néamine. Cette structure présente donc les deux états théoriquement rencontrés dans le ribosome en phase d'élongation.

2.6 Modélisation et cartographie de la région de l'élément constitutif de transport (CTE) de l'ARN du virus de Mason-Pfizer

Ce projet collaboratif consiste à caractériser l'interaction ARN-protéine observée entre une région de l'ARN génomique du rétrovirus simien de Mason-Pfizer (MPMV) et un facteur cellulaire de l'hôte, la protéine Tap (71). La région étudiée de l'ARN viral (CTE : Constitutive Transport Element) est en effet constitutivement reconnue par ce facteur qui est impliqué dans l'export du noyau vers le cytoplasme, étape requise pour la réplication rétrovirale qui repose sur l'expression des gènes Gag/Pol. Ce mécanisme n'est pas seulement important pour la réplication de ce virus mais également pour l'export d'ARNm qui doivent sortir du noyau muni de leur intron. C'est par exemple le cas de l'ARNm de Tap qui possède lui-même entre ses exons 10 et 11 un intron de 1800 nucléotides qui contient un CTE fonctionnel homologue au CTE de l'ARN viral du MPMV (72).

Mes collaborateurs, Michael Sattler et Elena Conti ont accumulé des données de cristallographie et de RMN sur la protéine libre (73) et sous forme complexée à l'ARN. Par contre, la construction d'ARN utilisée n'a été caractérisée sous forme libre et liée que par diffusion des neutrons aux petits angles (Small Angle Neutron Scattering : SANS) et demande donc une caractérisation plus fine au niveau structural. Notamment, nous souhaitons identifier les régions de l'ARN interagissant avec la protéine. J'ai donc été contacté pour construire dans un premier temps un modèle tridimensionnel que j'ai affiné en réalisant des expériences de cartographie enzymatique en solution.

Le modèle a été construit pour vérifier s'il pouvait s'inscrire dans l'enveloppe SANS obtenue pour l'ARN isolé et également pour vérifier s'il permettait d'interpréter les enveloppes SANS obtenues en présence du complexe ARN-protéine. J'ai ainsi pu mettre en évidence une courbure d'environ 20° qui semble due à la structure de la grande boucle interne de l'ARN. Cette dernière très riche en purines semble particulièrement structurée puisque l'enveloppe obtenue par SANS est bien définie (Figure 13). Par contre, sa séquence semble clairement exclure le fait qu'elle adopte une structure en boucle E. L'architecture de cette boucle présente donc un empilement optimisé par la construction d'un tandem de paires de bases G-A impliquant la face sucre de la G et la face Hoogsteen de la A. A ce tandem s'ajoute une paire A-C établie entre A(N3) et C(N4). Les trois nucléotides résiduels présents de

chaque coté de cet empilement confèrent au modèle la flexibilité nécessaire pour correctement s'inscrire dans l'enveloppe SANS tout en conservant de bonnes interactions d'empilement.

Bien que le modèle ait permis d'interpréter les données SANS obtenues avec l'ARN isolé, les données sur le complexe ne pouvaient pas l'être à cause du manque d'information sur la région de l'ARN en interaction avec la protéine. En effet, cette situation aboutissait à générer de trop nombreux modèles du complexe ARN-protéine, chacun devant être confronté aux données SANS. Le trop grand nombre de degrés de liberté du système compromettait alors l'obtention de résultats probants. Nous avons donc décidé qu'il convenait de caractériser la région de l'ARN en interaction avec la protéine en réalisant des expériences d'empreintes en solution (1). Les expériences menées avec les RNase T1 et V1 ont permis de vérifier dans un premier temps, que les données obtenues sur l'ARN isolé étaient conformes à celles déjà obtenues par une autre équipe (74). Les expériences d'empreintes en présence de Tap dans les conditions utilisées pour l'ARN isolé ont permis de mettre en évidence des changements fins dans la structure de l'ARN et plusieurs régions du CTE potentiellement en interaction avec Tap.

Les résultats de cartographie montre que la boucle interne riche en purines est très structurée dans l'ARN isolé (Figure 13). En effet des coupures V1 d'intensités variées sont observées sur chacun de ses brins. Le brin le plus touché est le brin 3' (entre C38 et A50). Le brin 5' montre également des coupures toutefois moins intenses (en 5' et en 3' de A14). Ce profil est modifié lorsque la protéine est ajoutée. D'une part, les clivages en 5' de A14 disparaissent et d'autre part toute la région comprise entre les résidus 36 et 50 subit une importante baisse d'intensité de clivage. Ceci va de pair avec l'intensification des coupures dans la région hélicoïdale proximale de la boucle apicale. Nous observons même un glissement de cette zone vers la boucle apicale ce qui tend à montrer que l'adjonction de la protéine aboutit à une meilleure structuration de la boucle apicale sans doute accompagnée par la formation de l'appariement C24-U36. Ces observations se confirment à l'analyse des données de clivage à la RNase T1 puisque C38 devient protégée et que les coupures dans la boucle s'intensifient également suite à l'ajout de Tap au CTE. Les nucléotides 11 à 13 ont été montrés importants dans cette interaction par des expériences dans un système triple-hybride (75). La matérialisation des résidus révélés impliqués dans l'interaction avec la protéine permet de mettre en évidence deux régions de l'ARN qui se concentrent d'un même coté du modèle. Ceci confirme que la grande boucle interne a un fort caractère structurant et

que la protéine tap semble interagir avec le sillon peu profond d'un côté de l'ARN. Cette étude permet donc pour la première fois de décrire l'interaction d'un CTE avec la protéine tap. Les études de cette interaction avec d'autres ARN permettront de vérifier ces conclusions.

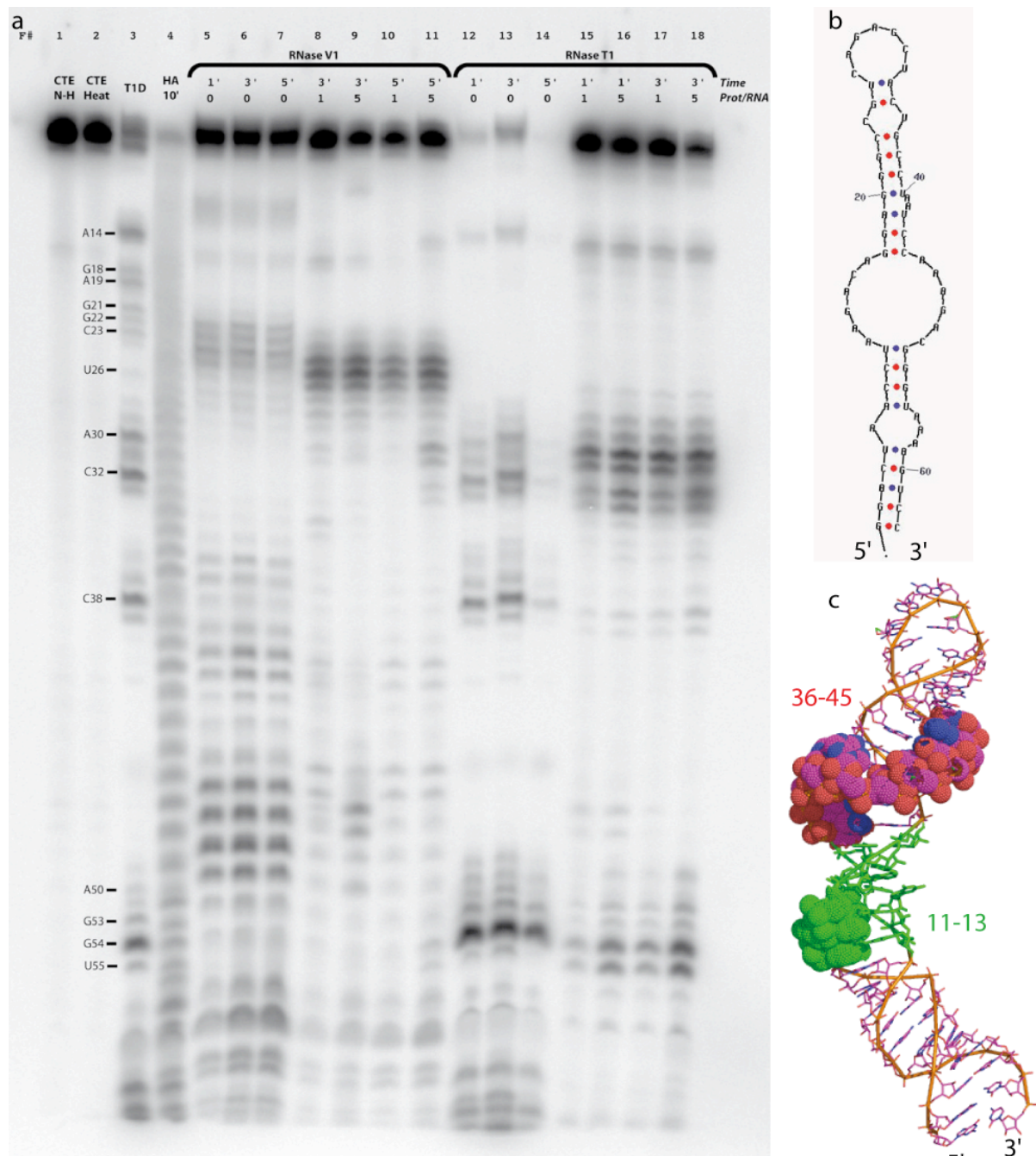


Figure 13 : Modélisation et cartographie du CTE du MPMV. (a) Structure secondaire de l'ARN CTE obtenue par MFOLD (76). (b) Gel de cartographie enzymatique aux RNases V1 et T1. (c) Modèle structural du CTE montrant la localisation de la boucle interne riche en purines (vert) et les résidus protégés par la protéine Tap de l'action de la RNase V1. A noter que les deux jeux de résidus se situent majoritairement du même côté du modèle.

3 Projets

Les ARN que j'ai étudié jusqu'à présent sont impliqués directement (ARN ribosomiques) ou indirectement (RNase P) dans la traduction ou dans la régulation transcriptionnelle (introns, ARN TAR de HIV) et traductionnelle (opérateurs présents dans les 5'-UTR). L'ensemble de mes travaux contribue à montrer que la structure des ARN joue un rôle primordial dans les mécanismes de traduction et de contrôle de l'expression génétique.

Ainsi, l'ensemble des motifs structuraux d'ARN semble récurrent. Ces motifs peuvent être retrouvés dans des ARN distincts pour être exploités suivant des caractéristiques diverses pour finalement présenter des fonctions différentes. Le motif A-minor (61) par exemple, est utilisé par l'intron de *Tetrahymena thermophila* pour favoriser le repliement du domaine P4-P6 (31) et il est également utilisé par le site A du ribosome pour vérifier la complémentarité de l'interaction codon-anticodon (62). Dans le premier cas, ce motif a un rôle structural alors que ses propriétés dynamiques sont mises à profit dans le deuxième exemple. De manière intéressante, l'introduction d'un grain de sable comme un aminoglycoside perturbe la dynamique du site A sans modifier sa structure et favorise les erreurs d'incorporation d'acides-amino pendant la traduction (4). Le motif boucle E qui peut être caractérisé par une succession particulière d'appariements non-canoniques (77) est capable de reconnaître la protéine ribosomique L25 (78) ou bien d'interagir en *cis* dans l'intron *td* du phage T4 (41) en fonction du contexte structural dans lequel ce motif est placé. De nombreux exemples de ce type pourraient être décrits ici.

Il est donc frappant de constater que, le brassage de motifs d'ARN récurrent intervient de manière à aboutir à la construction d'édifices moléculaires qui présentent une diversité de fonctions extrêmement étendue. Cette propriété de l'ARN véhiculée par la séquence stockée sous forme d'ADN pour être transcrite précisément en temps et en heure en fait jouer à cette molécule un rôle central dans de toujours plus nombreux mécanismes. Le fil conducteur de mon travail consiste à décortiquer la structure des ARN par des approches théoriques et expérimentales pour mettre en évidence les motifs récurrents qui les composent ainsi que leurs fonctions. Des questions fondamentales découlent de ce travail. Comment les séquences d'ARN évoluent-elles pour permettre le brassage de motifs structuraux ? Comment ces motifs sont-ils utilisés ? Ceci dans le but d'identifier des ARN encore inconnus dans les bases de données génomiques.

Mes travaux de thèse et de stages post-doctoraux, en couplant la modélisation moléculaire et la cristallographie, ont contribué à affiner la méthode de modélisation en développant des critères de comparaison de motifs structuraux d'ARN qui ont posé les jalons de règles de construction systématiques (5). Ces règles sont sans cesse affinées et développées dans l'équipe en particulier récemment par Neocles Leontis (79) (Professeur à l'Université de Bowling-green) et Aurélie Lescoute (80,81) (étudiante que j'ai encadrée en DEA). Mes travaux de thèse ont également permis de développer la cristallographie puisque j'y ai résolu la première structure cristallographique d'ARN au sein de notre équipe et formé l'étudiant qui a résolu les premières structures de complexes entre aminoglycosides et site de décodage. Mon projet de recherche s'inscrit donc résolument dans l'étude de la structure des ARN.

Mon projet de recherches s'organise autour de deux projets majeurs. Eric Westhof m'a encouragé à prendre en charge la poursuite d'une collaboration avec les équipes de Henrik Nielsen (Université de Copenhague, Danemark) et Steinar Johansen (Université de Tromsø, Norvège) pour étudier la structure d'un ribozyme siamois qui permet la régulation de l'expression du gène codant pour une endonucléase voyageuse (Homing Endonuclease) chez le protiste *Didymium iridis* (§3.3). Ce système est constitué d'un intron de groupe I dans lequel est insérée une phase de lecture ouverte codant pour l'endonucléase ainsi qu'un ribozyme dérivé d'un intron de groupe I mais qui catalyse une seule coupure aboutissant à la formation d'un lasso protecteur de quatre nucléotides à l'extrémité 5' du messager (82). Les réactions de ces ribozymes peuvent s'enchaîner dans des ordres différents en fonction des conditions physiologiques ou même *in vitro*, indiquant des mécanismes de régulation fins dont la compréhension passe par l'étude de la structure.

L'équipe a également une volonté forte de participer à l'exploration du monde de l'ARN tel que les découvertes récentes sur la régulation ARN-dépendante des gènes tant chez les eucaryotes (83) que chez les bactéries (84,85) le présente. Imaginer que de nombreux ARN sont encore inconnus tant au niveau structural que fonctionnel est forcément passionnant pour le scientifique qui étudie ces molécules depuis plusieurs années. C'est pourquoi je souhaite que mon projet de recherche contribue à identifier et à caractériser de nouveaux ARN qui pourraient être impliqués dans la régulation des gènes chez des organismes pathogènes comme *Staphylococcus aureus* et *Listeria monocytogenes* dans le cadre de collaborations avec des équipes de recherches spécialisées dans la microbiologie de ces organismes (86,87).

J'ai également développé deux projets collaboratifs plus ponctuels au sein même de l'IBMC. Le premier d'entre eux est un projet d'étude de la structure du complexe entre l'ARN TAR (Trans-Activation Responsive region) et la protéine Tat du VIH-1 avec l'UPR 9021 ICT (Immunochimie et Chimie Thérapeutiques, IBMC ; voir §2.3). Olivier Chaloin, ingénieur de recherches de l'UPR9021 (ICT), a réussi à synthétiser chimiquement en phase solide la protéine tat entière qui a donc pu être obtenue de manière extrêmement pure pour la première fois (37). Nous souhaitons ici obtenir des cristaux du complexe formé entre l'ARN TAR et la protéine Tat synthétique. Diverses approches sont envisagées parmi lesquelles l'utilisation de constructions synthétiques d'ARN ou le clonage de constructions dans lesquelles la boucle TAR s'insère dans un contexte ARN exogène, par exemple un ribozyme à tête de marteau. L'affinité de différentes constructions de TAR pour la protéine *tat* a été déterminée en utilisant la Résonance Plasmonique de Surface (RPS) pour valider les constructions. Ainsi nous avons pu démontrer que la cycline n'était pas requise pour observer la formation du complexe (37). Les premiers résultats montrent une grande susceptibilité de la protéine à l'oxydation. Pour pallier ce problème, nous avons développé un dispositif d'observation d'essais de cristallisation sous atmosphère inerte.

Un autre projet consiste à modéliser une région de l'ARNm de l'histone H4 en collaboration avec Frank Martin et Gilbert Eriani (UPR 9002) qui ont récemment démontré que l'ARNm de l'histone H4 est capable de contourner l'étape d'initiation de la traduction en agissant comme un IRES (Internal Ribosome Entry Site). Cette faculté semble faciliter la production massive d'histone H4 qui lors de la phase S du cycle cellulaire doit représenter 60 millions de molécules présentes par noyau. Mes partenaires ont également cartographié l'IRES à l'aide de sondes chimiques et enzymatiques afin d'en établir la structure secondaire. C'est principalement sur ces données que nous appuierons notre travail de modélisation.

3.1 Structure et évolution de l'intron Dir956-1 de *Didymium iridis*

L'intron Dir956-1, situé dans le gène extra-chromosomique codant pour l'ARN 18S du myxomycète *Didymium iridis* possède une organisation originale. Il est constitué d'un intron de groupe I classique dans lequel est inséré un gène encodant une endonucléase voyageuse (Homing Endonuclease). En amont de ce gène se trouve un ARN dénommé GIR1 dont la structure semble fortement apparentée à celle des introns de groupe I. Toutefois, il présente des différences topologiques qui ne sont pas du type des différences habituellement

rencontrées chez ces introns qui consistent généralement en des insertions ou des délétions de domaines périphériques additionnels qui préservent l'organisation du cœur catalytique (Figure 14).

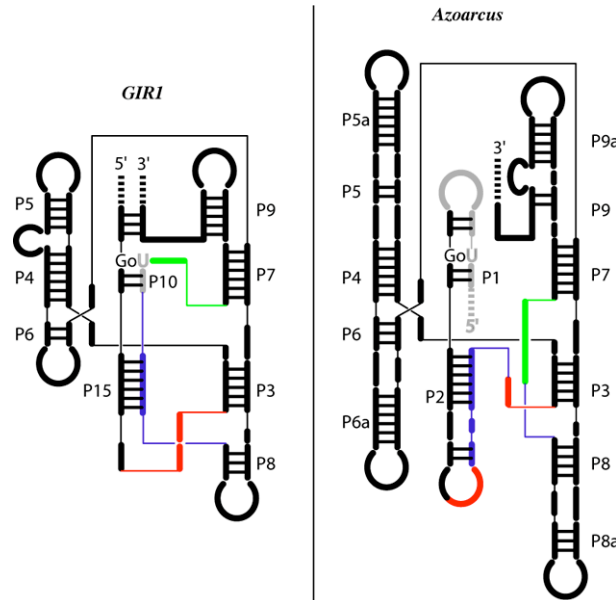


Figure 14 : Schéma comparatif entre les structures secondaires du ribozyme GIR1 de *Didymium iridis* et de l'intron IC3 rencontré chez la bactérie *Azoarcus* dans le pré-ARNt^{leu} (*Azo*). De manière générale, GIR1 possède des éléments de structure secondaire plus courts que chez *Azo*. Les régions de topologie différentes entre les deux ribozymes sont colorées.

GIR1 est un ribozyme qui catalyse la formation d'un lasso de quatre résidus (82) ce qui n'est pas sans rappeler le lasso formé lors de la première étape de l'épissage des introns de groupe II (88,89), toutefois dans un contexte structural qui semble très différent. De plus GIR1 présente un cœur catalytique qui se distingue de celui des introns de groupe I. Ces traits caractéristiques justifient à eux seuls l'étude que nous proposons de mener et qui consiste à proposer un modèle moléculaire de GIR1 qui devrait nous permettre de mieux comprendre les bases moléculaires de son activité et nous aider à concevoir des constructions qui nous serviront à résoudre sa structure cristallographique. Il nous semble en effet primordial de corroborer les études biochimiques par une structure cristallographique qui lèvera une grande partie des interrogations soulevées par la modélisation au niveau de l'organisation du site catalytique. Il faut également considérer que les distinctions topologiques entre les cœurs catalytiques de GIR1 et des introns de groupe I nous motive également à tenter de comprendre comment l'évolution de l'une de ces molécules vers l'autre a pu se faire et réciproquement.

Avec Bertrand Beckert (étudiant en thèse), nous avons construit un modèle de GIR1 en nous basant sur les connaissances les plus récentes sur la structure des introns de

groupe I, à savoir les structures cristallographiques de trois introns de la bactérie *Azoarcus* (23), du bactériophage *Twort* (42) et de *Tetrahymena thermophila* (90). Le modèle que nous proposons permet de concilier les données biochimiques et de mutagenèse accumulées par nos partenaires Henrik Nielsen et Steinar Johansen dans un modèle structural qui permet de mieux comprendre comment les changements topologiques observés dans la structure secondaire ont pu aboutir à un changement de nature de l'activité enzymatique de GIR1 (manuscrit en préparation). GIR1 présente la plus forte homologie structurale avec l'intron d'*Azoarcus* (*Azo*). C'est donc naturellement à *Azo* que nous comparons GIR1.

Tout comme les introns de groupe I, la structure de GIR1 s'organise autour de plusieurs segments hélicoïdaux (notés P pour l'anglais « paired ») numérotés de P1 à P10 (Figure 15). Bien que P1 et P2 n'existent pas dans GIR1, les hélices P3 à P10 s'y retrouvent. Globalement, ces segments s'organisent sous forme de trois domaines constitués par l'empilement des hélices P10-P15, P3-P7-P8-P9 et P4-P5-P6. Parmi ces éléments, certains sont importants pour la structuration de la molécule et d'autres pour la catalyse. La structuration est permise via la mise en place d'un jeu d'interactions tertiaires parmi lesquelles la formation d'un double pseudonœud, et les interactions entre la boucle L6 et P3 et entre la jonction J15/3 et P8. Le site catalytique comprend l'élément P7 qui reconnaît la guanine équivalente à la ω G de *Azo*, le résidu clivé au cours de la réaction, le résidu U de la paire GoU de P10 dont le groupement 2'-hydroxyl constitue le nucléophile et enfin la jonction J5/4 qui joue un rôle dans le positionnement de l'hélice substrat P10. Le site catalytique s'organise donc sur toute la largeur de la molécule dans un disque d'un diamètre d'une cinquantaine d'Ångströms environ. La caractéristique distinctive principale de GIR1 par rapport à *Azo* consiste en la présence d'un double pseudonœud établi entre les hélices P3, P7 et P15. Ce dernier est formé par l'appariement entre le brin 5' de P2 et le brin sortant de P8. Les brins 3' de P15 et P10 semblent donc correspondre à J8/7 de *Azo*. La disparition de P2 s'accompagne de la liaison directe entre le brin 5' de P3 et la boucle de P15 qui semble correspondre à une réminiscence de la boucle de P2 de *Azo*. Le double pseudonœud, en contraignant fortement la structure du cœur, compense la modification de l'interaction habituellement observée entre la boucle L2 et P8 chez *Azo*. Nous pensons également que le gain de stabilité dû au le double pseudonœud est à l'origine du raccourcissement des éléments périphériques comme P5 et P6, même si la boucle de ce dernier conserve l'interaction tertiaire établie avec P3.

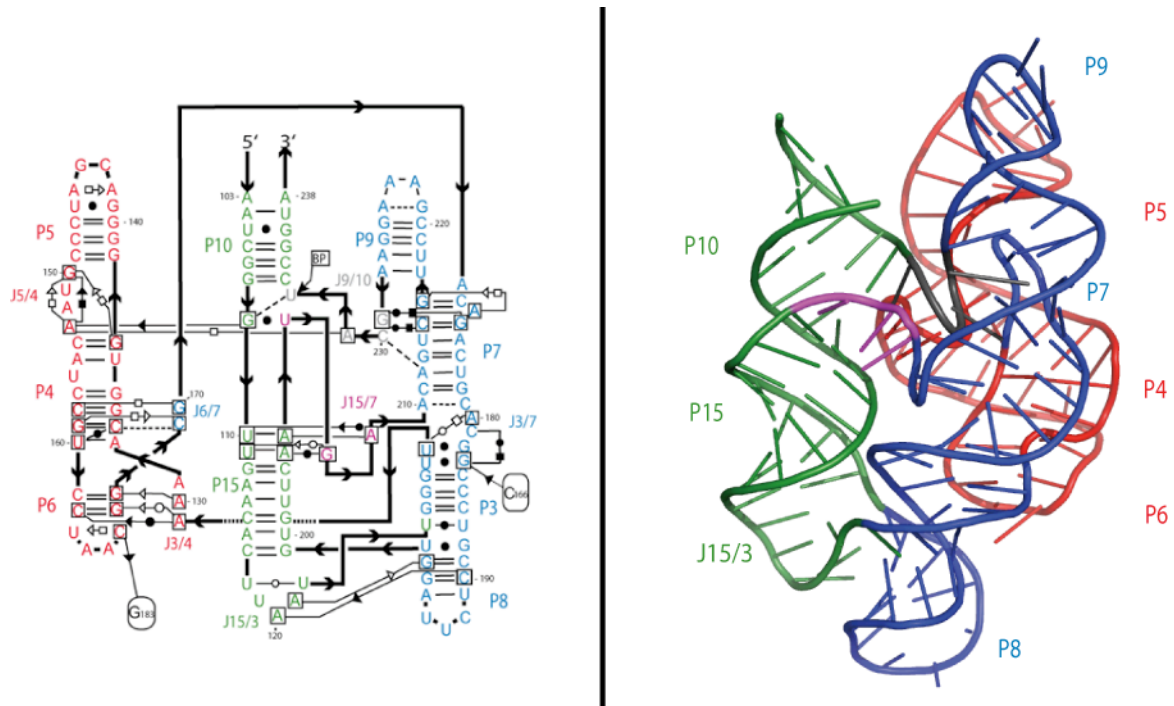


Figure 15 : Représentation de la structure secondaire et du modèle structural de GIR1. Les deux panneaux présentent les trois domaines sous forme 3d et sous forme de structure secondaire ainsi que les interactions non canoniques à l'aide de la nomenclature Leontis-Westhof (91).

Ces changements topologiques aux conséquences structurales importantes aboutissent à la disparition de la jonction J8/7 de *Azo*. J8/7 est le seul élément à ne pas trouver d'équivalent chez GIR1. Or J8/7 joue un rôle très important chez *Azo*. Cette jonction traverse le cœur catalytique de part en part en stabilisant les trois domaines par un jeu compliqué de liaisons hydrogène impliquant plusieurs magnésiums dont certains sont impliqués dans le mécanisme catalytique. Cette modification structurale profonde permet de mieux comprendre que GIR1 opère une catalyse différente de celle des introns de groupe I. En effet, la disparition de certains résidus de J8/7 a pu provoquer un remodelage du cœur catalytique. Le substrat a donc pu être reconnu de manière erronée ce qui a pu faire apparaître une nouvelle activité catalytique. En effet, la réaction de GIR1 correspond à l'étape 2 de la réaction d'épissage à la différence près que c'est le groupement 2'-hydroxyl qui constitue le nucléophile, ce qui autorise le groupement 3'-hydroxyl à déjà être lié à un autre résidu. Dans cette situation, il est possible d'imaginer que dans un contexte structural perturbé, l'étape 2 ait pu se produire avant l'étape 1.

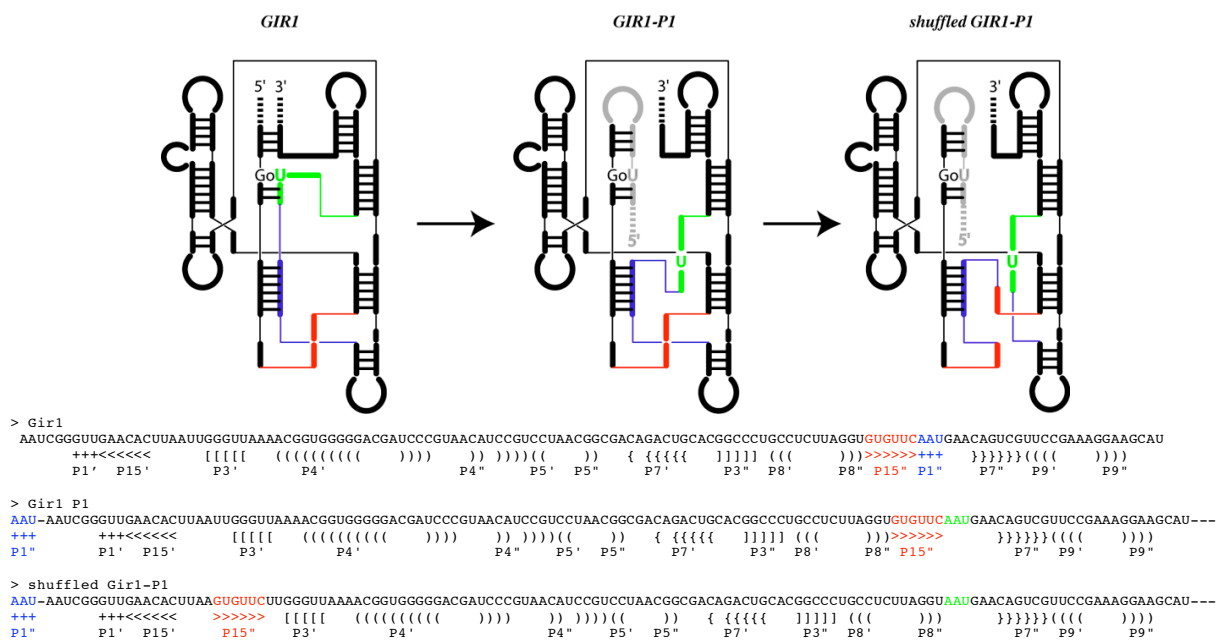


Figure 16 : Rétro-évolution de GIR1 vers un intron de groupe I de type IC3. Deux représentations complémentaires pour comprendre quels sont les changements à opérer pour faire basculer la topologie de GIR1 vers celle d'un intron de groupe I de type IC tel celui d'*Azoarcus*. Sur le panneau du haut, certains brins impliqués ou non dans des hélices de la structure secondaire sont colorés de manière uniforme (à droite). Sur le schéma central on voit que le désappariement des extrémités 5' et 3' de GIR1 permet la formation d'une structure secondaire semblable à P1 (gris) impliquant une région en amont du ribozyme. Ceci libère une partie de J8/7 (vert). Enfin le schéma de droite illustre qu'une fois P1 formé, des sauts de brins successifs interviennent pour restaurer la topologie d'un intron IC. Pour ce faire, le brin rouge est connecté au brin bleu, puis le brin rouge encore libre est connecté à l'extrémité 3' du brin bleu. Ainsi l'extrémité 5' libre du brin bleu peut être liée au brin vert libre. Le panneau du bas décrit le même événement au niveau 1D. La première séquence indique la séquence de GIR1 sauvage. La région en rouge correspond au brin 3' de P15 et celle en bleu correspond à la séquence de P15 qui mime l'hélice substrat P1. La deuxième séquence indique la présence additionnelle d'une séquence (en bleu) capable de mimer P1 à la place de la séquence verte. Enfin, la troisième séquence montre à quelle position la transposition de la séquence rouge doit être opérée pour rétablir une topologie identique à celle rencontrée dans les introns de groupe I de type IC.

La modélisation a également permis d'appréhender les aspects évolutifs de GIR1 par rapport aux introns classiques. Il semble en effet que GIR1 a évolué à partir d'un intron de groupe I (Figure 16). Se pose alors la question de savoir comment les changements topologiques observés dans GIR1 ont pu intervenir. Notre hypothèse de travail provient de l'observation suivante. L'inspection du modèle nous a permis de remarquer que les changements topologiques interviennent en des endroits où les squelettes ribose-phosphate des jonctions incriminées passent à de courtes distances les uns des autres semblant autoriser des sauts de brins les rationalisant. Pour aller plus loin, nous avons tenté de dériver un modèle permettant d'expliquer ces sauts de brins par un mécanisme simple. L'analyse de la séquence nous a alors révélé que la simple transposition d'une séquence de six nucléotides (5'-GUGUUC-3') appartenant au brin 3' de P15 dans la jonction J15/3 permettait de retrouver la topologie des introns de groupe I dans GIR1. En effet, cette transposition permet de rétablir un « vrai » P2 et une vraie « J8/7 ». Il est alors possible d'imaginer que GIR1-transposé puisse

à nouveau se comporter en intron de groupe I, quoique sans doute moins efficace puisque n'étant pas aussi bien stabilisé par des jeux d'interactions tertiaires. La modélisation de GIR1 nous motive donc aujourd'hui à essayer de convertir expérimentalement GIR1 en intron de groupe I. Par ailleurs, nous tenterons également de convertir l'intron de l'ARN^{tle} d'*Azoarcus* en un ribozyme réalisant une réaction de branchement en opérant une transposition opposée. Ces expériences ont déjà été initiées et auront une portée importante quant à notre compréhension de l'évolution de la structure des ARN en général et des ribozymes en particulier.

3.2 Identification et caractérisation de nouveaux ARN chez *S.aureus* et *L.monocytogenes*

Depuis quelques années, de nombreux types d'ARN non-codants contrôlant la régulation génétique en *trans* ou en *cis* ont été répertoriés chez les eucaryotes et chez les procaryotes. Ces différents types d'ARN que l'on peut baptiser riborégulateurs sont reliés par leur fonction mais se distinguent les uns des autres par leur mécanisme d'action, leur structure et leur taille. Cette diversité sera brièvement présentée dans les paragraphes suivants avant de présenter notre projet qui vise à identifier des ARN non-codant chez certaines bactéries *Gram*⁺ pathogènes.

3.2.1 L'interférence à ARN chez les eucaryotes

Chez les eucaryotes supérieurs, la régulation ARN-dépendante obéit au mécanisme mis à jour récemment d'interférence à ARN (RNAi) revu dans (92). Ce mécanisme peut être déclenché par deux types d'ARN suivant que leur provenance est exogène (infection virale par exemple) ou endogène. Dans le premier cas, de petits ARN interférants (siRNA) sont introduits dans les cellules sous forme de double-hélice. Dans le second cas, des épingles à cheveux d'ARN sont transcrites à partir du génome. Les micro ARN précurseurs (miRNA) qui en résultent sont alors clivés par Dicer en petits ARN de taille comprise entre 21 et 23 nucléotides puis pris en charge par le complexe RISC. Ce complexe utilise le miARN clivé pour reconnaître la région complémentaire de l'ARNm cible. RISC procède alors au clivage du messager dans la région d'hybridation (93) abolissant ainsi l'expression du gène.

3.2.2 Les petits ARN régulateurs chez *E. coli*

Chez les procaryotes, la régulation de l'expression des gènes par des mécanismes impliquant les ARN messagers peut être envisagé suivant différents modèles. Il existe dans les extrémités 5' non-traduites (5'-UTR) de certains messagers des régions structurées capables de reconnaître des protéines dont l'association assure le contrôle négatif ou positif de leur traduction. Des exemples bien connus au laboratoire regroupent le contrôle de l'expression de certaines protéines, comme la protéine ribosomique S15 (94,95) ou la thréonyl aminoacyl-ARNt synthétase (ThrRS) (52). Dans ce type de systèmes, la régulation de l'expression repose sur la mise en place de structures dans l'ARN capables d'entrer en compétition avec le ribosome pendant la phase d'initiation ou d'élongation de la traduction. Dans le cas de la régulation de l'expression de la ThrRS, la répression de la traduction découle de l'interaction entre une région du messager à cheval entre l'extrémité 5' non-traduite (5'-UTR) et la région codante ce qui permet à la synthétase de piéger la séquence Shine-Dalgarno dans un complexe ribonucléotéique (55).

Chez *E. coli*, de petits ARN régulateurs non-codants (souvent notés sRNA ou ncRNA) sont exprimés en corrélation avec les variations des conditions physiques et chimiques de leur environnement. Des études très récentes ont permis de mettre en évidence chez *E. coli* des ARN impliqués en *trans* dans la régulation de la synthèse d'un peptide signal de l'arrêt de la croissance bactérienne (96) et, de manière plus générale, environ quatre-vingts ARNnc sont maintenant répertoriés chez *E. coli*. Ils sont impliqués dans la réponse à divers stress comme les variations de pH, le stress oxydatif, la carence en fer, ou encore les changements de pression osmotique (97). Chez *Aquifex aeolicus*, une autre bactérie *Gram*⁻ mais thermostable, des données préliminaires obtenues par ARNomique montrent que des ARN antisens semblent dirigés contre des ARN de transfert (98). Pratiquement, la plupart de ces ARNnc fonctionnent comme des antisens. Ils s'apparient à de courtes régions de l'ARN messenger chevauchant la région Shine-Dalgarno et le codon initiateur AUG pour empêcher le ribosome de procéder à la traduction du messager (répression traductionnelle) (99). La reconnaissance du messager par l'ARN non-codant effecteur peut également s'accompagner de la fusion de structures inhibitrices de la traduction comme dans le cas de l'activation de la traduction du facteur de transcription σ_s par l'ARN DsrA (100). Ces modes d'action ressemblent fortement au système décrit au paragraphe précédent impliquant la formation d'un complexe ARN/protéine. Le système antisens présente toutefois quelques avantages. Il peut être mis en œuvre dès la transcription, assurant ainsi un contrôle temporel immédiat. De

plus, les ARN sont en général plus rapidement et plus facilement dégradés que les protéines ce qui peut être mis à profit pour contrôler finement l'arrêt du mécanisme. De plus, du point de vue chimie combinatoire, il est conceptuellement plus simple de synthétiser une séquence complémentaire à une autre plutôt que de construire une édifice protéique doué de spécificité pour un ARN structuré. Les ARN antisens pourrait donc avoir été utilisés dès l'apparition du mécanisme de synthèse protéique et ainsi assurer le majeure partie du contrôle de l'expression des gènes avant l'apparition de mécanismes de contrôle de l'expression génétique basés sur des interactions ARN/protéine.

3.2.3 La régulation de la virulence chez les Gram⁺ pathogènes

Le mécanisme basé sur la reconnaissance en *trans* par un ARN antisens est également utilisé pour réguler les gènes de virulence chez de nombreux pathogènes notamment par l'intermédiaire de la mesure de la densité cellulaire (quorum sensing). Chez *Staphylococcus aureus*, le système *agr* (101) active la transcription de l'ARNIII. Celui-ci code d'une part pour l'hémolysine δ (*hld*) et d'autre part contrôle la répression de l'expression de la protéine A (*spa*) en interagissant avec son messenger en phase exponentielle (Figure 17). De plus, il active l'expression de toxines extracellulaires et d'enzymes durant la phase post-exponentielle. Cette répression est suivie par la dégradation du messenger par la RNase III (102). Plusieurs messagers sont régulés de cette manière ce qui montre qu'une kyrielle de gènes est sous le contrôle simultané d'un même ARN non-codant.

Chez les Gram⁻, une autre enzyme potentiellement impliquée dans la dégradation est la RNase E qui semble favoriser, en présence de Hfq la dégradation de certains messagers chez *Vibrio cholerae* (103). De manière intéressante, la RNase III est homologue à Dicer (voire §3.2.1) sur le plan fonctionnel et structural (104). Ces observations permettent d'émettre l'hypothèse qu'un mécanisme apparenté à l'interférence à ARN eucaryotes puisse exister chez les procaryotes. Toutefois, les expériences testant cette hypothèse ont pour l'instant échoué à démontrer la nécessité de la dégradation de l'ARNm pour aboutir à l'inhibition de la traduction (105).

Enfin, chez *Listeria monocytogenes*, des ARN non-codant régulateurs potentiellement impliqués dans la virulence ont été identifiés par combinaison d'approches théoriques et expérimentales (106). Chez *Salmonella typhimurium*, Hfq semble jouer un rôle fondamental comme chez *E. coli* puisqu'une souche *hfq* montre un profil d'expression altéré pour certaines protéines régulées par des ARN non-codant (107). Les phénotypes identifiés

montrent des défauts de motilité cellulaire et de composition de la membrane, une adhésion réduite et l'arrêt de la sécrétion de protéines effectrices. Les phénotypes observés chez des mutants *hfq* de plusieurs espèces de bactéries peuvent être aujourd'hui reliés au fait que la régulation de l'expression génétique médiée par des ARN non-codant est dans ce cas perturbée (108).

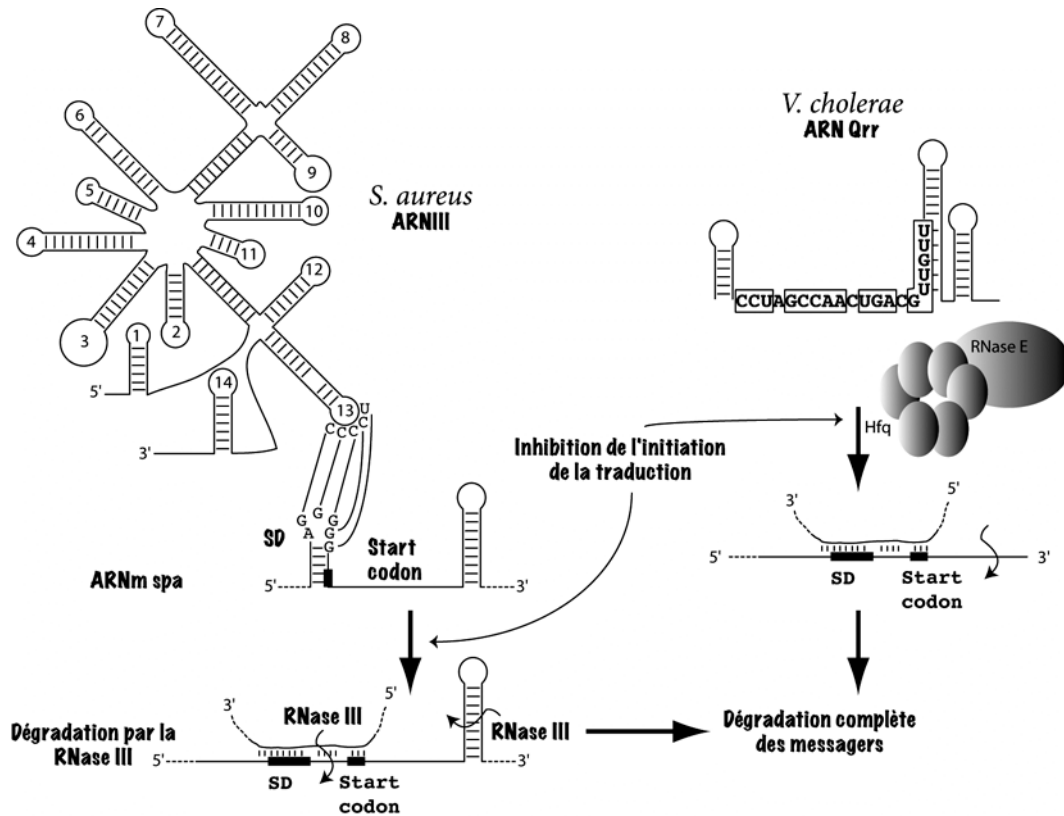


Figure 17 : Deux mécanismes de régulation de gènes de virulence chez des pathogènes (Figure adaptée de (109)). Chez *S. aureus* (*Gram*⁺), l'ARNIII régule l'expression de *spa* en formant des appariements entre la boucle 13 qui est complémentaire à la séquence Shine-Dalgarno (SD) de l'ARNm. Cette interaction se propage alors grâce à la présence de régions complémentaires dans les épingles à cheveux 13 et 12 et aboutit au masquage de la séquence SD et du codon initiateur AUG. La formation de ce complexe ARN/ARN favorise l'action de la RNase III qui coupe dans les régions structurées et permet ensuite la dégradation complète du messager. Chez *V. cholerae* (*Gram*⁻), le mécanisme de la régulation de l'expression de *hapR* par l'ARN Qrr (103) est similaire mais cette fois Hfq semble requise pour recruter la RNase E responsable de la dégradation initiale du messager.

3.2.4 Les interrupteurs à ARN

Les systèmes décrits aux paragraphes précédents permettent la régulation de l'expression des gènes au niveau post-transcriptionnel. Il nous reste donc à décrire un troisième système qui cette fois régule l'expression des gènes en *cis* au niveau transcriptionnel. Il repose sur la présence de régions structurées d'ARN dans les 5'-UTR qui reconnaissent un ligand organique tel un métabolite qui provoque des remaniements structuraux modifiant les structures impliquées dans la transcription. Ces ribo-interrupteurs

(riboswitch) sont de découverte récente et ont été principalement étudiés chez *Bacillus subtilis* (110,111). Ils sont constitués par l'enchaînement successif d'une région capable de fixer un ligand (un aptamère) suivie d'une plateforme d'expression qui module directement l'expression du gène en fonction de l'état lié ou non de l'aptamère (Figure 18). Ce mécanisme de régulation repose sur la formation de structures secondaires mutuellement exclusives et dépendantes de la reconnaissance du ligand par l'aptamère.

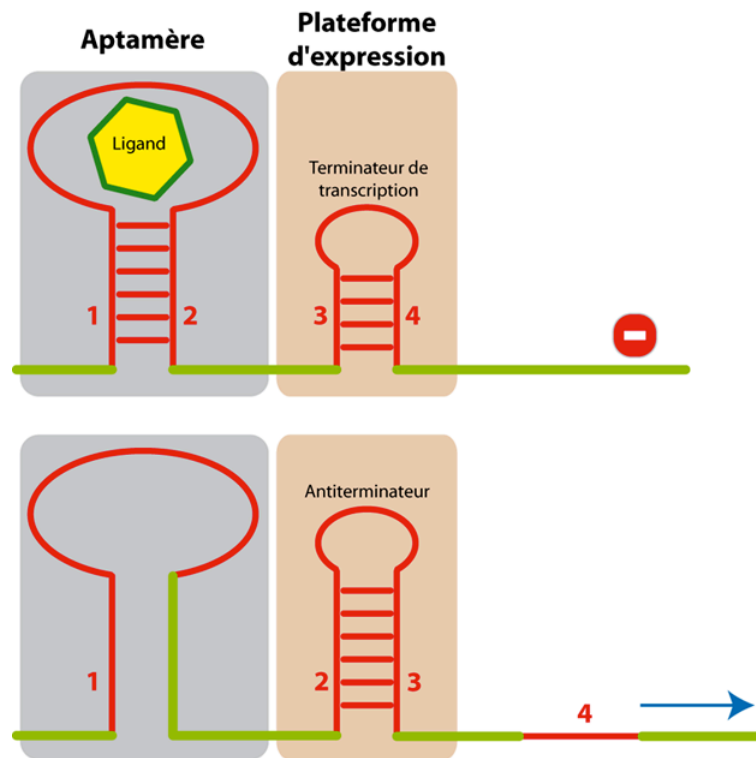


Figure 18 : Principe de régulation de l'expression génétique par un riboswitch. (haut) En présence du ligand, le riboswitch se replie et promeut la formation d'un terminateur de transcription qui va empêcher l'ARN polymérase de synthétiser les transcrits. (bas) Lorsque le ligand est absent, le riboswitch ne se replie pas et un antiterminateur se forme permettant à la polymérase de synthétiser des transcrits de pleine longueur qui peuvent être alors traduits. Cette dynamique est rendue possible par la possibilité de formation de deux structures secondaires mutuellement exclusives [(1,2) (3,4)] ou [1 (2,3) 4].

Dans le système le plus simple, la liaison du ligand structure la région aptamère et force la plateforme d'expression à former un terminateur de transcription. L'aptamère retient une séquence qui est complémentaire à la plateforme d'expression. Au contraire, l'absence du ligand favorise la formation d'une structure anti-terminatrice qui contient la séquence retenue par l'aptamère lorsque le ligand est en place. Dans le cas le plus simple, le ligand est un métabolite qui correspond au produit formé par la protéine dont l'expression est régulée. C'est un cas de rétro-inhibition de la transcription. Par exemple chez *B.subtilis*, la flavine mononucleotide (FMN) produite par l'opéron *ribDEAHT* se fixe dans la région 5'-UTR de l'opéron et y induit la formation d'un terminateur de transcription (112). Ces ARN-

interrupteurs semblent distribués à travers tout le règne bactérien et permettraient le contrôle d'environ 2% des gènes chez *B.subtilis* (113). Ils peuvent également être placés en homo ou hétéro-tandems pour réguler l'expression de gènes de manière coopérative (114) ou pour tirer parti de l'utilisation de deux substrats par une enzyme (115). Les riboswitchs peuvent avoir des ligands de taille variée allant de la vitamine B12 (116) jusqu'aux ions magnésiums (117). Leur nature les destine donc également à devenir les cibles potentielles de molécules thérapeutiques qui permettraient de cibler les voies métaboliques d'organismes pathogènes. Des amino-pyrimidines sont par exemple capables de se fixer au riboswitch qui contrôle l'expression du gène de l'adénosine désaminase (*add*) de *V. vulnificus* (118).

3.2.5 Identifier de nouveaux ARN chez des bactéries pathogènes

Les systèmes de régulation de l'expression des gènes décrits aux paragraphes précédents sont utilisés par tous les procaryotes. Il est donc particulièrement intéressant de mettre en évidence et de caractériser de tels systèmes chez des organismes pathogènes afin de mieux comprendre les mécanismes fins de la régulation de la virulence. Pascale Romby et son équipe à l'IBMC et Pascale Cossart à l'institut Pasteur ont déjà opéré des avancées significatives sur la compréhension de la régulation de l'expression de certains gènes de virulence chez *S.aureus* et *L. monocytogenes*, respectivement. En travaillant conjointement avec Pascale Romby et Christine Gaspin (INRA, Toulouse) sur la virulence chez *S. aureus*, nous souhaitons élargir le champ de recherches en essayant d'identifier les riboswitchs de *S.aureus* qui n'ont pas encore pu l'être et peut-être en identifier de nouveaux. Par exemple, il semble curieux qu'aucun riboswitch reconnaissant l'intermédiaire préQ1 de la biosynthèse de la queuosine ou l'adenosylcobalamine (précurseur de la vitamine B12) n'ait pu être identifié chez *S.aureus* (source Rfam : <http://www.sanger.ac.uk/Software/Rfam/>, (119)). De plus, les conditions de culture induisant la virulence présentent une signature physico-chimique détectable par la bactérie par l'intermédiaire de métabolites spécifiques. Il semble donc raisonnable d'émettre l'hypothèse que des riboswitchs puissent être impliqués à différentes étapes de la virulence chez certains organismes pathogènes.

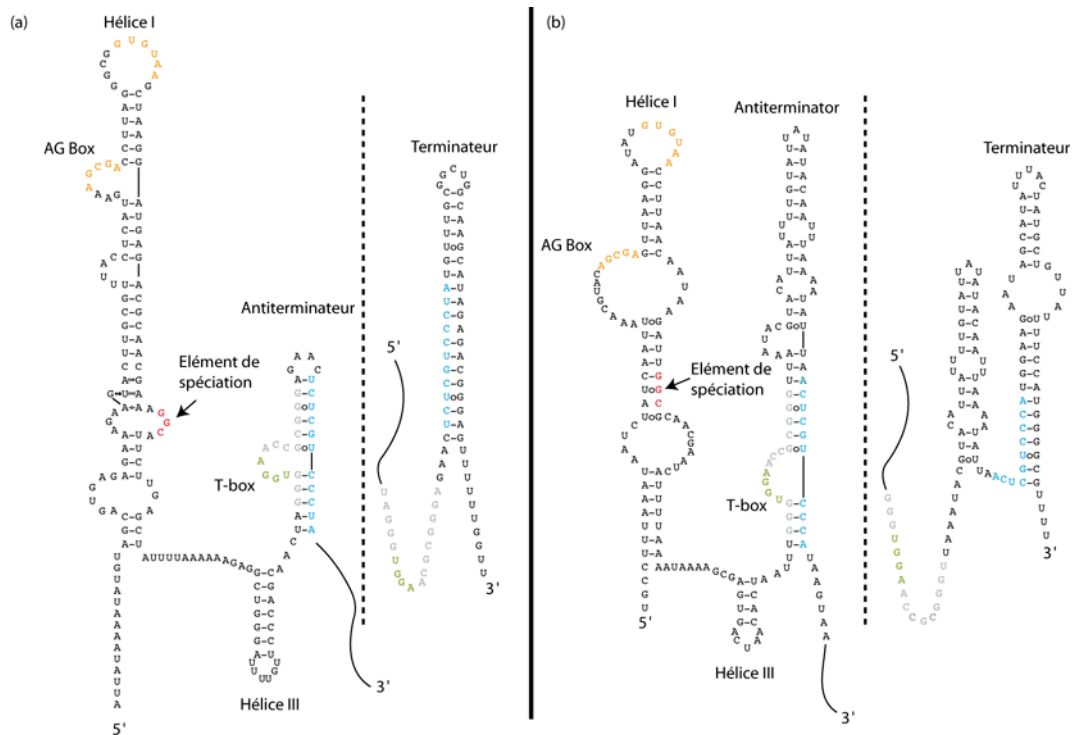


Figure 19 : Représentation des structures alternatives de deux T-box chez *B. subtilis* (a) et *S. aureus* (b). (a) L'expression du gène *glyQS* est sous le contrôle de la T-box capable d'interagir avec l'ARNt^{Gly} (120). La boucle anticodon de l'ARNt interagit avec la séquence correspondant au codon (élément de spéciation en rouge) présente dans la tige-boucle I. Celle-ci contient des éléments conservés, une AG box et une séquence spécifique dans la boucle apicale. La séquence de 14 nucléotides (grise) caractérisant les T-box et donnant son nom à l'élément régulateur complet se trouve dans l'antiterminateur qui ne se forme que lorsque l'extrémité acceptrice 5'-UCCA-3' libre de l'ARNt s'apparie à la région complémentaire (kaki). Si l'ARNt est amino-acylé, l'appariement avec la T-box ne se fait pas et un terminateur se forme au détriment de l'antiterminateur car le brin 3' de ce dernier s'apparie à une séquence complémentaire située en aval. (b) Les éléments de séquence décrits plus haut sont retrouvés dans la T-box potentielle que nous avons identifiée en amont du gène *glyS* chez *S. aureus*. La petite épingle à cheveux en amont du terminateur ne contient qu'une seule paire GC et ne devrait donc pas se former. Nous souhaitons caractériser *in vivo* et *in vitro* cette T-box.

Notre stratégie consiste à procéder à l'analyse comparative des différents génomes de *S. aureus* puis à analyser les régions 5'-UTR par des approches bioinformatiques qui consistent à analyser l'organisation des régions non traduites en y cherchant la présence de signaux caractéristiques de la présence de riboswitchs. Ces signaux regroupent (1) la longueur de la séquence (supérieure à environ 130 nucléotides), (2) un pourcentage de GC supérieur à la moyenne du génome (spécialement chez les organismes à faible pourcentage de GC comme *S. aureus*) ou l'énergie libre, la présence (3) d'épingles à cheveux mutuellement exclusives, (4) d'un terminateur de transcription en amont du gène considéré ou (5) de structures secondaires capables de séquestrer la région Shine-Dalgarno, (6) de promoteurs de transcription (recherche des boîtes -35/-10). L'utilisation de profils cachés de Markov (HMM, HMMER (121)) permet, à partir d'alignement de séquences, de générer des modèles statistiques qui sont mis à profit pour identifier les riboswitchs connus. L'analyse croisée de la présence de

ces signaux dans les séquences intergéniques permet alors d'identifier des candidats potentiellement intéressants. Nous tenterons également de déterminer la structure des ARN que nous aurons identifiés

Cette approche nous a tout d'abord permis de retrouver les riboswitchs connus chez *S. aureus*. La poursuite de l'analyse des candidats a récemment permis d'identifier en amont du gène de la glycyL-ARNt synthétase (*glyS*) une T-box potentielle qui semble adopter une structure secondaire légèrement atypique (Figure 19). Les T-box sont des riboswitchs dont le ligand est l'ARNt. Le fait qu'il soit aminoacylé ou non permet de modifier une structure secondaire jouant alternativement le rôle de terminateur ou d'anti-terminateur de transcription. Après avoir vérifié l'expression de l'ARN encodant la T-box, nous la caractériserons du point de vue fonctionnel (120). Chez *S. aureus*, il existe 5 ARNt iso-accepteurs potentiellement capables d'interagir avec cette T-box. De manière intéressante, la glycine entre dans la composition de la paroi bactérienne qui est synthétisée par un système de synthèse peptidique parallèle reposant sur l'activité de *femX* (122). L'expression de *glyS* pourrait donc être sous contrôle de la synthèse protéique et de la synthèse de la paroi bactérienne. L'identification d'autres éléments nouveaux ouvriront sans doute d'autres pistes de recherches aussi intéressantes et qui permettront de révéler des aspects encore inconnus de la régulation médiée par des ARN structurés.

4 Bibliographie

1. Ehresmann, C., Baudin, F., Mougél, M., Romby, P., Ebel, J.P. and Ehresmann, B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.*, 15, 53-71.
2. Masquida, B. and Westhof, E. (2005) In Hartmann, R. K., Bindereif, A., Schön, A. and Westhof, E. (eds.), *Handbook of RNA biochemistry*. Wiley VCH Verlag GmbH & Co, pp. 536-545.
3. Lescoute, A., Leontis, N.B., Massire, C. and Westhof, E. (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res*, 33, 2395-2409.
4. Lescoute, A. and Westhof, E. (2006) The A-minor motifs in the decoding recognition process. *Biochimie*, 88, 993-999.
5. Masquida, B., Felden, B. and Westhof, E. (1997) Context dependent RNA-RNA recognition in a three-dimensional model of the 16S rRNA core. *Bioorgan Med Chem*, 5, 1021-1035.
6. Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vornrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature*, 407, 327-339.
7. Carugo, O. and Pongor, S. (2001) A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci*, 10, 1470-1473.
8. Yusupova, G., Jenner, L., Rees, B., Moras, D. and Yusupov, M. (2006) Structural basis for messenger RNA movement on the ribosome. *Nature*, 444, 391-394.
9. Noller, H.F., Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Lancaster, L., Dallas, A., Fredrick, K., Earnest, T.N. and Cate, J.H. (2001) Structure of the ribosome at 5.5 Å resolution and its interactions with functional ligands. *Cold Spring Harb Symp Quant Biol*, 66, 57-66.
10. Agrawal, R.K., Spahn, C.M., Penczek, P., Grassucci, R.A., Nierhaus, K.H. and Frank, J. (2000) Visualization of tRNA movements on the Escherichia coli 70S ribosome during the elongation cycle. *J Cell Biol*, 150, 447-460.
11. Serganov, A.A., Masquida, B., Westhof, E., Cachia, C., Portier, C., Garber, M., Ehresmann, B. and Ehresmann, C. (1996) The 16S rRNA binding site of Thermus thermophilus ribosomal protein S15: Comparison with Escherichia coli S15, minimum site and structure. *RNA*, 2, 1124-1138.
12. Masquida, B., Earnshaw, D.J., Müller, S., Sigurdsson, S.T., Eckstein, F., Westhof, E. and Gait, M.J. (1997) Inter-domain cross-linking and molecular modelling of the hairpin ribozyme. *J. Mol. Biol.*, 274, 197-212.
13. Rupert, P.B. and Ferre-D'Amare, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, 410, 780-786.
14. Szewczak, A.A. and Moore, P.B. (1995) The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.*, 247, 81-98.
15. Szewczak, A.A., Moore, P.B., Chang, Y.L. and Wool, I.G. (1993) The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc. Natl. Acad. Sci. USA*, 90, 9581-9585.
16. Wimberly, B., Varani, G. and Tinoco, I., Jr. (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochem.*, 32, 1078-1087.
17. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science.*, 273, 1678-1684.

18. Chowrira, B.M., Berzal-Herranz, A., Keller, C.F. and Burke, J.E. (1993) Four ribose 2'-hydroxyl groups essential for catalytic function of the hairpin ribozyme. *J. Biol. Chem.*, 268, 19458-19462.
19. Masquida, B. and Westhof, E. (1999) In Neidle, S. (ed.), *Oxford handbook of nucleic acid structures*. Oxford University Press, Oxford, UK, pp. 533-565.
20. Masquida, B., Sauter, C. and Westhof, E. (1999) A sulfate pocket formed by three GoU pairs in the 0.97 Å resolution X-ray structure of a nonameric RNA. *RNA*, 5, 1384-1395.
21. Deng, J., Xiong, Y. and Sundaralingam, M. (2001) X-ray analysis of an RNA tetraplex (UGGGU)₄ with divalent Sr²⁺ ions at subatomic resolution (0.61 Å). *Proc Natl Acad Sci U S A*, 98, 13665-13670.
22. Su, L., Chen, L., Berger, J.M. and Rich, A. (1999) Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nature Struct. Biology*, 6, 285-292.
23. Adams, P.L., Stahley, M.R., Gill, M.L., Kosek, A.B., Wang, J. and Strobel, S.A. (2004) Crystal structure of a group I intron splicing intermediate. *Rna*, 10, 1867-1887.
24. Masquida, B. and Westhof, E. (2000) On the wobble GoU and related pairs. *RNA*, 6, 9-15.
25. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35, 849-857.
26. Loria, A. and Pan, T. (1996) Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA*, 2, 551-563.
27. Krasilnikov, A.S., Yang, X., Pan, T. and Mondragon, A. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature*, 421, 760-764.
28. Krasilnikov, A.S., Xiao, Y., Pan, T. and Mondragon, A. (2004) Basis for structural diversity in homologous RNAs. *Science*, 306, 104-107.
29. Torres-Larios, A., Swinger, K.K., Krasilnikov, A.S., Pan, T. and Mondragon, A. (2005) Crystal structure of the RNA component of bacterial ribonuclease P. *Nature*, 437, 584-587.
30. Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, 289, 905-920.
31. Doherty, E.A., Batey, R.T., Masquida, B. and Doudna, J.A. (2001) A universal mode of helix packing in RNA. *Nat Struct Biol*, 8, 339-343.
32. Milligan, J.F. and Uhlenbeck, O.C. (1990) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.*, 180, 51-62.
33. Helm, M., Brule, H., Giege, R. and Florentz, C. (1999) More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA*, 5, 618-621.
34. Collier, A.J., Gallego, J., Klinck, R., Cole, P.T., Harris, S.J., Harrison, G.P., Aboul-Ela, F., Varani, G. and Walker, S. (2002) A conserved RNA structure within the HCV IRES eIF3-binding site. *Nat Struct Biol*, 9, 375-380.
35. Kuzuya, A., Mizoguchi, R., Morisawa, F., Machida, K. and Komiyama, M. (2002) Metal-ion-induced site-selective RNA hydrolysis by use of acridine-bearing oligonucleotide as cofactor. *J. Am. Chem. Soc.*, 124, 6887-6894.
36. Francois, B., Russell, R.J., Murray, J.B., Aboul-ela, F., Masquida, B., Vicens, Q. and Westhof, E. (2005) Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res*, 33, 5677-5690.
37. Chaloin, O., Peter, J.C., Briand, J.P., Masquida, B., Desgranges, C., Muller, S. and Hoebeke, J. (2005) The N-terminus of HIV-1 Tat protein is essential for Tat-TAR RNA interaction. *Cell Mol Life Sci*, 62, 355-361.

38. Golden, B.L., Gooding, A.R., Podell, E.R. and Cech, T.R. (1998) A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science*, 282, 259-264.
39. Streicher, B., Westhof, E. and Schroeder, R. (1996) The environment of two metal ions surrounding the splice site of a group I intron. *Embo J*, 15, 2556-2564.
40. Clemons, W.M., Jr., Brodersen, D.E., McCutcheon, J.P., May, J.L., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T. and Ramakrishnan, V. (2001) Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: purification, crystallization and structure determination. *J Mol Biol*, 310, 827-843.
41. Waldsich, C., Masquida, B., Westhof, E. and Schroeder, R. (2002) Monitoring intermediate folding states of the td group I intron in vivo. *Embo J*, 21, 5281-5291.
42. Golden, B.L., Kim, H. and Chase, E. (2004) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol*.
43. Michel, F. and Westhof, E. (1990) Modelling of the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, 216, 585-610.
44. Lehnert, V., Jaeger, L., Michel, F. and Westhof, E. (1996) New loop-loop interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem. & Biol.*, 3, 993-1009.
45. Rangan, P., Masquida, B., Westhof, E. and Woodson, S.A. (2003) Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc Natl Acad Sci U S A*, 100, 1574-1579.
46. Rangan, P., Masquida, B., Westhof, E. and Woodson, S.A. (2004) Architecture and folding mechanism of the *Azoarcus* Group I Pre-tRNA. *J Mol Biol*, 339, 41-51.
47. Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J. and Strobel, S.A. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature*.
48. Massire, C., Jaeger, L. and Westhof, E. (1998) Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J Mol Biol*, 279, 773-793.
49. Biswas, R., Ledman, D.W., Fox, R.O., Altman, S. and Gopalan, V. (2000) Mapping RNA-protein interactions in ribonuclease P from *Escherichia coli* using disulfide-linked EDTA-Fe. *J Mol Biol*, 296, 19-31.
50. Whirl-Carrillo, M., Gabashvili, I.S., Bada, M., Banatao, D.R. and Altman, R.B. (2002) Mining Biochemical information: Lessons taught by the ribosome. *RNA*, 8, 279-289.
51. Tsai, H.Y., Masquida, B., Biswas, R., Westhof, E. and Gopalan, V. (2003) Molecular Modeling of the Three-dimensional Structure of the Bacterial RNase P Holoenzyme. *J Mol Biol*, 325, 661-675.
52. Romby, P., Caillet, J., Ebel, C., Sacerdot, C., Graffe, M., Eyermann, F., Brunel, C., Moine, H., Ehresmann, C., Ehresmann, B. *et al.* (1996) The expression of *E.coli* threonyl-tRNA synthetase is regulated at the translational level by symmetrical operator-repressor interactions. *Embo J*, 15, 5976-5987.
53. Sankaranarayanan, R., Dock-Bregeon, A.C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B. and Moras, D. (1999) The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell*, 97, 371-381.
54. Torres-Larios, A., Dock-Bregeon, A.C., Romby, P., Rees, B., Sankaranarayanan, R., Caillet, J., Springer, M., Ehresmann, C., Ehresmann, B. and Moras, D. (2002) Structural basis of translational control by *Escherichia coli* threonyl tRNA synthetase. *Nat Struct Biol*, 9, 343-347.
55. Caillet, J., Nogueira, T., Masquida, B., Winter, F., Graffe, M., Dock-Bregeon, A.C., Torres-Larios, A., Sankaranarayanan, R., Westhof, E., Ehresmann, B. *et al.* (2003) The

- modular structure of *Escherichia coli* threonyl-tRNA synthetase as both an enzyme and a regulator of gene expression. *Mol Microbiol*, 47, 961-974.
56. Ryckelynck, M., Masquida, B., Giege, R. and Frugier, M. (2005) An intricate RNA structure with two tRNA-derived motifs directs complex formation between yeast aspartyl-tRNA synthetase and its mRNA. *J Mol Biol*, 354, 614-629.
 57. Agrawal, R.K., Penczek, P., Grassucci, R.A., Li, Y., Leith, A., Nierhaus, K. and Franck, J. (1996) Direct visualization of A-, -P, and E-site transfer RNAs in the *Escherichia coli* ribosome. *Science*, 271, 1000-1002.
 58. Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr., Tarry, M.J., Carter, A.P. and Ramakrishnan, V. (2001) Recognition of Cognate Transfer RNA by the 30S Ribosomal Subunit. *Science* %R 10.1126/science.1060612, 292, 897-902.
 59. Foloppe, N., Chen, I.J., Davis, B., Hold, A., Morley, D. and Howes, R. (2004) A structure-based strategy to identify new molecular scaffolds targeting the bacterial ribosomal A-site. *Bioorganic & Medicinal Chemistry*, 12, 935.
 60. Rodnina, M.V., Daviter, T., Gromadski, K. and Wintermeyer, W. (2002) Structural dynamics of ribosomal RNA during decoding on the ribosome. *Biochimie*, 84, 745-754.
 61. Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. and Steitz, T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A*, 98, 4899-4903.
 62. Daviter, T., Gromadski, K.B. and Rodnina, M.V. (2006) The ribosome's response to codon-anticodon mismatches. *Biochimie*, 88, 1001-1011.
 63. Stern, S. and Purohit, P. (1995) An oligonucleotide analog approach to the decoding region of 16S rRNA. *Biochem. Cell. Biol.*, 73, 899-905.
 64. Vicens, Q. and Westhof, E. (2001) Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure (Camb)*, 9, 647-658.
 65. Vicens, Q. and Westhof, E. (2002) Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding a site. *Chem Biol*, 9, 747-755.
 66. Vicens, Q. and Westhof, E. (2003) Crystal Structure of Geneticin Bound to a Bacterial 16S Ribosomal RNA A Site Oligonucleotide. *J Mol Biol*, 326, 1175-1188.
 67. Francois, B., Szychowski, J., Adhikari, S.S., Pachamuthu, K., Swayze, E.E., Griffey, R.H., Migawa, M.T., Westhof, E. and Hanessian, S. (2004) Antibacterial aminoglycosides with a modified mode of binding to the ribosomal-RNA decoding site. *Angew Chem Int Ed Engl*, 43, 6735-6738.
 68. Kondo, J., Francois, B., Russell, R.J., Murray, J.B. and Westhof, E. (2006) Crystal structure of the bacterial ribosomal decoding site complexed with amikacin containing the gamma-amino-alpha-hydroxybutyryl (haba) group. *Biochimie*, 88, 1027-1031.
 69. Kondo, J., Francois, B., Urzhumtsev, A. and Westhof, E. (2006) Crystal Structure of the Homo sapiens Cytoplasmic Ribosomal Decoding Site Complexed with Apramycin. *Angew Chem Int Ed Engl*.
 70. Kondo, J., Urzhumtsev, A. and Westhof, E. (2006) Two conformational states in the crystal structure of the Homo sapiens cytoplasmic ribosomal decoding A site. *Nucleic Acids Res*, 34, 676-685.
 71. Braun, I.C., Rohrbach, E., Schmitt, C. and Izaurralde, E. (1999) TAP binds to the constitutive transport element (CTE) through a novel RNA-binding motif that is sufficient to promote CTE-dependent RNA export from the nucleus. *Embo J*, 18, 1953-1965.
 72. Li, Y., Bor, Y.-c., Misawa, Y., Xue, Y., Rekosh, D. and Hammarskjold, M.-L. (2006) An intron with a constitutive transport element is retained in a Tap messenger RNA. *Nature*, 443, 234.

73. Liker, E., Fernandez, E., Izaurrealde, E. and Conti, E. (2000) The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. *Embo J*, 19, 5587-5598.
74. Ernst, R.K., Bray, M., Rekosh, D. and Hammariskjold, M.L. (1997) Secondary structure and mutational analysis of the Mason-Pfizer monkey virus RNA constitutive transport element. *Rna*, 3, 210-222.
75. Kang, Y., Bogerd, H.P., Yang, J. and Cullen, B.R. (1999) Analysis of the RNA Binding Specificity of the Human Tap Protein, a Constitutive Transport Element-Specific Nuclear RNA Export Factor. *Virology*, 262, 200.
76. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 31, 3406-3415.
77. Leontis, N.B. and Westhof, E. (1998) The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA*, 4, 1134-1153.
78. Lu, M. and Steitz, T.A. (2000) Structure of Escherichia coli ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-Å resolution. *Proc Natl Acad Sci U S A*, 97, 2023-2028.
79. Leontis, N.B., Lescoute, A. and Westhof, E. (2006) The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol*, 16, 279-287.
80. Lescoute, A. and Westhof, E. (2006) Topology of three-way junctions in folded RNAs. *Rna*, 12, 83-93.
81. Lescoute, A. and Westhof, E. (2006) The interaction networks of structured RNAs. *Nucleic Acids Res*.
82. Nielsen, H., Westhof, E. and Johansen, S. (2005) An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. *Science*, 309, 1584-1587.
83. Ying, S.Y., Chang, D.C., Miller, J.D. and Lin, S.L. (2006) The microRNA: overview of the RNA gene that modulates gene functions. *Methods Mol Biol*, 342, 1-18.
84. Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 1, 7.
85. Winkler, W.C. and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol*, 59, 487-517.
86. Geissmann, T., Possedko, M., Huntzinger, E., Fechter, P., Ehresmann, C. and Romby, P. (2006) Regulatory RNAs as mediators of virulence gene expression in bacteria. *Handb Exp Pharmacol*, 9-43.
87. Hamon, M., Bierne, H. and Cossart, P. (2006) *Listeria monocytogenes*: a multifaceted model. *Nat Rev Microbiol*, 4, 423-434.
88. van der Veen, R., Arnberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F. and Grivell, L.A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*, 44, 225.
89. Peebles, C.L., Perlman, P.S., Mecklenburg, K.L., Petrillo, M.L., Tabor, J.H., Jarrell, K.A. and Cheng, H.L. (1986) A self-splicing RNA excises an intron lariat. *Cell*, 44, 213.
90. Guo, F., Gooding, A.R. and Cech, T.R. (2004) Structure of the tetrahymena ribozyme; base triple sandwich and metal ion at the active site. *Mol Cell*, 16, 351-362.
91. Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *Rna*, 7, 499-512.
92. Rana, T.M. (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol*, 8, 23.

93. Tomari, Y. and Zamore, P.D. (2005) Perspective: machines for RNAi. *Genes Dev*, 19, 517-529.
94. Philippe, C., Eyermann, F., Benard, L., Portier, C., Ehresmann, B. and Ehresmann, C. (1993) Ribosomal protein S15 from *Escherichia coli* modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc Natl Acad Sci U S A*, 90, 4394-4398.
95. Ehresmann, C., Philippe, C., Westhof, E., Benard, L., Portier, C. and Ehresmann, B. (1995) A pseudoknot is required for efficient translational initiation and regulation of the *Escherichia coli* rpsO gene coding for ribosomal protein S15. *Biochem Cell Biol*, 73, 1131-1140.
96. Vogel, J., Argaman, L., Wagner, E.G. and Altuvia, S. (2004) The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr Biol*, 14, 2271-2276.
97. Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jager, J.G., Huttenhofer, A. and Wagner, E.G. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res*, 31, 6435-6443.
98. Willkomm, D.K., Minnerup, J., Huttenhofer, A. and Hartmann, R.K. (2005) Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog. *Nucleic Acids Res*, 33, 1949-1960.
99. Argaman, L. and Altuvia, S. (2000) fhlA repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. *Journal of Molecular Biology*, 300, 1101.
100. Lease, R.A. and Belfort, M. (2000) Riboregulation by DsrA RNA: trans-actions for global economy. MicroReview. *Molecular Microbiology*, 38, 667-672.
101. Novick, R.P. (2003) Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Molecular Microbiology*, 48, 1429-1449.
102. Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C. *et al.* (2005) *Staphylococcus aureus* RNAlII and the endoribonuclease III coordinately regulate spa gene expression. *Embo J*, 24, 824-835.
103. Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S. and Bassler, B.L. (2004) The Small RNA Chaperone Hfq and Multiple Small RNAs Control Quorum Sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell*, 118, 69.
104. MacRae, I.J. and Doudna, J.A. (2007) Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Current Opinion in Structural Biology*, 17, 138.
105. Morita, T., Mochizuki, Y. and Aiba, H. (2006) Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *PNAS* %R 10.1073/pnas.0509638103, 103, 4858-4863.
106. Mandin, P., Repoila, F., Vergassola, M., Geissmann, T. and Cossart, P. (2007) Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucl. Acids Res.* %R 10.1093/nar/gkl1096, 35, 962-974.
107. Sittka, A., Pfeiffer, V., Tedin, K. and Vogel, J. (2007) The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Molecular Microbiology*, 63, 193-217.
108. Vogel, J. and Sharma, C.M. (2005) How to find small non-coding RNAs in bacteria. *Biol Chem*, 386, 1219-1238.
109. Romby, P., Vandenesch, F. and Wagner, E.G.H. (2006) The role of RNAs in the regulation of virulence-gene expression. *Current Opinion in Microbiology*, 9, 229.
110. Winkler, W., Nahvi, A. and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, 419, 952-956.

111. Epshtein, V., Mironov, A.S. and Nudler, E. (2003) The riboswitch-mediated control of sulfur metabolism in bacteria. *PNAS* %R 10.1073/pnas.0531307100, 100, 5052-5056.
112. Wickiser, J.K., Winkler, W.C., Breaker, R.R. and Crothers, D.M. (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell*, 18, 49-60.
113. Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I. *et al.* (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A*, 101, 6421-6426.
114. Welz, R. and Breaker, R.R. (2007) Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA* %R 10.1261/rna.407707, rna.407707.
115. Sudarsan, N., Hammond, M.C., Block, K.F., Welz, R., Barrick, J.E., Roth, A. and Breaker, R.R. (2006) Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. *Science* %R 10.1126/science.1130716, 314, 300-304.
116. Nahvi, A., Barrick, J.E. and Breaker, R.R. (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res*, 32, 143-150.
117. Cromie, M.J., Shi, Y., Latifi, T. and Groisman, E.A. (2006) An RNA Sensor for Intracellular Mg²⁺. *Cell*, 125, 71.
118. Gilbert, S.D., Mediatore, S.J. and Batey, R.T. (2006) Modified Pyrimidines Specifically Bind the Purine Riboswitch. *J. Am. Chem. Soc.*, 128, 14214-14215.
119. Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R. and Bateman, A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucl. Acids Res.* %R 10.1093/nar/gki081, 33, D121-124.
120. Yousef, M.R., Grundy, F.J. and Henkin, T.M. (2005) Structural Transitions Induced by the Interaction between tRNA^{Gly} and the *Bacillus subtilis* glyQS T Box Leader RNA. *Journal of Molecular Biology*, 349, 273.
121. Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Finn, R.D. and Sonnhammer, E.L. (1999) Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucl. Acids Res.* %R 10.1093/nar/27.1.260, 27, 260-262.
122. Biarrotte-Sorin, S., Maillard, A.P., Delettre, J., Sougakoff, W., Arthur, M. and Mayer, C. (2004) Crystal Structures of *Weissella viridescens* FemX and Its Complex with UDP-MurNAc-Pentapeptide: Insights into FemABX Family Substrates Recognition. *Structure*, 12, 257.

**Articles publiés dans des
revues à comité de lecture**