



**HAL**  
open science

# Etude structurale et fonctionnelle du facteur d'épissage alternatif tissu spécifique MEC-8 chez *C.elegans*

Heddy Soufari-Rouba

► **To cite this version:**

Heddy Soufari-Rouba. Etude structurale et fonctionnelle du facteur d'épissage alternatif tissu spécifique MEC-8 chez *C.elegans*. Biochimie [q-bio.BM]. Université de Bordeaux, 2015. Français. NNT : 2015BORD0406 . tel-01552173

**HAL Id: tel-01552173**

**<https://theses.hal.science/tel-01552173v1>**

Submitted on 1 Jul 2017

**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.

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

**DOCTEUR DE**

**L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE

SPÉCIALITÉ BIOCHIMIE

Par Heddy SOUFARI

**Etude structurale et fonctionnelle du facteur  
d'épissage alternatif tissu spécifique MEC-8 chez  
*C.elegans***

Sous la direction de : Dr .Cameron MACKERETH CR1 INSERM

Soutenue le 10 Décembre 2015

Membres du jury :

**Pr Ijsbrand Kramer**  
**Dr Sophie Zinn-Justin**  
**Dr Cyril Dominguez**  
**Dr Virginie Gervais**

**Professeur Université de Bordeaux**  
**Directeur de recherche CEA Saclay**  
**Lecturer University of Leicester**  
**CR1 IPBS Toulouse**

**Président**  
**Rapporteur**  
**Rapporteur**  
**Examineur**





## REMERCIEMENT

Je souhaite tout d'abord remercier ma mère qui a toujours été présente pour moi, qui ma soutenue tout au long de m'a vie et m'a supporté dans chacune de mes décisions. Tu m'as appris à me battre pour mes ambitions, j'espère un jour pouvoir te rembourser pour tes sacrifices.

Camila sans toi ces trois dernières années aurait été monotones et pénibles. Ton succès j'en suis certain, sera à la hauteur de ta folie. Un jour on mangera Bio c'est promis... Notre tour du monde ne fait que commencer.

Dursun, Kashyap et Adrien, j'espère qu'on continuera d'allonger la liste de nos voyages insolites.

Je tiens à saluer mes amis resté en Alsace, vous m'avez terriblement manqués, je rentre bientôt promis !

Cameron je te suis très reconnaissant pour l'accueil que tu m'as offert au sein de ton laboratoire, je te souhaite bonne chance pour la suite.

Je remercie également les personnes de l'Institut Européen de Chimie et Biologie avec lesquels j'ai travaillé pendant ces trois ans : Samir Amrane, Aurore Guedin, Julia Guegueniat, Rico cornes, Lionel Beaurepaire, Brice Kauffman, Axel Innis, Stephane Thore.

Enfin je souhaite remercier les Dr Sophie Zinn-Justin, Cyril Dominguez, Virginie Gervais et Ijsbrand Kramer de m'avoir accordé de leur temps précieux pour juger mon travail.

## Index

Index.....	4
Liste des illustrations.....	8
Liste des tables.....	10
Résumé.....	12
Objectif de Thèse.....	13
Introduction.....	14
I. L'épissage des pre-ARN messagers.....	16
II. L'épissage alternatif.....	19
1) Les séquences régulatrice de l'épissage.....	21
2) Les protéines de la famille SR.....	23
3) Les protéines de la famille hnRNP.....	24
III. Les protéines liant l'ARN.....	26
1) Les protéines liant l'ARN.....	26
2) Les domaines de liaison à l'ARN.....	27
a) Le domaine dsRBD.....	28
b) Le domaine KH.....	30
c) Le domaine RRM.....	32
Introduction de MEC-8.....	35
1) La protéine MEC-8.....	35
2) Les gènes cible de MEC-8.....	38
1) Le gène unc-52.....	38
2) Le gène mec-2.....	39
3) Les gènes supresseurs de MEC-8.....	40
4) Les gènes de la famille sym.....	42
5) Les protéines RBPMS et RBPMS 2.....	44
6) La protéine CPO.....	47
Chapitre I : Structures et fonction du RRM1 de MEC-8.....	49
I. Etude structurale du premier domaine de MEC-8.....	49
1) Production et purification des protéines recombinantes.....	49
2) L'assignement des résonances de MEC-8 RRM1 libre.....	51
3) L'oligométrie du RRM1.....	52
4) Structure cristallographique de MEC-8 RRM1.....	54
II. Etude interactionnel de MEC-8 RRM1.....	58
1) Recherche d'une séquence de liaison nucléotidique optimal.....	58
2) Interaction du domaine RRM1 avec des motifs en tandem.....	61
III. Structures cristallographiques du RRM1 en complexe avec un acides nucléiques.....	64

1) Structure du complexe protéine-ADN.....	64
2) Structure du complexe protéine-ARN.....	69
IV. Importance de la région C-terminal du RRM1.....	72
V. Etude structurale en solution du complexe RRM1-ARN.....	76
VI. Etude de la famille de protéine RBPMS.....	79
VII. Matériels et méthodes.....	82
1) Clonage.....	82
2) Production.....	82
3) Purification.....	83
4) Synthèse des acides nucléiques.....	83
5) Spectroscopie de résonance magnétique nucléaire.....	83
6) Spectroscopie des rayons X.....	84
7) Titration calorimétrique isotherme.....	86
Chapitre II : Structure et fonction du RRM2 de MEC-8.....	87
I. Etude structurale du second domaine de MEC-8.....	87
1) Obtention d'un domaine protéique soluble.....	87
2) Assignement des résonances de MEC-8 RRM2.....	90
3) Etude dynamique de MEC-8 RRM2.....	91
II. Etude interactionnelle de MEC-8 RRM2.....	92
1) Analyse de l'interaction inter-protéique entre les domaines RRM1 et RRM2.....	92
2) Exploration de l'interaction protéine-ARN du RRM2 de MEC-8.....	94
III. Structure de MEC-8 RRM2 en complexe avec un acide nucléique.....	98
IV. Analogies RRM1/RRM2.....	104
V. Matériels et méthodes.....	106
1) Clonage, Production, Purification.....	106
2) Spectroscopie de résonance magnétique nucléaire.....	107
3) Spectroscopie des rayons X.....	107
4) Titration calorimétrique isotherme.....	108
Chapitre III : Le rôle de MEC-8 en tant que facteur d'épissage alternatif.....	109
Chapitre IV : Discussion générale.....	112
Conclusion de l'étude in-vitro des deux domaines RRM de MEC-8.....	112
MEC-8 est très similaire à la famille de protéine U1A/U2B''/SNF.....	117
La protéine MEC-8 reconnaît-elle les motifs en tige-boucle.....	120
Couch Potato Protein <i>D.mel</i> comme pierre angulaire dans l'évolution de la famille RBPMS.....	123
Les sites de liaison in-vivo potentiel de MEC-8.....	126
Les modèles d'action hypothétique du facteur MEC-8.....	129
Conclusion.....	131

Références bibliographique ..... 133



## Liste des illustrations

Figure 1 . Flux de l'information génétique et ses régulations.....	15
Figure 2. L'épissage des ARN messagers .....	16
Figure 3. Les séquences conservées du pré-ARN messager.....	16
Figure 4. Les deux étapes de transesterification de la réaction d'épissage .....	18
Figure 5. L'épissage alternatif des ARN messagers.....	19
Figure 6. Les différents modes d'épissage alternatif .....	20
Figure 7. Les séquences cis régulatrices de l'épissage.....	22
Figure 8. La famille des protéines SR humaine.....	23
Figure 9. La famille des protéine hnRNP .....	24
Figure 10. Association en clusters des modules de liaison à l'ARN chez les RBP .....	26
Figure 11. Les principaux domaines liant l'ARN .....	27
Figure 12. Le domaine dsRBD .....	29
Figure 13. Le domain KH .....	31
Figure 14. Le domaine RRM .....	32
Figure 15. Le domaine RRM, un motif conservé avec de nombreuses variantes .....	33
Figure 16. L'exemple de la protéine PTB.....	34
Figure 17. La famille de protéine RBPMS .....	35
Figure 18. Alignement des séquences des protéines de la famille RBPMS.....	36
Figure 19. Alignement de séquence des domaines RRM N-terminaux des protéines de la famille RBPMS.....	36
Figure 20. Les motifs de reconnaissance présumés .....	37
Figure 21. Les isoformes de l'ARN messager de unc-52 induits par MEC-8.....	38
Figure 22. L'isoforme de l'ARN messager de mec-2 induit par MEC-8 .....	39
Figure 23. Fonction des protéine SMU-1 et SMU-2 .....	41
Figure 24. Organisations structurales des protéines de la famille sym .....	43
Figure 25. Les isoformes de l'ARN messager de RBPMS.....	44
Figure 26. Séquence consensus de liaison calculé pour RBPMS.....	46
Figure 27. Le motif d'homodimérisation de RBPMS 2.....	46
Figure 28. Les informations structurales issus d'un spectre 1H-15N HSQC.....	49
Figure 29. Exploration structurale du domaine RRM1 de MEC-8 .....	50
Figure 30. Spectre 1H-15N HSQC du domaine RRM1(28-117).....	51
Figure 31. Corrélacion entre temps de corrélation rotationnel et poids moléculaire .....	52
Figure 32. Etude dynamique du domaine RRM1 de MEC-8 .....	53
Figure 33 : Structure cristallographique de MEC-8 RRM1.....	54
Figure 34. Détails structuraux de l'interface de dimérisation RRM1-RRM1 .....	55
Figure 35. Interface de dimérisation du domaine RRM1 .....	56
Figure 36. Spectre 1H-15N HSQC du complexe RRM1-AGCACA .....	58
Figure 37. Affinité par ITC des motif AGCACA ADN et ARN avec le domaine RRM1.....	59
Figure 38. La position 5 est permissive à la substitution par une adénine .....	60
Figure 39. Motif de liaison optimal du domaine RRM1 .....	60
Figure 40. Apport d'une liaison coopérative du dimère RRM1 .....	61
Figure 41. Spectre 1H-15N HSQC du complexe RRM1 avec un motif ARN en tandem.....	63
Figure 42. Structure du complexe RRM1 avec le motif d-AGCAC.....	65
Figure 43. Détails atomique du complexe RRM1-ADN.....	67
Figure 44. Optimisation des cristaux du RRM1 en complexe avec ARN .....	69
Figure 45. Le domaine RRM1 en complexe avec un motif ARN.....	70

Figure 46. Le domaine RRM1(28-123).....	72
Figure 47. Etude dynamique de la région C-terminal du RRM1.....	73
Figure 48. Apport d'une construction élargie en C-terminal en termes d'affinité.....	74
Figure 49. Spectre 1H-15N HSQC de MEC-8 27-123 en complexe avec AGCACA .....	75
Figure 50. Comparaison des déplacements chimiques entre les complexes avec motifs simples et motifs en tandems.....	75
Figure 51. Spectres 1H-13C half filtered NOESY .....	76
Figure 52. Spectres 1H-13C double half filtered NOESY .....	78
Figure 53. Le domaine RRM de la famille RBPMS .....	79
Figure 54. Reconnaissance bipartite de la famille RBPMS .....	81
Figure 55. Le domaine RRM2 de la famille de protéine MEC-8 .....	87
Figure 56. Obtention d'un domaine protéique soluble pour le RRM2 de MEC-8.....	88
Figure 57: Spectre 15N-H HSQC de gb1::RRM2 .....	90
Figure 58. Etude dynamique de gb1 ::RRM2 .....	91
Figure 59. Les domaines RRM1 et RRM2 n'interagissent pas ensemble.....	93
Figure 60. Le domaine RRM2 de MEC-8 se lie au motif AGCACA .....	94
Figure 61. Le domaine RRM2 de MEC-8 se lie à l'ARN et à l'ADN.....	95
Figure 62. Le motif de liaison optimal du domaine RRM2 de MEC-8.....	96
Figure 63. Spécificité de liaison du RRM2 .....	97
Figure 64. Structure cristallographique du complexe RRM2-ADN.....	99
Figure 65. Analyse de la dégradation du domaine gb1 ::RRM2 .....	100
Figure 66. Détails atomique du RRM2 en complexe avec d-AGCACA.....	102
Figure 67. RRM en tandems liants des séquences d'ARN répétées.....	104
Figure 68: Analogie structurale entre le RRM 1 et le RRM2.....	105
Figure 69. Spectre 1H-15N HSQC de MEC-8.....	109
Figure 70. Un rapporteur fluorescent de l'activité de MEC-8 .....	111
Figure 71. Un mode d'homodimérisation conservé pour la famille RBPMS.....	112
Figure 72. Modèle de liaison hypothétique.....	113
Figure 73. Comparaison du mode de liaison des domaines RRM N-terminaux de MEC-8 et RBPMS aux acides nucléiques .....	115
Figure 74. Modèle structural de la protéine MEC-8 .....	116
Figure 75. Un mode de liaison conservé .....	118
Figure 76. Détails structuraux du mode de liaison .....	119
Figure 77. Les protéines U1A et U2B lient préférentiellement des motifs en tige-boucle.....	120
Figure 78. Caractéristiques de la reconnaissance de la structure en tige boucle des protéines U1A/U2B ' .....	121
Figure 79. Affinité des domaines RRM de MEC-8 avec un motif tige-boucle .....	122
Figure 80. CPO D.mel isoforme S .....	123
Figure 81. Arbre phylogénétique de la famille de protéine RBPMS .....	124
Figure 82. Les mutants phénotypiques de mec-8.....	125
Figure 83.Unc-52 et mec-2, deux pré-ARN messagers cibles de MEC-8.....	127
Figure 84. Zk180.5 un exemple de gène affecté par un dysfonctionnement de MEC-8.....	128
Figure 85. Les différents modèles de l'action de MEC .....	130

## Liste des tables

<b>Table 1. Statistiques de la structure du domaine RRM1 .....</b>	<b>57</b>
<b>Table 2. Les affinités de MEC-8 RRM1 avec les mutants du motif AGCACA par ITC .....</b>	<b>60</b>
<b>Table 3. Affinité par ITC du RRM1 avec des ligands ADN comportant deux motifs.....</b>	<b>62</b>
<b>Table 4. Statistiques de la structure du complexe RRM1-ARN.....</b>	<b>68</b>
<b>Table 5. Statistiques de la structure du complexe RRM1-ADN.....</b>	<b>71</b>
<b>Table 6. Effet sur l'affinité de liaison des mutations de MEC-8.....</b>	<b>77</b>
<b>Table 7. Affinité par ITC, des membres de la famille RBPMS ;.....</b>	<b>80</b>
<b>Table 8. Les affinités de MEC-8 RRM1 avec les mutants du motif AGCACA par ITC .....</b>	<b>96</b>
<b>Table 9. Statistiques préliminaires de la structure du RRM2-ADN.....</b>	<b>103</b>



## Résumé

### Résumé

#### **Etude structurale et fonctionnelle du facteur d'épissage alternatif tissu spécifique MEC-8 chez *C.elegans***

Chez les organismes multicellulaires la diversité protéique dans chaque cellule et chaque tissu est obtenue initialement en régulant l'expression d'une partie des gènes d'un génome. Ces gènes sélectionnés peuvent ensuite être soumis à un épissage alternatif de sorte que certains exons sont retenus ou exclus dans l'ARNm final. Nous étudions les détails moléculaires de la protéine MEC-8, un facteur d'épissage tissu spécifique chez *Caenorhabditis elegans*. Les mutants MEC-8 sont responsables d'un phénotype insensible au touché chez *Caenorhabditis elegans*. Plus précisément, MEC-8 lie le pré-ARNm de *mec-2* un composant des récepteurs mécanosensoriels afin de réguler la production d'un isoforme particulier nécessaire pour la transduction du signal mécanosensoriel. Des études portant sur le motif conservé de reconnaissance à l'ARN (RRM) chez des orthologues des vertébrés (RBPMS) et des insectes (couch potato, CPO) ont démontré la présence d'un motif d'homodimérisation dans le domaine RRM1 de MEC-8. Cependant MEC-8 contient aussi un second domaine RRM dans sa partie C-terminale, domaine qui n'est pas retrouvé dans les protéines RBPMS et CPO. Nous avons donc exprimé chaque domaine RRM de MEC-8 indépendamment ainsi que la protéine entière et ces constructions ont été utilisées pour diverses expériences biophysiques. Nous avons ainsi identifié la séquence de liaison optimale pour les deux domaines RRM1 et RRM2. Ces analyses ont aussi été menées sur les domaines homologues issus des protéines RBPMS et CPO qui présentent une forte affinité pour la même séquence d'ARN. Nous avons donc découvert que malgré des différences de fonction et de localisation les membres de la famille RBPMS lient tous le même motif d'ARN. Les détails atomiques des deux RRM en complexe avec leur motif de liaison ont été obtenus en utilisant de la spectroscopie RMN et de la cristallographie des rayons X. Les deux complexes RRM-ligand de MEC-8 présentent de surprenantes similarités dans leur architecture.

Mots clés : Epissage alternatif, Biologie structurale, ARN

### Abstract

#### **Structural and functional study of the tissue specific alternative splicing factor MEC-8 from *C.elegans***

In multicellular organisms, proteomic diversity in each cell and tissue is provided initially by selective expression of gene subsets from the total genome, which are further subjected to alternative splicing, such that a different pattern of exons can be retained or excluded in the final protein coding mRNA. We are investigating the molecular details of the tissue-specific splicing factor protein MEC-8 from the worm *Caenorhabditis elegans*. The MEC-8 mutant protein is responsible for a touch insensitive phenotype in *Caenorhabditis elegans*, relating to its role as an alternative splicing factor. More precisely, MEC-8 can bind to the *mec-2* pre-mRNA, a component of mechanosensory receptor, to regulate the production of a certain isoform required for transducing the touch signal. Previous studies of the conserved RNA Recognition Motif (RRM) domain in orthologues from vertebrate (RBPMS) and insect (couch potato; CPO) have demonstrated a homodimerization motif in MEC-8 RRM1. However, MEC-8 also contains a second RRM domain in the C-terminus that is not found in the characterized RBPMS and CPO proteins. We have therefore expressed the independent RNA-binding domains of MEC-8 as well as the full-length protein and have used these constructs in a variety of biophysical assays. We identified the optimal RNA binding sequence for both the RRM1 and RRM2, and quantified the penalty of sequence variations. The investigation has also been extended to the homologous domains from human RBPMS and *Drosophila* CPO, which show a high affinity to the same RNA sequence. We therefore find that despite differences in function and localization, the members of the RBPMS protein family all bind to the same RNA motif. Atomic details of binding have also been obtained by using a combination of NMR spectroscopy and X-ray crystallography. The ligand-bound complexes reveal a surprising similarity in the architecture of the bound ligand for the first and second RRM domains from MEC-8.

Key words : Alternative splicing, Structural biology, RNA

## Objectif de Thèse

L'objectif de ce travail de thèse est de caractériser les structures tridimensionnelles des deux domaines structurés du facteur d'épissage alternatif MEC-8 de *C.elegans*. La résolution des structures à l'état libre et en complexe avec leurs partenaires biologique permettra une compréhension approfondie des mécanismes moléculaires impliqués dans la fonction de cette protéine.

Ce projet commencera donc par l'obtention de constructions solubles et correctement repliées de chacun des domaines de MEC-8. Il s'en suivra une caractérisation structurale de ceux-ci à l'état libre en solution.

Une seconde étape sera la définition des éventuelles séquences nucléiques reconnues par chaque domaine, ainsi que la caractérisation d'une éventuelle interaction des RRM entre eux. Ce travail sera principalement réalisé à l'aide des techniques de spectroscopie de Résonance Magnétique nucléaire (RMN) et de Titration Isotherme Calorimétrique (ITC).

La dernière étape de l'étude structurale de la protéine MEC-8 sera le calcul des structures de chacun des RRM en complexe avec leurs ligands. Ce travail fera appel à la spectroscopie RMN et des rayons X.

La finalité de ce travail sera d'élaborer un système rapporteur *in-vivo* et de l'exploiter afin d'étudier la fonction de MEC-8 chez l'organisme modèle *C.elegans*.

## Introduction

La cellule eucaryote est une unité très complexe. Elle a la capacité de se spécifier de façon intensive ainsi que de s'organiser en tissus élaborés pour participer à l'organisation physiologique complexe que peuvent arborer les organismes eucaryotes.

L'intégralité des caractéristiques que les cellules d'un organisme peuvent adopter est issue de la traduction de l'information contenue dans son génome. La différenciation entre deux tissus est le résultat d'une régulation spécifique et pointue du produit de l'expression de ce génome à plusieurs niveaux du flux de l'expression génique (*Figure 1*).

La régulation post-transcriptionnelle est un mécanisme important qui peut être divisé en plusieurs catégories : l'ajout de la coiffe 5', la poly-adénylation, l'épissage et épissage alternatif ainsi que l'édition des ARN. Ces mécanismes agissent directement sur la complexité et la diversité des protéomes eucaryotes (Maydanovych and Beal 2006, Keren, Lev-Maor et al. 2010).

Les protéines liant l'ARN jouent un rôle primordial dans les mécanismes de régulation post-transcriptionnelle. Elles fonctionnent en interagissant directement avec les molécules d'ARN, à différents niveaux spatiaux (noyaux, cytoplasme), temporels (différenciation, variations du milieu..) et fonctionnels (localisation, stabilisation, dégradation, modification...). L'étude de ces interactions protéine-ARN est donc cruciale pour la bonne compréhension des réseaux de régulations agissant dans la cellule.

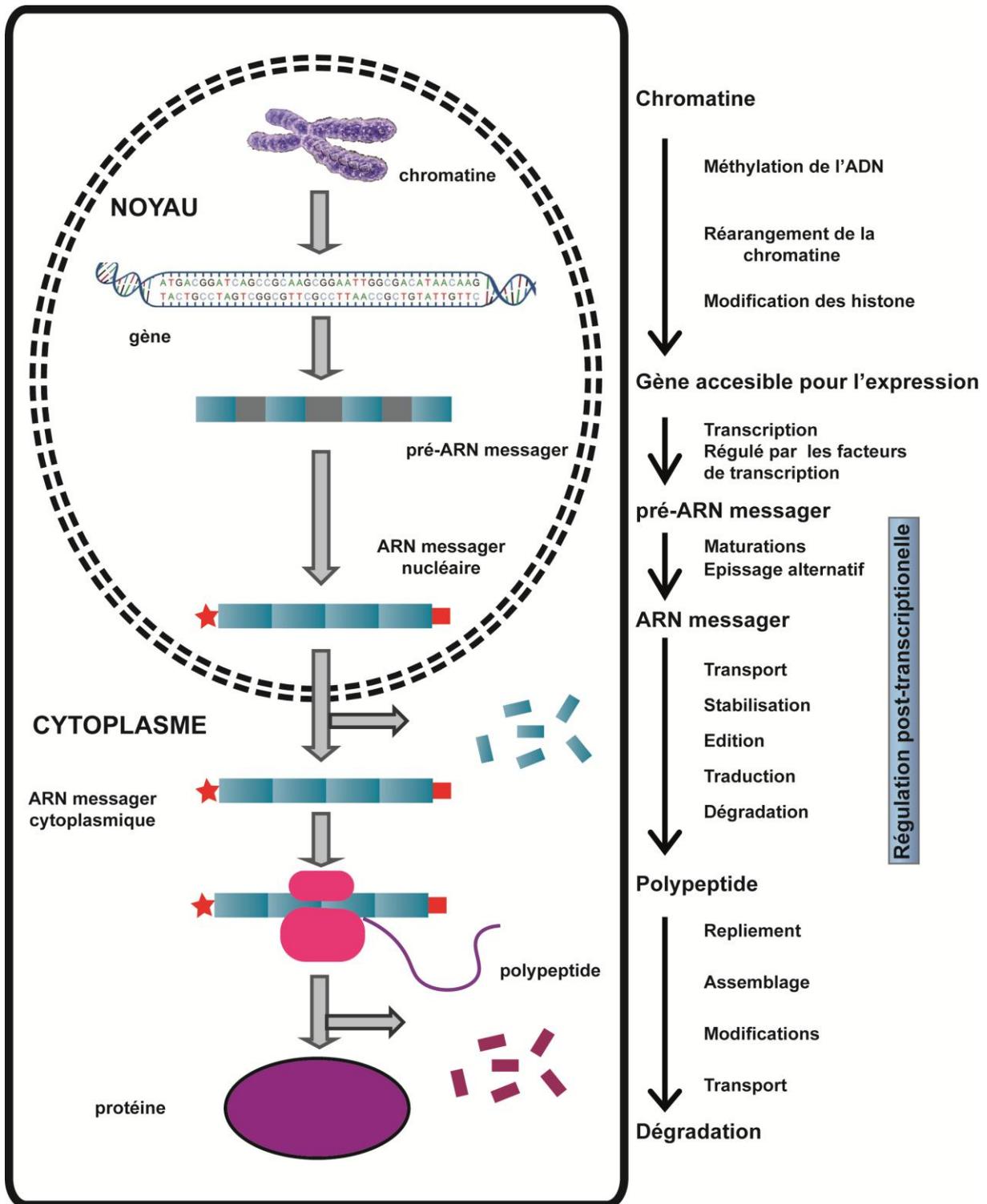
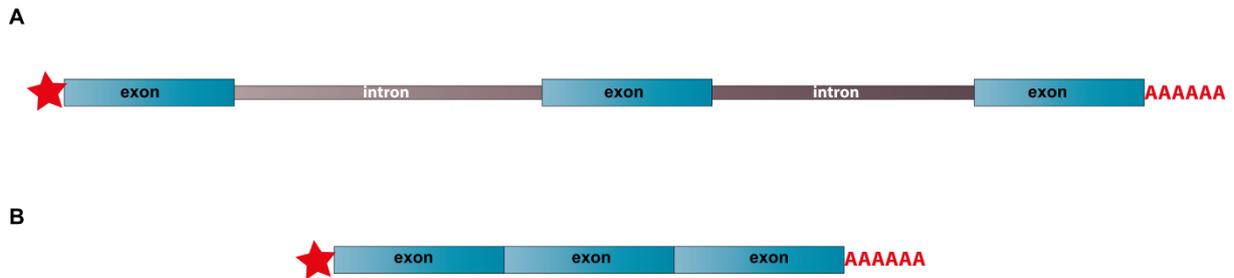


Figure 1 . Flux de l'information génétique et ses régulations

## I. L'épissage des pre-ARN messagers

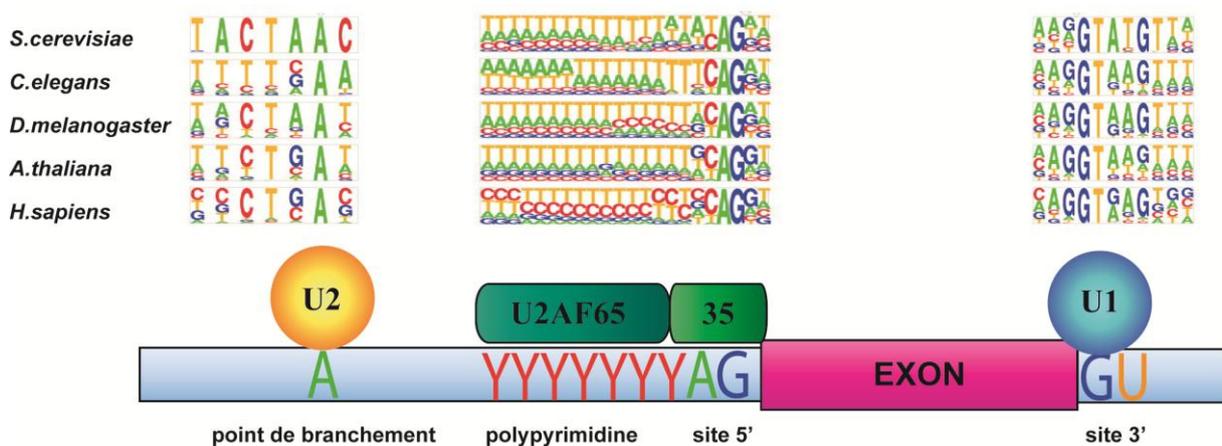
Les ARN messagers eucaryotiques sont synthétisés sous une forme dite de précurseurs, intégrant à la fois des séquences codantes (exons) et non codantes (introns). Le processus d'épissage des ARNm consiste à exciser les introns non codants tout en liant les exons les uns aux autres afin de produire un ARN messager dit mature (*Figure 2*).



**Figure 2. L'épissage des ARN messagers**

(A) Pré-ARN messenger comportant des sections codantes et non codantes. (B) ARN messenger mature débarrassé de ses introns.

La définition de la nature codante ou non des séquences du pré-ARNm est réalisée par la reconnaissance de séquences conservées aux jonctions exons/introns/exons par les sous unités du spliceosome (Mount 1982)(*Figure 3*).



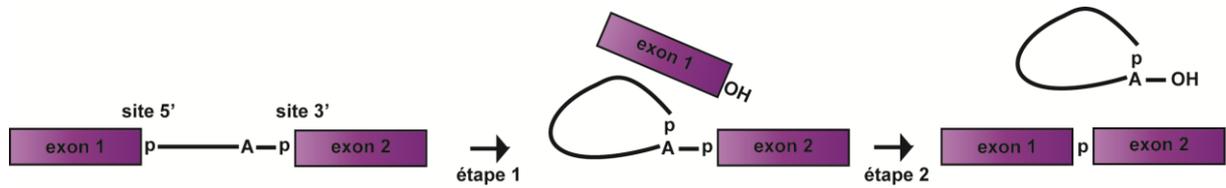
**Figure 3. Les séquences conservées du pré-ARN messenger**

Séquences consensus des principaux éléments conservés permettant la reconnaissance et l'épissage du pré-ARN messenger par les protéines du spliceosome. Inspiré de (Lim and Burge 2001)

Le spliceosome est un complexe ribonucléoprotéique impliqué à la fois dans l'excision de l'intron et la jonction des exons adjacents (Maniatis and Reed 1987). La reconnaissance des séquences conservées commence par la liaison du complexe U1snRNP au motif GU en 5' de l'intron. Dans un second temps la protéine SF1 se fixe à une adénine dite point de branchement, permettant le recrutement des protéines U2AF65 et 35 qui viendront se fixer à la région riche en pyrimidine et le site en 3' de l'intron caractérisé par la séquence conservée AG. Le second complexe snRNP à venir se fixer est U2 qui reconnaît le point de branchement de l'intron. Cela marque la fin de la reconnaissance des bornes de l'intron par le spliceosome, les étapes suivantes consistent en un réarrangement spatiotemporel de cet intron afin que la réaction d'épissage soit catalysée.

Pour cette étape, trois autres snRNP sont recrutées dans le complexe spliceosomal. Tout d'abord U4/U6, U6 prend la place de U1 et se lie à U2 (Stark and Luhrmann 2006). L'interaction entre U6 et U2 va engendrer un réarrangement structural qui permettra de rapprocher l'extrémité 5' à l'adénine du point de branchement ainsi que de recruter le dernier partenaire, le complexe snRNP U5.

Ce réarrangement structural engendre le départ des facteurs U1 et U4, conduisant à la première réaction de transestérification. Le 2'OH libre de l'adénine du point de branchement attaque la liaison Phosphodiester en 3' du G de l'exon1 (*Figure 4*). Le résultat est la formation d'un Phosphotriester 2',3',5' P, un arrangement peu stable qui va rapidement se cliver du côté 3' de la guanine de l'exon 1, le 2'OH de l'adénine reste lié au premier G de l'intron. Il s'en suit la formation d'une structure en forme de lasso de cet intron.



**Figure 4. Les deux étapes de transesterification de la réaction d'épissage**

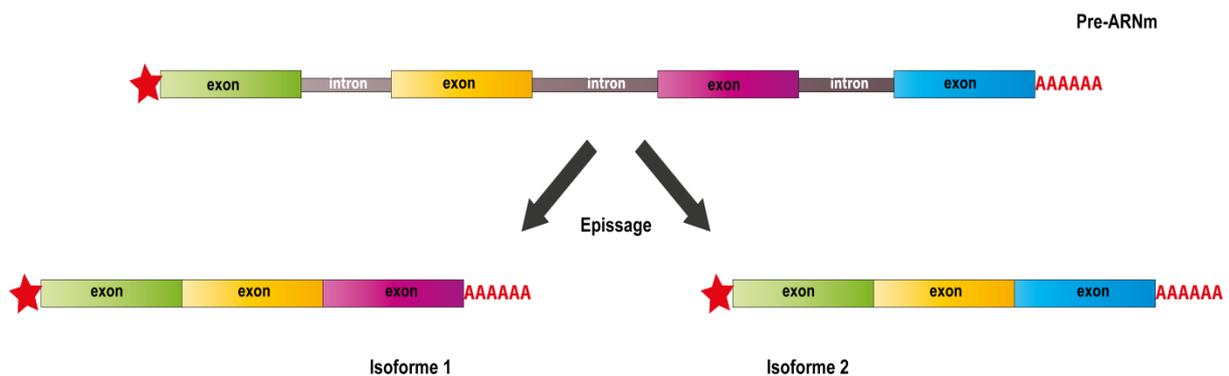
Les deux étapes de transesterification permettant l'exclusion des introns et la jonction des exons du pré-ARN messager.

Pour que l'épissage de l'intron soit complet une seconde réaction de transesterification est nécessaire. Cette réaction a pour but de joindre l'extrémité 3' de l'exon 1 avec l'extrémité 5' de l'exon 2 tout en libérant l'intron (Padgett, Grabowski et al. 1986, Ruby and Abelson 1991) (**Figure 4**). Pour cela il est nécessaire d'avoir une proximité entre les exons 1 et 2, ceci est réalisé par le complexe snRNP U5.

## II. L'épissage alternatif

L'épissage alternatif est un processus qui introduit des variants d'ARN messager à partir d'un même pré-ARNm (*Figure 5*). Cela a pour conséquence une augmentation importante de la diversité protéique au sein de l'organisme, démultipliant la capacité codante pour un génome de taille relativement faible.

L'épissage alternatif est un mécanisme très important qui touche plus de 90% des gènes chez l'humain (Pan, Shai et al. 2008) et n'a pas pour seul but d'introduire de la diversité génétique. Il permet également la régulation de l'expression génique, en effet un nombre important des ARN messagers ayant subi un épissage alternatif ne sont jamais traduits en protéine mais sont dirigés vers la voie NMD (Non-sense RNA Mediated Decay) pour être dégradés (Nagy and Maquat 1998, Stamm, Ben-Ari et al. 2005).



**Figure 5. L'épissage alternatif des ARN messagers**

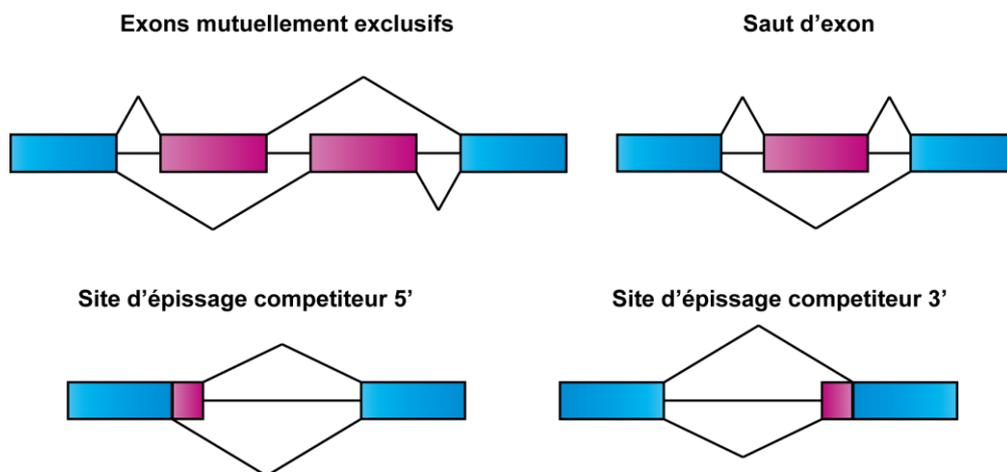
Deux potentielles isoformes d'ARN messager suite à l'épissage alternatif d'un même pré-ARNm.

La différenciation cellulaire est un autre exemple de processus dans lequel l'épissage alternatif joue un rôle majeur (Black 2003). Une certaine isoforme d'un gène peut être nécessaire au développement d'un tissu particulier, en effet un réarrangement de la séquence d'une protéine peut avoir pour conséquence une modification de sa structure, ce qui peut engendrer des modifications de sa fonction (spécificité de liaison, localisation,...). Un facteur

d'épissage alternatif sera alors exprimé dans ces tissus précis afin de réguler cette expression spécifique.

Les facteurs d'épissage alternatif sont des protéines qui vont influencer un événement d'épissage donné en venant se lier sur des sites particuliers du pré-ARNm. Ces sites sont des séquences situées aussi bien sur les exons que les introns, ils peuvent avoir un effet de régulation positif ou négatif sur l'épissage (activateur/répresseur) (**Figure 6**).

L'épissage alternatif d'un gène est classé de différentes façons selon le mécanisme par lequel il est achevé (**Figure 6**)



**Figure 6. Les différents modes d'épissage alternatif**

Représentation schématique des différents modes d'épissage alternatif. En rose sont représentés les exons pouvant subir un épissage alternatif.

## 1) Les séquences régulatrices de l'épissage

Les séquences régulatrices de l'épissage appelées aussi éléments cis, sont des séquences courtes (6-12nt) reconnues par les domaines de liaison à l'ARN des facteurs d'épissage. Ces séquences peuvent être promotrices ou inhibitrices de l'épissage, on les trouve aussi bien dans les exons que dans les introns (*Figure 7*). On distingue donc quatre types de séquences selon leurs spécificités :

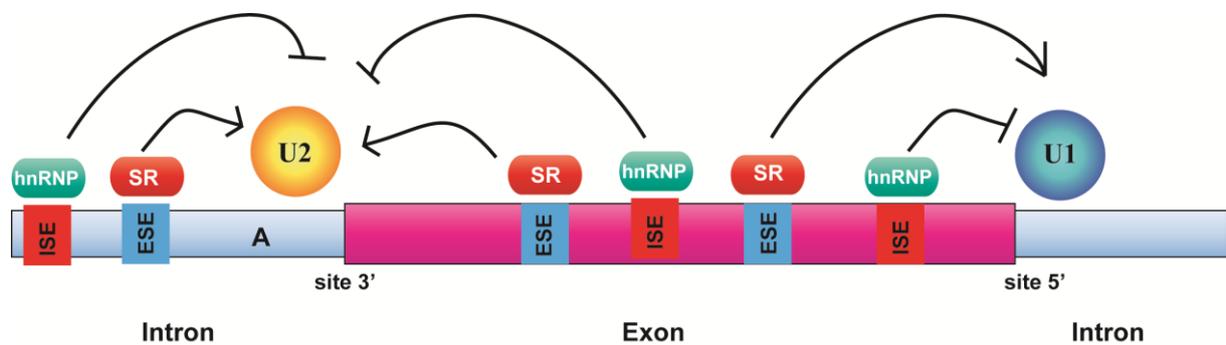
- Les ESE (Exonic Splicing Enhancer) sont des régions présentes dans la plupart voir la totalité des exons chez l'humain (Wang and Burge 2008). La majorité des ESE fonctionnent en recrutant des protéines de la famille SR qui s'associent à ces séquences par le biais de leur domaine RRM (Graveley and Maniatis 1998). Cette association des protéines SR engendre le recrutement de la machinerie d'épissage.
- Les ESS (Exonic Splicing Silencer) sont comme les ESE très répandus dans les séquences exoniques. Elles recrutent le plus souvent des protéines de la famille des hnRNP qui interagissent par le biais de leurs domaines RRM (parfois KH) avec ces séquences régulatrices (Pozzoli and Sironi 2005). L'interaction des hnRNP avec les ESS a pour conséquence d'inhiber l'association du spliceosome par différents mécanismes (Sharma, Falick et al. 2005, Wang and Burge 2008)

Les ESE et ESS jouent un rôle important dans l'épissage alternatif mais aussi dans l'épissage canonique où ils aident à la définition des exons.

- Les ISE (Intronic Splicing Enhancer) et ISS (Intronic Splicing Silencer) sont eux principalement régulateurs de l'épissage alternatif. Les séquences de ces sites sont beaucoup moins bien répertoriées que pour les ESE et ESS, ainsi que les protéines qui

les reconnaissent. Les ISE peuvent être liées par des protéines SR (Hastings, Wilson et al. 2001) ou des protéines hnRNP (Hui, Hung et al. 2005), ces séquences ont pour rôle de promouvoir l'épissage.

Les séquences promotrices et inhibitrices ont souvent des rôles antagonistes, la définition finale du pattern d'épissage est donc une addition d'événements complexes, faisant intervenir différents partenaires, mécanismes et conditions.

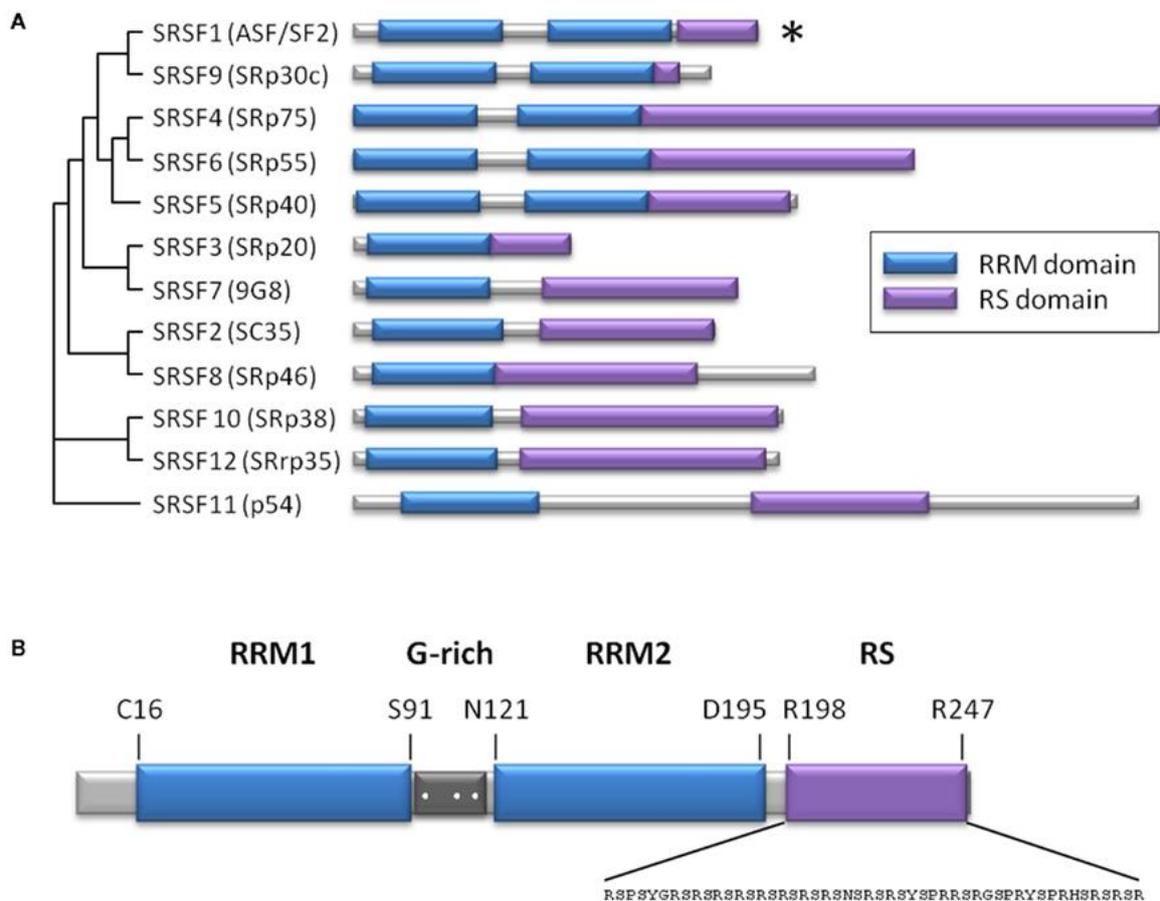


**Figure 7. Les séquences cis régulatrices de l'épissage**

## 2) Les protéines de la famille SR

Les protéines de la famille SR humaines sont des protéines liant l'ARN (RBP) qui ont la particularité de posséder en plus d'un ou plusieurs domaines RRM un domaine riche en résidus sérine et arginine (SR). Il existe 12 protéines SR bien caractérisées à ce jour (Manley and Krainer 2010) (*Figure 8*).

Les protéines SR jouent un rôle important dans le mécanisme d'épissage canonique en interagissant avec des éléments de régulations situés sur les exons (ESE), par exemple la protéine SRSF1 (ASF/SF2) recrute le complexe U1 sur le site de reconnaissance 5' (Cho, Hoang et al. 2011) et la protéine SRSF2 (SC35) induit le recrutement du complexe U2SNRNP au point de branchement (Fu and Maniatis 1992).



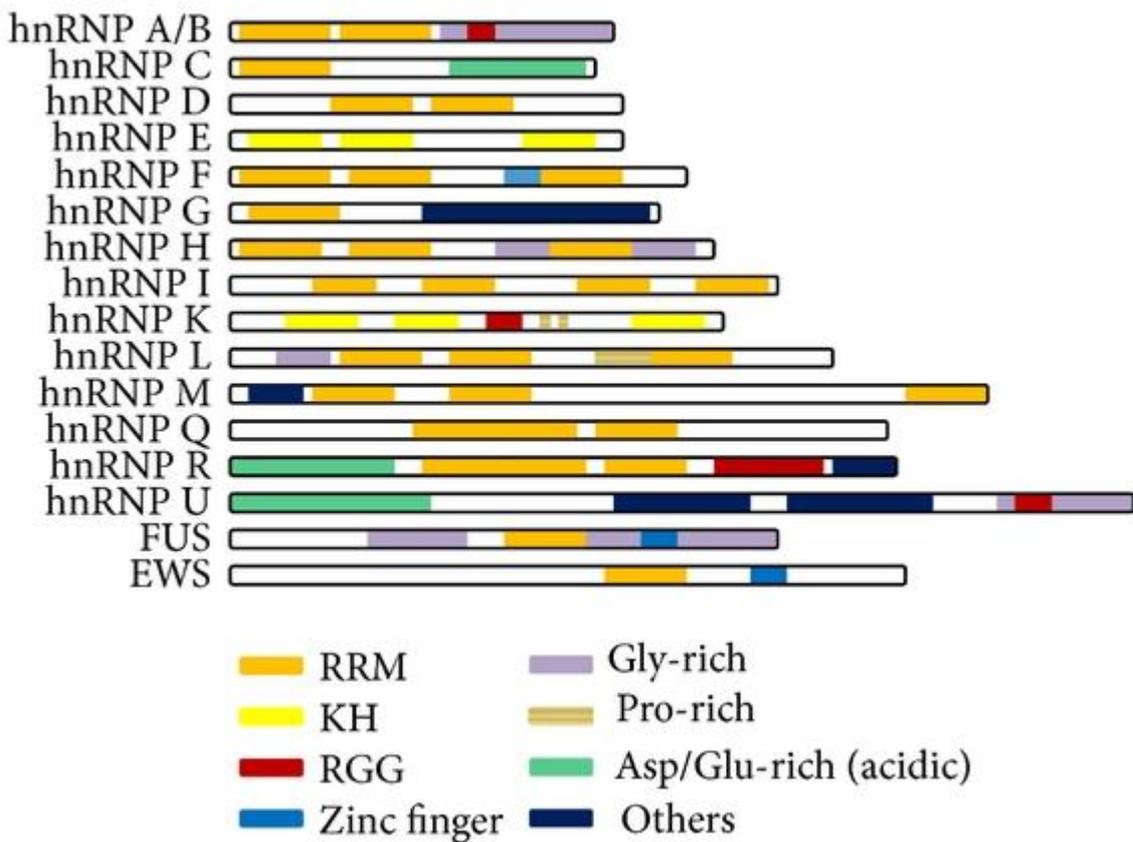
**Figure 8. La famille des protéines SR humaine.**

*Organisation structurale des protéines de la famille SR humaines. Issu de (Blanco and Bernabeu 2012)*

### 3) Les protéines de la famille hnRNP

La famille hnRNP est constituée de nombreuses protéines (*Figure 9*) qui ont toute la particularité de posséder un ou plusieurs domaines de liaison à l'ARN ainsi qu'un domaine riche en glycine et arginine (RGG) qui peut lui aussi interagir avec l'ARN ou un partenaire protéique (Chabot, LeBel et al. 2003).

Le rôle primaire de cette famille de protéine est d'interagir avec les molécules d'ARN à tout moment et à chaque étape de leur cycle de vie.



**Figure 9. La famille des protéine hnRNP**

*Organisation structurale des protéines de la famille hnRNP humaines. Issu de (Chen, Jin et al. 2013).*

Les protéines hnRNP sont plus connues pour leur rôle dans la régulation de l'épissage alternatif où elles se lient le plus souvent à des séquences inhibitrices de l'épissage (ESS, ISS) afin de bloquer l'action du spliceosome.

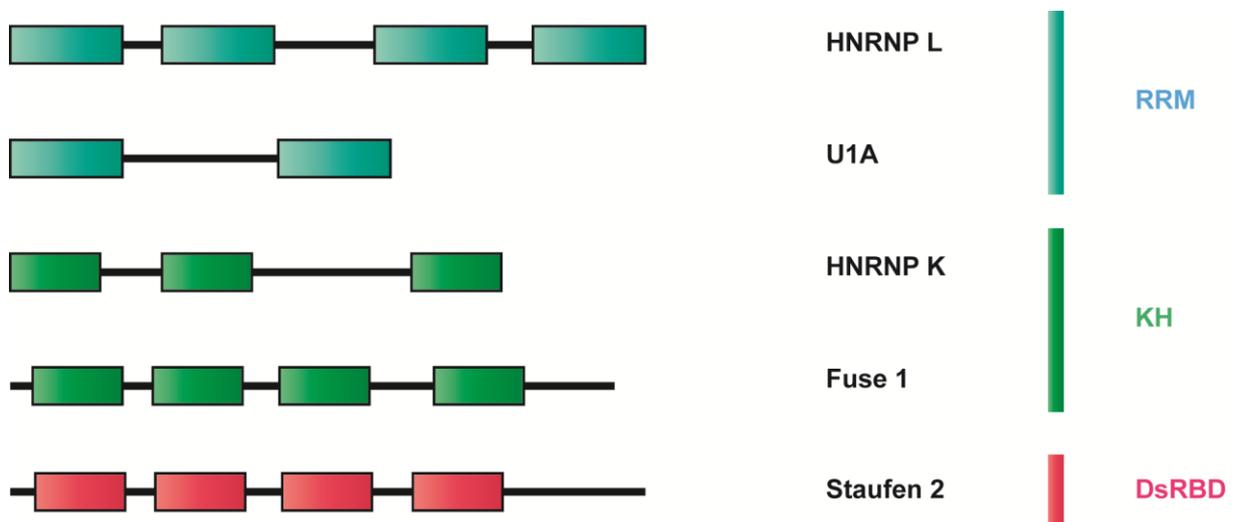
Un exemple d'action de régulation de l'épissage par une protéine hnRNP est le cas de la protéine PTB (Polypyrimidine Tract Binding). La protéine PTB (hnRNP I) a été très étudiée, elle possède quatre domaines RRM (Oh, Hahm et al. 1998) qui lient des régions riches en pyrimidine (Garcia-Blanco, Jamison et al. 1989), elle a de nombreuses fonctions dans différents mécanismes cellulaires incluant la régulation de l'épissage alternatif.

PTB est plus particulièrement connue pour son rôle de répresseur de l'épissage de nombreux gènes où elle agit de différentes façons. Un premier modèle d'action de PTB est sa liaison à la région riche en pyrimidine normalement reconnue par le facteur U2AF empêchant *de facto* l'assemblage du spliceosome (Lin and Patton 1995). Un autre modèle propose la formation par la protéine PTB d'une boucle dans l'ARN séquestrant une partie de celui-ci provoquant soit l'inhibition de la définition d'un exon pour Fas (Izquierdo, Majos et al. 2005) ou de l'intron pour c-src (Sharma, Falick et al. 2005)

### III. Les protéines liant l'ARN

#### 1) Les protéines liant l'ARN

Les molécules d'ARN présentes dans un organisme s'organisent sous forme de complexes avec une grande variété de protéines. Les protéines liant l'ARN (RBP :RNA Binding Protein) reconnaissent les molécules d'ARN à tout instant de leurs vies, et sous leurs différentes formes. Ces interactions sont impliquées dans de nombreuses fonctions (Stabilisation, localisation, transport, dégradation...) (Dreyfuss, Kim et al. 2002). Une telle diversité de fonctions et de cibles moléculaires devrait être synonyme d'une importante diversité structurale de ces RBP. Cependant la grande majorité des protéines liant l'ARN est composée d'un nombre réduit de modules de liaison à l'ARN (Burd and Dreyfuss 1994, Lunde, Moore et al. 2007). La spécificité des RBP pour leur substrat est assurée par la répétition de ces domaines de liaison à l'ARN de façons particulières (*Figure 10*).

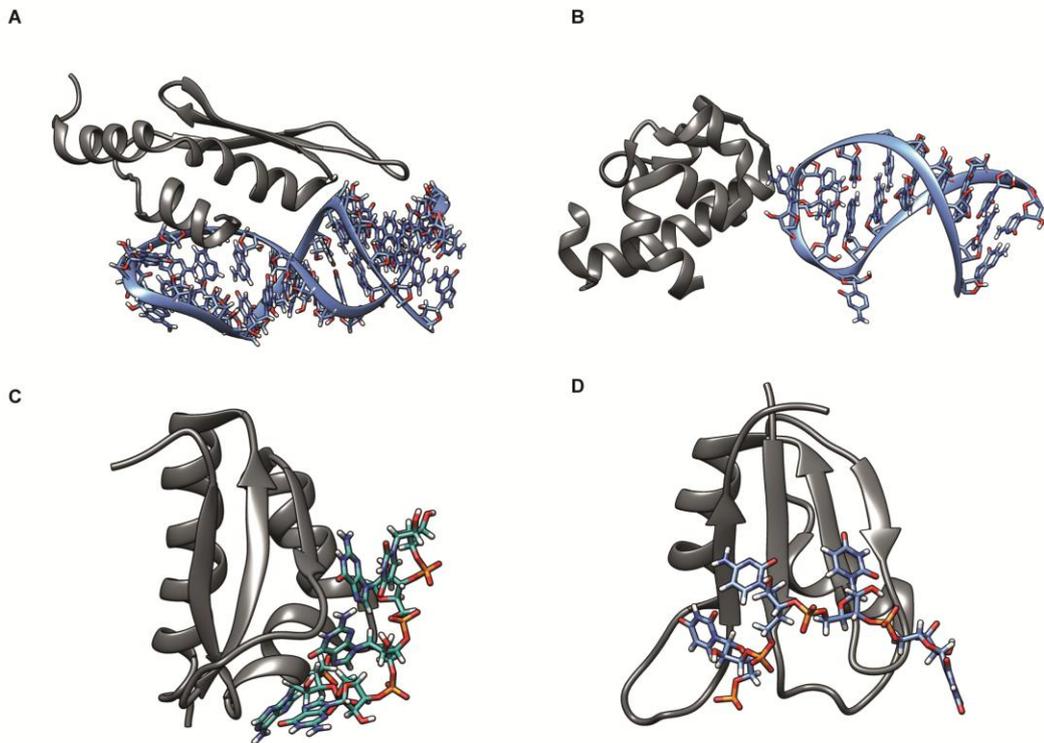


**Figure 10. Association en clusters des modules de liaison à l'ARN chez les RBP**

Représentation de protéines humaines arborant de multiples copies de domaines liant l'ARN. RRM : RNA Recognition Motif. KH : K Homology domain. DsRBD : Double stranded RNA Binding Domain.

## 2) Les domaines de liaison à l'ARN

Les domaines de liaison à l'ARN ont des repliements et des tailles qui leur sont propres mais ont tous la particularité d'interagir de façon très spécifique avec leurs cibles. On peut classer ces domaines en deux catégories différentes selon qu'ils reconnaissent leurs cibles en fonction de leurs structures (DsRBD, doigts de zinc), ou de leurs séquences nucléotidiques (RRM, KH) (*Figure 11*).



**Figure 11. Les principaux domaines liant l'ARN**

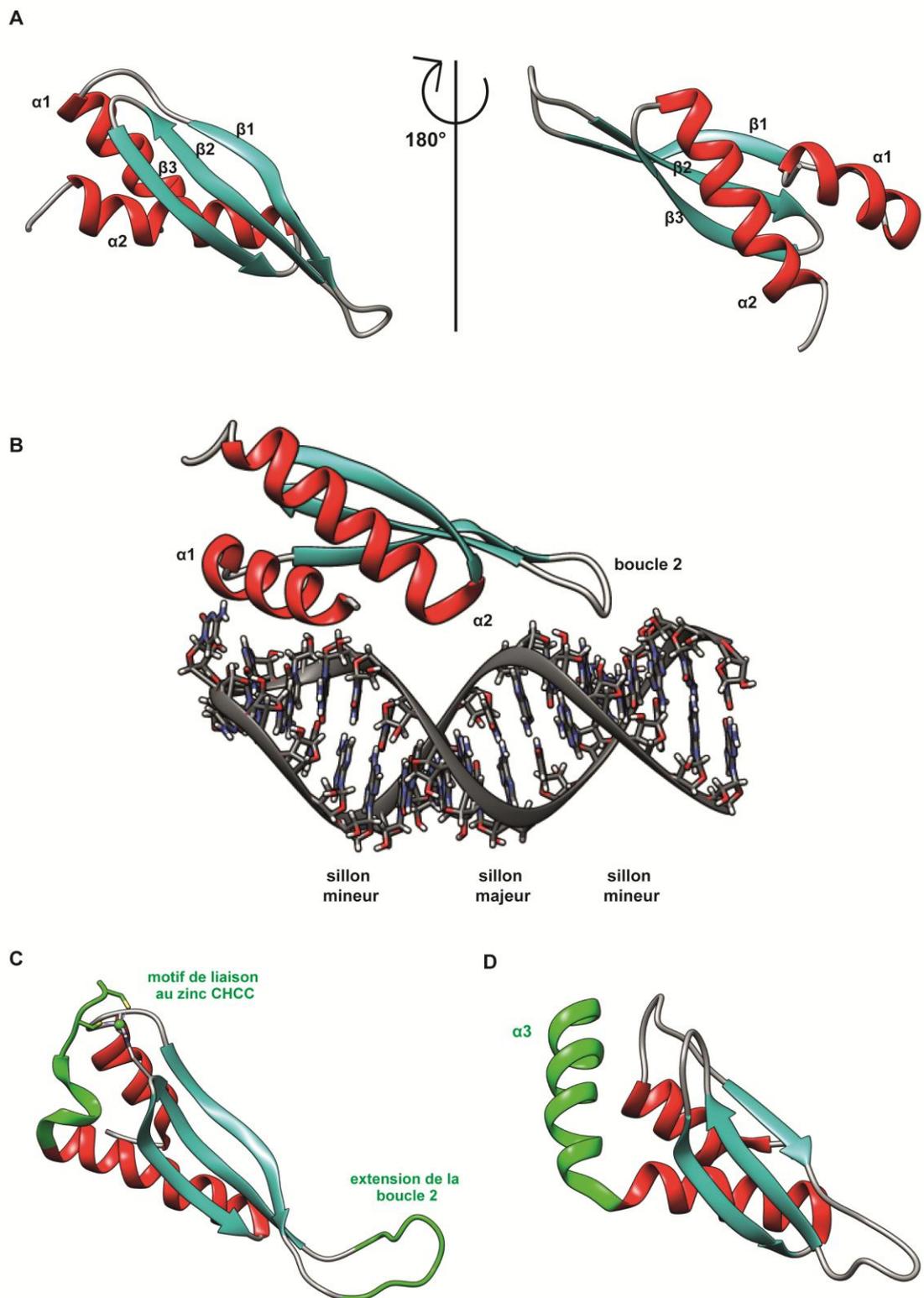
(A) *Double stranded RNA binding domain en complexe avec un ARN double brin (pdb : 2LBS).* (B) *SAM domain en complexe avec un ARN en tige-boucle (pdb : 2B6G).* (C) *Domaine KH en complexe avec un ARN simple brin riche en G (pdb : 4B8T).* (D) *Domaine RRM de Nab3 en complexe avec un ARN simple brin UCUU (pdb : 2L41).*

### a) *Le domaine dsRBD*

Le domaine dsRBD (double stranded RNA Binding Domain) est un domaine protéique qui est composé d'environ 70 acides aminés adoptant un repliement  $\alpha\beta\beta\alpha$ , où les deux hélices font face au feuillet beta (**Figure 12.A**). Sa fonction principale est de lier spécifiquement les ARN sous forme double brins. On le retrouve dans la plupart des processus biologiques comme l'édition (Stefl, Oberstrass et al. 2010) ou le silencing des ARN (Gredell, Dittmer et al. 2010).

La reconnaissance spécifique des hélices double brins d'ARN par le domaine dsRBD est effectuée par la liaison du sillon majeur par l'hélice 1 et la boucle 2, le sillon mineur est lié par l'hélice 2 (**Figure 12.B**). La discrimination entre hélices ARN et ADN est basée sur la différence de taille des sillons de chaque espèce, permettant l'interaction avec des bases du sillon mineur par le dsRBD sous la forme ARN, ainsi que la présence de groupement 2'O-H dans le sillon majeur absent dans les hélices ADN .

Plusieurs variantes du repliement du dsRBD ont été découvertes à ce jour, la boucle 2 entre les brins  $\beta_1$  et  $\beta_2$  peut avoir une taille plus importante dans certains cas (**Figure 12.C**). La région C-terminale peut accommoder une troisième hélice (Leulliot, Quevillon-Cheruel et al. 2004) voir un motif de liaison au zinc dans le cas de la protéine Dcr1 *K. polysporu* (Weinberg, Nakanishi et al. 2011) (**Figure 12.D**). Ces variations ont montré un intérêt dans la stabilité des domaines dsRBD et aident ainsi à l'interaction avec leurs ARN.



**Figure 12. Le domaine dsRBD**

(A) représentations du domaine dsRBD canonique de la protéine TAR rbp2 (pdb :2CPN). (B) Complexe entre le dsRBD de ADAR2 et l'ARN lsl (pdb : 2L3C). (C) dsRBD c-terminal de Dcr1(pdb : 2L6M). (D)dsRBD de Rnt1p (pdb : 1T4N).

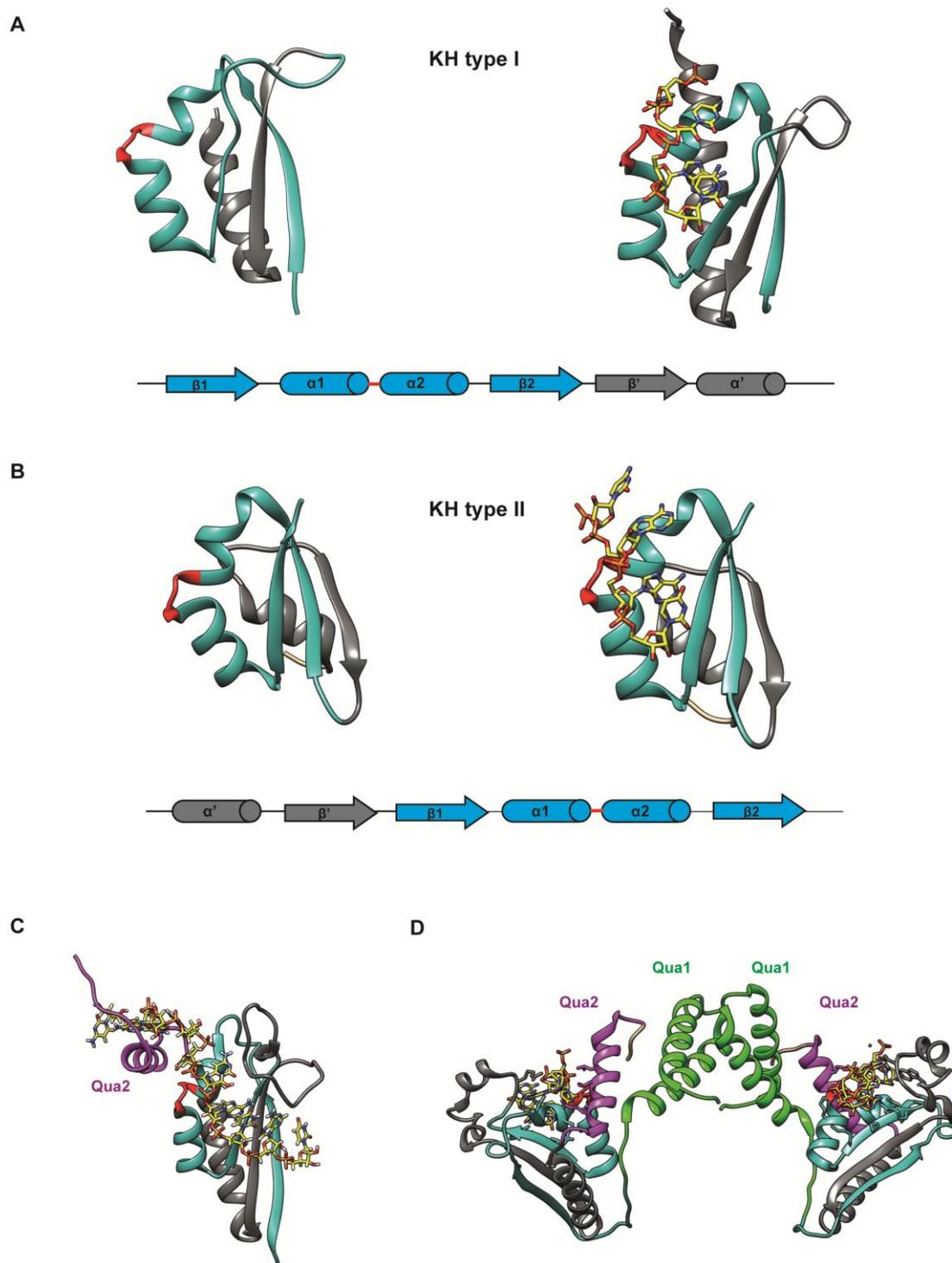
### *b) Le domaine KH*

Le K Homology domain (KH) est composé d'environ 70 acides aminés. Une distinction est faite entre deux types selon leur repliement. Le type I est sous la forme  $\beta\alpha\alpha\beta\beta\alpha$  (*Figure 13.A*), le type II est lui sous la forme  $\alpha\beta\beta\alpha\alpha\beta$  (*Figure 13.B*).

Les domaines KH lient quatre acides nucléiques par le biais d'une poche hydrophobe créée par les hélices 1 et 2 et la boucle conservée GxxG qui les sépare d'un côté et le brin  $\beta 2$  de l'autre. Malgré un repliement différent les deux types de domaines lient donc les ARN ou ADN avec le même motif canonique  $\beta\alpha\alpha\beta$  (*Figure 13*).

Les protéines possédant des domaines KH augmentent la taille de leurs séquences de reconnaissance de deux manières. Premièrement en multipliant leur nombre de domaine KH allant jusqu'à 15 copies chez la protéine vigilin. Le second moyen particulièrement utilisé par la famille des protéines STAR est d'ajouter des motifs structuraux supplémentaires au domaine KH.

Ainsi on peut observer en plus du domaine KH un domaine N-terminal Qua1 (Chen, Damaj et al. 1997, Teplova, Hafner et al. 2013) qui permet la dimérisation du domaine et une hélice C-terminale Qua2 impliqué dans la reconnaissance à l'ARN permettant d'accommoder deux nucléotides supplémentaire (Daubner, Brummer et al. 2014) (*Figure 13.C.D*)

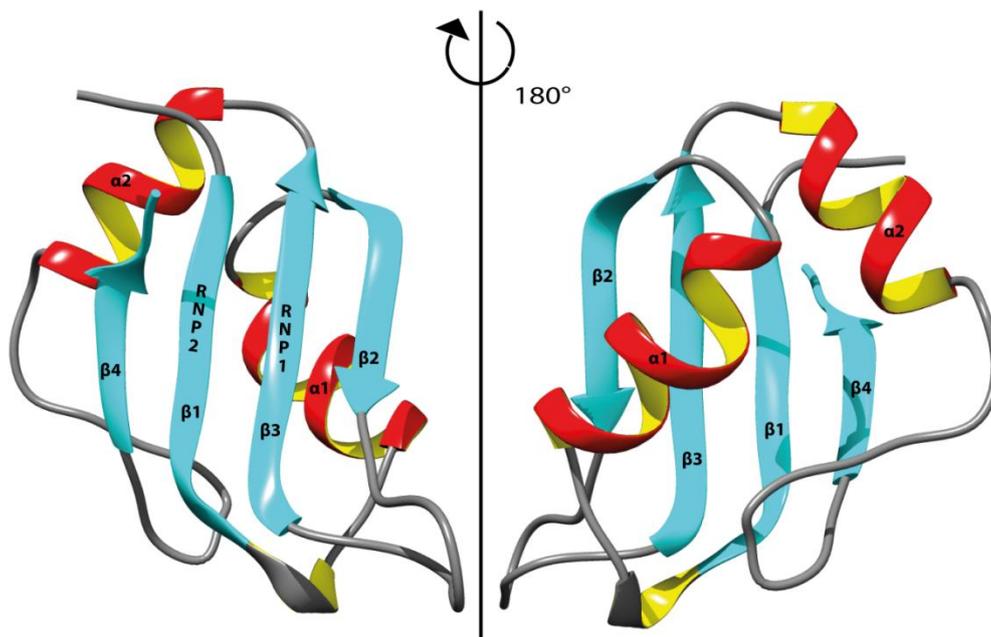


**Figure 13. Le domaine KH**

(A) . (B) Dmaine KH 2 de Nusa libre et en complexe avec ARN. (C) Domaine KH de Gld1 en complexe avec ARN, le motif Qua2 est en mauve. (D) Protéine Gld1 en complexe avec ARN sous forme de dimère avec le motif Qua2 en mauve et Qua1 en vert.

### c) Le domaine RRM

Le domaine RRM (RNA Recognition Motif) est d'une taille d'environ 90 acides aminés qui forme un repliement très conservé de type  $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$  où les brins beta forment un feuillet faisant face aux deux hélices (Maris, Dominguez et al. 2005). La séquence peptidique d'un RRM peut-être très variable mais deux éléments consensus impliqués dans la liaison à l'ARN restent conservés sur l'ensemble de la famille, les motifs RNP1 et RNP2 (**Figure 14**). Ces deux motifs sont situés respectivement sur le brin  $\beta 1$  pour RNP2 et  $\beta 3$  pour RNP1.



RNP2 : ([ILV]-[FY]-[ILV]-X-N-L)    RNP1 : ([KR]-G-[FY]-[GA]- [FY]-[ILV]-X-[FY])

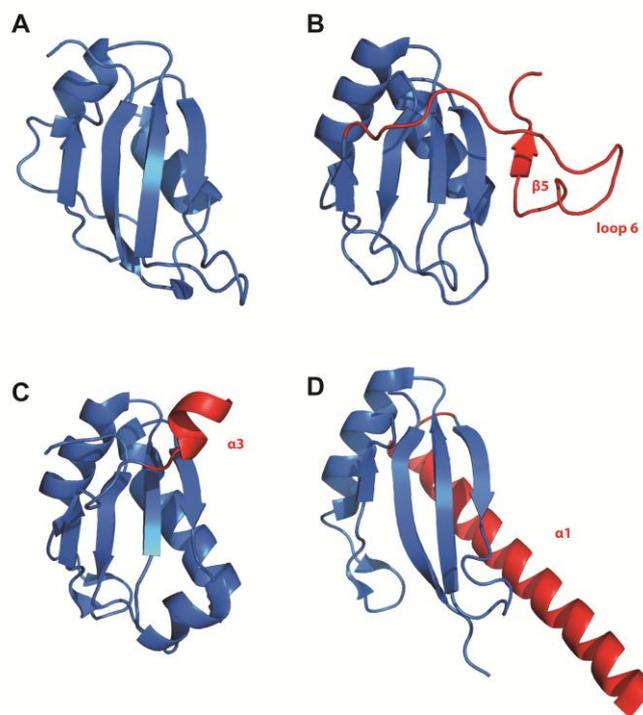
**Figure 14. Le domaine RRM**

*Le domaine RRM canonique vue de face (feuillet beta) et de dos (hélices alpha). Les brins  $\beta$  sont couleur cyan, les hélices rouges à l'extérieur et intérieur jaune, enfin les boucles sont en gris. Les séquences consensus des deux motifs RNP sont représentées en dessous du domaine*

Parmi les résidus conservés des motifs RNP deux d'entre eux ont été décrit comme éléments clés pour la reconnaissance des ARN simple brin par le RRM. Il s'agit des deux résidus hydrophobes en position 5 du RNP1 et 2 du RNP2 sur lesquels on observe un empilement aromatique de deux bases sur les deux cycles aromatiques de ces résidus. Ces résidus sont

quasiment toujours présents chez les membres de la famille des RRM. Dans la plupart des structures de complexe RRM-ARN, le premier nucléotide interagit avec le brin  $\beta 4$ , les deux suivants interagissent avec les résidus hydrophobes des deux motifs RNP et un dernier vient se lier au brin  $\beta 2$  (Auweter, Oberstrass et al. 2006). Ces caractéristiques communes sont donc la base de la reconnaissance de l'ARN par le RRM, cependant la spécificité de liaison par le RRM en termes de séquence nucléotidique provient de résidus non conservés. Ces résidus sont souvent trouvés dans les boucles et la région C-terminale du domaine.

Une interface pouvant couvrir une séquence entre 3 et 7 nucléotides peut être limitante en termes de spécificité notamment pour des processus biologique cruciaux demandant une régulation précise. Pour cela plusieurs éléments peuvent venir s'ajouter au motif principal conservé du domaine RRM.



**Figure 15. Le domaine RRM, un motif conservé avec de nombreuses variantes**

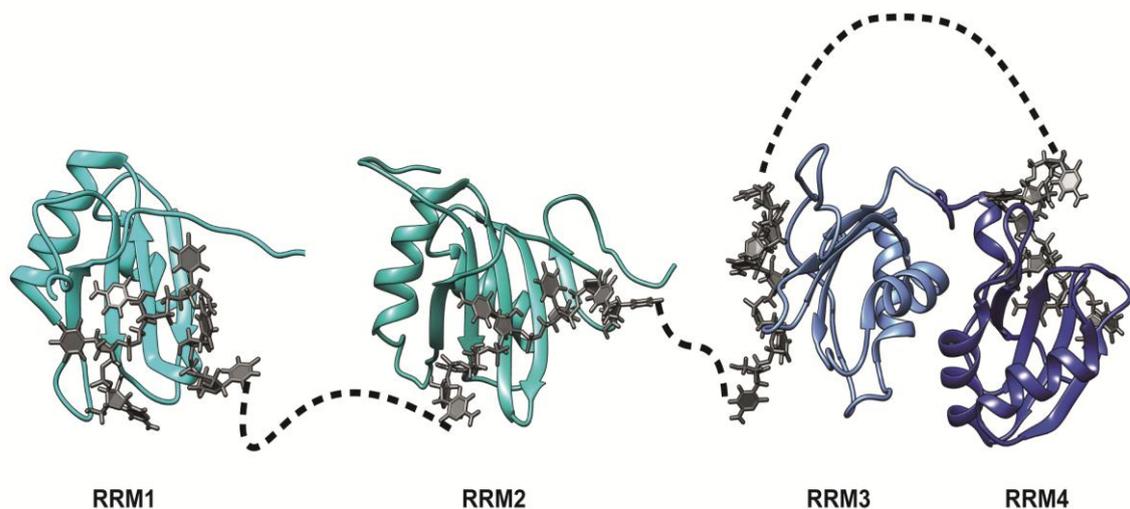
En bleu le repliement canonique du domaine RRM avec en rouge quelques exemples de particularités structurales pouvant s'ajouter à ce domaine. (A) RRM2 de la protéine HNRNP A1 . (B) RRM 3 de la protéine PTB. (C) RRM1 de la protéine U1A. (D) RRM1 de la protéine U2AF-35.

Plusieurs variants du domaine RRM ont été répertoriés dans la littérature. Ainsi on observe des structures avec un brin  $\beta 5$  supplémentaire, une hélice  $\alpha 3$  en c-terminale ou une hélice  $\alpha 1$

de taille étendue (*Figure 15*). Tous ces variants ont pour conséquences d'élargir l'interface de reconnaissance de l'ARN.

Un autre mécanisme permettant d'assurer la reconnaissance de séquences de taille plus importante est l'association de plusieurs domaines RRM avec une même cible.

Ces associations peuvent se faire de plusieurs façons différentes. En tandem, sans association directe entre les domaines RRM, soit avec des domaines s'associant en dimère. La protéine PTB est un exemple de ces deux cas avec ses 4 RRM (Oberstrass, Auweter et al. 2005) (*Figure 16*). Dans les deux cas la séquence nucléotidique reconnue peut être continue ou composé de deux motifs distants.



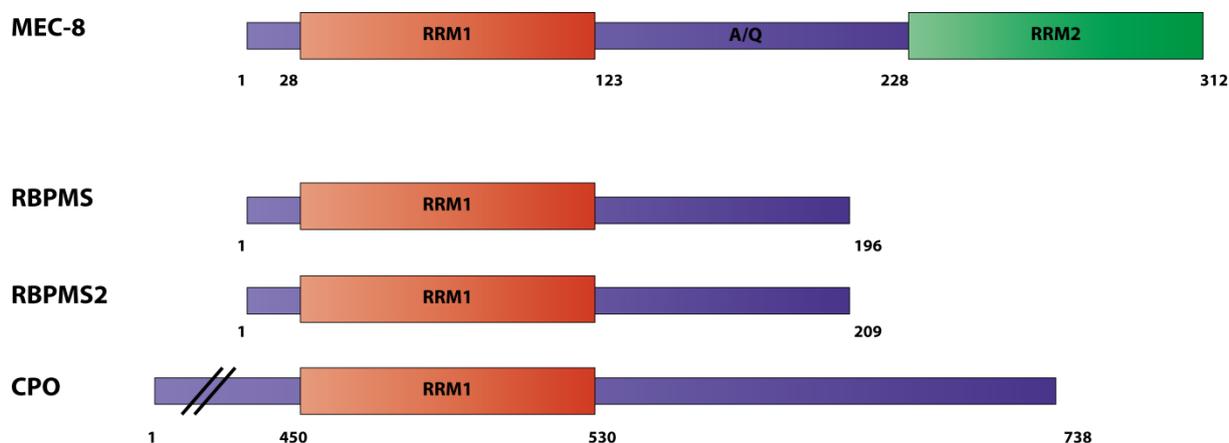
**Figure 16. L'exemple de la protéine PTB**

*Structure des quatre domaines RRM présent dans la protéine PTB en complexe avec la séquence d'ARN CUCUCU. Les identifiants pdb RRM1 :2AD9, RRM2 :2ADB, RRM3 2ADC*

## Introduction de MEC-8

### 1) La protéine MEC-8

La protéine MEC-8 est d'une taille de 33 kDa présente chez *C. elegans*. Elle possède deux domaines RRM séparés par une région non structurée riche en résidus A/Q (Lundquist, Herman et al. 1996) (*Figure 17*).

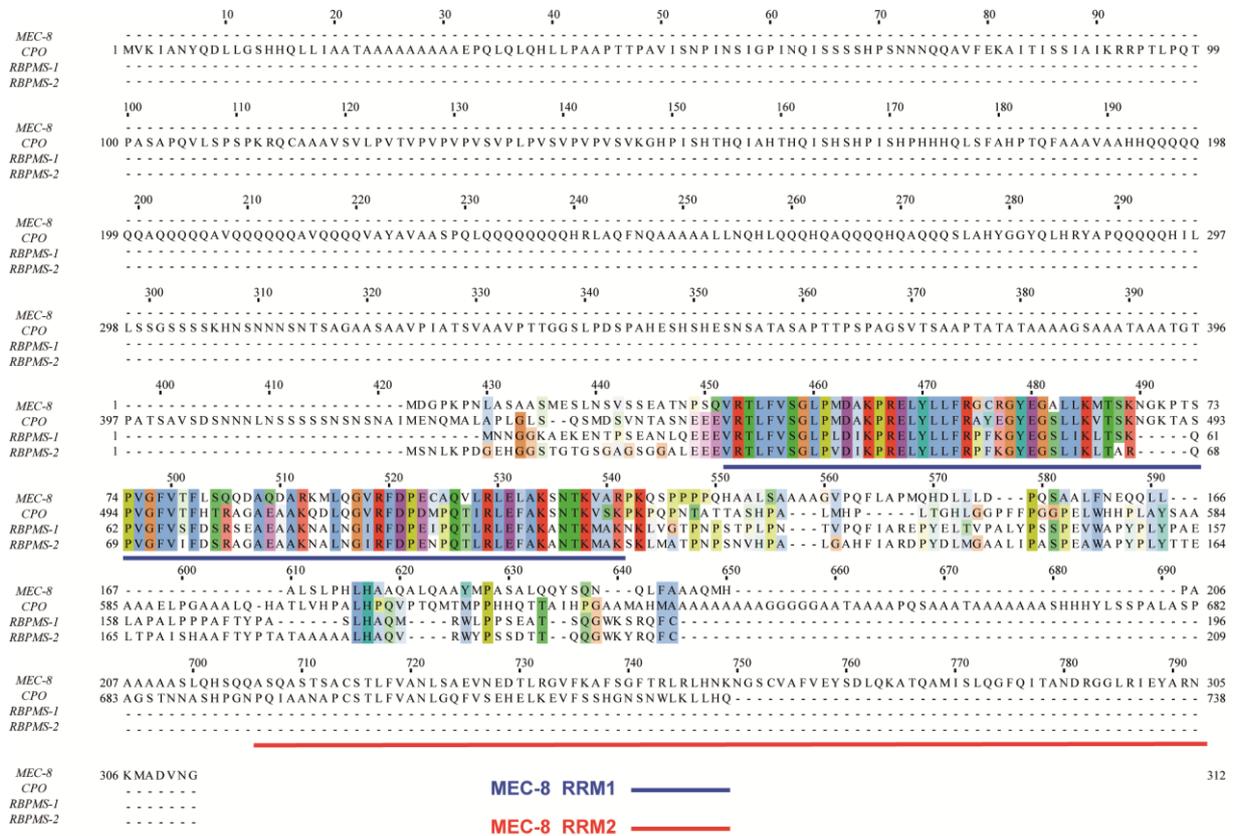


**Figure 17. La famille de protéine RBPMS**

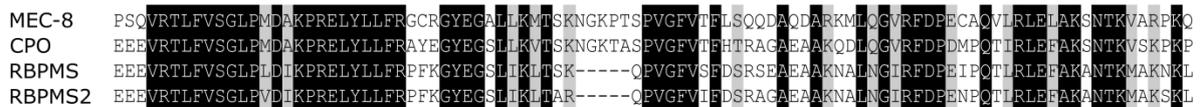
MEC-8 fait partie de la famille de protéine RBPMS comptant les protéines humaine RBPMS (RNA Binding Protein With Multiple Splicing) et RBPMS2 ainsi que la protéine CPO (Couch Potato Protein) de *D. melanogaster*. Les membres de cette famille ne partagent que leur premier domaine RRM en commun (*Figure 18*). Celui-ci est très conservé d'un point de vue de séquence (~60% identité) (*Figure 19*).

MEC-8 est une protéine ubiquitaire durant le stade embryonnaire de *C. elegans*, puis elle persiste dans l'hypoderme et dans les 6 neurones mécanosensoriels (Lundquist and Herman 1994). MEC-8 est nécessaire au développement embryonnaire et larvaire, à la fasciculation des neurones et à la mécanosensation. MEC-8 a un rôle de facteur d'épissage alternatif, il a

été démontré qu'elle interagissait directement dans la régulation de deux gènes, unc-52 et mec-2 (Spike, Davies et al. 2002, Calixto, Ma et al. 2010).



**Figure 18. Alignement des séquences des protéines de la famille RBPMS**



**Figure 19. Alignement de séquence des domaines RRM N-terminaux des protéines de la famille RBPMS.**

MEC-8 est donc impliquée dans l'épissage alternatif au sein de tissus différents, ayant un rôle à la fois sur le développement embryonnaire, musculaire et neuronal. Cependant bien qu'étudiée depuis plus de 30 ans très peu d'éléments sont connus sur le fonctionnement de cette protéine. Les homologues présentés ne sont pas beaucoup plus connus, leurs activités et

localisations sont différentes. Les protéines RBPMS et RBPMS2 sont citées comme liant des partenaires protéiques ainsi que des ARN (Furukawa, Sakamoto et al. 2015). Mais aucune séquence de reconnaissance ARN n'est connue pour le moment. Les seules informations au sujet du mécanisme de liaison à l'ARN de la famille RBPMS sont une expérience de « RNA compete » (Ray, Kazan et al. 2009) pour les protéines MEC-8 et CPO (Ray, Kazan et al. 2013) et une analyse informatique pour la protéine RBPMS (Farazi, Leonhardt et al. 2014).



**Figure 20. Les motifs de reconnaissance présumés**

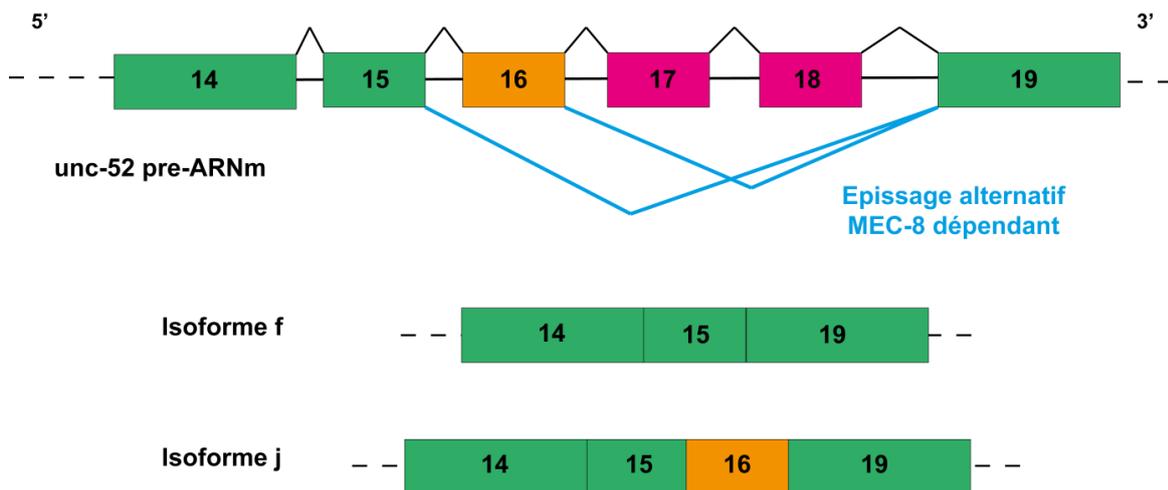
(A) Motif représentant la fréquence d'apparition de chaque base à une position spécifique dans les oligonucléotides liés par la protéine MEC-8 entière. (B) Même motif pour le domaine RRM de la protéine CPO. (C) Résultat issu de l'analyse computationnelle d'une expérience de PAR-Clip pour le domaine RRM de la protéine RBPMS.

## 2) Les gènes cible de MEC-8

### 1) Le gène *unc-52*

Unc-52 est un gène codant pour la protéine UNC-52, une protéine homologue à la protéine humaine PERLECAN (Rogalski, Williams et al. 1993). Elle est exprimée dans l'hypoderme et excrétée dans la matrice extracellulaire. Elle a un rôle dans l'attachement des tissus musculaire à la membrane basale et leur développement (Rogalski, Gilchrist et al. 1995).

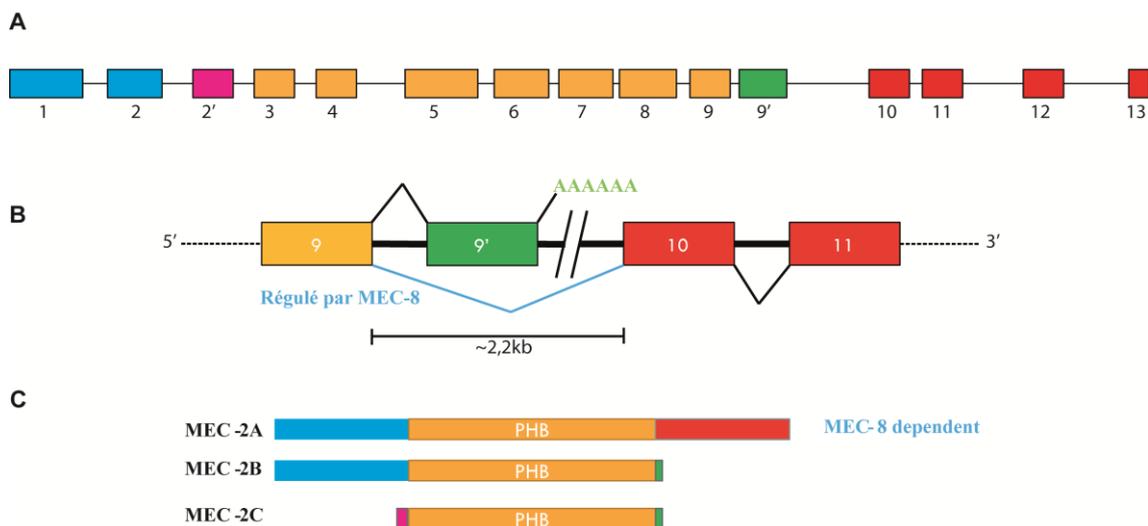
Unc-52 compte 14 isoformes, dont deux ont pour origine l'action de MEC-8. MEC-8 induit le saut des exons 16, 17 et 18 dans le cas de l'isoforme f ou simplement des exons 17 et 18 dans le cas de l'isoforme j.



**Figure 21. Les isoformes de l'ARN messenger de *unc-52* induits par MEC-8**

## 2) Le gène mec-2

Mec-2 est le gène codant pour la protéine du même nom. C'est l'homologue de la Stomatin chez l'humain (Mannsfeldt, Carroll et al. 1999), elle régule positivement l'activité du complexe de transduction du signal mechanosensoriel (Zhang, Arnadottir et al. 2004) sans en être un partenaire constitutionnel. Les mutants mec-2 présentent un phénotype insensible au toucher.



**Figure 22. L'isoforme de l'ARN messager de mec-2 induit par MEC-8**

(A) schéma représentant le pré-ARNm du gène mec-2. (B) Zone du pré-ARNm où a lieu l'épissage alternatif induit par MEC-8. (C) Les isoformes obtenues après épissage alternatif

Le facteur MEC-8 induit le saut de l'exon 9' (*Figure 22.B*) ce qui a pour conséquence la formation de l'isoforme MEC-2A comportant un domaine C-terminal supplémentaire (*Figure 22C*). Il a été récemment démontré que pour que ce saut d'exon soit induit par MEC-8, seule la séquence entre les exons 9 et 10 longue de 2.2kb était nécessaire (Calixto, Ma et al. 2010).

### 3) Les gènes suppresseurs de MEC-8

Lors de l'étude des gènes *mec-8* et *unc-52* il a été découvert que deux gènes étaient suppresseur de leurs phénotypes lorsque ceux-ci étaient mutés en même temps que *mec-8* ou *unc-52* (Lundquist and Herman 1994). Ces gènes ont été nommés *smu-1* et *smu-2* (Suppressor of Mec-8 and Unc-52).

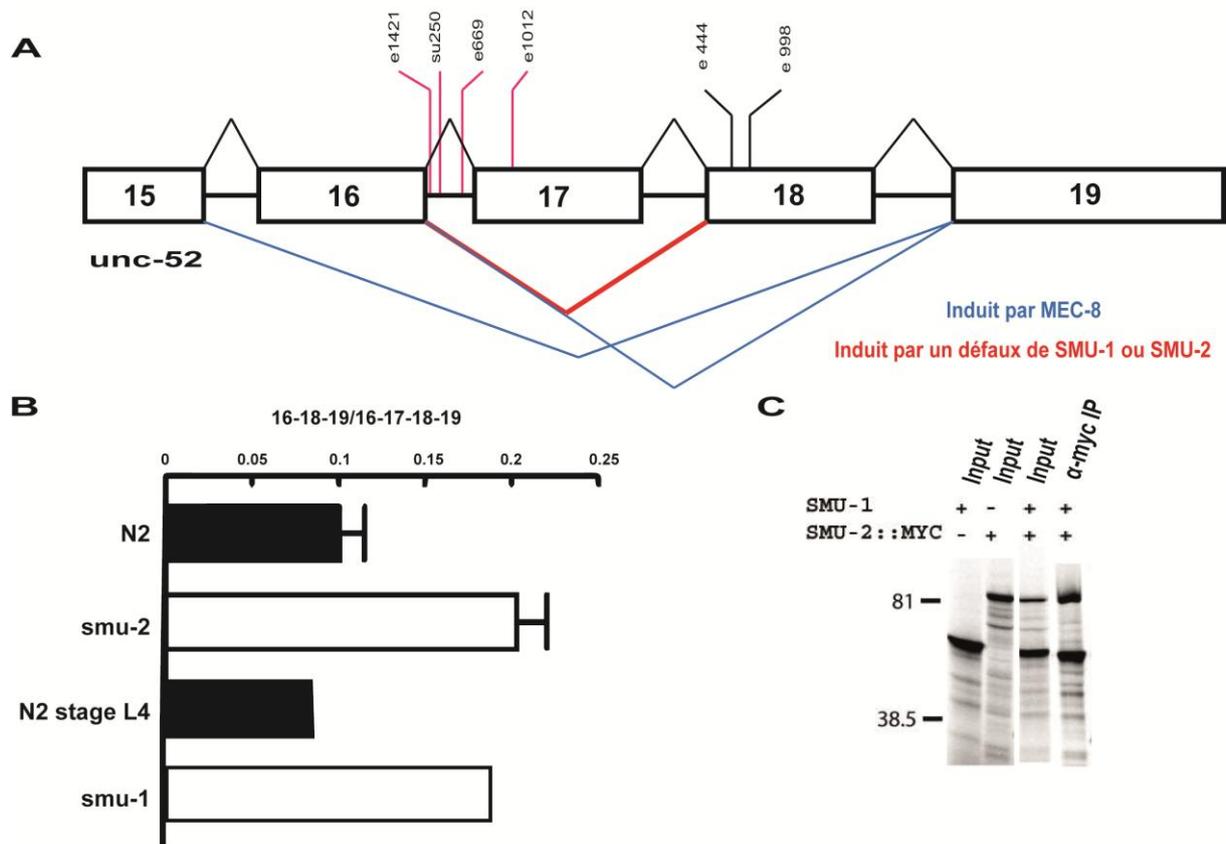
Une mutation chez les gènes *smu-1* ou *2* entraîne des anomalies de l'épissage de certains gènes (Spartz, Herman et al. 2004). Parmi eux *unc-52*, le saut de l'exon 17 de celui-ci est augmenté par un facteur deux lorsque la fonction de *smu-1* ou *smu-2* est affectée (Spike, Shaw et al. 2001, Spartz, Herman et al. 2004) (*Figure 23.B*).

Les gènes *smu-1* et *smu-2* codent pour des protéines contenant de nombreuses répétitions de domaine WD (Spike, Shaw et al. 2001, Spartz, Herman et al. 2004). Les protéines SMU-1 et SMU-2 interagissent physiquement l'une avec l'autre (*Figure 23.C*) et les mutants nuls de *smu-1* et *smu-2* ont exactement les mêmes phénotypes.

L'augmentation de l'isoforme du gène *unc-52* manquant l'exon 17 provoquée par un défaut des protéines *smu-1* ou *smu-2*, permet d'expliquer de façon directe la suppression des phénotypes causés par des mutations présentes dans l'exon 17 et non celles présentes dans l'exon 18 (*Figure 23.A*).

La suppression de tous les phénotypes de MEC-8 incluant le défaut de mécanosensation lié à l'activité de celui-ci sur le gène *mec-2* par des mutants de *smu-1* et *smu-2* est plus compliquée à analyser. En effet cela implique une activité des protéines SMU-1 et SMU-2 sur un panel de gènes potentiellement important, ce qui est en accord avec l'expression ubiquitaire de ceux-ci.

Les protéines SMU-1 et SMU-2 sont les homologues des protéines RED et fSAP57 chez l'humain (Spartz, Herman et al. 2004). Celle-ci ont récemment été citées comme composantes du spliceosome. On peut donc faire l'hypothèse que les protéines SMU-1 et SMU-2 sont des membres du spliceosome chez *C.elegans* permettant la régulation de certains isoformes chez un panel de gènes en apportant une augmentation de la fidélité pour certains sites d'épissage. Les gènes *smu-1* et *smu-2* ne codent cependant pas pour des protéines contenant des domaines de liaison à l'ARN. Le mécanisme d'action de ceux-ci est donc pour le moment inconnu.



**Figure 23. Fonction des protéines SMU-1 et SMU-2**

(A) représentation de la partie du gène de *unc-52* affecté par MEC-8, SMU-1 et SMU-2. Les mutations en rouge voient leurs phénotypes supprimés par un défaut des protéines SMU. (B) Effet sur le taux d'expression de l'isoforme de *unc-52* 16-18-19 induit par un défaut de SMU-1 ou SMU-2. (C) Co-IP de SMU-1 avec SMU-2 adapté de (Spartz, Herman et al. 2004).

#### 4) Les gènes de la famille sym

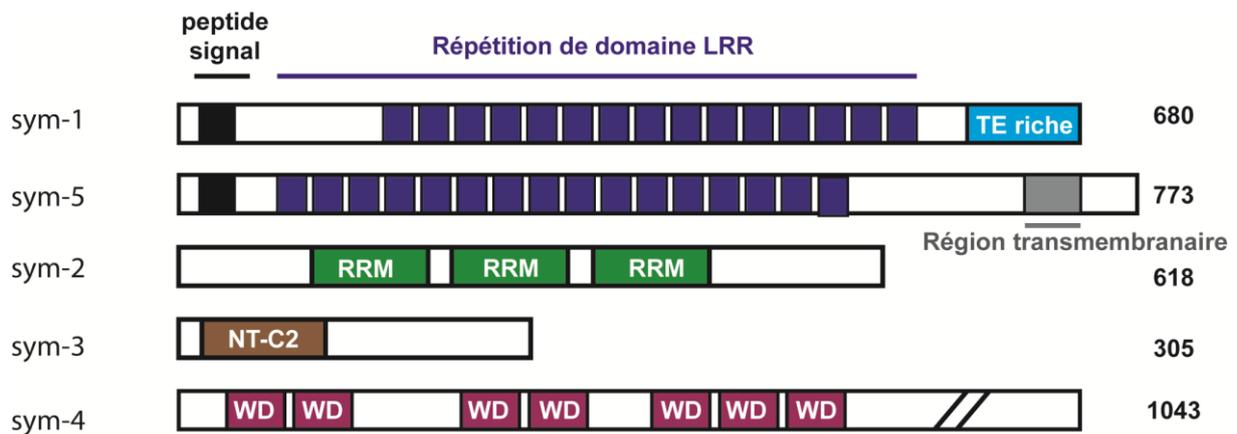
Les gènes de la famille sym (synthetically lethal with mec-8) nommés sym-1 à 5 sont des gènes dont les mutations induisent une mortalité au stade embryonnaire lorsqu'elles apparaissent chez un organisme mec-8 muté (Davies, Spike et al. 1999). Les mutants de ces gènes ne montrent pas de phénotypes particulier lorsque présent seul.

Les protéines SYM-1 et SYM-5 sont des homologues très proches, elles sont les seules à induire une létalité lorsque mutées ensemble, dans la famille des protéines SYM (Davies, Spike et al. 1999). Elles ont donc une fonction présumée redondante. Ce sont des protéines avec des répétitions de domaine riche en leucine, elles sont excrétées dans la matrice extracellulaire des cellules épithéliales où elles jouent un rôle dans la jonction cellulaire permettant l'intégrité des tissus (Mancuso, Parry et al. 2012).

Le gène sym-2 doublement muté avec mec-8 engendre un arrêt embryonnaire au stade « 2-fold » (Davies, Spike et al. 1999). Cet arrêt est caractéristique des organismes présentant des mutations dans les gènes régulant l'attachement des cellules musculaire à la cuticule (Williams and Waterston 1994). Sym-2 code pour une protéine possédant trois domaines quasi-RRM (Dominguez, Fisette et al. 2010) (*Figure 24*) homologue à la famille de protéines HNRNP F/H. Plusieurs hypothèses sont alors possible, soit les protéines SYM-2 et MEC-8 régulent l'épissage de deux gènes codant chacune pour une protéine impliquée dans le développement musculaire et la double mutation a un effet d'addition renforcent le phénotype de chacun. Soit MEC-8 agit sur le gène de sym-2 et permet de rendre ses mutations non phénotypiques.

Les gènes sym-3 et sym-4 donnent tous les deux le même phénotype très particulier lorsque muté simultanément avec mec-8 (Yochem, Bell et al. 2004). Le phénotype résulte en un pharynx correctement développé mais non attaché à la cuticule et une cavité buccale en forme de bulbe, les larves finissent par mourir au stade L1. La protéine SYM-3 comporte un

domaine N-terminal C2 lorsque la protéine SYM-4 contient 7 domaines WD (*Figure 24*). Il est donc très vraisemblable que ces protéines agissent à un niveau différent dans la même voie de signalisation.

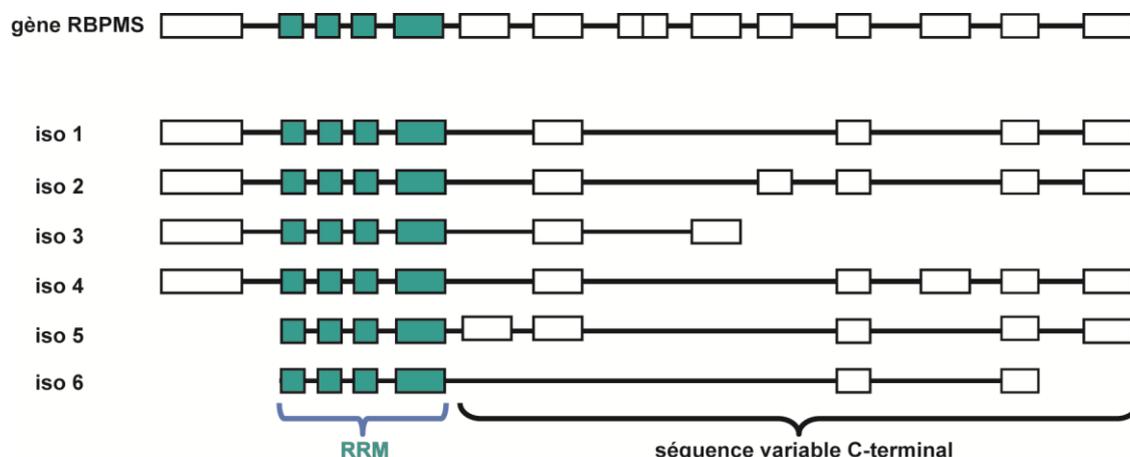


**Figure 24. Organisations structurales des protéines de la famille sym**

## 5) Les protéines RBPMS et RBPMS 2

La protéine humaine RBPMS/HERMES (RNA Binding Protein with Multiple Splicing / HEart, RRM Expressed Sequence) a une séquence de 197 acides aminés comprenant un domaine RRM en N-terminal (Shimamoto, Kitao et al. 1996). RBPMS est exprimé en très grande quantité dans les tissus du cœur et dans les ganglions des cellules rétinales lors du développement (Shimamoto, Kitao et al. 1996, Gerber, Yatskievych et al. 1999, Kwong, Caprioli et al. 2010).

La protéine RBPMS comme son nom l'indique est soumise à un épissage alternatif générant un nombre important d'isoformes (Shimamoto, Kitao et al. 1996). Tous les isoformes conservent le domaine RRM N-terminal et ne divergent que par leur domaine C-terminal (Shimamoto, Kitao et al. 1996)(*Figure 25*).



**Figure 25. Les isoformes de l'ARN messager de RBPMS**

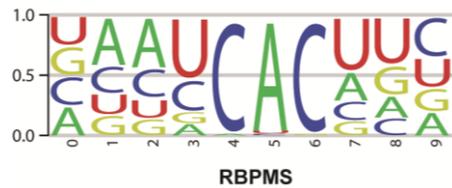
RBPMS est impliqué dans diverses fonctions comme le développement des axones (Hornberg, Wollerton-van Horck et al. 2013), le développement du cœur (Gerber, Vokes et al. 2002) et la régulation de la prolifération et la migration de cellules cancéreuses (Fu, Cheng et al. 2015).

Ces différentes fonctions l'impliquent à la fois dans des interactions protéine-protéine (Furukawa, Sakamoto et al. 2015) et protéine-ARN (Gerber, Vokes et al. 2002). Le domaine RRM permet l'interaction avec des ARN ainsi qu'avec des protéines. Cependant les différentes isoformes de RBPMS étant spécifiques à certaines interactions protéiques (Fu, Cheng et al. 2015), le rôle permettant l'interaction protéine-protéine est très probablement détenue par le domaine C-terminale de RBPMS et non par son domaine RRM.

La protéine RBPMS 2 est une protéine de 209 acides aminés comportant un domaine RRM en N-terminal identique à plus de 95% à celui de RBPMS. RBPMS2 n'a qu'un second isoforme, celui-ci manque la partie N-terminal comprenant le domaine RRM.

Les fonctions et localisation de RBPMS2 sont différentes de RBPMS. RBPMS2 est exprimé dans les muscle lisse gastro-intestinaux (Notarnicola, Rouleau et al. 2012) ou elle joue un rôle dans le développement de ceux-ci en interagissant avec l'ARN messager *noggin*. Elle est aussi trouvée en grande quantité dans les tumeurs du stroma gastro-intestinal (Hapkova, Skarda et al. 2013). Chez le poisson zèbre l'homologue de RBPMS 2 est exprimé dans les oocytes ou il interagit physiquement avec la protéine Buc ainsi qu'avec l'ARN messager de cette même protéine (Heim, Hartung et al. 2014).

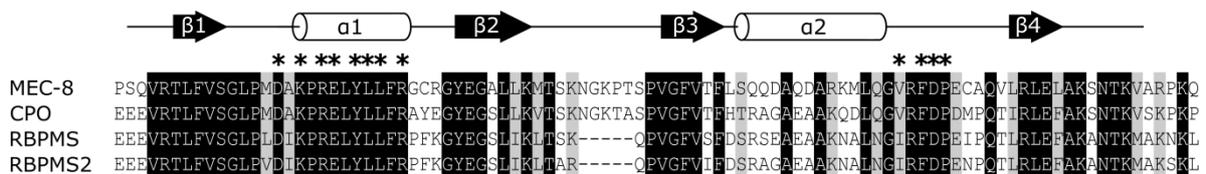
Le domaine RRM de RBPMS a fait l'objet d'une étude à grande échelle afin d'identifier ses différentes cibles au niveau du transcriptome grâce à une expérience de PAR-CLIP (Farazi, Leonhardt et al. 2014). Une étude bio-informatique de ces résultats a permis de mettre en évidence une hypothétique séquence de reconnaissance (Farazi, Leonhardt et al. 2014) (*Figure 26.*) Ce motif est présumé être reconnu en tandem.



**Figure 26. Séquence consensus de liaison calculé pour RBPMS**

Les protéines RBPMS et RBPMS2 sont donc des protéines liant les ARN ainsi que diverses protéines. Il semble que le domaine C-terminal soit un moyen de réguler la localisation intracellulaire de la protéine en exprimant des isoformes ayant chacun une spécificité pour un partenaire propres. Le domaine RRM quant à lui est conservé et bien qu'il lie divers ARN cibles cela se fait par la liaison à une séquence de reconnaissance identique.

Récemment une structure en solution de la protéine RBPMS 2 a été publiée, elle met en évidence un motif d'homodimérisation (*Figure 27*) présent dans le domaine RRM de celle-ci. Ce motif conservé dans toute la famille RBPMS (Sagnol, Yang et al. 2014).



**Figure 27. Le motif d'homodimérisation de RBPMS 2**

*Alignement des séquences des protéines de la famille RBPMS, les étoiles représente l'interface de dimérisation décrite dans (Sagnol, Yang et al. 2014).*

## 6) La protéine CPO

La protéine Couch POTato (CPO) de *D.melanogaster* est exprimée dans les cellules mère sensoriel, dans le système nerveux périphérique et central (Bellen, Vaessin et al. 1992) et localisé dans les noyaux de ces cellules. L'expression du gène *cpo* est observable chez les embryons, les larves et les adultes, cependant on observe des pics d'expression important durant les stades L1 et P1 du développement (Zhang and Denlinger 2011).

Les mutations affectant le gène *cpo* induisent un comportement hypoactif, une incapacité partielle voire totale à voler, ainsi que des défauts dans la coordination musculaire et le développement neuronal (Glasscock and Tanouye 2005).

Le gène *cpo* est très complexe, il s'étend sur 100 kb et est la cible d'épissage alternatif provoquant l'expression de nombreux messagers différents. Tous les isoformes de CPO contiennent un domaine de reconnaissance de l'ARN RRM à l'exception de l'isoforme CPO-RH (Bellen, Kooyer et al. 1992)

CPO est impliqué dans la diapause chez les drosophiladea (Kankare, Salminen et al. 2010), le moustique *Culex pipiens* (Zhang and Denlinger 2011) et le zooplancton *Calanus finmarchicus* (Christie, Roncalli et al. 2013). La diapause est un phénomène très important chez les insectes et les crustacés engendrant une pause dans le développement suite à des stimuli extérieurs (changement de saisons, température, manque de nourriture,...) (Allen 2007). Une mutation présente chez CPO-RH le seul isoforme à ne pas contenir de domaine RRM a été caractérisée comme prolongeant significativement ce phénomène de diapause. Il est donc envisageable que l'activité de CPO dans ce processus ne fasse pas intervenir la maturation d'un ARN de façon directe (Lee, Sgro et al. 2011) (Schmidt, Zhu et al. 2008).

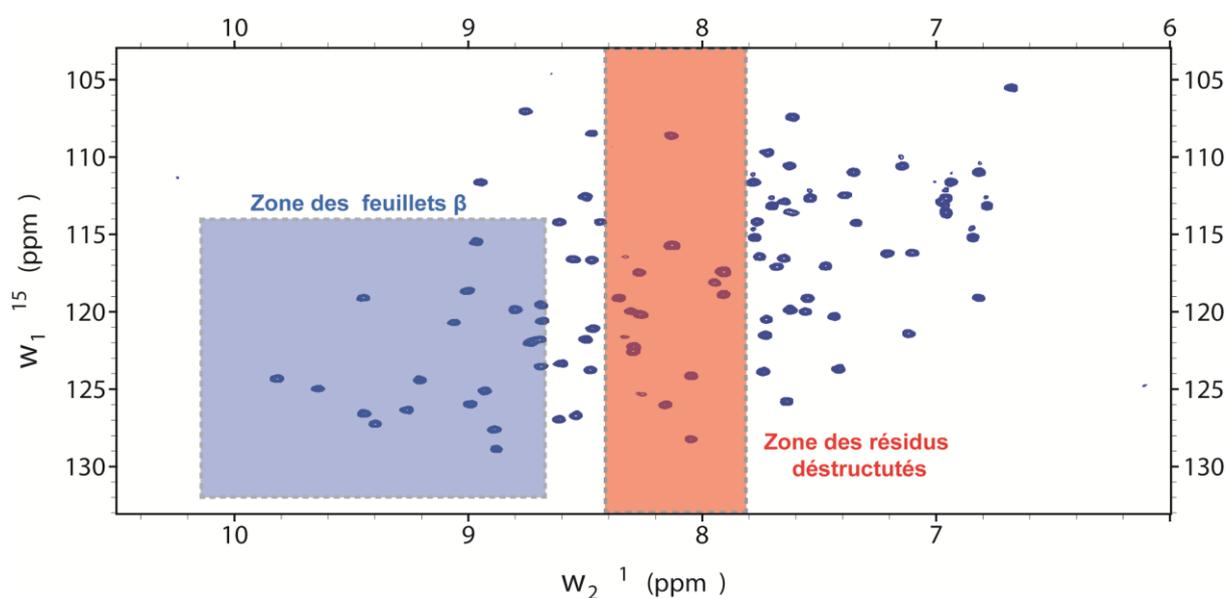


## Chapitre I : Structures et fonction du RRM1 de MEC-8

### I. Etude structurale du premier domaine de MEC-8

#### 1) Production et purification des protéines recombinantes

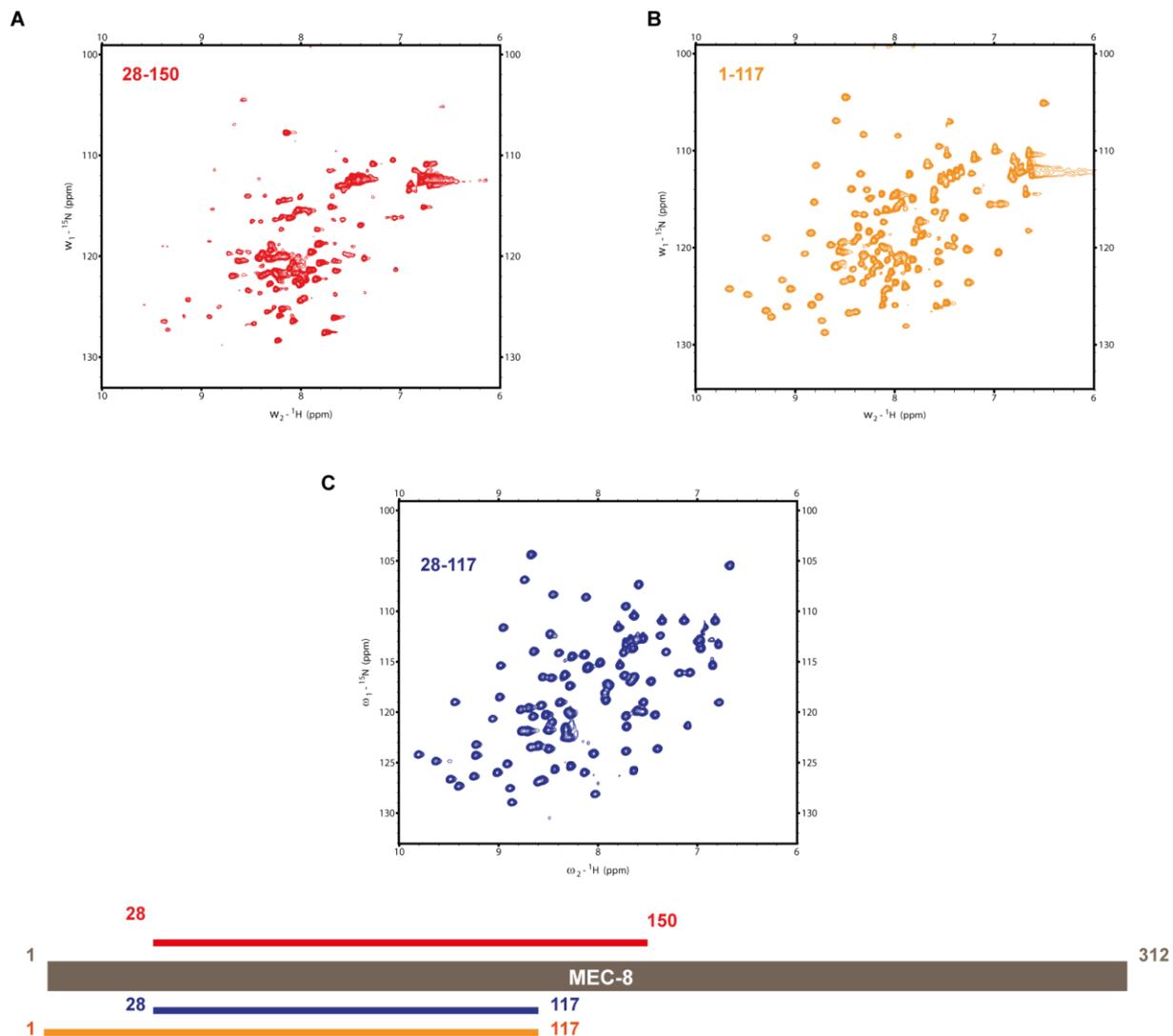
Comme il l'a été rappelé dans l'introduction, le domaine RRM est défini par un cœur structural très conservé tout en arborant dans un nombre important de cas une spécificité structurale additionnelle. L'investigation de telles spécificités structurales dans le cas du domaine RRM1 de MEC-8 a donc été entreprise. Plusieurs constructions avec des extensions en N et C-terminal ont été clonées, produites et analysées par spectroscopie RMN.



**Figure 28. Les informations structurales issus d'un spectre 1H-15N HSQC**

*La zone bleue représente la région des résidus impliqués dans un repliement en feuillet  $\beta$ . La zone orange représente la région des résidus présents dans les parties peu structurées de la protéine.*

La caractérisation de la nature structurée ou non des extensions N et C-terminales se fait par l'analyse des spectres 1H-15N HSQC. En effet ce spectre nous permet d'identifier si les nouveaux pics qui apparaissent font partie d'une région de la protéine structurée ou non (*Figure 28*).



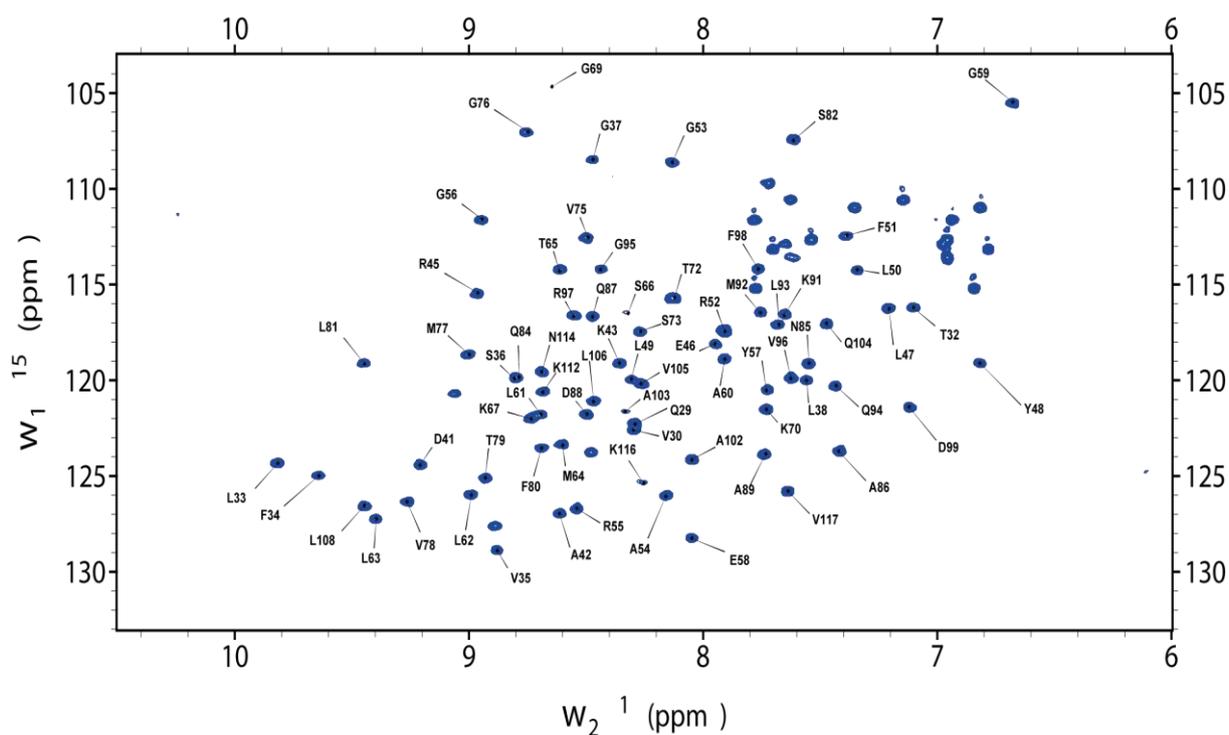
**Figure 29. Exploration structurale du domaine RRM1 de MEC-8**

(A) Spectre 1H-15N HSQC de MEC-8(28-150). (B) Spectre 1H-15N HSQC de MEC-8(1-117). (C) Spectre 1H-15N HSQC de MEC-8(28-117).

Après production des constructions protéique avec des extensions en C et N-terminale marquées  $^{15}\text{N}$ , les échantillons sont analysés par RMN. Dans les deux cas de figure on observe nettement sur les spectres 1H-15N HSQC que les résidus supplémentaires restent dans la région des résidus désordonnés (*Figure 29*). On en déduit que le domaine RRM1 de MEC-8 (28-117) est canonique, ne comportant pas de spécificité structurale additionnelle.

## 2) L'assignement des résonances de MEC-8 RRM1 libre

L'attribution du squelette peptidique a été déterminée à partir des spectres 2D  $^1\text{H}^{15}\text{N}$ -HSQC, 2D  $^1\text{H}^{13}\text{C}$ -HSQC, 3D HNCACB, 3D HN(CO)CACB, 3D CBCA(CO)NH, 3D HNCA, 3D HNCO, 3D HN(CA)CO selon la méthode décrite par (Sattler, Schleucher et al. 1999).



**Figure 30. Spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC du domaine RRM1(28-117)**

Les chaînes latérales des résidus aliphatiques et aromatiques ont été attribuées à partir de spectres enregistrés dans 100 % et 10 % D<sub>2</sub>O. Les séquences 2D  $^1\text{H}^{13}\text{C}$ -HSQC, 3D HC(C)H-TOCSY, (H)CCH-TOCSY et H(CCO)NH ont été acquises pour attribuer les résonances des chaînes latérales aliphatiques. Les séquences 2D  $^1\text{H}^{13}\text{C}$ -HSQC, 2D  $^1\text{H}^1\text{H}$ -TOCSY (24 ms), 2D  $^1\text{H}^1\text{H}$ -NOESY (120 ms),  $^1\text{H}^{13}\text{C}$ -HSQC-NOESY et TROSY- $^1\text{H}^{13}\text{C}$ -HSQC ont été acquises pour attribuer les aromatiques. Les chaînes latérales des résidus asparagine et glutamine ont été attribuées grâce aux expériences 2D  $^1\text{H}^{15}\text{N}$ -HSQC, 3D CBCA(CO)NH, HNCACB et HNCO

### 3) L'oligométrie du RRM1

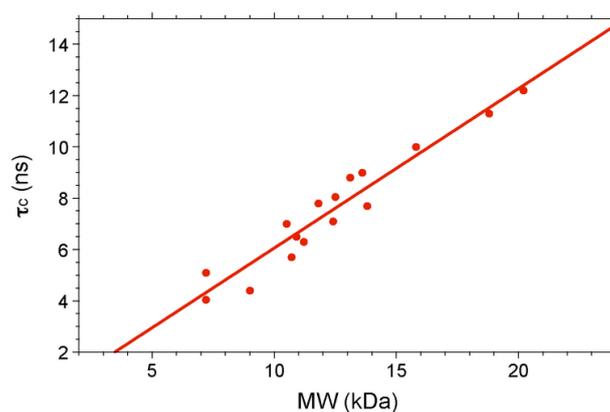
La spectroscopie RMN du liquide comme son nom l'indique permet d'étudier un échantillon de protéine en solution. L'étude de la dynamique d'une protéine en solution permet l'obtention de paramètres directement lié au temps de corrélation rotationnel ( $T_c$ ). La valeur du  $T_c$  d'une protéine représente le temps que prend cette protéine pour effectuer une rotation de 1 radian lors d'une rotation de diffusion brownienne. Ce  $T_c$  est donc directement lié au rayon de giration de la protéine étudiée, ce qui nous permet pour des molécules rigides et sphériques une estimation du poids moléculaire de celle-ci.

$$\tau_c \approx \frac{1}{4\pi\nu_N} \sqrt{6\frac{T_1}{T_2} - 7}$$

#### Équation 1. Approximation de la valeur de $T_c$ en fonction des constante de relaxation $T_1$ et $T_2$

*$\nu_N$  est la fréquence de la résonance  $^{15}N$ .*

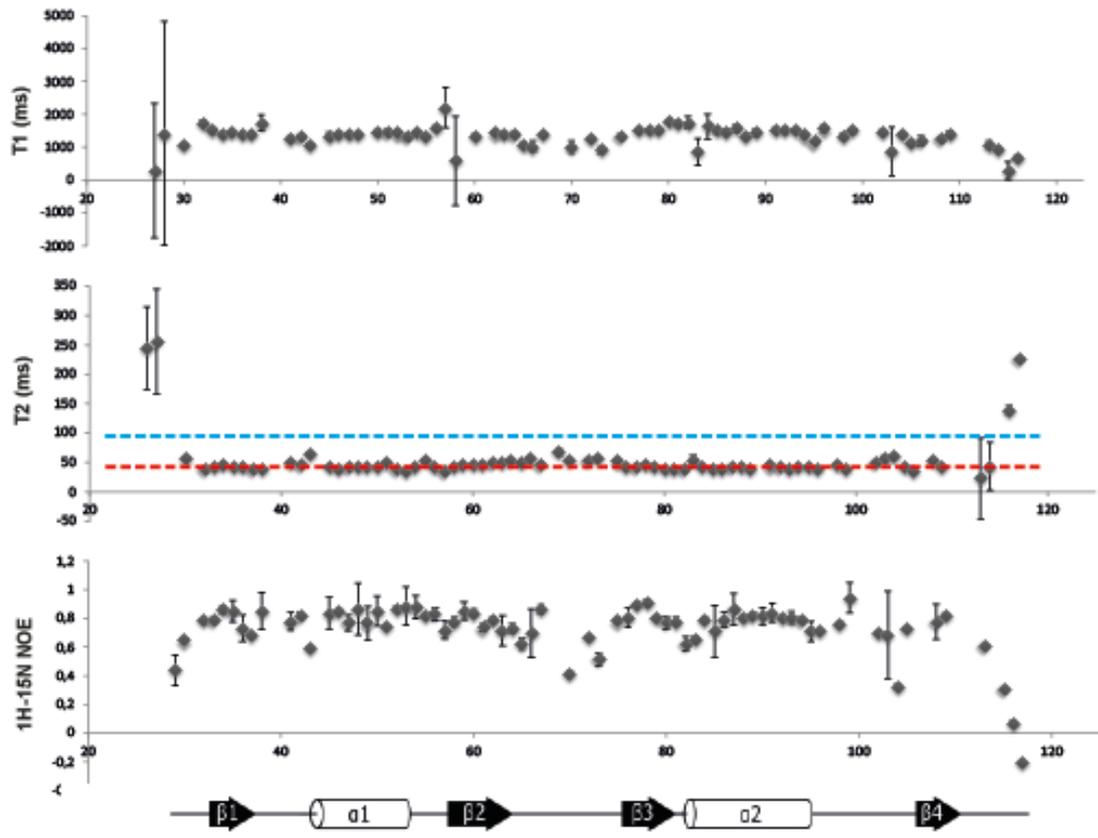
Nous avons donc un moyen simple d'estimer l'oligométrie du domaine RRM1 de MEC-8 en mesurant les valeurs de relaxation  $^{15}N$   $T_1$  et  $T_2$  de celui-ci (Équation 1).



**Figure 31. Corrélation entre temps de corrélation rotationnel et poids moléculaire**

*Données issues d'expérience à 298K sur un spectrometre Bruker 600Mhz.*

En effet une fois le temps de corrélation rotationnelle calculé, une comparaison peut être réalisée avec un échantillon de protéine dont la masse moléculaire et le Tc sont connue (*Figure 31.*). Cela permet alors une estimation fiable du poids moléculaire d'une protéine globulaire.



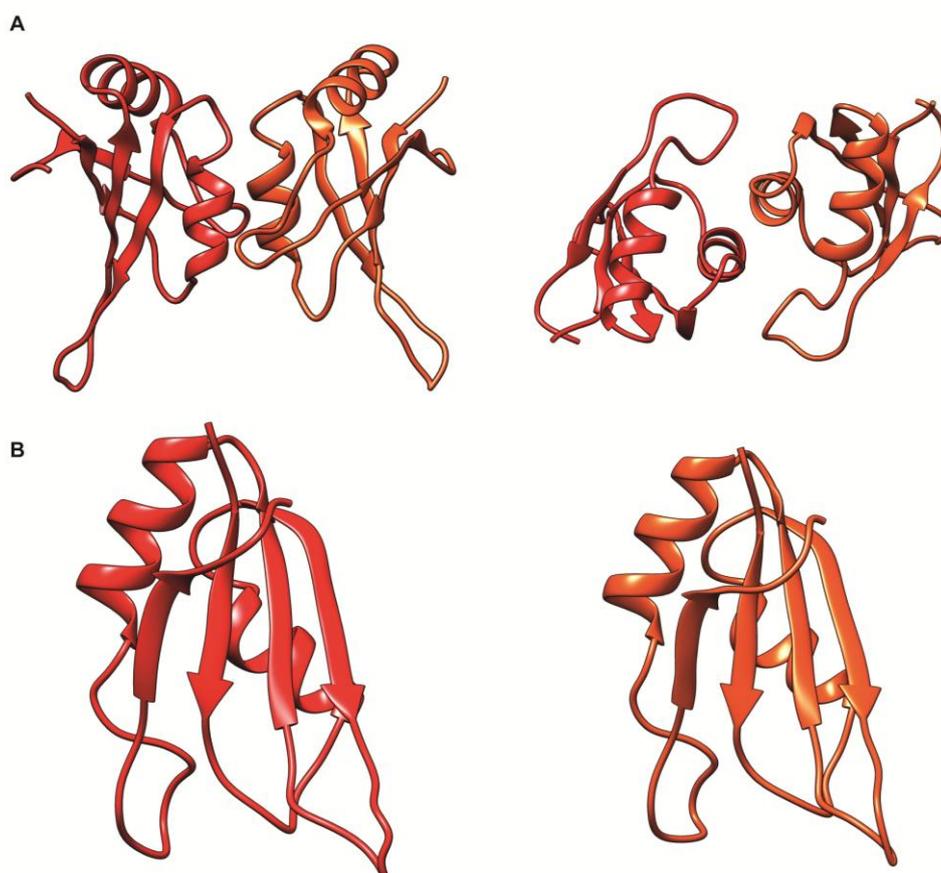
**Figure 32. Etude dynamique du domaine RRM1 de MEC-8**

*La ligne pointillée bleue représente les valeurs attendues pour un monomère, en rouge sont représentées les distributions moyennes des valeurs pour MEC-8(28-117).*

Les résultats pour le domaine RRM1 sont une valeur moyenne du Tc de 12 ns. Cela correspond à un domaine d'un poids moléculaire d'environ 20 kDa, soit exactement deux fois la taille attendu pour ce domaine. On peut donc conclure que le domaine RRM1 de la protéine MEC-8 est à l'origine de la dimérisation de celle-ci. Des analyses par chromatographie d'exclusion de taille ont permis de valider ces résultats.

#### 4) Structure cristallographique de MEC-8 RRM1

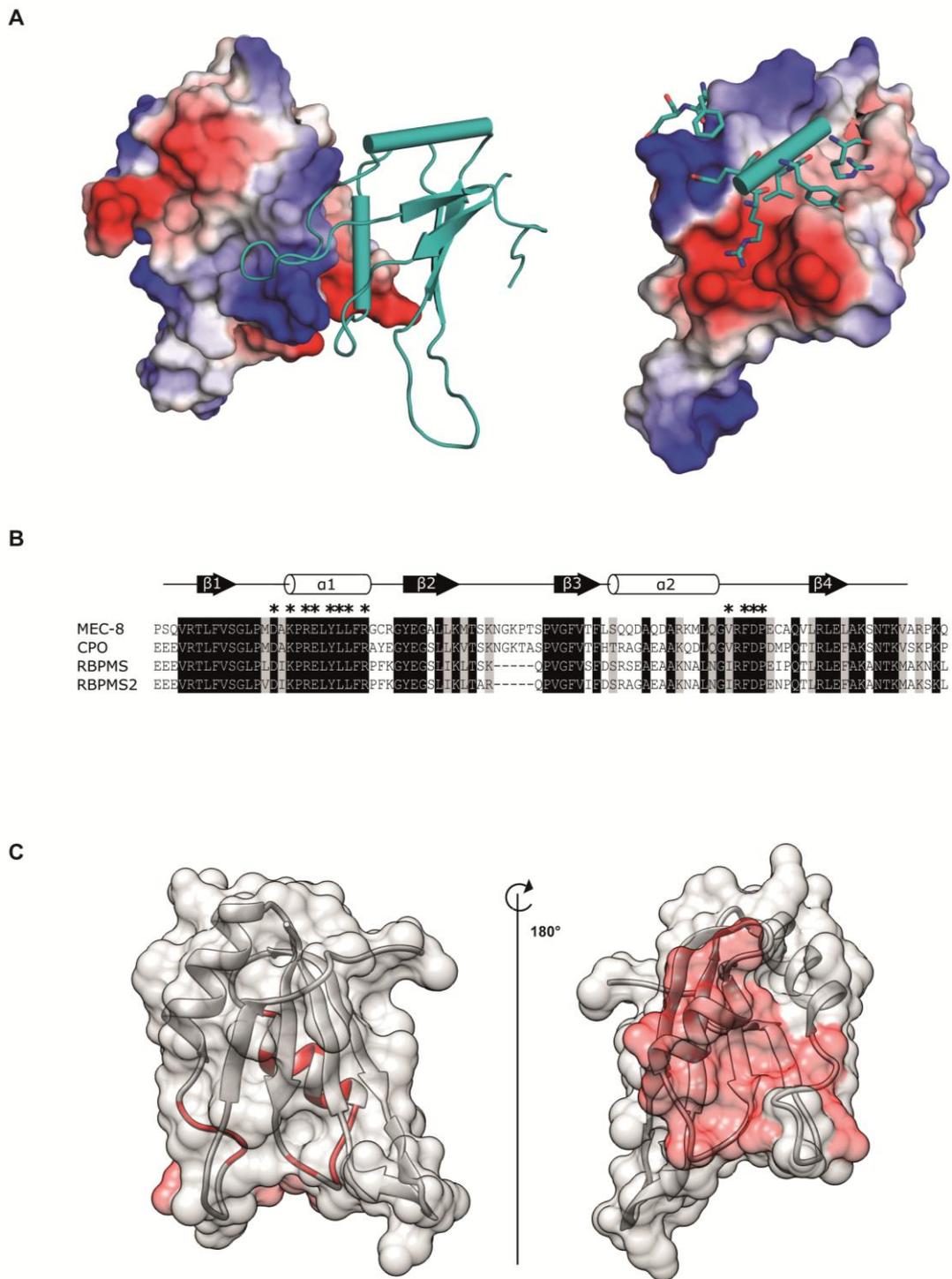
En parallèle de l'assignement des spectres RMN de MEC-8 RRM1, nous avons entrepris des essais de cristallogénèse qui ont rapidement donnés des cristaux. Les cristaux en forme de plaque longitudinale sont apparus dans un crible commercial 96 puits et n'ont jamais pu être reproduit. Néanmoins leurs tailles étaient suffisantes pour être analysés sur l'appareil de diffraction présent dans notre institut. Ces analyses ont permis l'acquisition de données d'une résolution d'un maximum de 3,8 Å. Cette résolution est descendue à 2,7 Å sur une source synchrotron, cela nous a permis de résoudre la structure cristallographique grâce au remplacement moléculaire.



**Figure 33 : Structure cristallographique de MEC-8 RRM1**

(A) Les deux dimères présent dans l'unité asymétrique, en rouge la chaîne A et en orange la chaîne B. (B) représentation des monomères.





**Figure 35. Interface de dimérisation du domaine RRM1.**

(A) Représentation de la surface électrostatique de la chaîne A, la chaîne B est illustrée en mode cartoon. (B) Alignement de la séquence de MEC-8 RRM1 avec les RRM de CPO, RBPMS et RBPMS2. Les étoiles représentent les résidus impliqués dans la dimérisation du RRM. (C) Représentation de la surface de dimérisation en rouge, sur un monomère.

**Table 1. Statistiques de la structure du domaine RRM1****Data collection**

Resolution range ( $\text{\AA}$ ...)	47.48 - 2.761 (2.859 - 2.76)
Space group	I 2 2 2
Unit cell	52.632 96.407 110.063 90 90 90
Total reflections	37551 (5150)
Unique reflections	7438 (1032)
Multiplicity	5.1 (5.0)
Completeness (%)	99.19 (96.51)
Mean I/sigma(I)	8.53 (1.72)
R-merge	0.172 (0.550)
R-meas	0.192 (0.613)

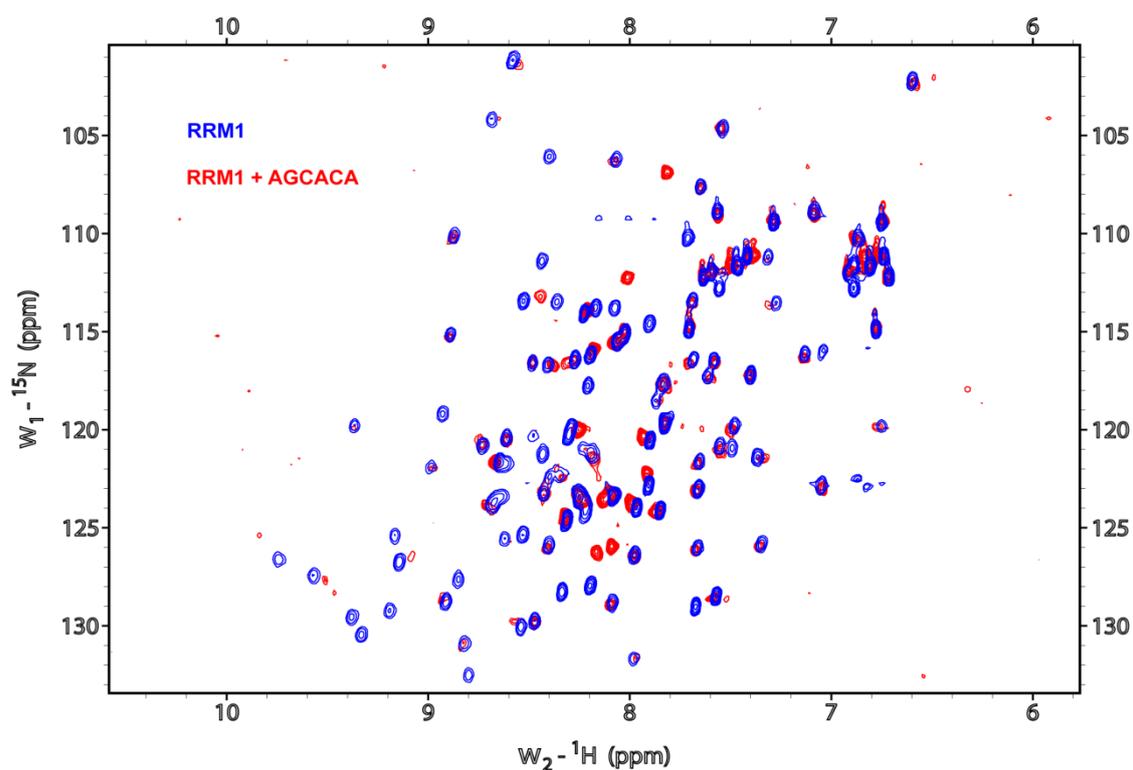
**Refinement**

R-work	0.2385 (0.3902)
R-free	0.2823 (0.3691)
Number of non-hydrogen atoms	1329
macromolecules	1316
water	13
Protein residues	170
RMS(bonds)	0.018
RMS(angles)	3.20
Ramachandran favored (%)	94.67
Ramachandran allowed (%)	5.33
Ramachandran outliers (%)	0
Average B-factor	63.40
macromolecules	63.50
solvent	45.20

## II. Etude interactionnel de MEC-8 RRM1

### 1) Recherche d'une séquence de liaison nucléotidique optimal

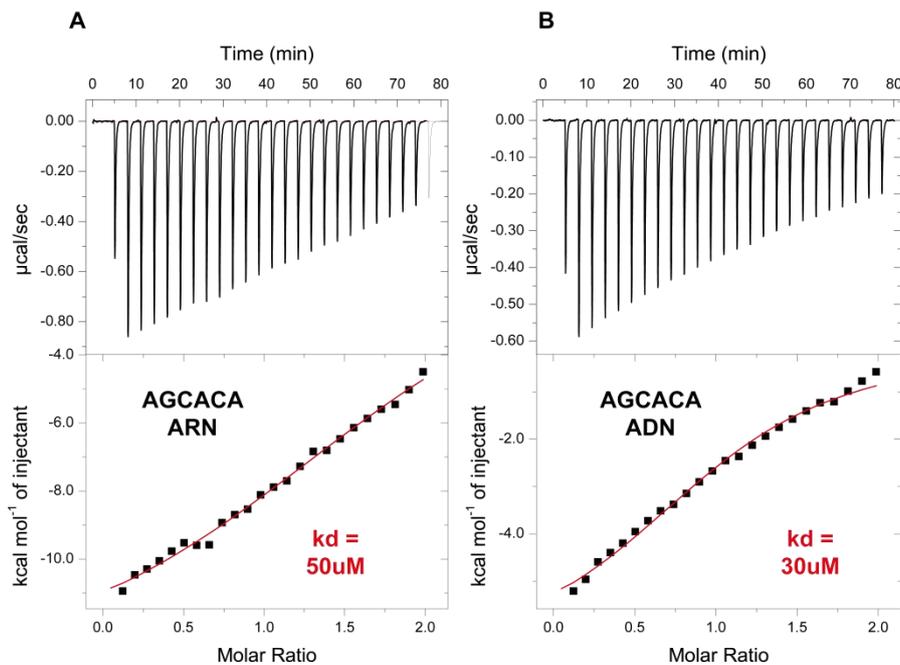
Une première étape pour l'étude fonctionnelle du domaine RRM1 est d'identifier la séquence de liaison nucléotidique qu'il reconnaît. Un motif a été précédemment décrit dans la littérature ((Ray, Kazan et al. 2013)) et nous avons donc commencé par étudier l'interaction avec cette séquence.



**Figure 36. Spectre 1H-15N HSQC du complexe RRM1-AGCACA**

*Spectre 1H-15N HSQC du domaine RRM1 seul en bleu et en complexe avec la séquence ARN AGCACA à un ratio 1/1.5 en rouge.*

Le spectre 15N-HSQC du domaine RRM1 avec la séquence d'ARN AGCACA permet de confirmer qu'il y a bien formation d'un complexe protéine-ARN. Cependant le complexe est en régime d'échange intermédiaire entraînant une disparition de la plupart des pics du spectre (*Figure 36*) empêchant toute étude structurale en solution de celui-ci.



**Figure 37. Affinité par ITC des motif AGCACA ADN et ARN avec le domaine RRM1**

(A) ITC de la protéine MEC-8 (28-117) avec la séquence d'ARN AGCACA. (B) ITC de la protéine MEC-8 (28-117) avec la séquence d'ADN.

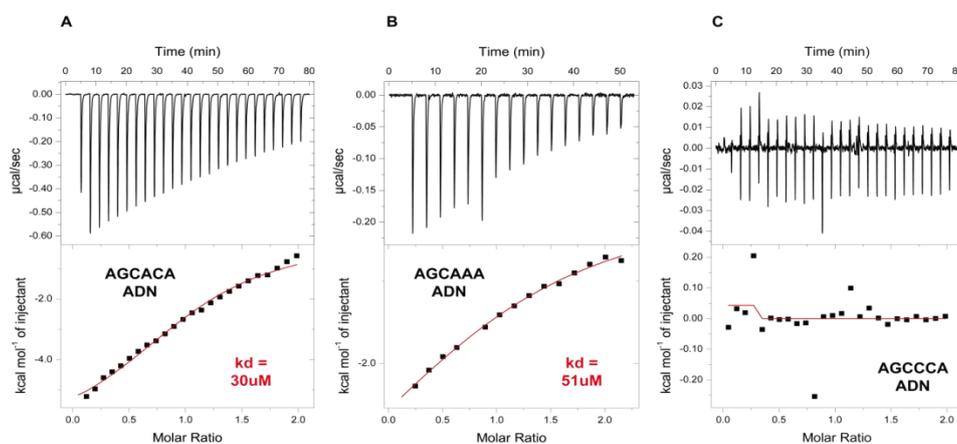
Il n'est pas inhabituel que les prédictions de motif nucléotidique résultant d'expériences *in vitro* comme le SELEX ou RNACompete ne soit pas l'unique séquence reconnue par une protéine (Foot, Feracci et al. 2014) et donc par la même occasion pas la séquence optimale ou « naturelle ».

Nous avons donc profité de la relativement bonne affinité du ligand ADN par rapport à l'ARN (Figure 37) pour faire une étude systématique de toutes les mutations possibles en chaque position du motif à moindre coûts. Cela nous permet d'identifier à chaque position, la base avec l'affinité la plus élevée, pour finalement obtenir une séquence nucléotidique optimale.

Cette expérience s'est conclue par un résultat similaire à celui de la littérature à l'exception du nucléotide en position 5 qui se montre assez permissif à la substitution d'une cytidine par une adénine (Figure 38). Néanmoins le motif AGCACA est bel et bien celui avec l'affinité la plus importante, la spécificité se concentrant sur le cœur GCAC (Figure 39).

**Table 2. Les affinités de MEC-8 RRM1 avec les mutants du motif AGCACA par ITC**

ligand	protein	Kd ± err (µm)
AGCACA (ARN)	MEC-8 28-117	53 ± 6
d-AGCACA	MEC-8 28-117	30 ± 5
d-AGCACC	MEC-8 28-117	49,5 ± 4,5
d-AGCACU	MEC-8 28-117	76,3 ± 6
d-AGCACG	MEC-8 28-117	46 ± 5,3
d-AGCAUA	MEC-8 28-117	NA
d-AGCAGA	MEC-8 28-117	NA
d-AGCAAA	MEC-8 28-117	41 ± 9,7
d-AGCCCA	MEC-8 28-117	NA
d-AGCUCA	MEC-8 28-117	1400 ± 11000
d-AGCGCA	MEC-8 28-117	245 ± 232
d-AGAACA	MEC-8 28-117	NA
d-AGUACA	MEC-8 28-117	346 ± 231
d-AGGACA	MEC-8 28-117	NA
d-ACCACA	MEC-8 28-117	223 ± 104,3
d-AUCACA	MEC-8 28-117	271 ± 96,9
d-AACACA	MEC-8 28-117	1200 ± 2700
d-CGCACA	MEC-8 28-117	70 ± 6
d-UGCACA	MEC-8 28-117	106 ± 13
d-GGCACA	MEC-8 28-117	89,3 ± 16,6



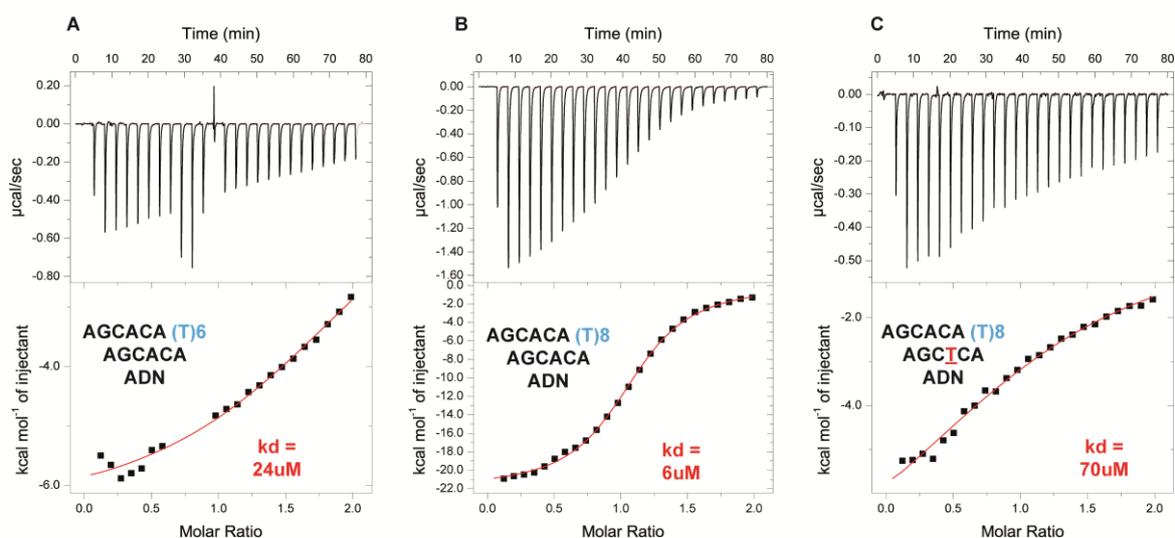
**Figure 38. La position 5 est permissive à la substitution par une adénine**



**Figure 39. Motif de liaison optimal du domaine RRM1**

## 2) Interaction du domaine RRM1 avec des motifs en tandem

Nous avons identifié une séquence de liaison ARN longue de 6 nucléotides pour le domaine RRM1. La séquence « GCAC » est la seule partie reconnue spécifiquement et son affinité est relativement faible empêchant toute étude structurale du complexe par RMN. Le domaine RRM N-terminal de MEC-8 est par ailleurs un homodimère en solution, cette caractéristique peut nous aider à créer une séquence nucléotidique comprenant deux motifs de reconnaissances séparés par une séquence de longueur suffisante pour permettre une liaison coopérative des deux monomères protéiques du complexe. Une liaison coopérative aura théoriquement une affinité beaucoup plus importante et nous permettrait de résoudre une structure du RRM1 en complexe avec son motif ARN.



**Figure 40. Apport d'une liaison coopérative du dimère RRM1**

*ITC entre le RRM1 et les séquences d'ADN tandem avec des linker de 6 (A) et 8 (B) nucléotides. (C) la séquence avec le tandem de motif AGCACA séparé par 8T comporte une mutation A4T dans le second motif.*

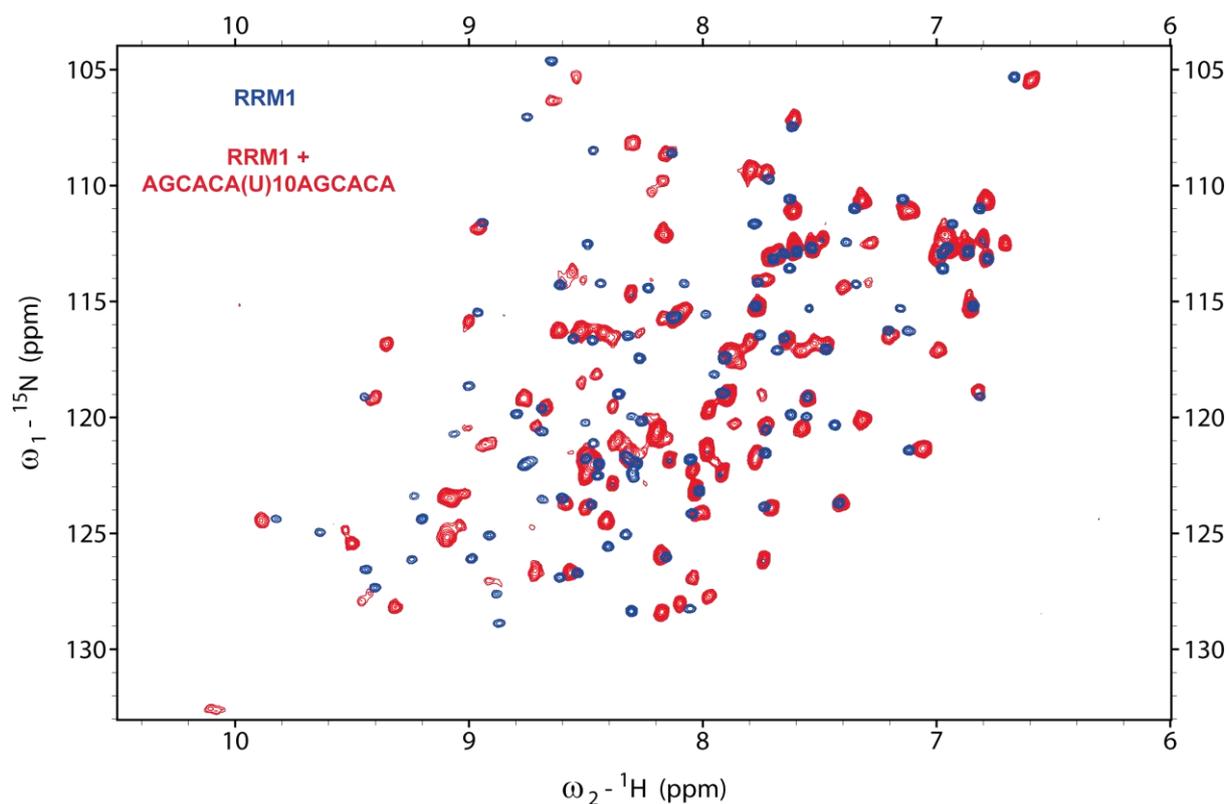
Nous avons utilisé la structure du RRRM1 précédemment calculée pour estimer le nombre de nucléotide suffisant pour que les deux motifs de reconnaissances puissent se lier au feuillet béta de chaque monomère. Ce nombre semble se situer entre 6 et 10 nucléotides. Pour discriminer sans ambiguïté un phénomène de liaison 2 :1(2 monomère pour 1 ligand) d'une liaison 1 :1 (1 monomère pour 1 ligand) nous avons testé l'affinité par ITC de ligands avec un espace de 4, 6, 8, 10 et 12 nucléotides entre les deux motifs.

Les résultats montrent une différence d'affinité à partir de 8 nucléotides d'espacement entre les motifs AGCACA (*Figure 40*). En effet l'affinité est multipliée par 5 pour les ligands avec 8,10 et 12 nucléotides d'espacement (*Table 3*).

**Table 3. Affinité par ITC du RRM1 avec des ligands ADN comportant deux motifs**

<b>ligand</b>	<b>protéine</b>	<b>Kd ± err (uM)</b>
AGCACA	RRM1	<b>30 ± 5</b>
AGCACA(4T)AGCACA	RRM1	<b>24 ± 2.8</b>
AGCACA(6T)AGCACA	RRM1	<b>24 ± 5</b>
AGCACA(8T)AGCACA	RRM1	<b>6.0 ± 0.3</b>
AGCACA(10T)AGCACA	RRM1	<b>5.2 ± 0.2</b>
AGCACA(8T)AGCTCA	RRM1	<b>70 ± 12</b>
TTCAC(T) <sub>11</sub> CACT	RRM1	<b>33 ± 2</b>
AC(6)	RRM1	<b>58 ± 4</b>
AC(9)	RRM1	<b>13.8 ± 0.7</b>

L'analyse du complexe RRM1 avec un tandem de motif par RMN bidimensionnel laisse apparaître un complexe en échange rapide contrairement au complexe avec un motif de 6 nucléotides. Ainsi tous les résidus sont observables sur les spectres HSQC (*Figure 41*) offrant la possibilité d'étudier la structure du complexe en solution.



**Figure 41. Spectre 1H-15N HSQC du complexe RRM1 avec un motif ARN en tandem**

En bleu la protéine MEC-8 (28-117). En rouge la protéine MEC-8 (28-117) avec la séquence d'ARN AGCACA(10U)AGCACA à un ratio 1/1.2. Les deux échantillons sont dans un tampon Tris 50 mM, 150 mM NaCl à pH 7.5.

Les perturbations des déplacements sont de telle sorte qu'il est impossible de suivre le déplacement de la majorité des pics. L'assignement du spectre de MEC-8 RRM1 en complexe nécessite donc l'acquisition de nouveaux spectres 3D (HNCA/HNCO). Malheureusement à ce jour toutes tentatives pour obtenir de tels spectres de qualité suffisante pour l'assignement des résonances du complexe sont restées infructueuses.

### III. Structures cristallographiques du RRM1 en complexe avec un acides nucléiques

#### 1) Structure du complexe protéine-ADN

Les résultats obtenus lors de l'analyse interactionnelle du premier domaine de MEC-8 privilégient l'étude structurale de MEC-8 RRM1 en complexe avec un tandem de motifs séparés par 8 Thymidines.

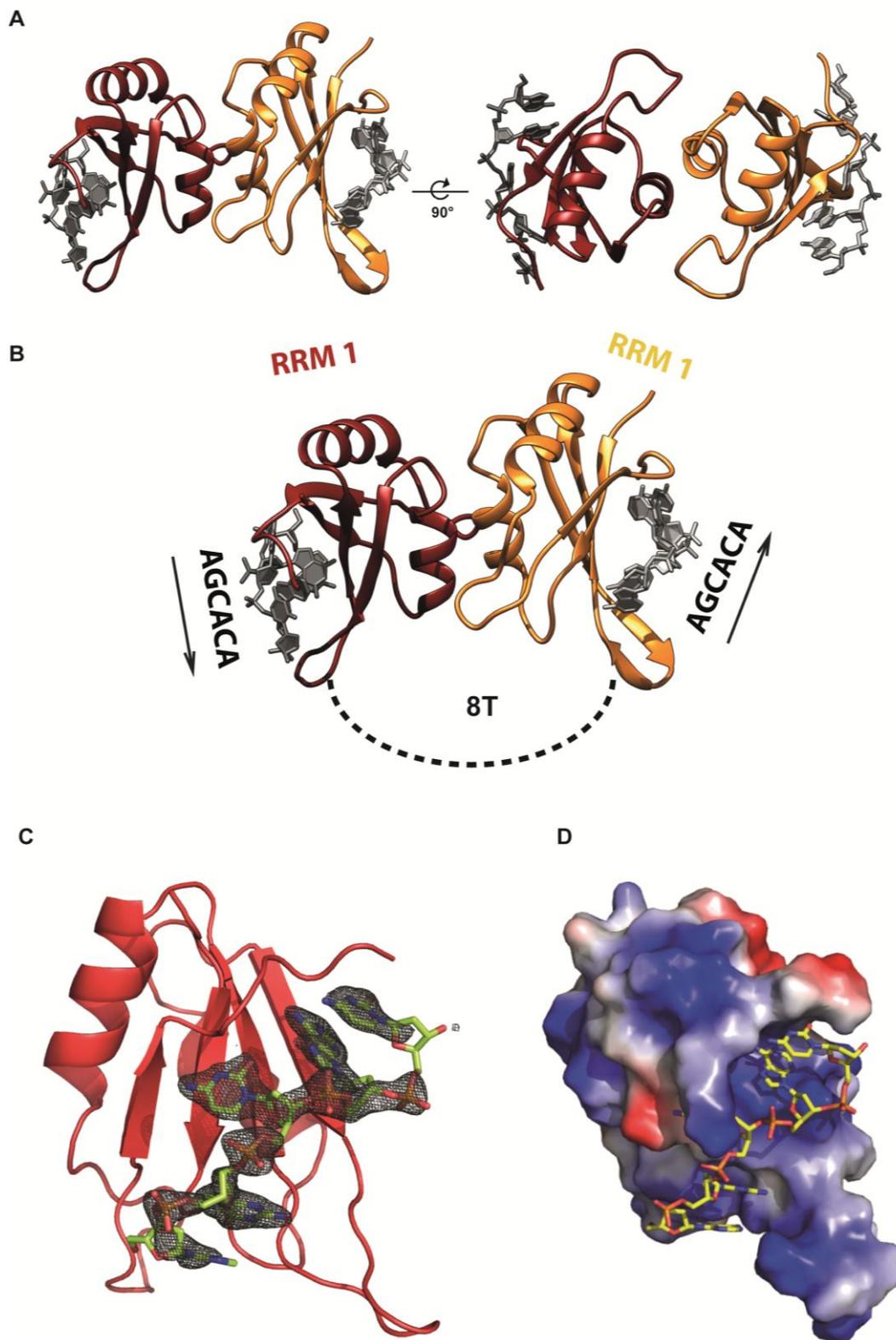


Contrairement au domaine à l'état libre, le domaine RRM1 complexé à un tandem de motif est très cristalogénique. Les gouttes de cristallisation produisent une quantité très importante de cristaux sur un spectre de conditions relativement large en un court délai (24-48h). Les cristaux sont tous sous la forme d'un X, cependant leur taille est hétérogène et très peu d'entre eux sont suffisamment grand pour être analysés sur une source synchrotron (>30µm).

Le ligand AGCACA(10T)AGCACA donne les mêmes résultats, cependant les cristaux diffractés pour ce ligand n'ont pas atteint des résolutions équivalentes à celle du ligand 8T, 4Å contre 1,5Å respectivement. Ces résultats sont vraisemblablement à imputer à des raisons de tailles des cristaux analysés et non aux propriétés intrinsèques du complexe.

La densité pour le ligand apparaît uniquement pour les cinq premiers résidus (AGCAC) du motif pour chaque RRM du dimère. Il manque donc la dernière Adénine de chaque motif ainsi que les 8 thymidines composant le linker entre les deux motifs. Néanmoins l'empilement du dimère dans le cristal ne laisse pas de doute quant à l'orientation du ligand dans l'unité assymétrique (*Figure 42.B*).

Nous avons donc une structure composée d'un homodimère dont chaque monomère lie un motif AGCAC. Chaque domaine lie la séquence d'ADN de façon canonique (*Figure 42.C*)



**Figure 42. Structure du complexe RRM1 avec le motif d-AGCACA**

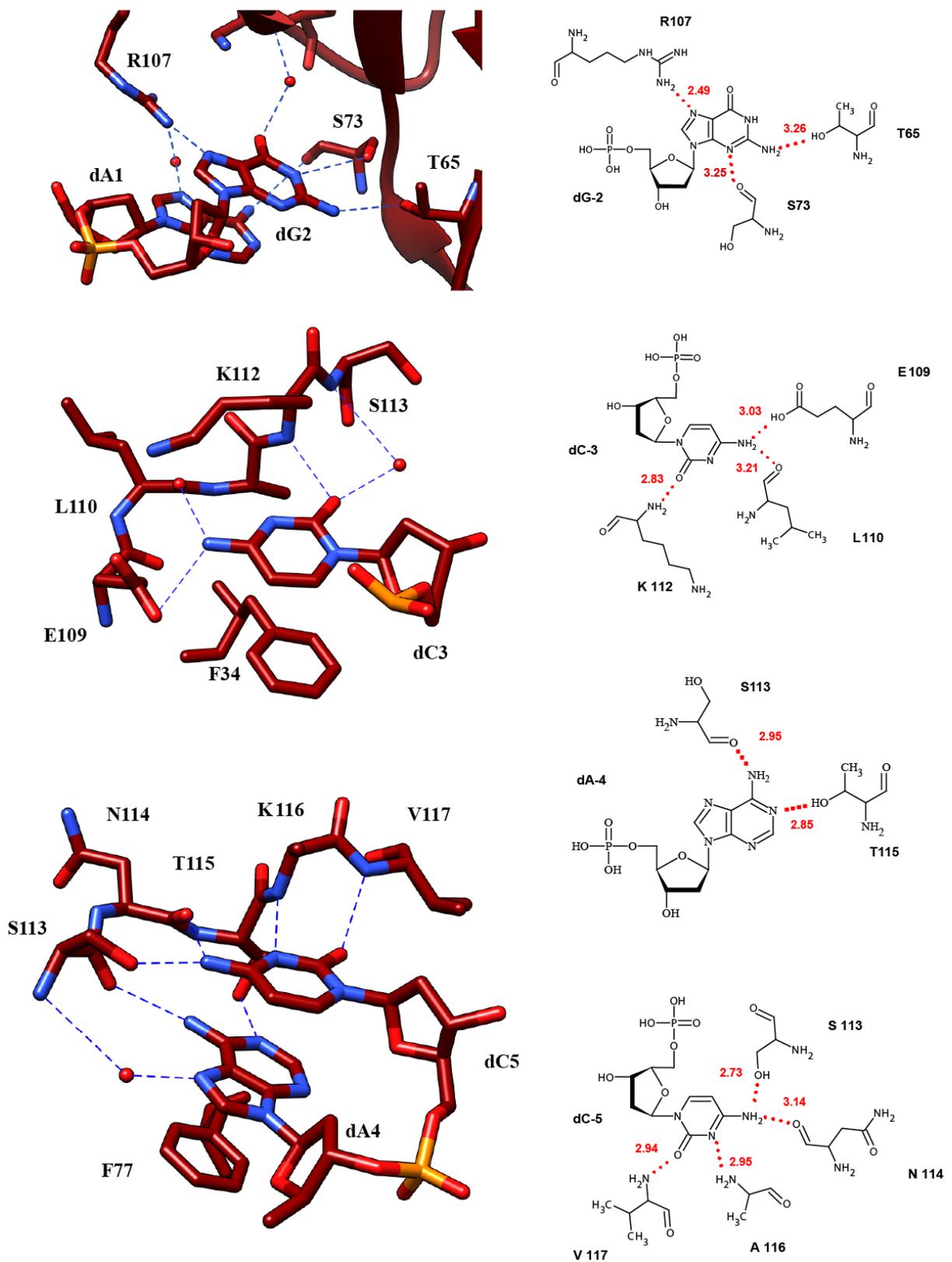
(A) Représentation cartoon vue de face et de haut du dimère MEC-8 RRM1 en complexe avec le motif d'ADN AGCACA(8T)AGCACA. Seules les bases AGCACA sont visibles (ici représentées en gris). (B) L'empilement cristallin laisse deviner l'orientation du ligand, chaîne A en rouge et chaîne B en orange. (C) Densité électronique du motif AGCACA ( $2FO-Fc$   $2\sigma$ ). (D) Surface électrostatique d'un monomère (chaîne A), le ligand est en jaune.

Les résidus sont disposés sur le feuillet beta de manière transversale sous une forme d'escalier caractéristique. Les bases A1 et G2 sont en orientation inverse du reste du motif, situées vers le bas du domaine RRM elles interagissent principalement avec la boucle  $\beta 2/\beta 3$  (*Figure 42.D*).

La base C3 effectue un empilement aromatique avec la phénylalanine 34 et est reconnue par trois résidus de la queue C-terminale (E109, L110 et K112).

De façon similaire la base A4 effectue un empilement aromatique avec la phénylalanine 77, la base C5 vient à son tour s'empiler sur l'adénine 4 et toute les deux interagissent de manière extensive avec la queue C-terminal du RRM1 (*Figure 43*).

La partie C-terminale joue donc un rôle important dans la spécificité de la reconnaissance du motif ADN. Cependant une grande partie des interactions se font à travers des liaisons entre les bases de l'ADN et des atomes du squelette peptidique.



**Figure 43. Détails atomique du complexe RRM1-ADN.**

*Représentation des interactions entre le motif d'ADN AGCAC et la protéine MEC-8 (28-117). A gauche une représentation sous forme stick du complexe cristallographique, à droite une représentation schématique. Les distances sont en rouge et l'unité de mesure est l'Angström.*

**Table 4. Statistiques de la structure du complexe RRM1-ARN****Collection Datas**

Resolution range ( $\text{\AA}$ ...)	25.74 - 1.529 (1.584 - 1.529)
Space group	P 32 2 1
Unit cell	52.188 52.188 156.534 90 90 120
Total reflections	76451 (7524)
Unique reflections	38386 (3768)
Multiplicity	2.0 (2.0)
Completeness (%)	99.99 (100.00)
Mean I/sigma(I)	8.03 (2.26)
R-merge	0.03396 (0.3595)
R-meas	0.04802

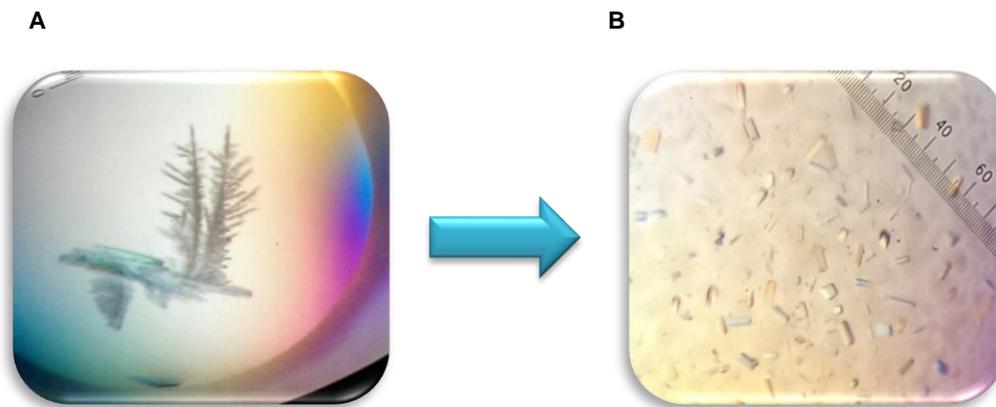
**Refinement**

R-work	0.2419 (0.4900)
R-free	0.2674 (0.4920)
Number of non-hydrogen atoms	1733
macromolecules	1581
water	152
Protein residues	188
RMS(bonds)	0.019
RMS(angles)	1.56
Ramachandran favored (%)	98.26
Ramachandran allowed (%)	1.74
Ramachandran outliers (%)	0
Average B-factor	49.50
macromolecules	49.40
solvent	50.50

## 2) Structure du complexe protéine-ARN

Les bons résultats obtenus avec le complexe RRM1-ADN nous ont donc naturellement orientés vers des essais de cristallisations avec le même ligand sous forme d'ARN. La différence entre les deux complexes étant à priori très mince voire inexistante, les premiers essais ont été réalisés dans des conditions identiques.

Une grande quantité de cristaux a été obtenue. Cependant ces cristaux ne sont jamais apparus sous forme de monocristaux de qualité suffisante pour la diffraction (*Figure 44*).



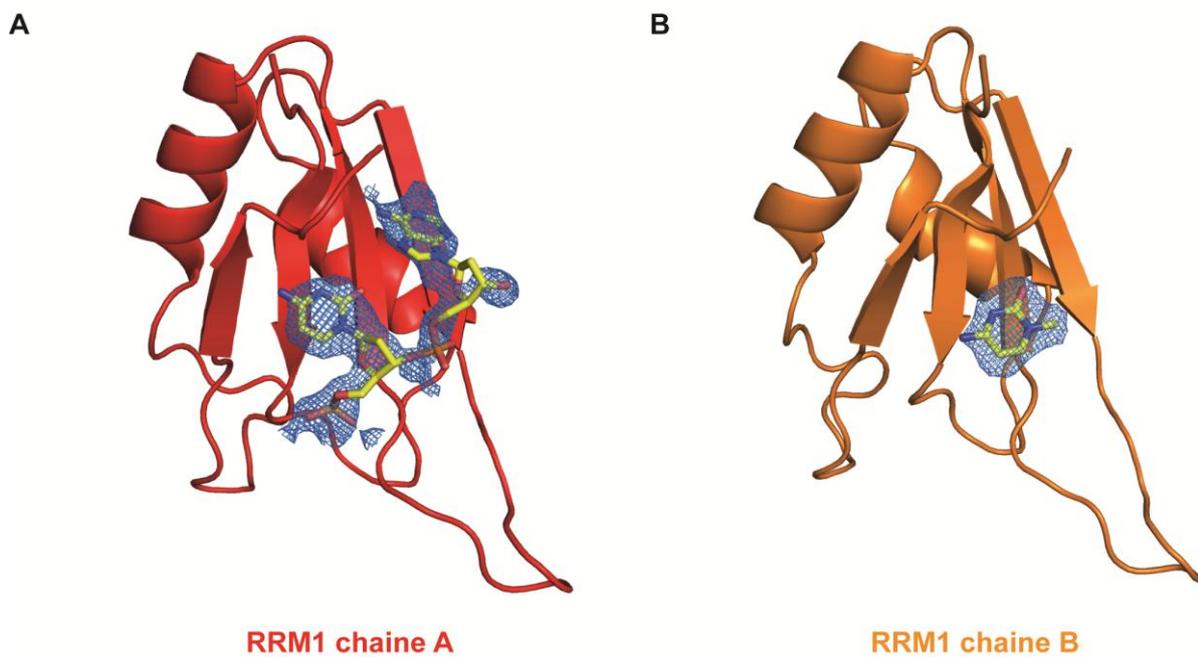
**Figure 44. Optimisation des cristaux du RRM1 en complexe avec ARN**

(A) cristal de MEC-8 (28-117) en complexe avec AAGCACA(8U)AGCACAA. Les conditions de cristallisation sont 10% isopropanol, 10% PEG(4000) en goutte assise avec un réservoir de 1 ml (B) Les conditions sont 13% isopropanol, 10 % PEG(4000) en goutte assise avec un réservoir de 1 ml.

Un criblage autour de ces conditions a été réalisé sur plaque 24 puits afin d'obtenir des cristaux de meilleures qualités. Après un temps relativement long (~1 mois) des cristaux de forme régulière et de taille importante ont finis par apparaître dans un puits du crible (*Figure 44.B*).

Une fois analysés sur une source synchrotron, les cristaux sous une forme cristalline différente (I4) ont permis la résolution d'une structure à 1.75Å.

Dans les deux jeux de données la densité électronique pour le ligand est de mauvaise qualité et discontinue. Il est uniquement possible d'observer deux bases C3 et A4 pour la chaîne A et la base A4 pour la chaîne B du dimère RRM1 (*Figure 45*).



**Figure 45. Le domaine RRM1 en complexe avec un motif ARN**

(A) La chaîne A du dimère en complexe avec deux bases ARN. (B) La chaîne B du dimère en complexe avec une base ARN.

La densité électronique pour l'ARN n'est pas de qualité comparable à celle du ligand ADN. Néanmoins le remplacement moléculaire et l'affinement de la structure ont été réalisés sans présence de ligand dans le modèle. Il n'y a donc pas d'ambiguïtés quant à la présence de l'ARN dans le cristal. De plus le peu d'informations que nous pouvons tirer de ces résultats concordent avec la structure du complexe avec ADN.

**Table 5. Statistiques de la structure du complexe RRM1-ADN****Collection datas**

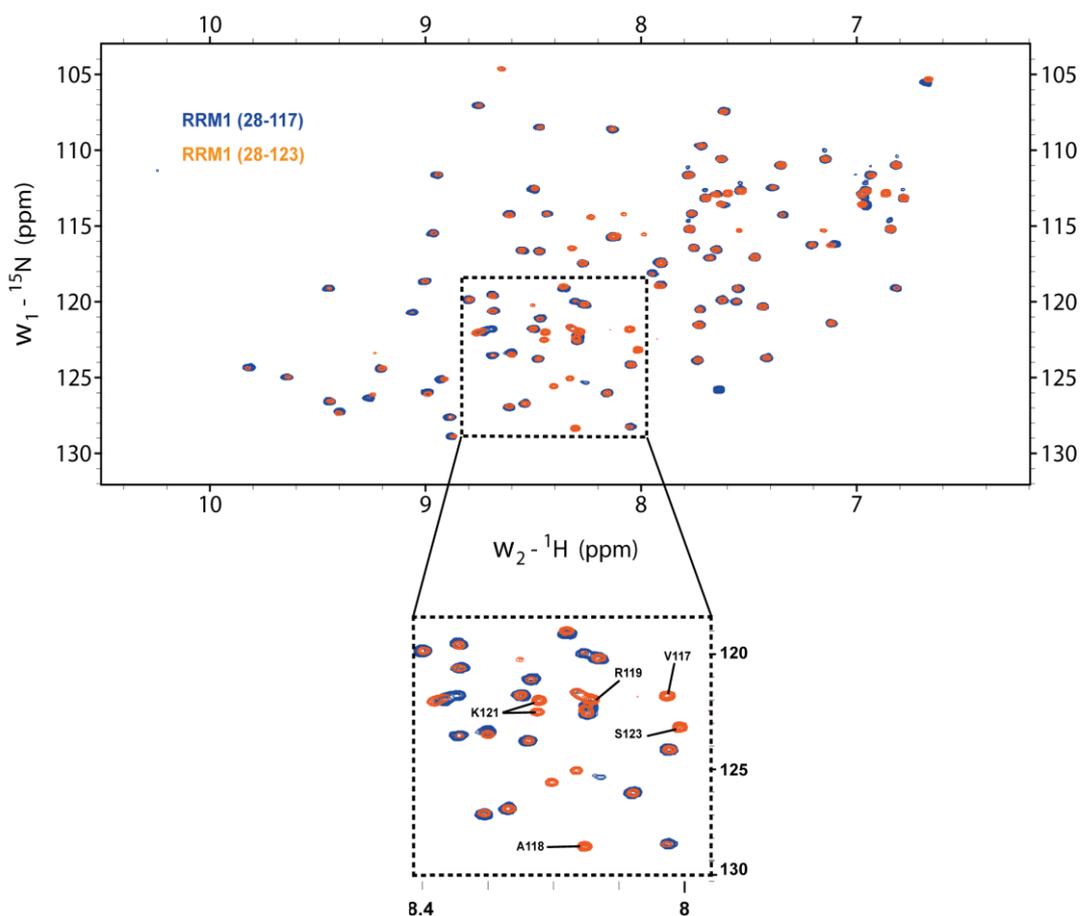
Resolution range (Å...)	34.62 - 1.797 (1.861 - 1.797)
Space group	I 4
Unit cell	103.252 103.252 52.334 90 90 90
Total reflections	188053 (17854)
Unique reflections	25744 (2489)
Multiplicity	7.3 (7.2)
Completeness (%)	99.61 (97.34)
Mean I/sigma(I)	18.05 (2.45)
R-merge	0.06132 (0.6225)
R-meas	0.06618

**Refinement**

R-work	0.1928
R-free	0.2196
Number of non-hydrogen atoms	1528
macromolecules	1381
water	147
Protein residues	175
RMS(bonds)	0.005
RMS(angles)	0.86
Ramachandran favored (%)	100
Ramachandran allowed (%)	0
Ramachandran outliers (%)	0
Average B-factor	40.20
macromolecules	39.40
solvent	46.90

#### IV. Importance de la région C-terminal du RRM1

L'analyse du complexe cristallographique du domaine RRM1 (28-117) nous laisse donc observer un domaine liant un motif de cinq acides nucléiques. La partie C-terminale de la protéine est entièrement impliquée dans la liaison avec le motif, la valine 117 interagissant avec la cytidine 5.

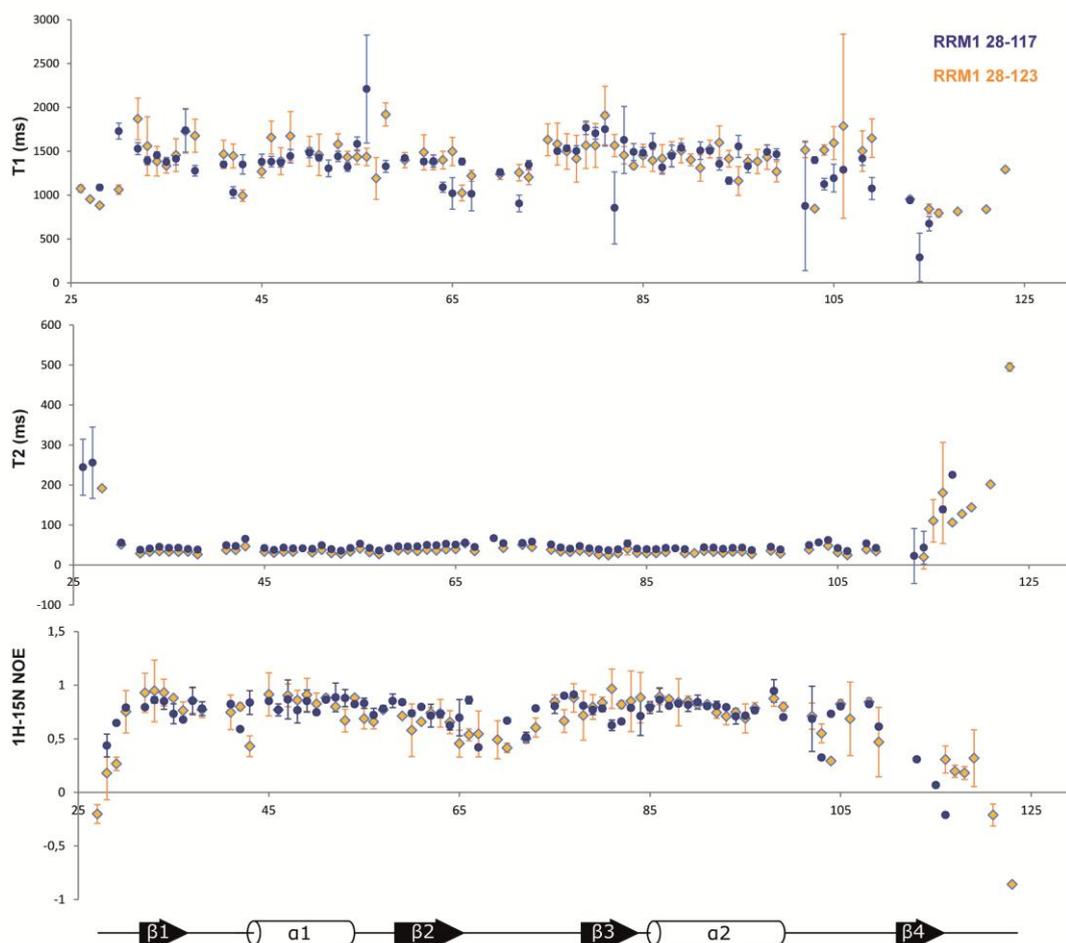


**Figure 46. Le domaine RRM1(28-123).**

*En bleu la construction 28-117 du domaine RRM1, en orange la construction élargie en c-terminal. Les pics orange supplémentaires non assignés de moindre intensité sont un reste de his-tag clivé.*

Au regard de ces résultats nous avons décidé d'aller plus loin dans l'exploration de la partie C-terminale du domaine protéique. Une construction étendue de cinq acides aminés (RRM1 28-123) a donc été produite.

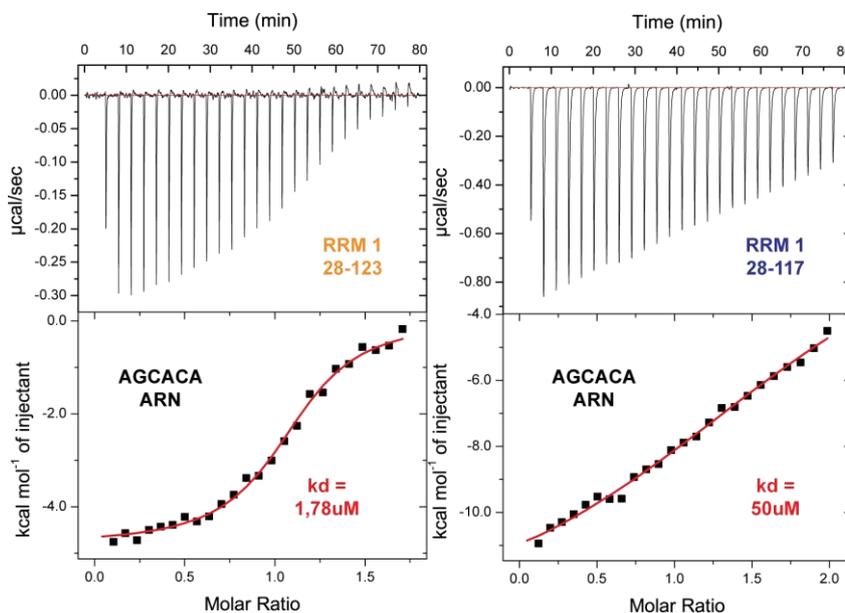
L'acquisition d'un spectre 15-N HSQC nous renseigne sur la nature non structurée de cette extension C-terminale. En effet les pics correspondant à ces acides aminés sont dans une région caractéristique des résidus appartenant à des parties faiblement structurées des protéines (*Figure 46*).



**Figure 47. Etude dynamique de la région C-terminal du RRM1**

Ce désordre conformationel a été vérifié en effectuant des expériences de relaxation par RMN. On observe que la queue C-terminale arbore des valeurs différentes de celle du cœur du domaine RRM (*Figure 47*), cette queue désordonnée commence au résidu 110. La région

110 à 117 qui ne présente pas de densité électronique sous forme cristalline lorsque la protéine est à l'état libre car trop dynamique se rigidifie et apparait comme participant activement au complexe dans le cristal de la protéine avec son ligand.

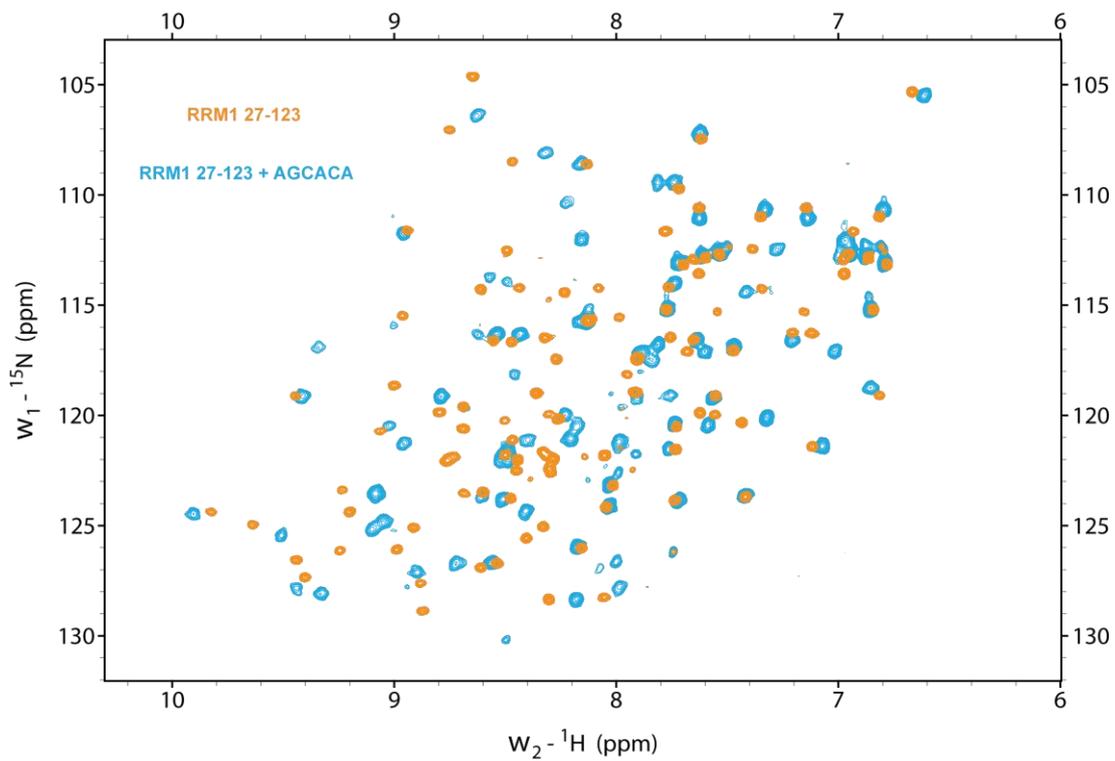


**Figure 48. Apport d'une construction élargie en C-terminal en termes d'affinité.**

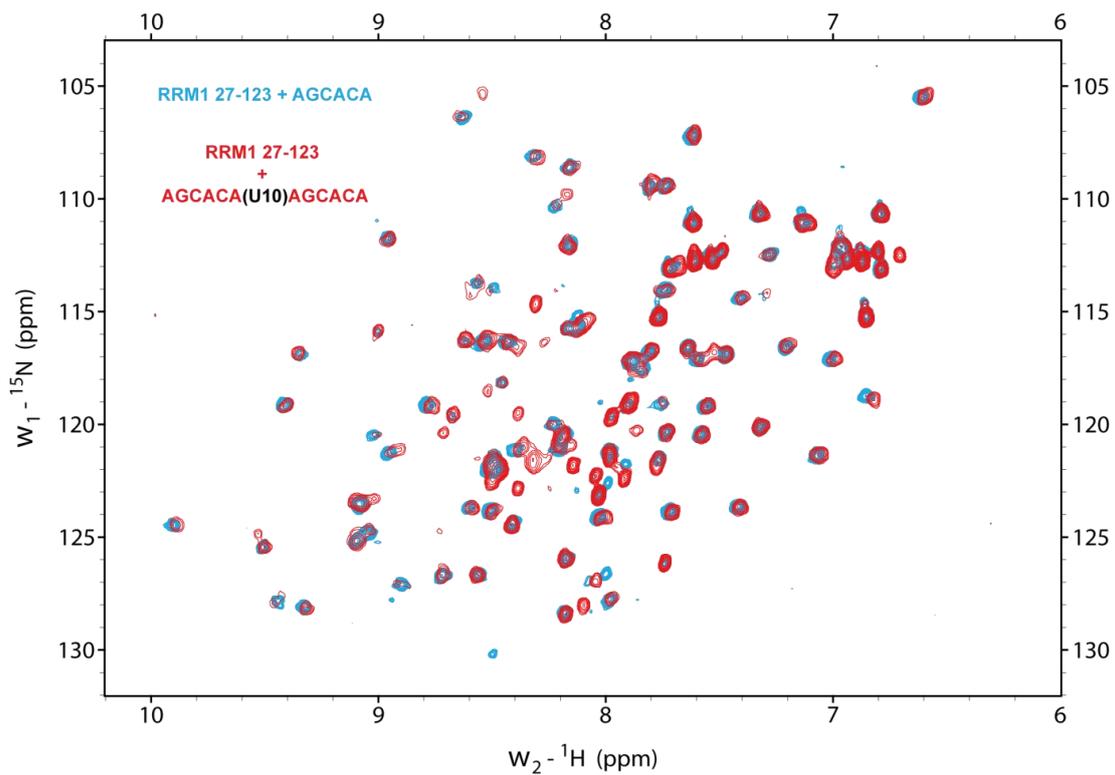
Les études d'interaction entreprises avec la construction protéique 28-123 permettent de conclure que celle-ci apporte une importante augmentation de l'affinité (~30 fois) par rapport à la construction 28-117 (*Figure 48*). Enfin contrairement au spectre 15N-HSQC du RRM1(28-117)-AGCACA celui du RRM1(28-123)-AGCACA est cette fois ci en échange rapide et de qualité suffisante pour l'analyse par RMN (*Figure 49*).

De plus si l'on superpose les spectres du RRM1(28-123) en complexe avec AGCACA(10U)AGCACA et AGCACA on observe des perturbations des déplacements chimiques équivalentes pour la quasi-totalité des résidus (*Figure 50*).

Cela confirme que les résidus impliqués dans la liaison du motif AGCACA sont les même que ceux impliqués dans la liaison du tandem de AGCACA et donc que les 5 résidus supplémentaires stabilisent uniquement les bases du motif et n'interagissent pas avec le linker du tandem.



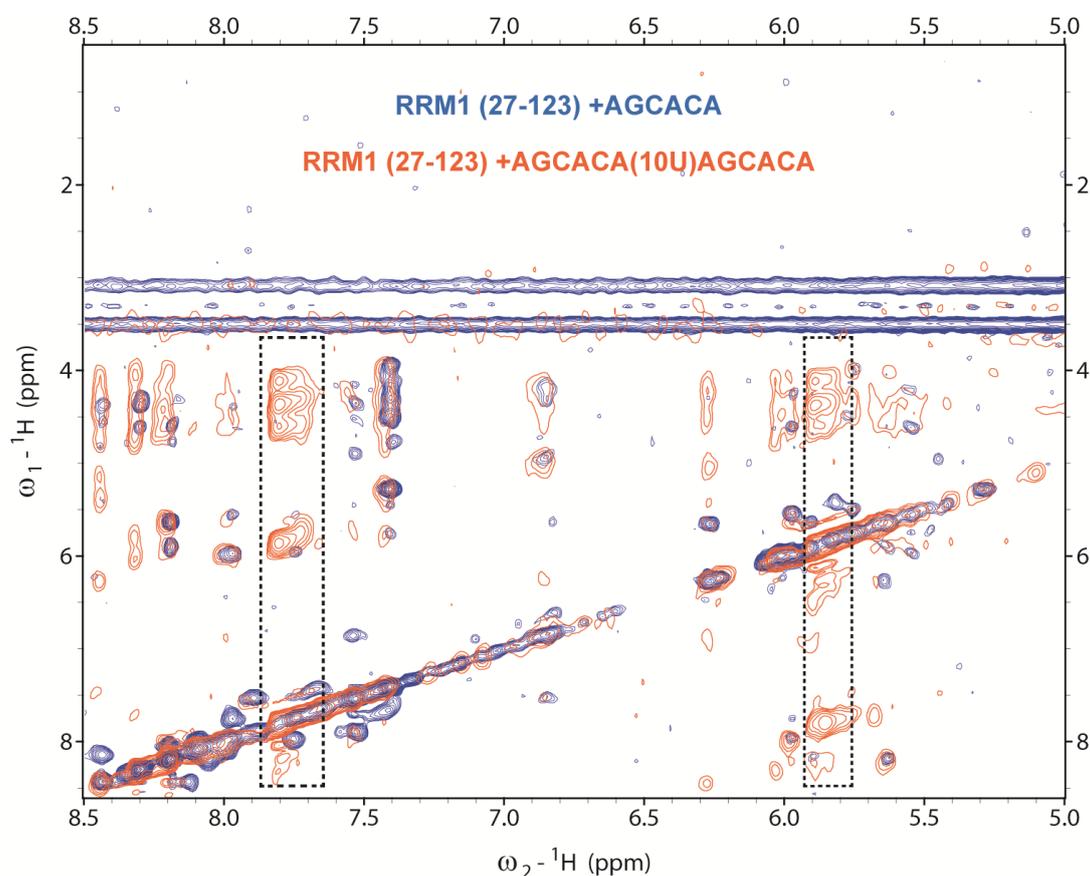
**Figure 49.** Spectre  $1\text{H}-15\text{N}$  HSQC de MEC-8 27-123 en complexe avec AGCACA



**Figure 50.** Comparaison des déplacements chimiques entre les complexes avec motifs simples et motifs en tandems

## V. Etude structurale en solution du complexe RRM1-ARN

L'utilisation d'une construction rallongée en C-terminal pour le domaine RRM1 nous a permis d'augmenter de façon importante l'affinité de celui-ci avec son motif ARN (*Figure 48, Figure 54.B*). Nous avons donc analysé les différents complexes RRM (28-123)-AGCACA et RRM (28-123)-AGCACA(10U)AGCACA par spectroscopie RMN. Le spectre 1H-13C half filtered NOESY laisse apparaître dans les deux cas une structuration du motif ARN. Pour le motif en tandem on identifie aisément les 10 Uridines (*Figure 51*) manquante pour le motif court.



**Figure 51. Spectres 1H-13C half filtered NOESY**

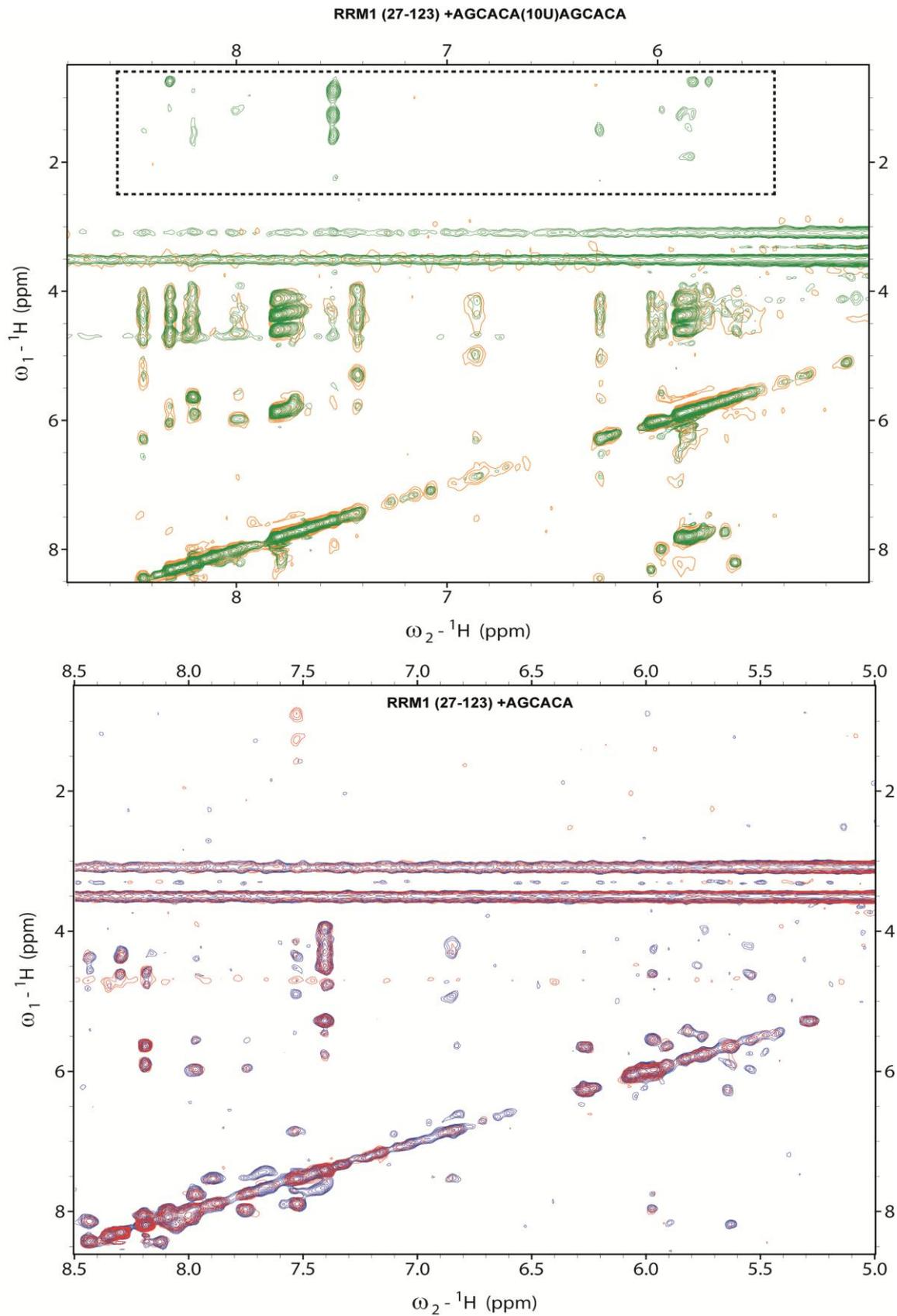
Réalisé dans 100% D2O à 295K. En rouge le complexe RRM1-AGCACA(U10)AGCACA, en bleu le complexe RRM1-AGCACA. En pointillé sont encadrés les résonances appartenant aux 10 uridines du linker pour le motif en tandem.

L'analyse des spectres 1H-13C double half filtered NOESY des complexe RRM-ARN montre des pics de corrélations Protéine-ARN de bonne qualité et en nombre important pour le motif en tandem. Le complexe avec un motif AGCACA laisse apparaitre un spectre avec très peu d'informations (*Figure 52*).

L'assignement de la protéine en complexe étant manquant, ces résultats ne sont pas exploitables pour le moment. Cependant l'étude préliminaire du complexe RRM-ARN donne des résultats très prometteurs qui feront l'objet d'une étude plus approfondie menant éventuellement à la résolution d'une structure protéine-ARN en solution.

**Table 6. Effet sur l'affinité de liaison des mutations de MEC-8**

Protéine	Mutation	Ligand	Kd (M)
Mec-8 (28-123)	R107K	AGCACA	$3.7 \times 10^{-5}$
Mec-8 (28-123)	K63A	AGCACA	$1.88 \times 10^{-5}$
Mec-8 (28-123)	F34A	AGCACA	$5.95 \times 10^{-5}$
Mec-8 (28-123)	E109Q	AGCACA	$1.33 \times 10^{-5}$
Mec-8 (28-123)	-	AGCACA	$1.7 \times 10^{-6}$
Mec-8 (28-117)	$\Delta$ (117-123)	AGCACA	$3 \times 10^{-5}$

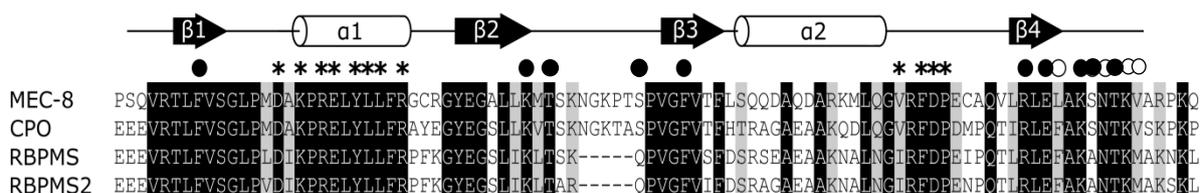


**Figure 52. Spectres 1H-13C double half filtered NOESY**

L'encadré pointillé représente les NOE inter moléculaire entre la protein et le ligans. En vert et rouge sont représenté les double et en bleu et orange les simple filtered 1H13C-NOESY. Réalisé dans 100% D2O à 295K.

## VI. Etude de la famille de protéine RBPMS

Le domaine RRM1 de MEC-8 possède des homologues chez l'humain, RBPMS et RPMS2 ainsi que chez *D. melanogaster* avec la protéine CPO. Les domaines RRM 1 de ces quatre protéines son très proches d'un point de vue de séquence avec une identité d'environ 60% (*Figure 53*).



**Figure 53. Le domaine RRM de la famille RBPMS**

Alignement des séquences des domaines RRM de la famille de protéine RBPMS. Ici la séquence de MEC-8 27-123 est annotée avec des étoiles pour représenter les résidus impliqués dans l'homodimérisation du domaine. Les ronds pleins représentent les résidus interagissant avec le motif AGCAC avec leurs chaînes latérales, les ronds vides représentent les résidus interagissant avec le squelette peptidique.

Les régions d'homodimérisation ainsi que les résidus impliqués dans la liaison à l'ARN du RRM1 de MEC8 sont conservés. Une étude interactionnelle a été menée afin de confirmer la spécificité de ces protéines pour les séquences GCAC.

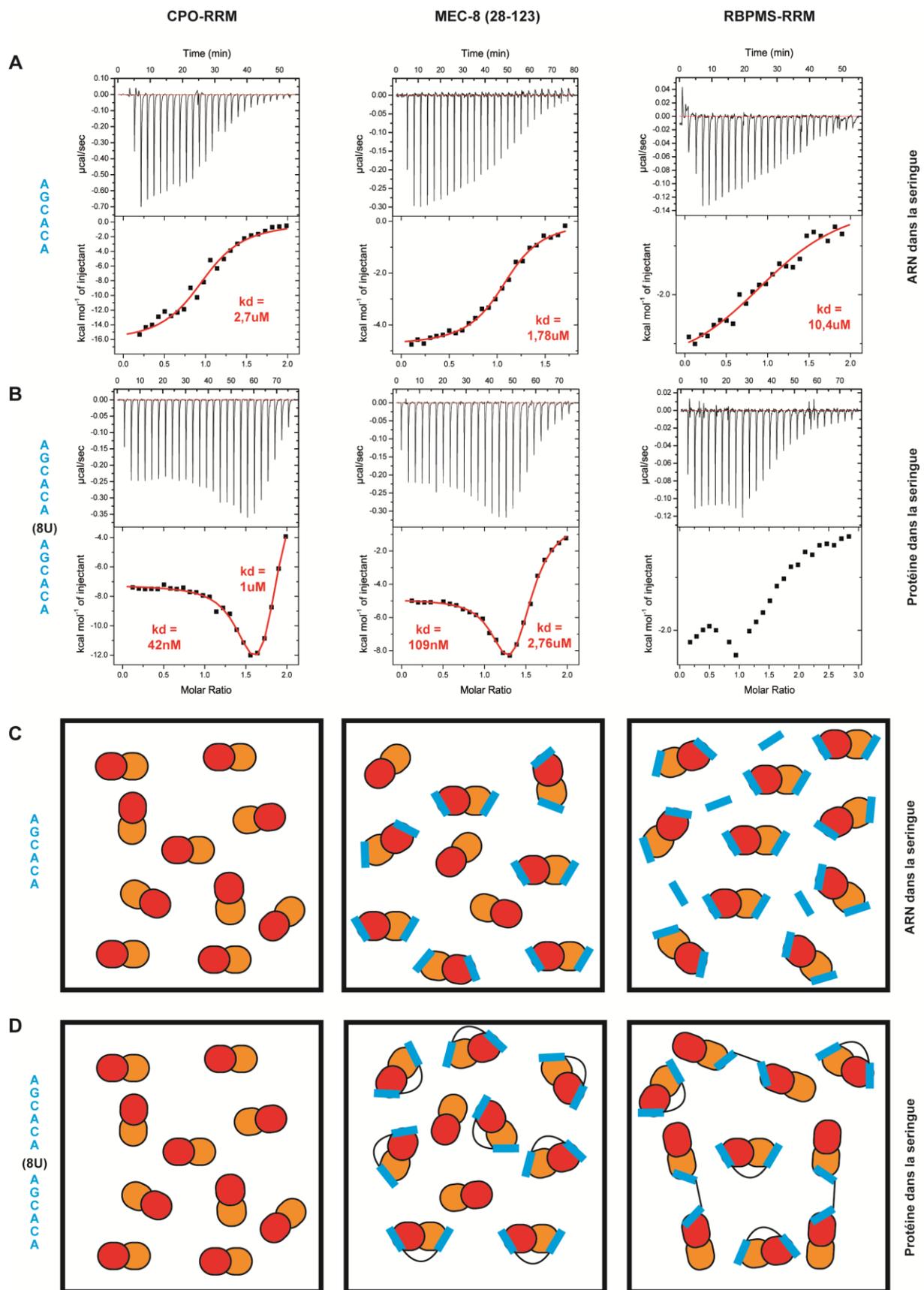
Les expériences d'ITC avec des motifs ARN court de 6 bases (*Figure 54*) donnent des résultats équivalents pour les RRM des protéines MEC-8 et CPO. Le RRM de RBPMS a une affinité un peu moindre mais celle-ci reste raisonnable. Ce domaine a la particularité d'avoir une boucle β2β3 réduite de cinq résidus par rapport à CPO et MEC-8 RRM1. Or cette région du domaine RRM a été décrite comme interagissant directement avec la base G du motif ARN (*Figure 43*). La protéine RBPMS ne semble pas non plus avoir la spécificité de reconnaître la séquence AGCAAA comme le RRM1 de MEC-8 (*Table 7*). Il est donc possible que le domaine RRM de RBPMS ai une séquence de liaison optimale un peu différente de celle de MEC-8, une étude bio-informatique a prédit un ligand CAC (Farazi, Leonhardt et al. 2014).

Des investigations plus poussées seront conduites pour élucider la potentielle divergence de reconnaissance entre MEC-8 et ses homologues humains RBPMS et RBPMS2.

**Table 7. Affinité par ITC, des membres de la famille RBPMS ;**

Ligand	Protéine	Kd (M)
AGCACA	MEC-8 (28-123)	$1.8 \times 10^{-6}$
AGCAUA	MEC-8 (28-123)	NA
AGCAAA	MEC-8 (28-123)	$2.95 \times 10^{-5}$
AGCACA	CPO-RRM	$2.7 \times 10^{-6}$
AGCAUA	CPO-RRM	NA
AGCACA	RBPMS-RRM	$1 \times 10^{-5}$
AGCAAA	RBPMS-RRM	NA
AGCAUA	RBPMS-RRM	NA

L'analyse de l'interaction de ces trois RRM avec des séquences ARN comportant des motifs en tandem montre des affinités élevées pour ceux-ci, de l'ordre du nano-molaire (*Figure 54.B*). Ces expériences effectuées avec la protéine dans la seringue de l'appareil ITC montre un pattern caractéristique d'un phénomène de liaison à deux sites (*Figure 54.C*). Cela marque la dichotomie entre une liaison bipartite coordonnée de deux domaines RRM d'un même dimère et la liaison de deux RRM appartenant à deux dimères différents qui intervient lorsque la protéine est en saturation.



**Figure 54. Reconnaissance bipartite de la famille RBPMS**

(A) ITC avec court motifs (ARN dans la seringue). (B) ITC avec motifs en tandem (protéine dans la seringue). (C) (D) Modèles de liaison avec motifs court et en tandem.

## VII. Matériels et méthodes

### 1) Clonage

Les différentes constructions du premier RRM de la protéine MEC-8 sont clonées par amplification de la séquence nucléotidique souhaitée par PCR. Les oligo-nucléotides utilisés contiennent les sites de coupure des enzymes de restriction NcoI et Acc65I pour l'insertion dans un plasmide dérivant de la famille pET-9d contenant en position N-terminal une séquence poly-histidine et un site de clivage pour la protéase du virus de la gravure du tabac. Les mutants sont produits de la même façon, les mutations sont introduites avec des amorces comportant les mutations voulues.

### 2) Production

Les vecteurs issus du clonage sont introduits dans des bactéries BL21 LysY (New England Biolabs) par le biais d'une réaction de transformation. Ces bactéries sont cultivées en milieu Luria Broth (LB) ou en milieu minimum M9 en présence de kanamycine supplémenté avec du  $^{15}\text{NH}_4\text{Cl}$  ( $1\text{g}\cdot\text{L}^{-1}$ ) et du  $^{13}\text{C}_6$ -glucose ( $2\text{g}\cdot\text{L}^{-1}$ ) afin d'obtenir des échantillons isotopiquement marqués pour la spectroscopie RMN. La culture cellulaire est réalisée à  $37^\circ\text{C}$  jusqu'à atteindre une densité optique (à 600 nm) comprise entre 0.6 et 0.8. L'expression des protéines par les bactéries est alors induite avec 0.5 mM de  $\beta$ -D-thiogalactopyranoside (IPTG), pour la nuit à  $25^\circ\text{C}$ . Nous récoltons les cellules par centrifugation (15 min à 5000 g) et le culot est resuspendu dans un tampon de lyse 50 mM Tris (pH7.5), 500 mM NaCl, avant la lyse bactérienne par digestion (lysozyme) et sonication. La solution est alors nettoyée par centrifugation (25 min, 12000 g).

### 3) Purification

La purification est réalisée par chromatographie d'affinité ( $\text{Ni}^{2+}$ ) avec les solutions composées de 50 mM Tris (pH7.5), 500 mM NaCl, et 5, 25 ou 500 mM d'imidazole pour le chargement, le lavage et l'élution respectivement. La queue poly-histidine est éliminée à température ambiante sur la nuit par digestion à la TEV, après un changement de solution vers une solution compatible avec les expériences de RMN i.e. 20 mM phosphate avec 150 mM de NaCl à pH 6.5. La concentration de l'échantillon et l'élimination du peptide poly-histidine sont réalisées grâce à un filtre Vivaspin 20 ayant un cutoff de 5000 kDa (Sartorius Stedim Biotech). La concentration est déterminée par la mesure de la  $\text{DO}_{280}$  avec un coefficient d'extinction molaire de  $2980 \text{ M}^{-1} \cdot \text{cm}^{-1}$ . Après la digestion, les acides aminés Gly-Ala-Met-Ala sont laissés en amont du domaine protéique.

### 4) Synthèse des acides nucléiques

Les séquences ADN et les séquences ARN de 6bases ont été ont été commandées chez Eurogentech. Les séquences ARN de tailles supérieures à 6 nucléotides ont été synthétisées sur une ÄKTA oligopilot plus 10 (GE healthcare) et dessalé à l'aide d'une colonne NAP-5 (GE healthcare).

### 5) Spectroscopie de résonance magnétique nucléaire

Les échantillons RMN contiennent environ 1mM de protéine dans 20 mM de tampon phosphate (pH 6.5) et 50 mM de NaCl, supplémenté par 10 % ou 99 %  $^2\text{H}_2\text{O}$ . Les spectres sont enregistrés à 298K avec les spectromètres Bruker Avance III à 800 MHz ou 700 MHz, équipés d'une sonde triple résonance cryoréfrigéré ou standard respectivement. Les attributions du squelette peptidique et des chaînes latérales aliphatiques furent déterminées à

partir des cartes 2D  $^{15}\text{N}$ -HSQC, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCA, 3D HNCO, 3D HCACO enregistrées dans 10 %  $^2\text{H}_2\text{O}$  et  $^{13}\text{C}$ -HSQC, 3D HC(C)H-TOCSY, 3D (H)CCH-TOCSY enregistrées dans 99 %  $^2\text{H}_2\text{O}$ .

Les mesures de relaxation  $^{15}\text{N}$  ont toutes été réalisées à 700 MHz et 298 K. Les spectres  $^1\text{H}$ - $^{15}\text{N}$  HSQC Hétéronucléaire ont été enregistré avec et sans 3 secondes de saturation du  $^1\text{H}$ .

Tous les spectres ont été traité avec le logiciel NMRpipe (Delaglio, Grzesiek et al. 1995), l'assignement des raies a été réalisé avec le logiciel CCPNmr.

## 6) Spectroscopie des rayons X

La protéine recombinante MEC-8 RRM1 est concentrée à environ  $10\text{mg}\cdot\text{ml}^{-1}$  dans un tampon 150 mM NaCl, 50mM Tris (pH 7.5).

- **RRM1 (28-117) libre**

Le domaine protéique seul a été cristallisé par diffusion de vapeur en goutte assise avec un réservoir de 70  $\mu\text{l}$  contenant 0.1M de sodium acétate (pH 6.5), 0.2 M ammonium sulfate et 30% polyéthylène glycol (PEG) 4000. La goutte est constitué de 0,2  $\mu\text{l}$  d'échantillon protéique mélangé au même volume de solution du réservoir.

Les cristaux sont apparus après 48 h en forme de longue plaque. Ils n'ont pas été cryoprotégé avant diffraction.

Les données de diffraction des cristaux de MEC-8 RRM1 libre ont été récoltées à l'ESRF (European Synchrotron Radiation Facility) sur la ligne ID23-2.

- **RRM1 (28-117) + d-AGCACA(8T)AGCACA**

Le ligand d-AGCACA(8T)AGCACA a été ajouté au domaine RRM1 (28-117) à un ratio de deux ligand pour un RRM afin de former un complexe stable.

Les conditions de cristallisation sont 10% isopropanol et 10% PEG4000.

Le complexe a été cristallisé par diffusion de vapeur en goutte assise avec un réservoir de 1ml.

La goutte est constituée de 1ul de complexe supplémenté de 1ul de solution du réservoir.

Les cristaux ont été cryoprotégé pendant 1minute dans une solution composée de 6.7 % isopropanol, 6.7 % PEG4000 et 25 % glycérol avant congélation.

Les données de diffraction des cristaux de MEC-8 RRM1 avec ADN ont été récoltées à l'ESRF (European Synchrotron Radiation Facility) sur la ligne ID23-2.

- **RRM1 (28-117) + AGCACA(8U)AGCACA**

Le ligand AGCACA(8U)AGCACA a été ajouté au domaine RRM1 (28-117) à un ratio de deux ligand pour un RRM afin de former un complexe stable.

Les conditions de cristallisation sont 13% isopropanol et 10% PEG4000.

Le complexe a été cristallisé par diffusion de vapeur en goutte suspendu avec un réservoir de 1ml. La goutte est constituée de 1ul de complexe supplémenté de 1ul de solution du réservoir.

Les cristaux ont été cryoprotégés pendant 1minute dans une solution composée de 8.7 % isopropanol, 6.7 % PEG4000 et 25 % glycérol avant congélation.

Les données de diffractions du cristal de MEC-8 RRM1 avec ARN ont été collectées au synchrotron SOLEIL sur la ligne Proxima I.

Tous les jeux de données ont été indexés et réduit avec le programme XDS et la suite CCP4.

Le phasage a été réalisé par remplacement moléculaire avec la protéine  $\gamma$ -Pub1(pdb : 3MD3),

pour la structure RRM1 libre. Les structures du RRM1 en complexe ont été résolues en utilisant la structure du RRM1 libre comme modèle.

La construction et l'affinement de chaque modèle a été réalisé en utilisant le programmes Coot et la suite logicielle Phenix.

#### **7) Titration calorimétrique isotherme**

Les mesures de titration calorimétrique isotherme ont été réalisées sur un appareil ITC200 microcal (Malvern Instruments Ltd, UK) à 25 °C. Les protéines ont été dialysées dans un tampon 50 mM Tris, pH 7.5, 150 mM NaCl. 25 injections successives ont été réalisées à 500rpm avec un espacement de 180 secondes entre chaque injection. Pour les échantillons testés avec des ligands ADN, la protéine est dans la cellule. Pour les échantillons testés avec des ligands ARN, la protéine est dans la seringue. Les données ont été analysées avec le programme Origin version 5.0.

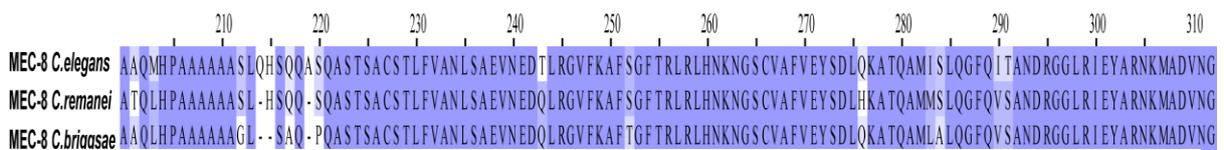
## Chapitre II : Structure et fonction du RRM2 de MEC-8

A la différence des autres protéines de la famille RBPMS, MEC-8 possède un second domaine RRM en position C-terminal. Ce domaine n'a aucun homologue proche en dehors des protéines MEC-8 présentes dans la famille des nématodes. Cependant les études génétiques sur MEC-8 qui ont permis de découvrir des phénotypes et donc une fonction pour cette protéine, utilisent toutes des organismes avec des mutations localisées dans la séquence codante pour le domaine RRM2. Il est donc très important d'étudier le rôle qu'a ce domaine dans le mécanisme d'épissage alternatif de MEC-8. Pour cela nous avons procédé à une étude fonctionnelle et structurale de ce domaine.

### I. Etude structurale du second domaine de MEC-8

#### 1) Obtention d'un domaine protéique soluble

Comme pour le RRM1 la première étape pour l'étude en solution du RRM2 de MEC-8 est l'obtention d'une construction protéique soluble. Dans ce cas il y a peu d'ambiguïtés quant à la nature structurée ou non de la partie en N-terminal du domaine RRM prédit tant la séquence est de pauvre complexité (**Figure 55**).



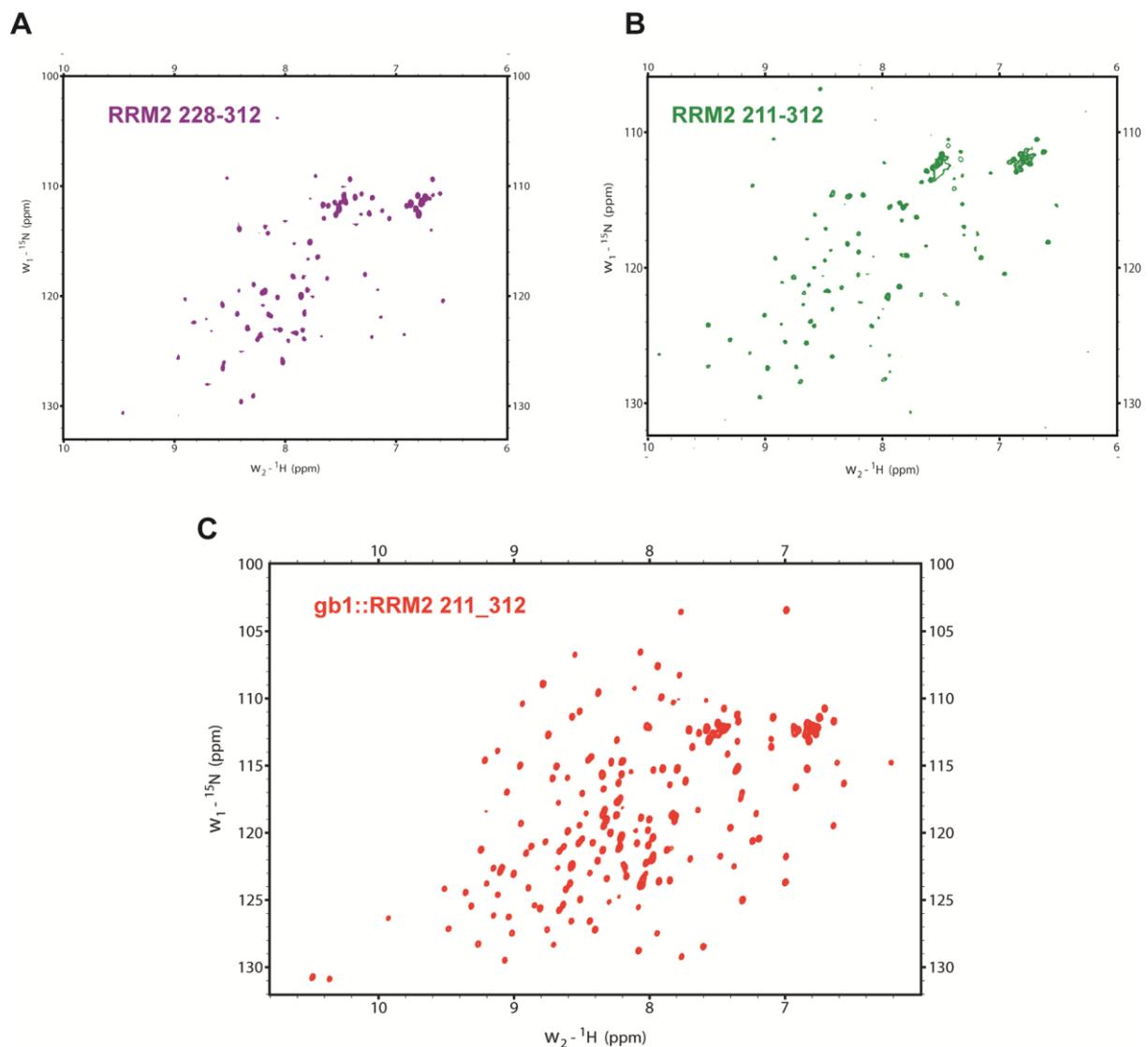
**Figure 55. Le domaine RRM2 de la famille de protéine MEC-8**

*Alignement des RRM C-terminaux des protéines MEC-8 de C.elegans, C.remanei et C.brissae*

Nous avons donc commencé par produire le domaine MEC-8 (228-312). La protéine très soluble (~ 0,3 mM) montre cependant un profil 1H15N HSQC de mauvaise qualité, il manque une quantité importante de pics, ceux présents ont des intensités différentes selon la

région du spectre et tendent à se concentrer dans la région des résidus déstructurés (*Figure 56.A*).

L'échantillon ne montre aucune trace de dégradation, cela laisse donc penser que la protéine n'est pas entièrement repliée du fait d'une construction trop courte. Nous avons donc décidé d'étendre la construction protéique en N-terminal pour produire le domaine MEC-8 (211-312). Cette fois-ci la protéine est très peu soluble, seule une partie minime se retrouve dans le lysat cellulaire et ne permet d'avoir un échantillon final d'une concentration de quelque dizaine de micro molaire. Cependant le spectre HSQC présente un domaine entièrement replié cette fois-ci (*Figure 56.B*).

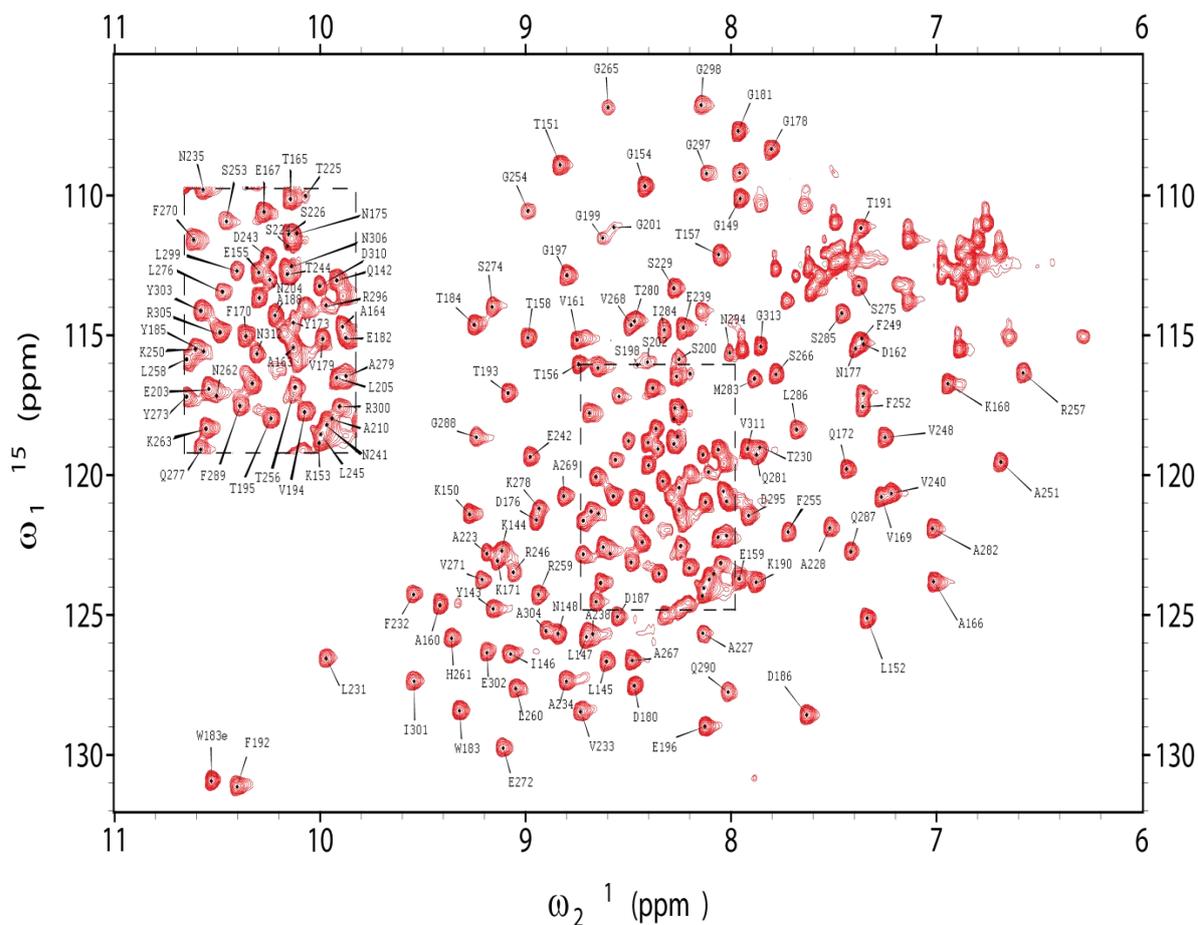


**Figure 56. Obtention d'un domaine protéique soluble pour le RRM2 de MEC-8**

Au regard de ces résultats, plusieurs vecteurs avec différentes protéines de fusion (GST, gb1, NusA, MBP, SUMO) ont été testé afin d'optimiser la solubilité du domaine MEC-8 211-312.

La plupart des protéines de fusion améliorent la solubilité de MEC-8 RRM2, cependant après clivage, celui-ci est très rapidement dégradé. Le domaine B1 de la protéine G (gb1) du fait de sa petite taille et de son expression importante a donc été retenu comme protéine de fusion pour l'étude du RRM2 de MEC-8.

## 2) Assignement des résonances de MEC-8 RRM2



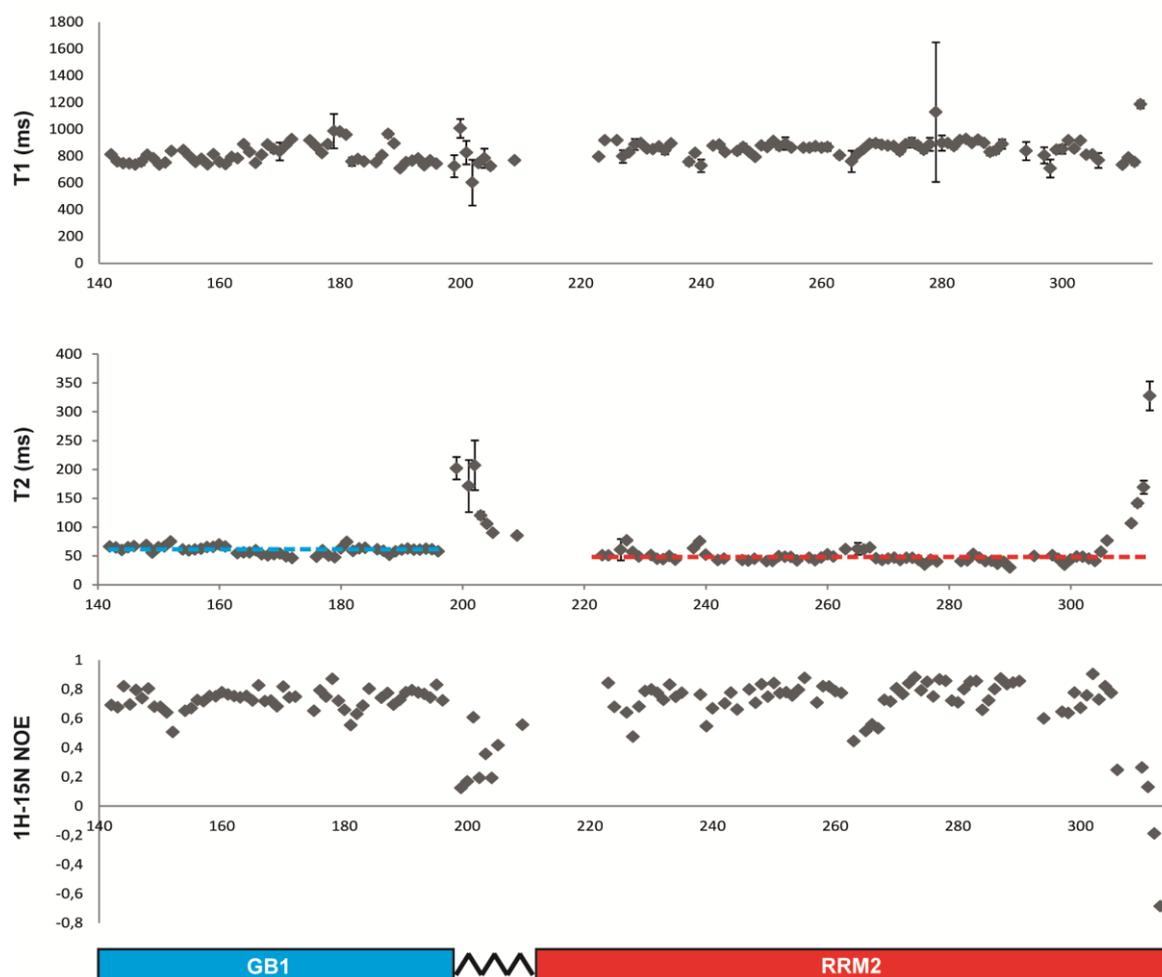
**Figure 57: Spectre 15N-H HSQC de gb1::RRM2**

Assignement des résonance du spectre 15N-H HSQC de la molécule de fusion gb1 RRM2(211-312) à 298K dans 50mM Tris 150 mM NaCl 10% D2O à pH 7.5.

L'assignement des résonances pour le domaine RRM2 a été effectué en suivant la même stratégie que pour le RRM1 (Sattler, Schleucher et al. 1999). La séquence entre les résidus 210 et 220 n'a pas pu être assigné car la majorité des pics de cette régions n'apparaissent pas sur les spectres. Aucun calcul de structure n'a été entreprise pour le moment.

### 3) Etude dynamique de MEC-8 RRM2

L'étude dynamique de la construction gb1 ::RRM2 par expériences de relaxations et NOE heteronucléaire montre deux domaines séparés par un court linker mobile. La région du RRM2 comprenant les résidus 210 à 220 n'est pas représentée car n'a pas pu être assignée. Il est cependant raisonnable de penser que cette région est déstructurée et très mobile. Les deux domaines ont une différence minimale dans leurs valeurs de T1 et T2, cela est cohérent avec leur assez faible écart de poids moléculaire (8.8 contre 9.9 kDa). On peut néanmoins conclure sur un comportement de ces deux domaines indépendants l'un de l'autre.



**Figure 58. Etude dynamique de gb1 ::RRM2**

*Mesures réalisées à 700 Mhz pour une température de 298K.*

## II. Etude interactionnelle de MEC-8 RRM2

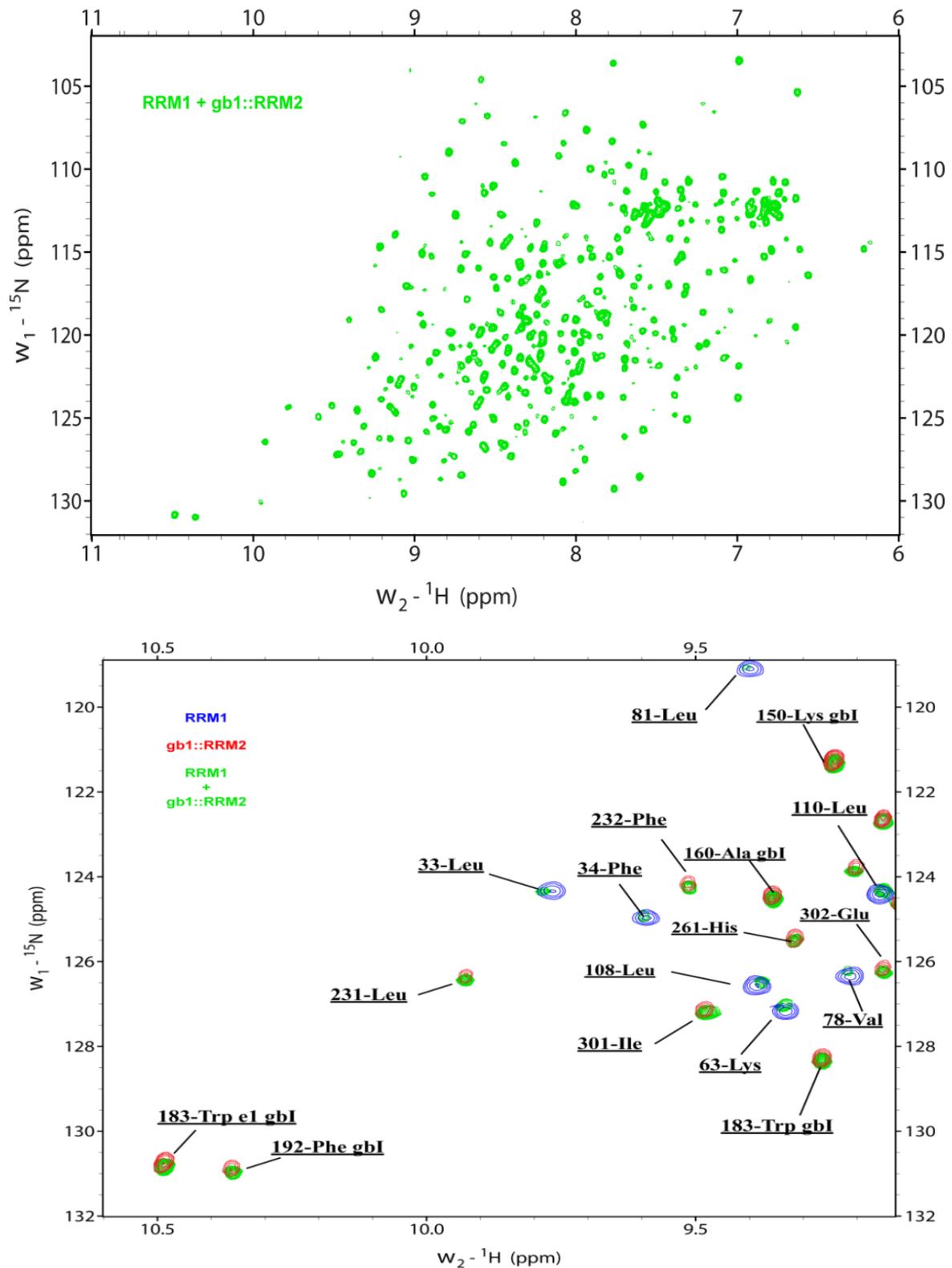
Le second domaine RRM de MEC-8 n'est pas conservé dans la famille de protéine RBPMS/CPO. Il est donc raisonnable de penser que celui-ci agit de manière indépendante du RRM1 à la fonction de MEC-8. Néanmoins il est très fréquent que des protéines contenant un tandem de RRM interagissent avec leur ligand par le biais d'une coopération entre ceux-ci (Mackereth, Madl et al. 2011, Afroz, Skrisovska et al. 2014)

Nous avons donc décidé de commencer l'investigation de la fonction du RRM2 de MEC-8 en étudiant l'éventuelle interaction d'une part entre le RRM2 et le RRM1 et d'autre part entre RRM2 et la séquence ARN reconnue par le RRM1.

### 1) Analyse de l'interaction inter-protéique entre les domaines RRM1 et RRM2

Nous avons déjà assigné les résonances H-N des deux domaines RRM de MEC-8. Dans ce cas la spectroscopie RMN nous offre un moyen simple de découvrir une éventuelle interaction inter-domaine, mais nous permet aussi d'avoir des informations à un niveau atomique sur l'interface d'interaction.

La titration du domaine RRM1 non marqué isotopiquement avec le domaine gb1::RRM2 marqué  $^{15}\text{N}$  n'a engendré aucune perturbation des déplacement chimique de ce dernier. On peut donc en conclure que les domaines RRM1 et RRM2 n'interagissent pas l'un avec l'autre. Le spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC des deux domaines marqué  $^{15}\text{N}$  illustre l'indépendance des deux domaines en solution (*Figure 59*).

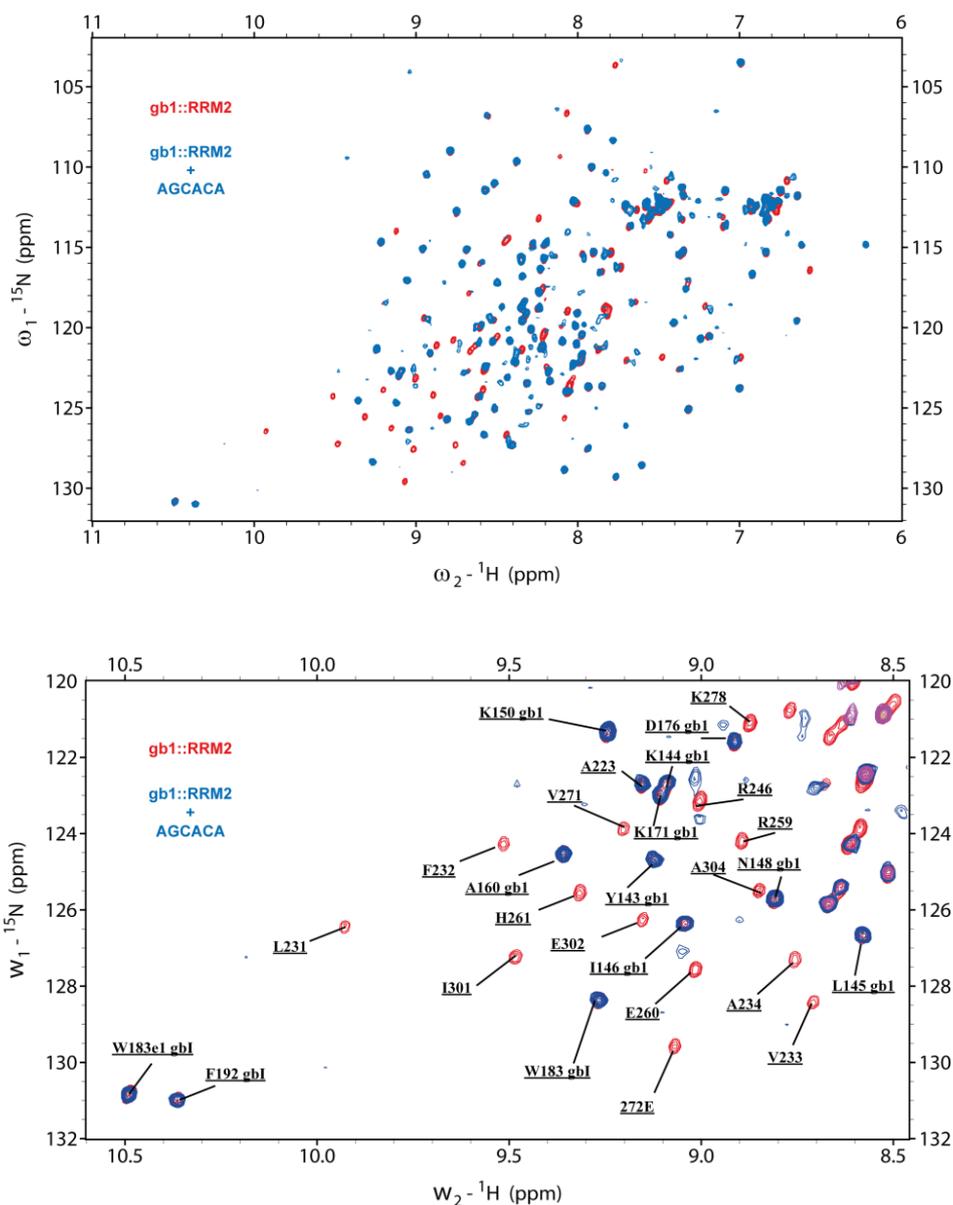


**Figure 59. Les domaines RRM1 et RRM2 n'interagissent pas ensemble**

Spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC du RRM1 avec le RRM2 en fusion avec le domaine gb1. En vert les deux domaines sont marqués  $^{15}\text{N}$ , en rouge et bleu sont les échantillons avec le domaine gb1::RRM2 et le domaine RRM1 marqués respectivement.

## 2) Exploration de l'interaction protéine-ARN du RRM2 de MEC-8

La spectroscopie RMN a encore une fois été privilégiée pour étudier l'interaction RRM2-AGCACA car elle offre deux avantages par rapport à l'ITC. D'une part elle permet d'observer des affinités de l'ordre du mM et d'autre part elle nécessite peu d'échantillon.

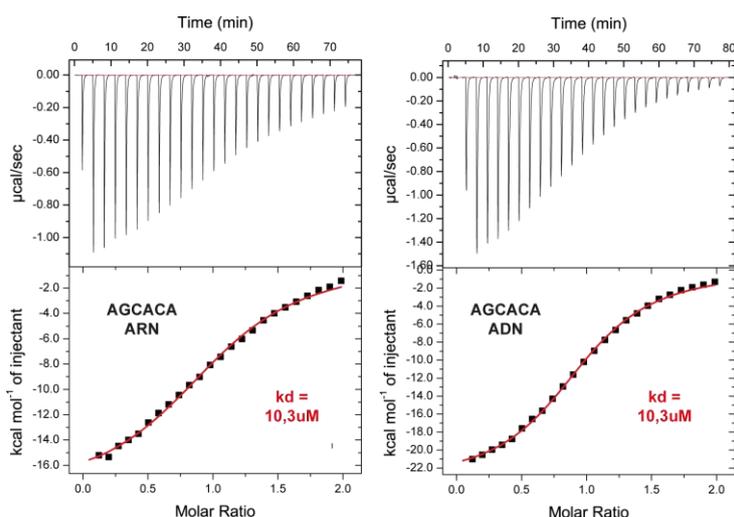


**Figure 60. Le domaine RRM2 de MEC-8 se lie au motif AGCACA**

Spectres  $1H-15N$  HSQC de  $gb1::RRM2$  en rouge et  $gb1::RRM2+AGCACA$  (ratio 1/1.2) en bleue. Le spectre du bas est un zoom ce focalisant sur une partie du spectre du haut. L'assignement spécifie les résidus appartenant au domaine  $gb1$ . Dans les deux cas la protéine est à une concentration de 50  $\mu M$  à 295K.

L'ajout de la séquence ARN AGCACA à la protéine de fusion gb1:RRM2 à un ratio de l'ordre de 1,2/1 induit une disparition totale des résonances du domaine RRM2 du spectre 1H 15N-HSQC (*Figure 60*). Cela est significatif de la transition en échange intermédiaire de la protéine, on peut donc en conclure que le RRM2 interagit avec la séquence ARN AGCACA. Sur ce même spectre on remarque que le domaine de fusion gb1 n'est quant à lui pas du tout affecté par l'ajout de cet ARN (*Figure 60*).

Le domaine RRM2 reconnaît donc bien la même séquence nucléotidique que le RRM1. Afin de quantifier l'affinité de cette interaction nous avons entrepris des expériences de titration isotherme calorimétrique qui nous permettront de juger la nature significative ou non de cette interaction.



**Figure 61. Le domaine RRM2 de MEC-8 se lie à l'ARN et à l'ADN**

Les affinités calculées par ITC de gb1 :RRM2 avec le motif AGCACA sous forme d'ARN à gauche et d'ADN à droite.

La titration du domaine RRM2 avec la séquence d'ARN AGCACA résulte en une affinité proche de celle du RRM1 ( $K_d = 10,3\mu\text{M}$  contre  $\sim 30\mu\text{M}$ ) (*Figure 61*). Cependant nous ne pouvons pas nous baser sur cette relativement bonne affinité pour conclure sur la fonction du

RRM2. En effet il est possible que le RRM2 lie une séquence différente avec une meilleure affinité. Nous avons donc encore une fois utilisé le fait que le motif ADN ait une affinité similaire à l'ARN pour faire une analyse systématique, en calculant l'affinité des mutations de chaque position du ligand.

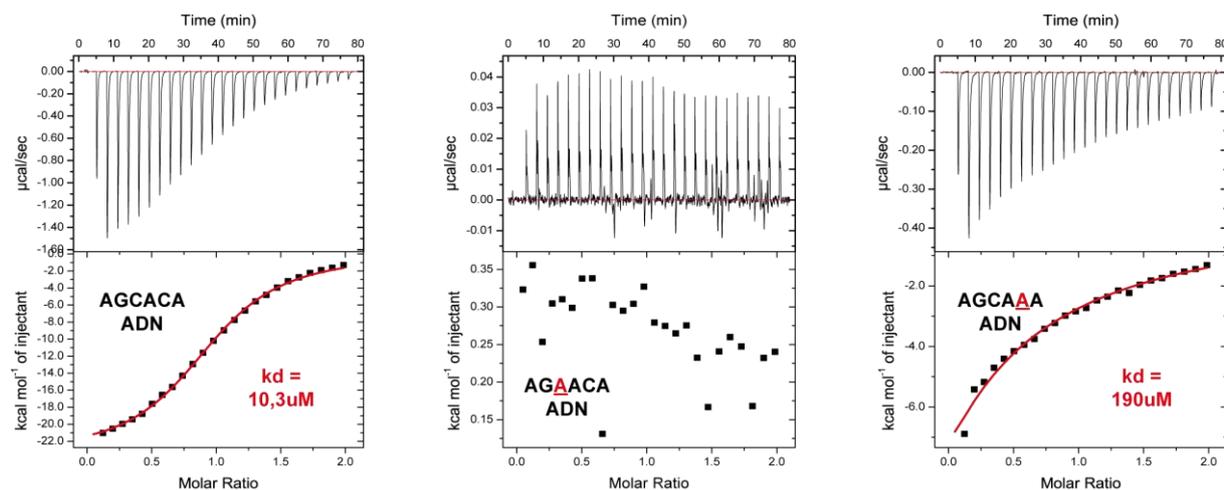


Figure 62. Le motif de liaison optimal du domaine RRM2 de MEC-8

Table 8. Les affinités de MEC-8 RRM1 avec les mutants du motif AGCACA par ITC

ligand	protéine	Kd ( $\mu\text{m}$ )
AGCACA (ARN)	gb1::RRM2	10,3
AGCACA	gb1::RRM2	10,3
AGCACC	gb1::RRM2	8,4
AGCACU	gb1::RRM2	9
AGCACG	gb1::RRM2	9,5
AGCAUA	gb1::RRM2	60
AGCAGA	gb1::RRM2	NA
AGCAAA	gb1::RRM2	190
AGCCCA	gb1::RRM2	77
AGCUCA	gb1::RRM2	NA
AGCGCA	gb1::RRM2	NA
AGAACA	gb1::RRM2	NA
AGUACA	gb1::RRM2	378
AGGACA	gb1::RRM2	NA
ACCACA	gb1::RRM2	80
AUCACA	gb1::RRM2	85,5
AACACA	gb1::RRM2	163
CGCACA	gb1::RRM2	20
UGCACA	gb1::RRM2	15,6
GGCACA	gb1::RRM2	16,6

Le résultat de cette analyse systématique est comparable à celui effectué pour le RRM1, le motif GCAC est la séquence de liaison optimale (*Figure 62*). Le RRM2 a la particularité d'être plus spécifique, en effet aucune des positions de ce motif n'accepte de substitution sans un effet drastique sur l'affinité.

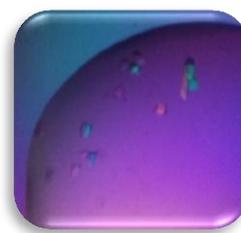


**Figure 63. Spécificité de liaison du RRM2**

*Courbe ITC à 25°C de différentes séquences ARN testées.*

### III. Structure de MEC-8 RRM2 en complexe avec un acide nucléique

La protéine de fusion gb1 : RRM2 a été mélangée à la séquence ADN AGCACA à un ratio de 1 pour 2. Des tests de cristallisation ont été conduits avec des cribles commerciaux. Deux conditions ont donné des cristaux. Dans les deux cas les cristaux sont de forme et de taille identique. Leur qualité est suffisante pour être analysés.

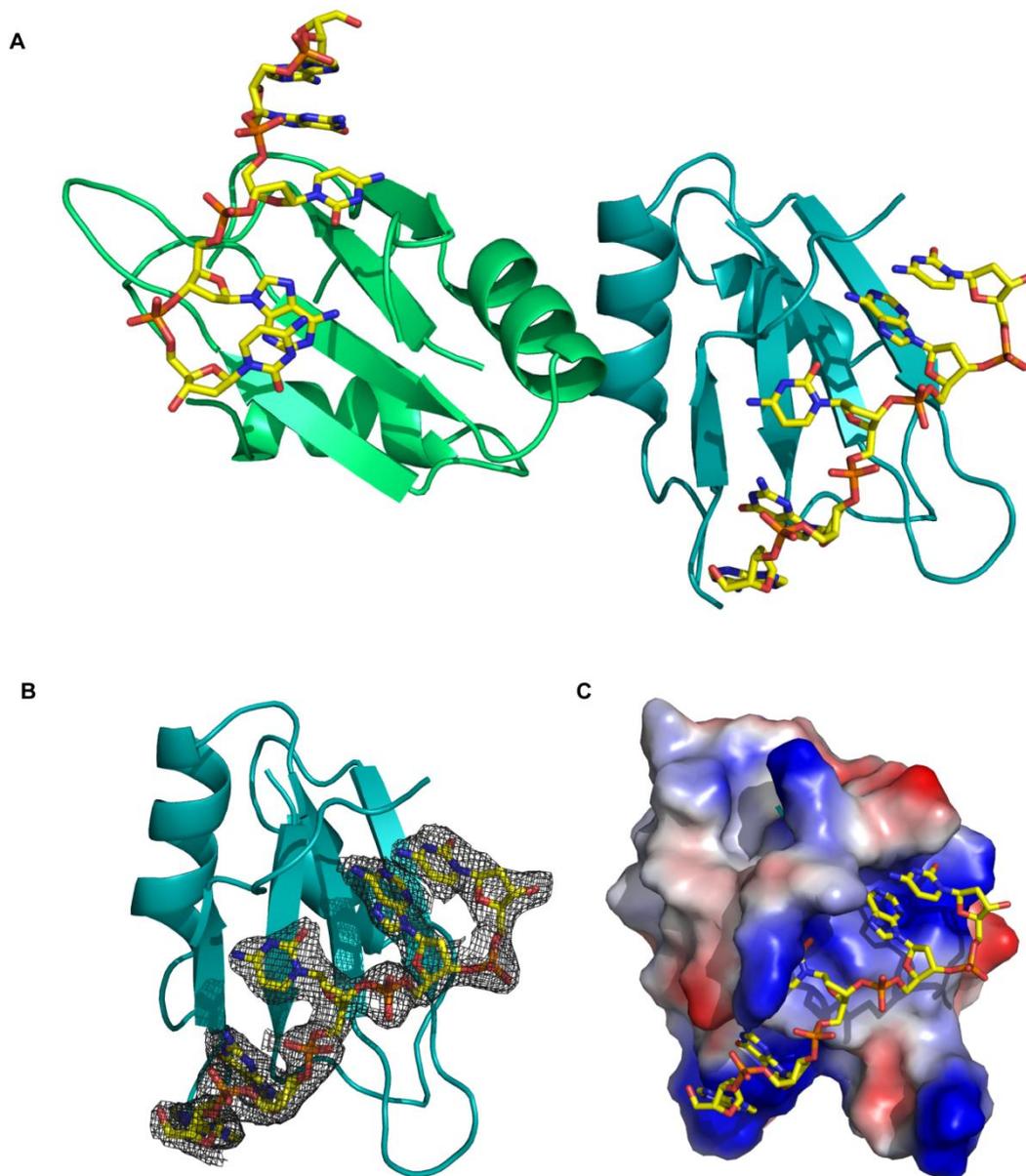


L'analyse sur la machine de diffraction présente au laboratoire résulte en un jeu de données réduit à 2.45Å de résolution. Le volume de l'unité asymétrique est en accord avec une molécule de fusion gb1::RRM2 par unité. La structure du domaine gb1 étant déjà résolue par cristallographie, le remplacement moléculaire a été entrepris à l'aide de différentes structures de celui-ci. Cette entreprise n'a cependant donné aucun résultat positif.

Une autre approche a donc été choisie, nous avons utilisé la suite logicielle BALBES afin de rechercher automatiquement des modèles adéquats pour le remplacement moléculaire. Les options pour 1 molécule de fusion gb1 ::RRM2, 2 molécules gb1 et 2 molécules RRM2 ont été testées. Un résultat positif fut obtenu avec l'option favorisant deux molécules de RRM2 par unité asymétrique. Finalement nous avons utilisé le modèle proposé par BALBES pour résoudre la structure du domaine RRM2 à 2.45Å.

Après reconstruction et affinement de la structure on observe bien un dimère de RRM2 sans aucune trace du domaine gb1 (*Figure 64. A*). La protéine de fusion a donc été clivée entre les deux domaines gb1 et RRM2 au cours des 10 jours qui précède l'apparition des cristaux.

L'échantillon utilisé avait la particularité de ne pas avoir subi une seconde purification par exclusion de taille comparé aux échantillons utilisés pour les études RMN beaucoup plus stables.

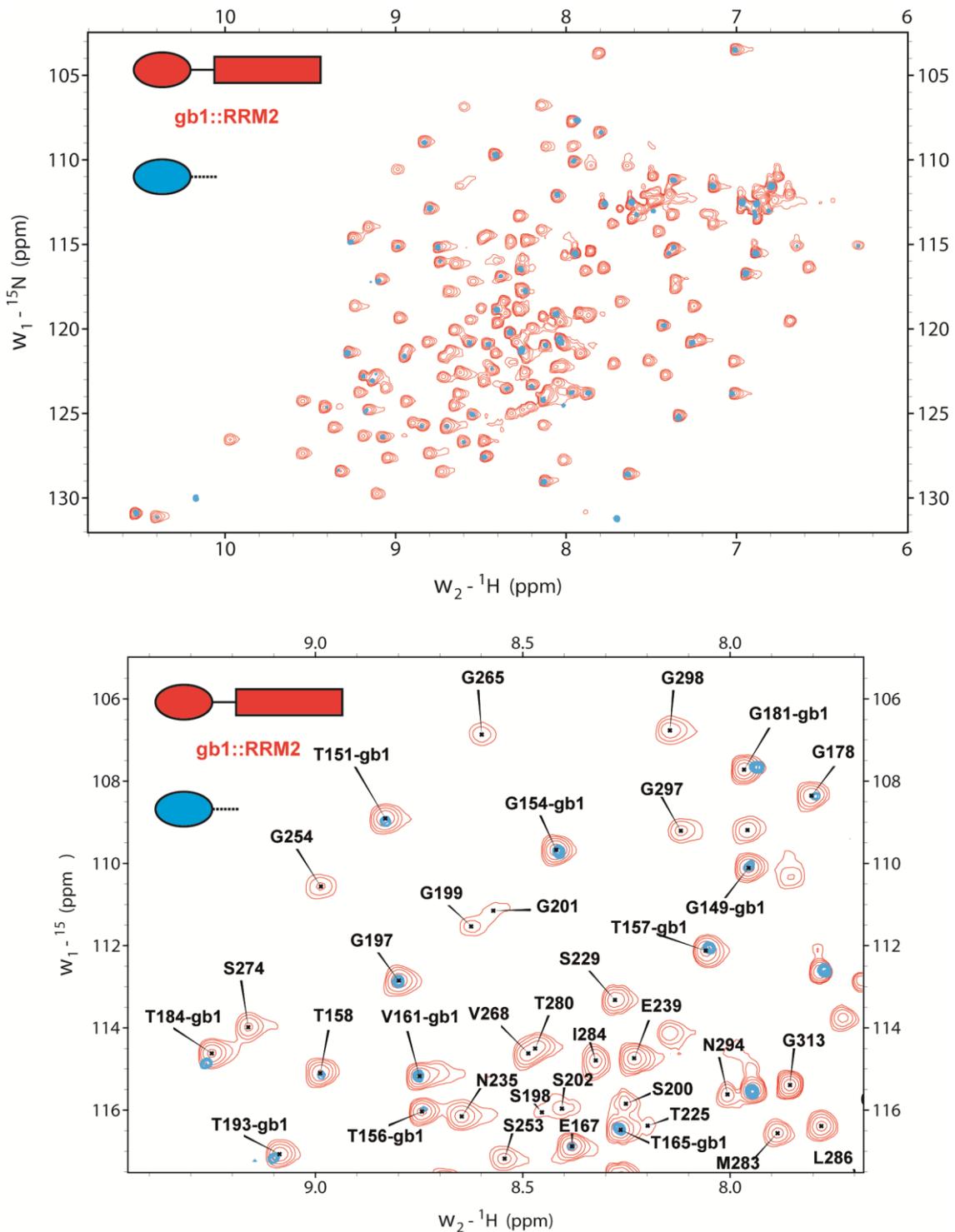


**Figure 64. Structure cristallographique du complexe RRM2-ADN.**

(A) Les deux monomères présents dans l'unité asymétrique en complexe avec le motif d'ADN AGCACA. (B) Densité électronique des bases AGCAC ( $2F_o - F_C$   $2\sigma$ ). (C) Surface électrostatique du domaine RRM2.

Nous avons répété ce protocole de purification avec un échantillon marqué  $^{15}\text{N}$  cette fois-ci afin d'observer ce processus de dégradation. L'échantillon laissé à température ambiante précipite après 48H. Une fois centrifugé, la solution débarrassée du culot de précipitât est observée sur un spectre  $^1\text{H}$ -  $^{15}\text{N}$  HSQC. On y observe un nombre de pics beaucoup moins

important que pour le domaine gb1::RRM2 (*Figure 65*). Une fois superposé au spectre de la protéine n'ayant pas subi de précipitation on remarque que seul le domaine gb1 est resté en solution. Cela explique le résultat obtenu par cristallographie.



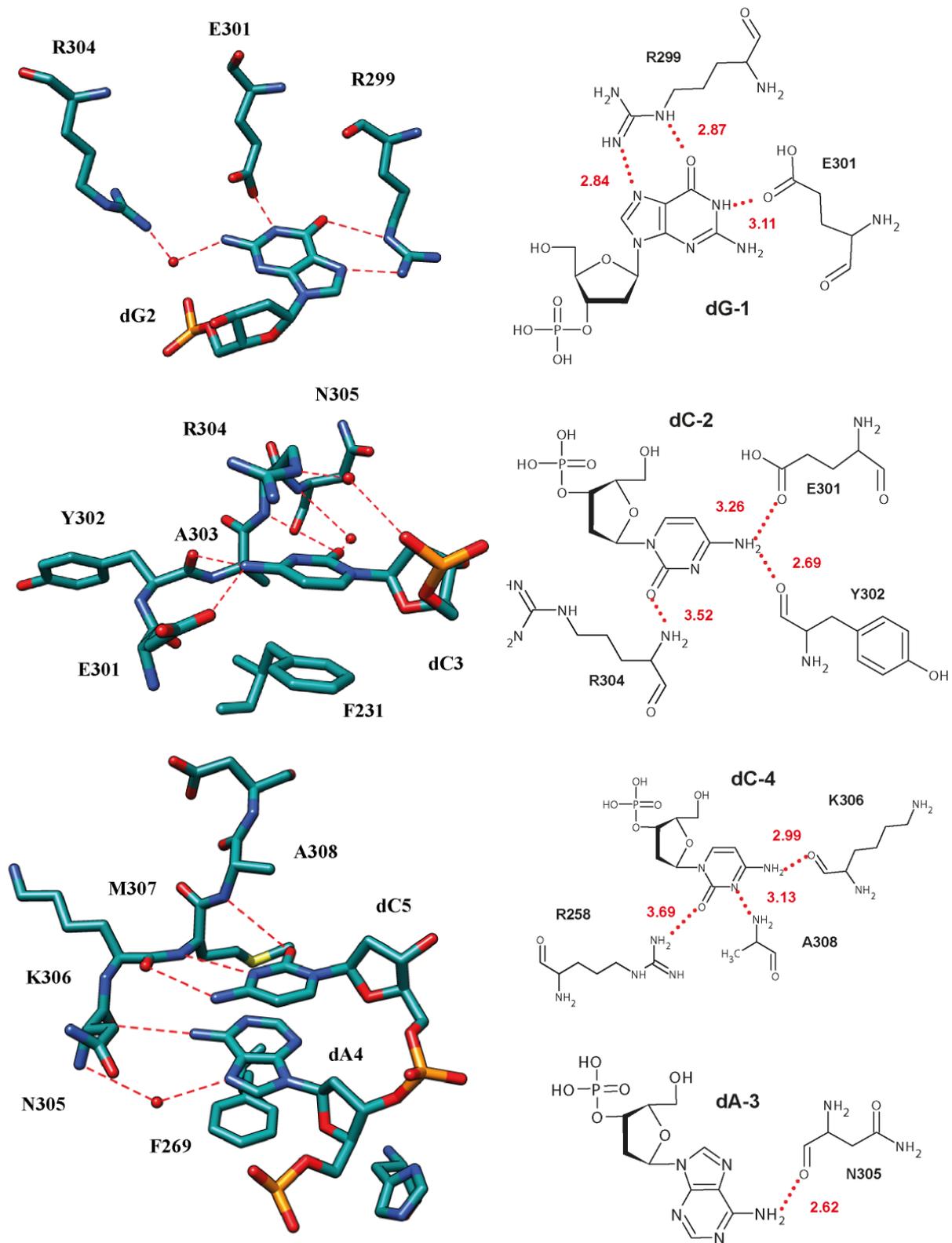
**Figure 65. Analyse de la dégradation du domaine gb1 ::RRM2**

*En rouge le spectre de la protéine gb1::RRM2 ayant subi une étape de chromatographie d'exclusion de taille. En bleu un échantillon de la même protéine sans cette étape de purification après 3 jours à 20°C. La totalité des pics résiduels (bleus) appartient au domaine gb1, le domaine RRM2 est entièrement précipité.*

Les deux monomères présents dans l'unité asymétrique ont une surface d'interaction très réduite, il s'agit indubitablement d'un dimère cristallographique et non le reflet d'un comportement biologique en solution.

Le ligand ADN interagit avec le feuillet beta du domaine RRM d'une façon similaire au RRM1. Les bases C3 et A4 effectuent un empilement aromatique sur les deux phénylalanines des motifs RNP1 et RNP2 et la base C5 se superpose à l'adénine 4.

Encore une fois la partie C-terminal non structurée est principalement responsable de la reconnaissance de la séquence ADN, avec une grande partie des interactions impliquant des atomes du squelette peptidique (*Figure 66*).



**Figure 66. Détails atomique du RRM2 en complexe avec d-AGCACA**

*Représentation des interactions entre le motif d'ADN AGCAC et le domaine RRM2 de la protéine MEC-8. A gauche une représentation sous forme stick du complexe cristallographique, à droite une représentation schématique. Les distances sont en rouge et l'unité de mesure est l'Angström.*

**Table 9. Statistiques préliminaires de la structure du RRM2-ADN****Collection datas**

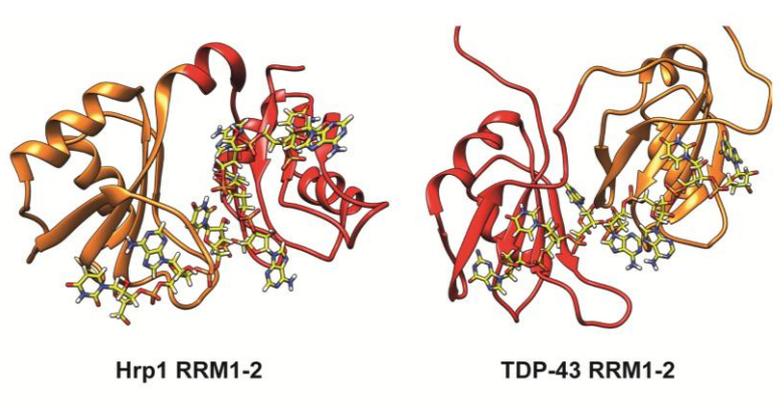
Resolution range ( $\text{\AA}\dots$ )	28.08 - 2.457 (2.544 - 2.457)
Space group	C 1 2 1
Unit cell	72.075 68.714 57.531 90 122.4 90
Unique reflections	8244 (658)
Completeness (%)	94.42 (76.33)
Mean I/sigma(I)	12.77 (3.04)

**Refinement**

R-work	0.2861 (0.3217)
R-free	0.3441 (0.3957)
Number of non-hydrogen atoms	1470
macromolecules	1387
water	83
Protein residues	165
RMS(bonds)	0.005
RMS(angles)	1.01
Ramachandran favored (%)	84
Ramachandran allowed (%)	12.5
Ramachandran outliers (%)	3.5
Clashscore	19.19
Average B-factor	35.30
macromolecules	35.20
solvent	37.00

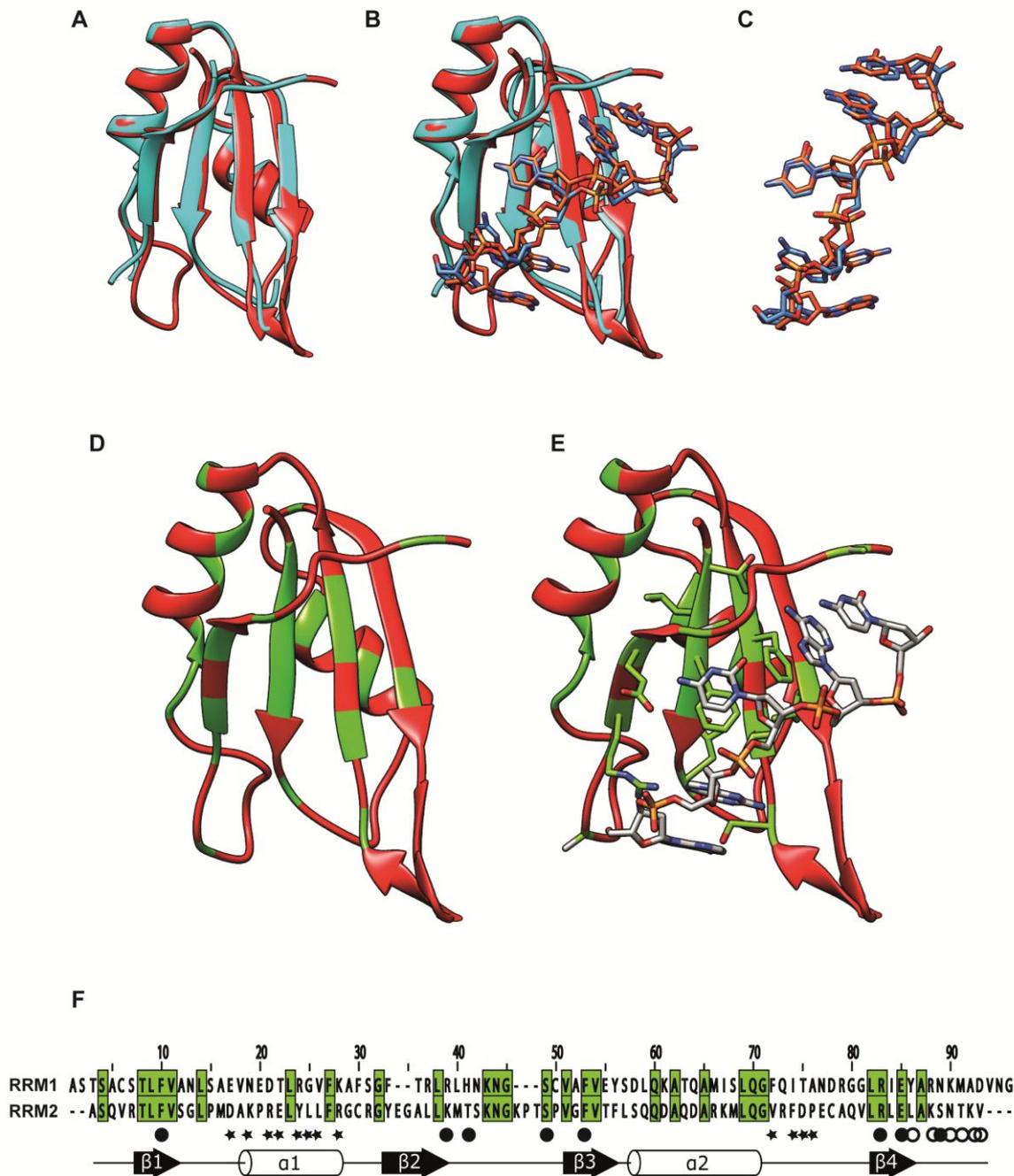
#### IV. Analogies RRM1/RRM2

Il n'est pas inhabituel que des protéines étant composées de multiples domaines RRM voient ceux-ci interagir avec une séquence similaire (PTB, HNRNP L, Hrp1, TDP-43). Dans la majorité des cas reportés dans la littérature, les RRM interagissent de manières différentes avec leurs séquences de reconnaissances. Il est aussi très fréquent que ceux-ci interagissent entre eux (*Figure 67*). Cependant, dans le cas des deux domaines de MEC-8 on observe deux particularités. Premièrement les domaines RRM1 et RRM2 de MEC-8 lient très vraisemblablement leurs séquences indépendamment l'un de l'autre. Deuxièmement, malgré la faible homologie entre ceux-ci (30%), les deux RRM de MEC-8 arborent non seulement des particularités structurales très similaires mais se complexent à leurs séquences de reconnaissance de manière quasi identique (*Figure 68*).



**Figure 67. RRM en tandems liants des séquences d'ARN répétées**

Les bases CAC ont une structure conservée, les nucléotides AG quant à eux ne sont plus fixés par la boucle  $\beta 2/\beta 3$  qui est plus courte chez le RRM2 ce qui engendre une différence dans leurs orientations (*Figure 68.B*). En alignant la séquence du RRM1 à celle du RRM2 on remarque que l'essentiel des résidus impliqués dans la liaison à l'ARN sont parmi les rare conservés entre les deux domaines (*Figure 68.F*).



**Figure 68: Analogie structurale entre le RRM 1 et le RRM2**

(A) Superposition des structures cristallographiques des domaines RRM1 (monomérique) en rouge et RRM2 en bleue. (B) Superposition des structures cristallographique des domaines RRM1 (monomérique) en rouge et RRM2 en bleue en complexe avec leurs ligand ADN. (C) Zoom sur la superposition des structures des motifs GCAC le motif rouge appartient au complexe RRM1 et le bleue au RRM2. (D) Structure du RRM1 monomérique en rouge, les résidus colorés en vert sont ceux conservés avec le RRM2. (E) Structure du RRM1 monomérique en rouge, les résidus colorés en vert sont ceux conservés avec le RRM2. Le ligand GCAC est représenté sous forme de stick en blanc, les résidus de la protéine sous forme de stick sont ceux impliqués dans la reconnaissance du motif nucléique. (F) Alignement des séquences du domaine RRM1 et du domaine RRM2 de MEC-8, les résidus surlignés en vert sont les résidus conservés Les étoile représente les résidus impliqué dans la dimérisation du RRM1, les rond plein désigne les résidus interagissant avec l'ARN par le biais de leurs chaîne latérale, les rond vide interagissent avec le squelette peptidique.

## V. Matériels et méthodes

### 1) Clonage, Production, Purification

Les deux constructions protéiques utilisées pour ce domaine sont MEC-8 (228-312) et MEC-8 (211-312).

MEC-8 (228-312) a été clonée dans le vecteur pet-His 1a.

MEC-8 211-312 a été clonée dans les vecteurs :

-Pet-His 1a

-Pet-GST 1a

-Pet-GB1 1a

La construction gb1::MEC-8(211-312) sélectionnée pour les analyses structurales et interactionnel a été transformée dans des bactéries E.Coli BL21 Lys Y (NEB) de la même façon que pour le RRM1.

Les bactéries ont été cultivées dans des volumes de 1L de milieu LB et M9 15N13C pour l'expression de protéines marquées comme pour le RRM1.

La purification de la protéine de fusion gb1::MEC-8(211-312) a été réalisée en deux étapes. La première est une chromatographie d'affinité sur ions Ni<sup>2+</sup>, le chargement, le lavage et l'élution sont fait avec des solutions composées de 50 mM Tris (pH7.5), 500 mM NaCl, et 5, 25 ou 500 mM d'imidazole respectivement.

La seconde étape est une chromatographie d'exclusion stérique avec une colonne Superdex 75 de GE Healthcare dans un tampons NaPO<sub>4</sub> 20mM NaCl 50mM pH6.5 pour les expériences de RMN et Tris 50mM NaCl 150mM pH 7.5 pour L'ITC. Les fractions composant le pic de la

protéine de fusion sont récoltées et concentré avec un amicon Ultra (Millipore) avec un seuil de coupure de 10 kDa.

Les échantillons utilisés pour les essais de cristallogenèses n'ont pas subi de purification par exclusion stérique.

## 2) Spectroscopie de résonance magnétique nucléaire

Les échantillons RMN contiennent entre 1 et 2 mM de gb1::MEC-8(211-312) dans 20 mM de tampon phosphate (pH 6.5) et 50 mM de NaCl, supplémenté par 10 % ou 99 % 2H<sub>2</sub>O. Les spectres sont enregistrés à 293K avec les spectromètres Bruker Avance III à 800 MHz ou 900 MHz, équipés d'une sonde triple résonance cryoréfrigéré.

L'étude de l'interaction gb1::MEC-8 (211-312) par RMN ne peut être réalisée que dans un tampon Tris 50mM NaCl 150mM à pH 7.5. La protéine est à une concentration de 150 uM et les acides nucléiques ajoutés à un ratio final de 2 pour 1.

Les mesures de relaxation 15N ont toutes été réalisées à 700 MHz et 298 K. Les spectres

## 3) Spectroscopie des rayons X

Les échantillons ont été concentrés à 10 mg.ml dans un tampon 50mM Tris 150mM NaCl à pH7.5 puis mélangés à la séquence d'ADN AGCACA pour un ratio 1 : 2. Les essais de cristallogenèse ont été effectués à l'aide de kits commerciaux PEGI et PEGII de Quiagen ainsi que Natrix Hampton sur plaque 96 puits en gouttes assises. Deux conditions du kit PEGI ont données des résultats, les cristaux ont pris 10 jours à 20°C pour apparaitre dans les deux cas.

Condition 1 : 1mM HEPES ph7.5 PEG 8000 25%

Condition 2 : 1mM MES ph6.5 PEG 8000 25%

Les cristaux étant de bonnes qualités d'un point de vue de la taille et de la forme, les conditions ont été répétées en plaque 24 puits dans des volumes supérieurs (2ul par goutte) sans affiner le screening des différents composants. Les cristaux sont reproductibles de façons constantes, ils ont été analysés sur l'appareil de diffraction présent à l'institut (Anode tournante Rigaku FRX 3kW à microfoyer, détecteur pixel hybride Dectris Pilatus 200 K).

La résolution maximum atteint les 2,45 Å après réductions des données avec le logiciel XDS. Le remplacement moléculaire a été réalisé avec le logiciel Phaser-MR en utilisant la structure cristallographique du premier domaine RRM de U1A (pdb :1URN) .

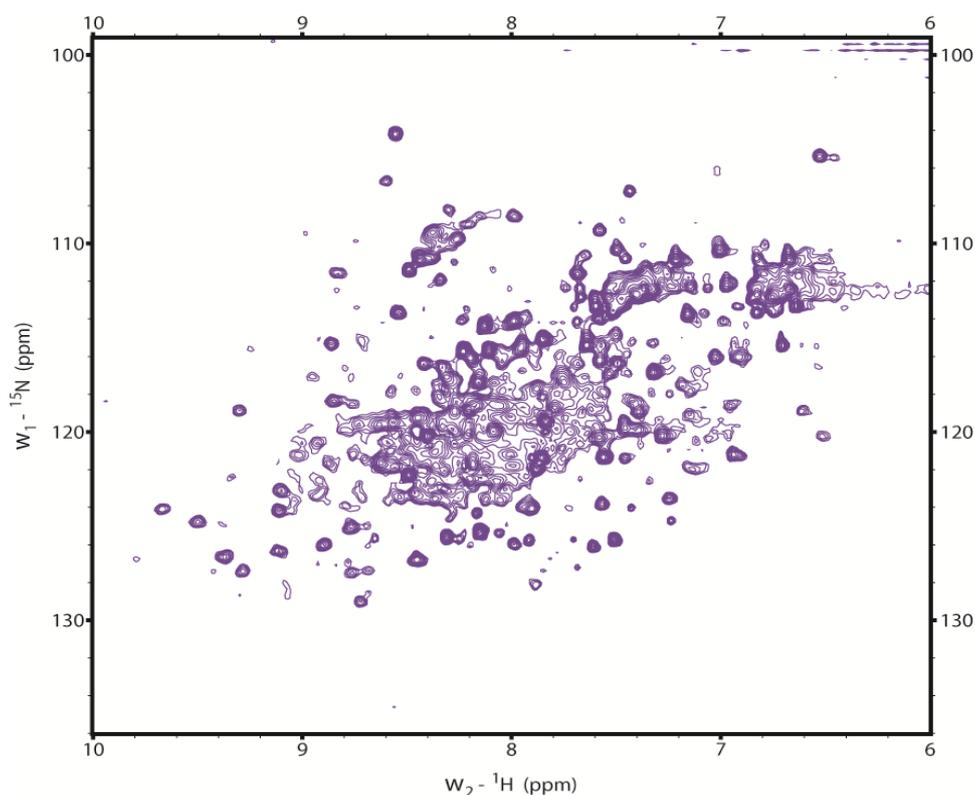
La construction et l'affinement du modèle a été réalisé en utilisant le programmes Coot et la suite Phenix.

#### **4) Titration calorimétrique isotherme**

Tous les échantillons ont été analysés dans 50mM Tris et 150 mM NaCl à pH7.5 à 25°C et 500rpm avec un appareil Microcal200. Pour les mesures protéine-ADN la protéine est présente dans la cellule à une concentration de 100uM et les oligos d'ADN dans la seringue à une concentration de 1mM. Dans le cas des mesures protéine ARN, l'ARN est dans la cellule à une concentration de 40uM et la protéine dans la seringue à 400uM. Les expériences consistent en 25 injections successives avec un espacement de 180 secondes entre chaque injection. L'analyse des données a été réalisée avec le logiciel OriginLab.

### Chapitre III : Le rôle de MEC-8 en tant que facteur d'épissage alternatif

L'étude de la protéine MEC-8 entière est une étape incontournable pour la pleine compréhension du mécanisme d'action de celle-ci *in-vivo*. Tout comme pour le domaine RMM2 la protéine entière est très peu soluble. La production de MEC-8 nécessite une protéine de fusion afin d'améliorer sa solubilité. Les deux protéines de fusion permettant une expression et purification de MEC-8 dans des quantités élevées (~mM) sont les protéines MBP et NusA.



**Figure 69. Spectre 1H-15N HSQC de MEC-8**

Ces deux constructions ont donc été utilisées pour produire la protéine MEC-8 afin de pouvoir réaliser une étude *in-vitro* de celle-ci. La protéine avec son tag en N-terminale reste soluble et stable à température ambiante. La protéine après clivage de son tag de solubilité est un dimère d'environ 64 kDa. Celui-ci est stable à 4°C cependant il tend à se dégrader après quelques heures à température ambiante.

Un spectre HSQC à 25°C de MEC-8 (*Figure 69*) montre une quantité très importante de résidus déstructurés au centre du spectre appartenant très vraisemblablement à la région désordonnée séparant les deux domaines RRM. Sur ce spectre on repère facilement les pics appartenant au domaine RRM1 d'intensité plus importante que le reste des résidus. Les pics du domaine RRM2 sont également présents mais d'une intensité moindre.

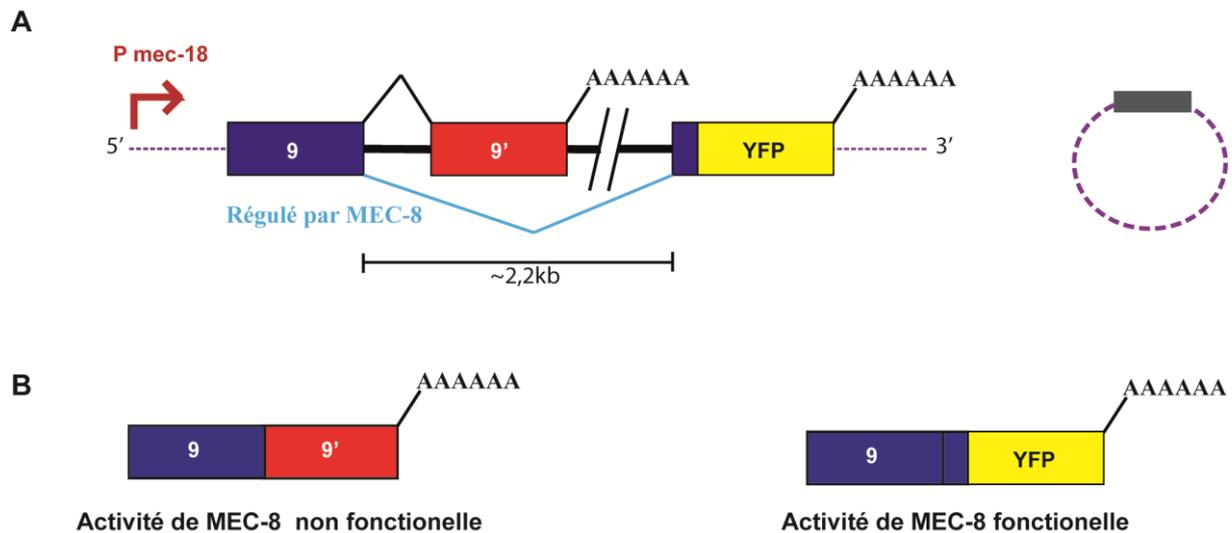
L'échantillon, après une journée à température ambiante laissait apparaître une quantité importante de précipitât. Il est par ailleurs très plausible que le spectre HSQC fut enregistré sur un échantillon comportant une protéine déjà partiellement clivée. Cela expliquerait la très bonne qualité des résonances des résidus du RRM1 pour une protéine de 65kDa.

Sans une optimisation des conditions permettant à la protéine MEC-8 de rester stable à température ambiante, il n'est pas envisageable pour le moment de mener des expériences RMN sur celle-ci.

Cependant des expériences d'interaction comme l'EMSA ou l'ITC sont potentiellement réalisables.

La seconde façon d'étudier la protéine MEC-8 entière est de manipuler celle-ci dans un environnement *in-vivo*. En effet l'organisme *C.elegans* est un organisme modèle de choix pour les études génétiques car il a de très nombreux avantages (cycle de vie court, reproduction importante, corps transparent, recombinaison génétique, facilité de culture, données génétique, transcriptomique et protéomique).

Nous avons en notre possession un plasmide permettant le suivi de l'activité de MEC-8 dans les neurones mécano-sensoriels grâce à un rapporteur fluorescent (*Figure 70*). Il s'agit d'une construction contenant la séquence minimale nécessaire au saut de l'exon 9' chez *mec-2* (Goodman, Ernstrom et al. 2002). Cette séquence de 2.2kb contient les exons 9 et 9' ainsi que le début de l'exon 10 fusionné à la séquence de la protéine fluorescente YFP.



**Figure 70. Un rapporteur fluorescent de l'activité de MEC-8**

(A) Représentation schématique du plasmide rapporteur de l'activité de MEC-8. (B) Représentation des deux différent ARNm produits par le rapporteur en fonction de la présence ou non d'une activité du facteur MEC-8.

Dans ces conditions si la fonction de la protéine MEC-8 est conservée, le saut de l'exon 9' se fera et une protéine chimérique émettant de la fluorescence sera produite. Si l'activité n'est pas conservée aucune fluorescence ne sera émise.

Il est prévu d'utiliser ce rapporteur de deux façons différentes. Tout d'abord dans un organisme *C.elegans* WT en effectuant des mutations ciblées sur les motifs GCAC présent dans la séquence minimale de 2.2kb exogène. Ces motifs sont au nombre de 4, cette expérience nous permettra d'évaluer la ou les séquences ciblées *in-vivo* par MEC-8 en suivant la baisse ou disparition de fluorescence sur une population d'organisme.

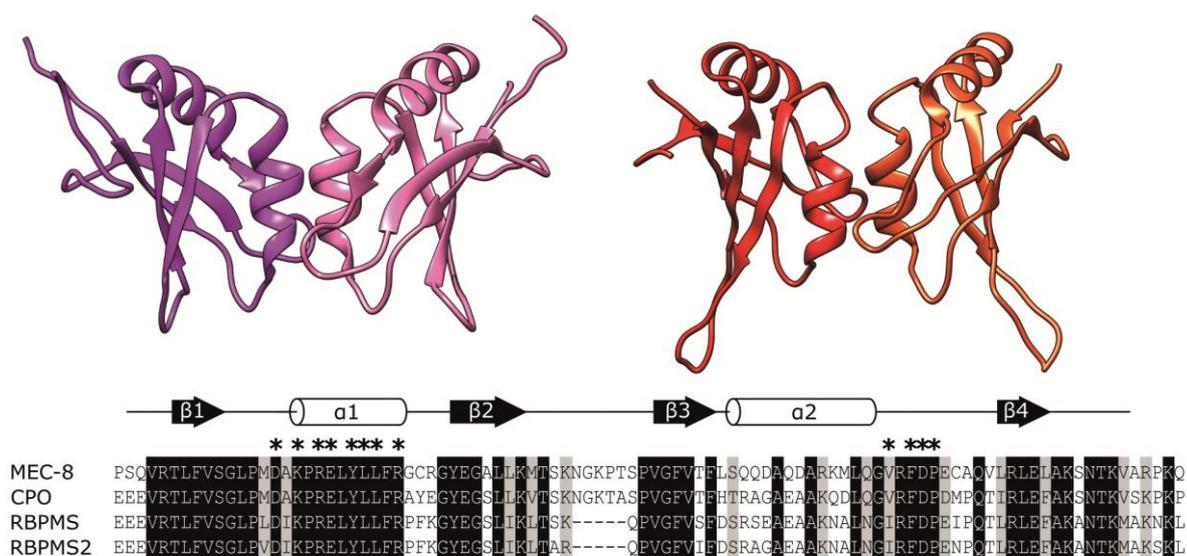
Une autre approche consistera à utiliser un organisme mutant ayant perdu la fonction de MEC-8 pour y injecter en plus du plasmide rapporteur, un plasmide d'expression de MEC-8. Différentes mutations de la protéine MEC-8 seront réalisées sur ce plasmide d'expression afin d'observer les résidus essentiels à la fonction de celle-ci encore une fois en suivant la baisse de fluorescence par rapport au vecteur d'expression WT.

## Chapitre IV : Discussion générale

### Conclusion de l'étude in-vitro des deux domaines RRM de MEC-8

Les études structurales et fonctionnelles pour les deux domaines RRM présent chez la protéine MEC-8 apportent beaucoup d'informations.

Le domaine RRM1 est un homodimère, chacun des monomères représente une structure canonique de RRM (**Figure 33**) Une étude publiée récemment (Sagnol, Yang et al. 2014) présentant la structure en solution du domaine RRM de RBPMS2 confirme la fonction d'oligomérisation de ce dernier. La séquence et la conformation de la région impliquée est conservée, on retrouve le motif de dimérisation chez toute la famille des protéines RBPMS.



**Figure 71. Un mode d'homodimérisation conservé pour la famille RBPMS**

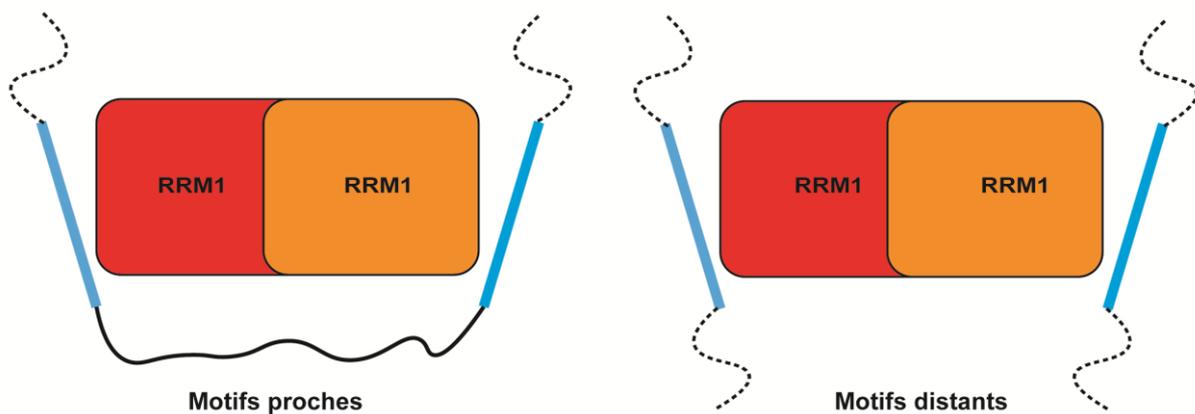
En violet et rose les deux chaînes du dimère de RBPMS2, en rouge et orange sont représenté celles de MEC-8.

Le domaine RRM1 reconnaît spécifiquement la séquence d'ARN GCAC avec une affinité de l'ordre du micro molaire (**Figure 48**) caractéristique de cette famille de protéine.

La liaison coopérative des deux domaines du dimère augmente de manière significative l'affinité de celui-ci avec son ligand (*Figure 54*).

Une structure en complexe avec un tandem de motif GCAC a pu être résolue (*Figure 42*). La structure ne contient pas les 8 bases entre ces deux motifs, cela est probablement engendré par une mobilité de cette partie du ligand laissant penser qu'elle est indépendante du complexe RRM1-AGCACA. Des expériences de RMN ont pu confirmer que ces résidus n'interagissaient pas avec le domaine RRM1 (*Figure 50*).

Cela laisse donc la possibilité au domaine RRM1 d'interagir avec son ARN messager de deux manières différentes. Soit avec deux motifs répétés proches, soit avec deux motifs très éloignés l'un par rapport à l'autre sur l'ARN messager et se rapprochant chacun l'un de l'autre grâce à un arrangement structural tridimensionnel de celui-ci (*Figure 72*).



**Figure 72. Modèle de liaison hypothétique**

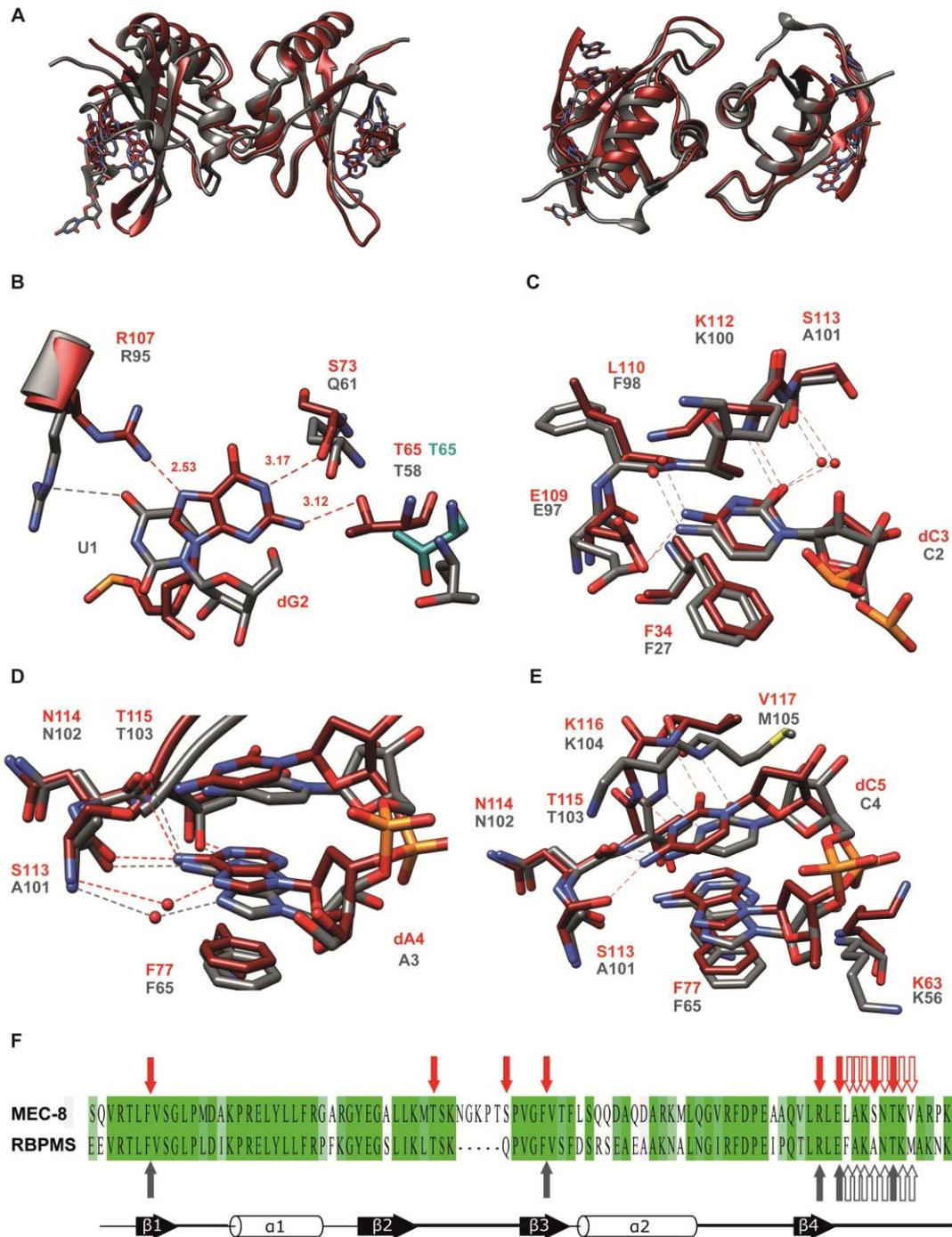
*Représentés en rouge et orange les deux domaines RRM 1 du dimère associé aux motifs GCAC en bleus.*

Comme indiqué en introduction, des études interactionnelles menées sur la protéine humaine RBPMS ont mis en évidence un élément de reconnaissance ARN pour cette protéine (Farazi, Leonhardt et al. 2014). Le motif présumé était alors similaire à celui proposé lors de notre étude à l'exception qu'il est plus court d'une base excluant la guanine du motif 'GCAC'. Très récemment une structure de la protéine RBPMS en complexe avec un ARN a été publiée

proposant les bases de la reconnaissance du motif 'CAC' par la protéine RBPMS (Teplova, Farazi et al. 2015).

Une comparaison des deux structures MEC-8 RRM1-AGCACA et RBPMS RRM-UCACU montrent de très grande similarités ce qui est cohérent avec la forte identité de séquence entre les deux protéines (~60%) (*Figure 73.F*).

Le motif CAC est lié de manière identique (*Figure 73.D*, *Figure 73E*), la différence majeure vient de la guanine (dG2) de MEC-8 RRM1-AGCACA qui est remplacée par une uridine (U1) dans le complexe RBPMS RRM-UCACU. La guanine dG2 est liée de façon importante par les chaînes latérales des résidus R107 et T65 ainsi que par le squelette peptidique du résidu S73 (*Figure 73.B*). U1 n'est lié que par un seul résidu, R95 pour la chaîne A (*Figure 73.B*) et Q61 pour la chaîne B, il est intéressant de noter que ces résidus sont à la même position que les résidus R107 et S73 chez MEC-8 respectivement. Il est donc très vraisemblable que RBPMS est la capacité de lier le motif GCAC de la même façon que MEC-8 RRM1. La thréonine 58 de RBPMS équivalente de T65 pour MEC-8 est éloignée du potentiel site de liaison, cependant c'est également le cas dans la structure de MEC-8 RRM1 sans ARN. La boucle  $\beta 2\beta 3$  effectue une rotation en direction de la base dG2 afin d'interagir avec celle-ci.



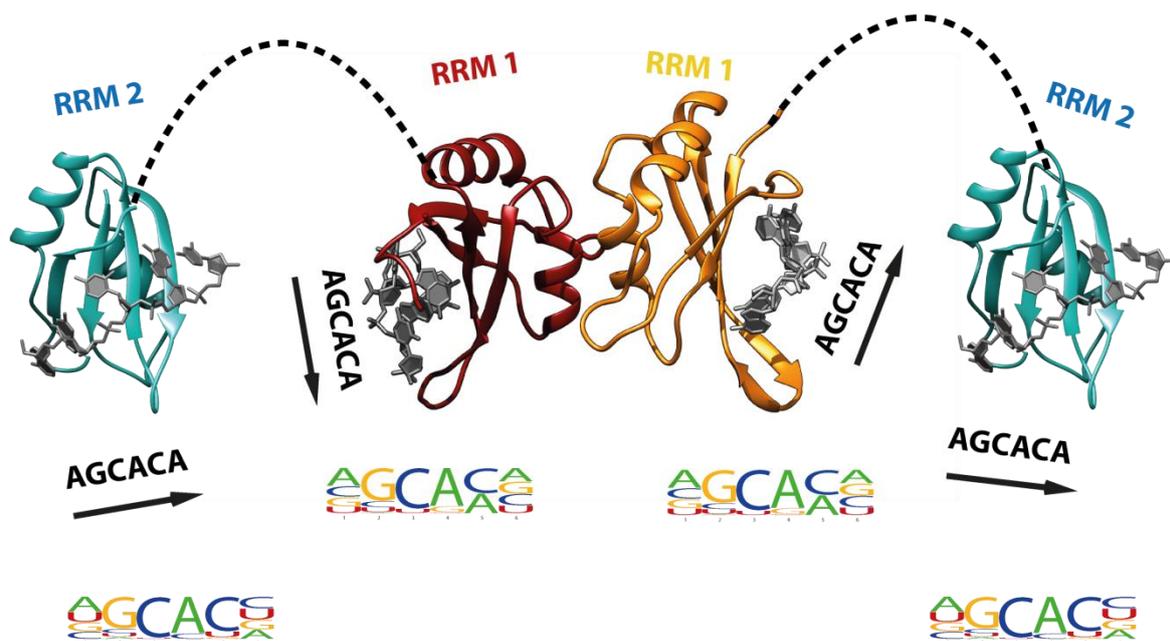
**Figure 73. Comparaison du mode de liaison des domaines RRM N-terminaux de MEC-8 et RBPMS aux acides nucléiques**

(A) Superposition des complexes RBPMS-ARN en gris et MEC-8 RRM1-ADN en rouge. (B) focus sur les détails de la reconnaissance de la guanine 2 comparé à l'uridine 1. (C)(D)(E) mode de liaison des deux RRM au motif CAC. (F) alignement des RRM 1 de MEC-8 et RBPMS. Les flèches représentent les résidus impliqués dans une interaction, elles sont pleines pour les résidus impliquant une chaîne latérale et creuses pour les chaînes principales.

Le domaine RRM2 est lui aussi un RRM canonique, très peu soluble il a néanmoins pu être analysé à l'aide de la protéine de fusion gb1 (Figure 56).

Les études d'interaction ont permis de déterminer que les domaines RRM1 et RRM2 n'interagissent pas l'un avec l'autre à l'état libre (*Figure 59*). L'analyse de l'interaction du domaine RRM2 avec des ARN par ITC conduit à la découverte d'un motif de liaison identique à celui du RRM1 (*Figure 62*).

Les domaines RRM1 et 2 interagissent donc avec la même séquence nucléotidique et ce d'une manière quasi identique. Les séquences des gènes affectés par une mutation nulle de MEC-8 (Kelley, Yochem et al. 2015) ne possèdent jamais de séquences GCAC qui se succèdent de manière répétitive. Il est alors possible de conclure que les domaines RRM1 et 2 n'interagissent non seulement pas à l'état libre mais non plus lors de la liaison à leurs séquences de reconnaissances.



**Figure 74. Modèle structural de la protéine MEC-8**

*Représentation des différents domaines composant la protéine MEC-8 entière avec leurs motifs de liaison.*

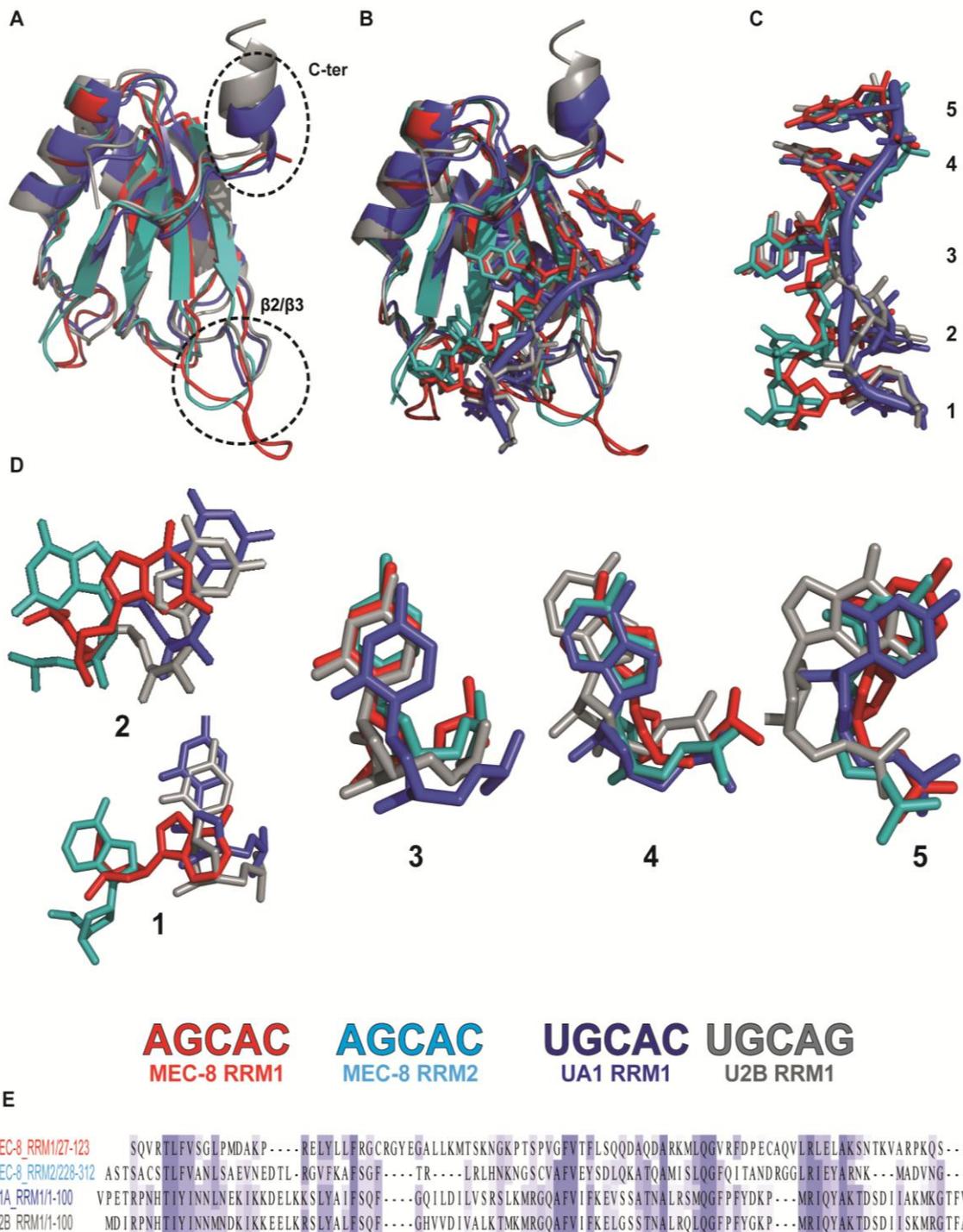
## MEC-8 est très similaire à la famille de protéine U1A/U2B''/SNF

Une autre structure de RRM en liaison avec un ARN de séquence similaire à MEC-8 est connue, il s'agit de la structure de la protéine humaine U1A en complexe avec la tige-boucle SLII du complexe U1 (Nagai, Oubridge et al. 1995). U1A reconnaît spécifiquement la séquence GCAC. La protéine U1A est très similaire à la protéine U2B'' chez l'humain (Delaney, Williams et al. 2014) dont la structure a elle aussi été résolue en complexe avec un ARN sous forme de tige boucle (Price, Evans et al. 1998).

En comparant les structures des RRM1 des protéines U1A et U2B'' on remarque une similarité importante du mode de liaison ainsi que de la structure que prennent les protéines et les ARN dans le complexe (*Figure 75*) (*Figure 76*). La similarité se concentre sur les résidus CAC à noter que U2B'' reconnaît la séquence GCAG de la tige boucle SLIV du complexe U2.

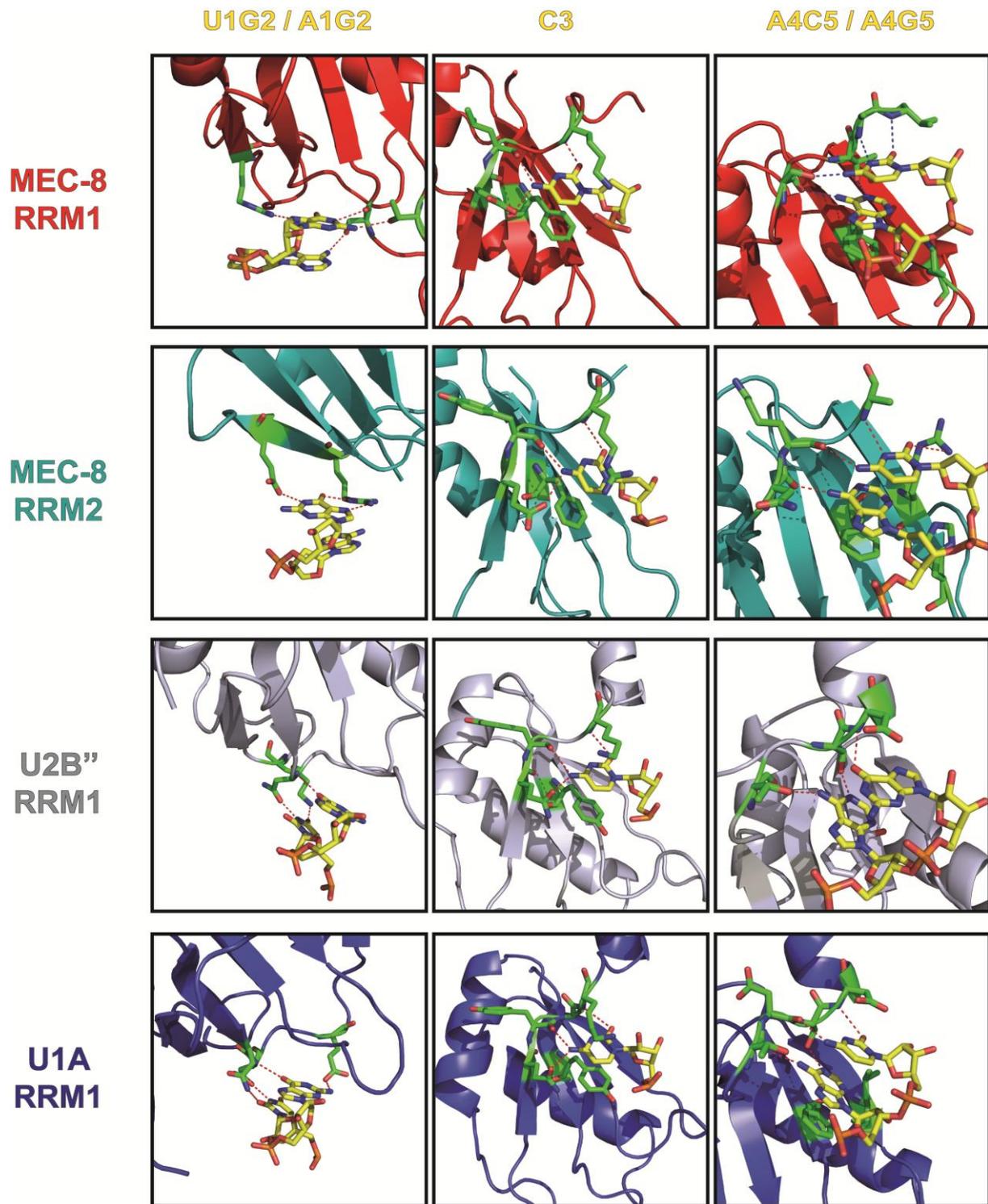
Quelques différences structurales existent néanmoins entre ces protéines. La boucle  $\beta 2/\beta 3$  est d'une taille supérieure chez le RRM1 de MEC-8. Les nucléotides 1 et 2 du motif A/U GCA C/G sont coordonnés par la boucle  $\beta 2/\beta 3$  chez U1A, U2B'' et MEC-8 RRM1 mais pas chez le RRM2 (*Figure 75*). Cela provoque une orientation complètement décalée de ceux-ci comme discuté auparavant (*Figure 68*).

Une autre différence entre ces structures est l'hélice C-terminale présente chez U1A et U2B, celle-ci a été reporté comme augmentant de façon significative l'affinité de U1A à son ligand (Zeng and Hall 1997). En parallèle on peut penser que la partie C-terminale (118-123) du RRM1 de MEC-8 adopte la même structure secondaire.



**Figure 75. Un mode de liaison conservé**

(A), (B), (C) Différentes représentation de la superposition des RRM1 N-terminaux de U1A et U2B''(bleu et gris) avec les deux RRM de MEC-8(en rouge le RRM1 et en cyan le RRM2). Tous les RRM sont en complexe avec leurs ligands. (D) Représentation de la superposition de chaque nucléotide du motif U/AGCAC/G lié par les différents domaines RRM. (E) Alignement des séquences des RRM représentés.

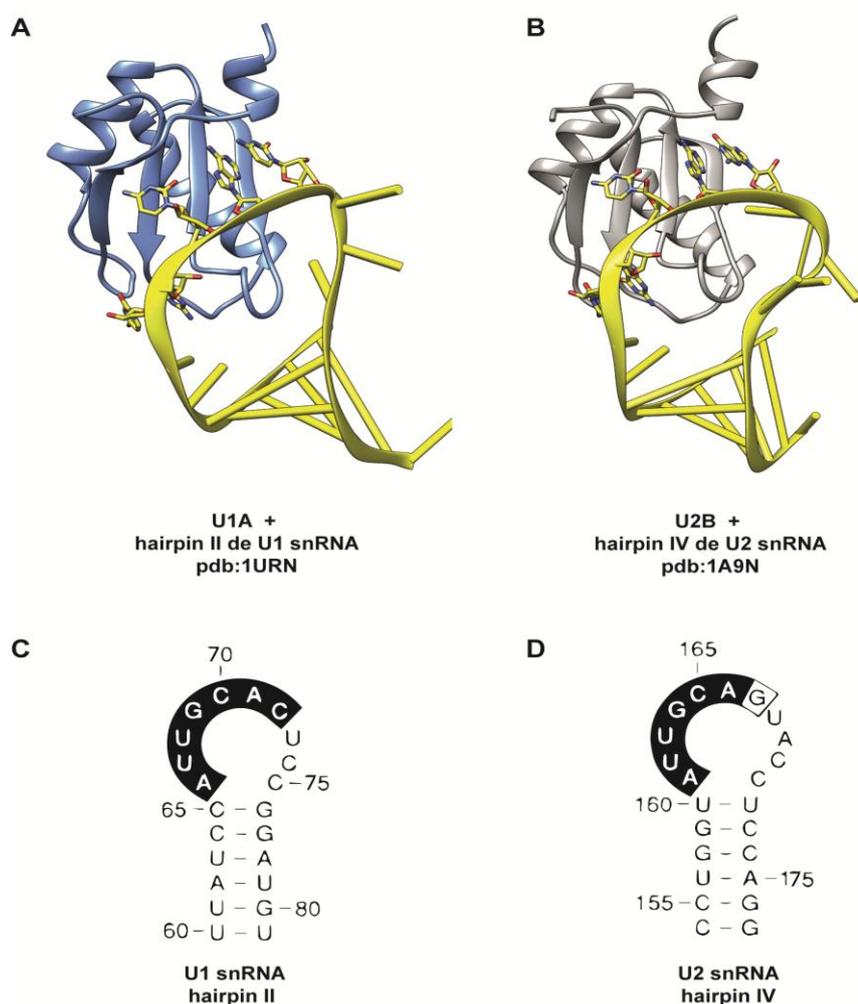


**Figure 76. Détails structuraux du mode de liaison**

*Comparaison des structures des protéines U1A, U2B'' et des domaines RRM de MEC-8 en complexe.*

## La protéine MEC-8 reconnaît-elle les motifs en tige-boucle

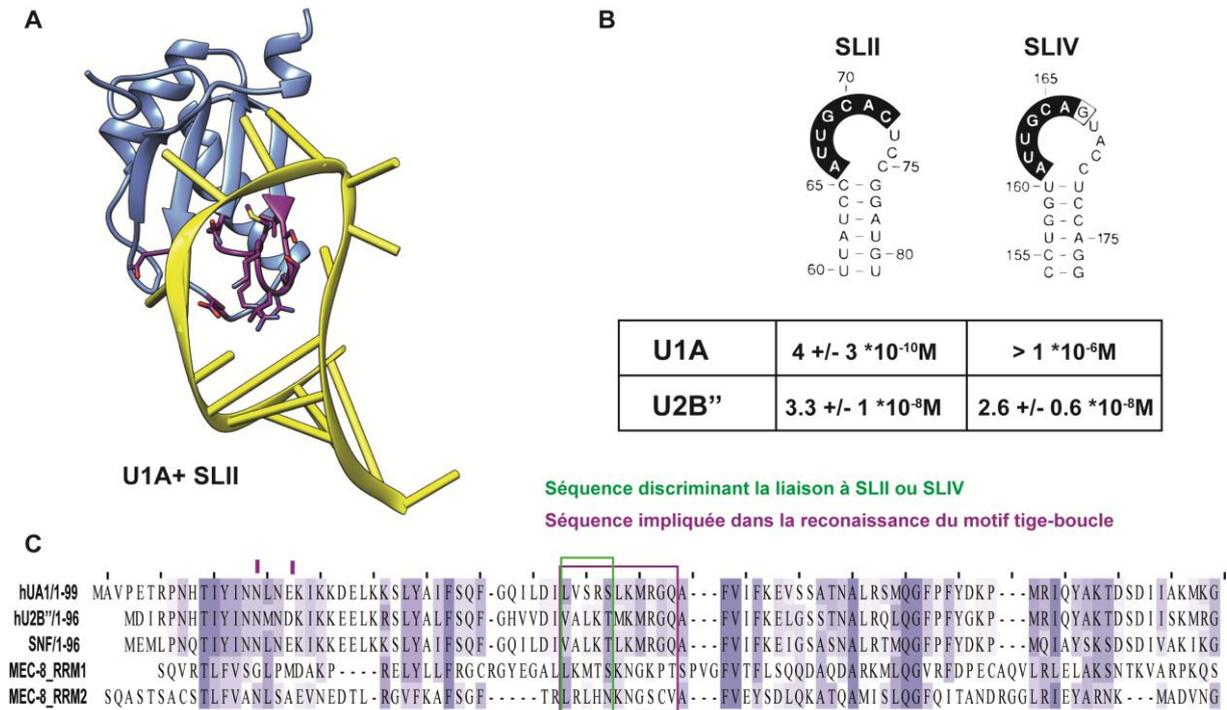
Une comparaison structurale de la liaison des motifs GCAC/G entre U1A/U2B'' et les deux RRM de MEC-8 a permis de mettre en évidence que le mécanisme de liaison de ces protéines est très similaire (*Figure 76*). Cependant bien que le motif GCAC/G soit le seul élément reconnu spécifiquement par la famille de protéine U1A/U2B'' les ligands *in-vivo* de ces protéines sont en réalité sous la forme de tige-boucle (*Figure 77*).



**Figure 77. Les protéines U1A et U2B lient préférentiellement des motifs en tige-boucle**

(A) Structure du RRM1 de U1A avec la tige boucle SLII. (B) Structure du RRM1 de U2B'' avec la tige-boucle SLIV. (C) Représentation 2D de la tige-boucle SLII. (D) Représentation 2D de la tige-boucle SLIV.

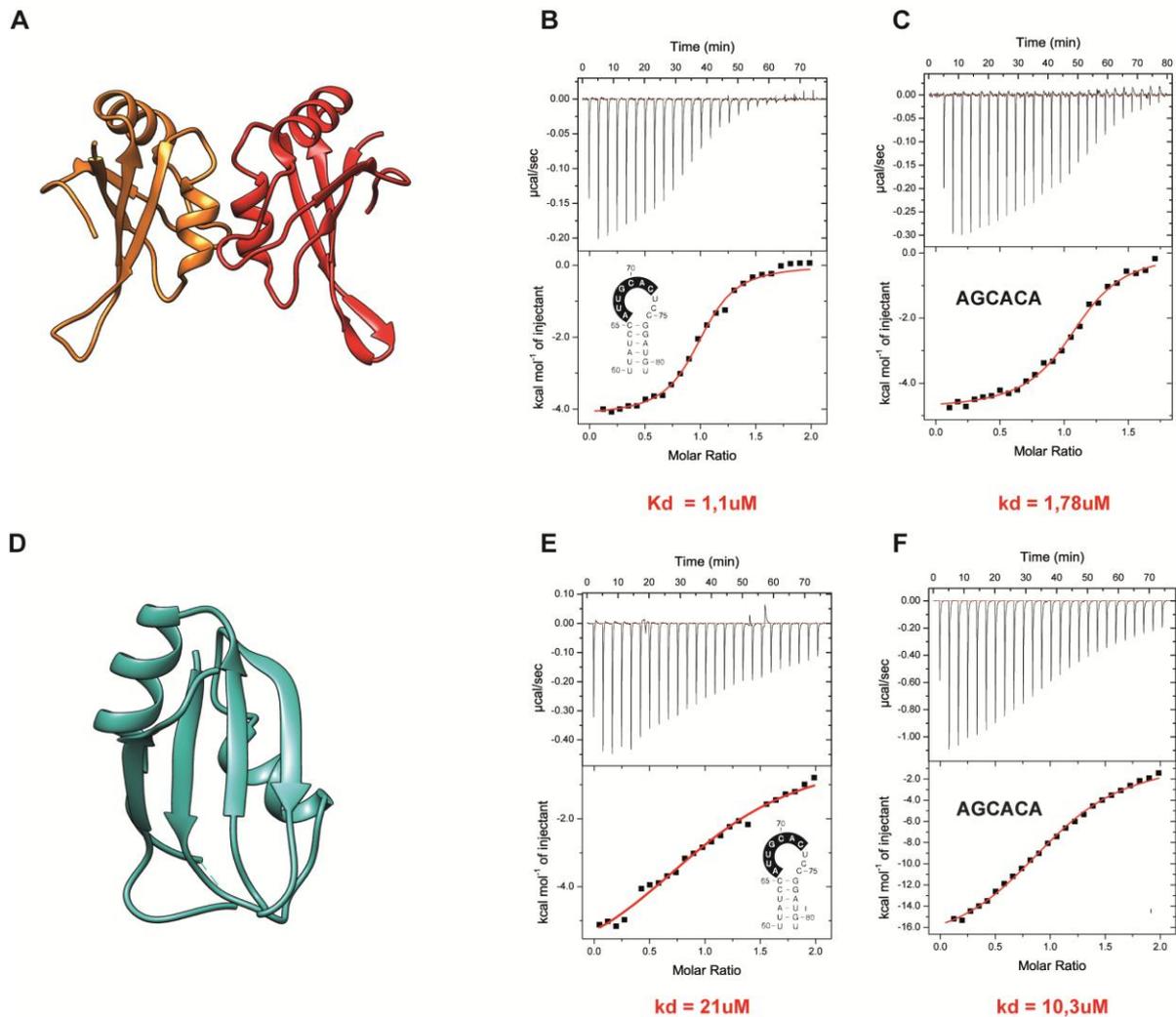
La forme tige-boucle de ces ligands permet aux protéines d'interagir de manière aspécifique avec les sucres et phosphates des nucléotides adjacents au motif de reconnaissance GCAC/G. La liaison des protéines à des motifs en tige-boucle a pour effet d'augmenter considérablement l'affinité de liaison à leurs cibles par rapport à un motif simple brin.



**Figure 78. Caractéristiques de la reconnaissance de la structure en tige boucle des protéines U1A/U2B''**

(A) Structure de U1A avec SLII, en mauve la région de la protéine impliquée dans la reconnaissance de la forme tige-boucle. (B) Affinités des protéines U1A et U2B'' avec les tige-boucle SLII et SLIV. (C) Alignement des séquences de U1A, U2B'', SNF, MEC-8 RRM1 et 2.

L'éventualité que les domaines RRM de MEC-8 lient un ARN sous forme de tige-boucle n'avait pas été envisagée lors de l'étude interactionnelle de ceux-ci. L'alignement des séquences des RRM N-terminaux de la famille U1A/U2B'' avec les domaines RRM1 et 2 de MEC-8 ne montre pas de conservation au niveau de la région des résidus impliqués dans l'interaction avec la structure secondaire tige-boucle (Figure 78). Néanmoins dans le cas des RRM1 et 2, cette région contient de nombreux résidus chargés positivement qui pourraient interagir avec des résidus ARN proches, comme dans le cas d'une tige-boucle.



**Figure 79. Affinité des domaines RRM de MEC-8 avec un motif tige-boucle**

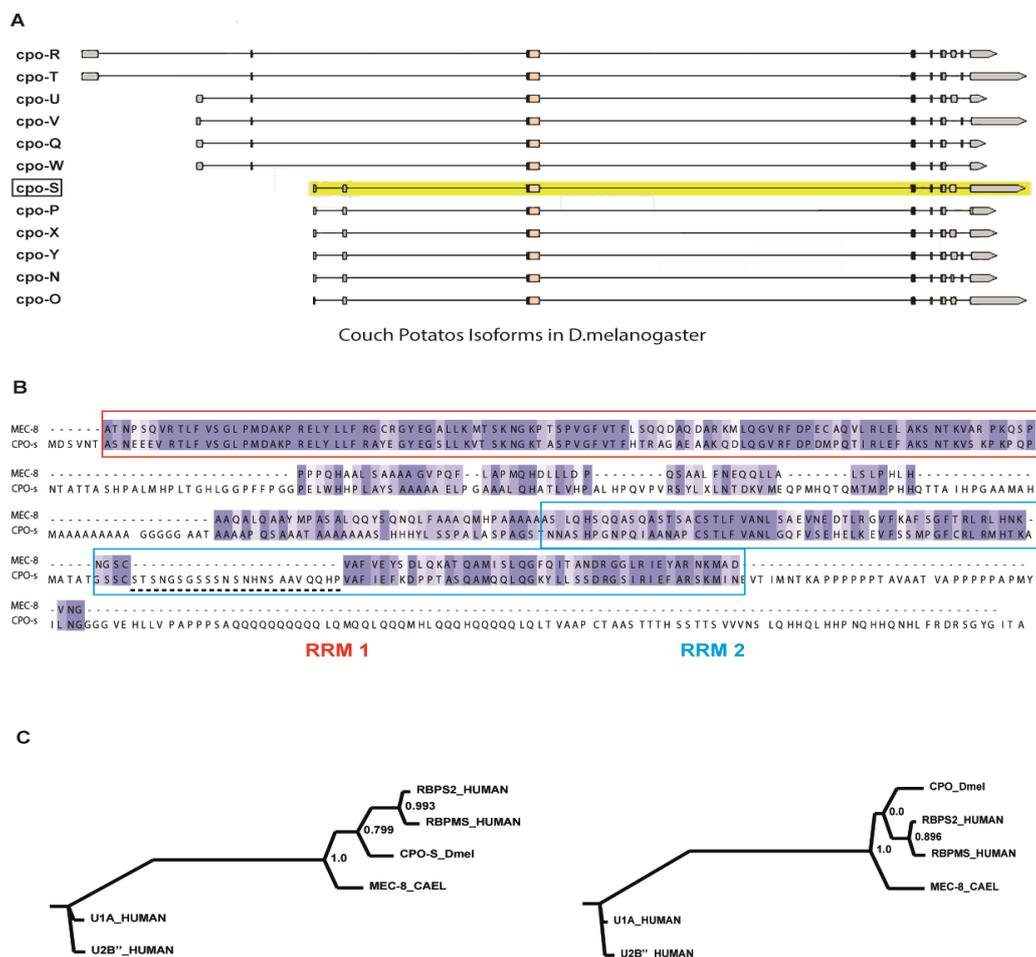
(A) domaine RRM1 de MEC-8. (B) ITC du RRM1 avec la tige boucle SLII. (C) ITC du RRM1 avec la séquence GCACA. (D) domaine RRM2 de MEC-8. (E) ITC du RRM2 avec la tige-boucle SLII. (F) ITC du RRM2 avec le motif GCACA.

Nous avons donc tenté de déterminer l'éventuelle reconnaissance de ce motif tige-boucle chez chacun des RRM de MEC-8. Pour cela nous avons fait des tests d'interaction par ITC en utilisant comme ligand modèle la tige-boucle SLII présente dans le snARN U1 humain qui est le ligand naturel du domaine RRM1 de hU1A (*Figure 77*).

Les résultats de ces tests montrent pour la première fois une différence en termes d'interaction entre le RRM1 et le RRM2. En effet le RRM1 voit son affinité augmenter d'un facteur 2 avec un motif en tige-boucle comparé à un motif simple brin alors que le RRM2 voit son affinité diminuer du même ordre (*Figure 79*).

## Couch Potato Protein *D.mel* comme pierre angulaire dans l'évolution de la famille RBPMS

La protéine CPO représente l'homologue direct de MEC-8 chez l'organisme *D.melanogaster*. CPO possèdent 12 isoformes (**Figure 80.A**) dont 4 répertoriés et annotés dans la base de données UNIPROT. C'est donc une protéine qui comme pour les homologues humains RBPMS et RBPMS2 comporte un polymorphisme de séquence. Le second point commun entre ces protéines est l'absence d'un second domaine RRM en position C-terminal (**Figure 18**), celui-ci est remplacé par des séquences variables selon les isoformes.

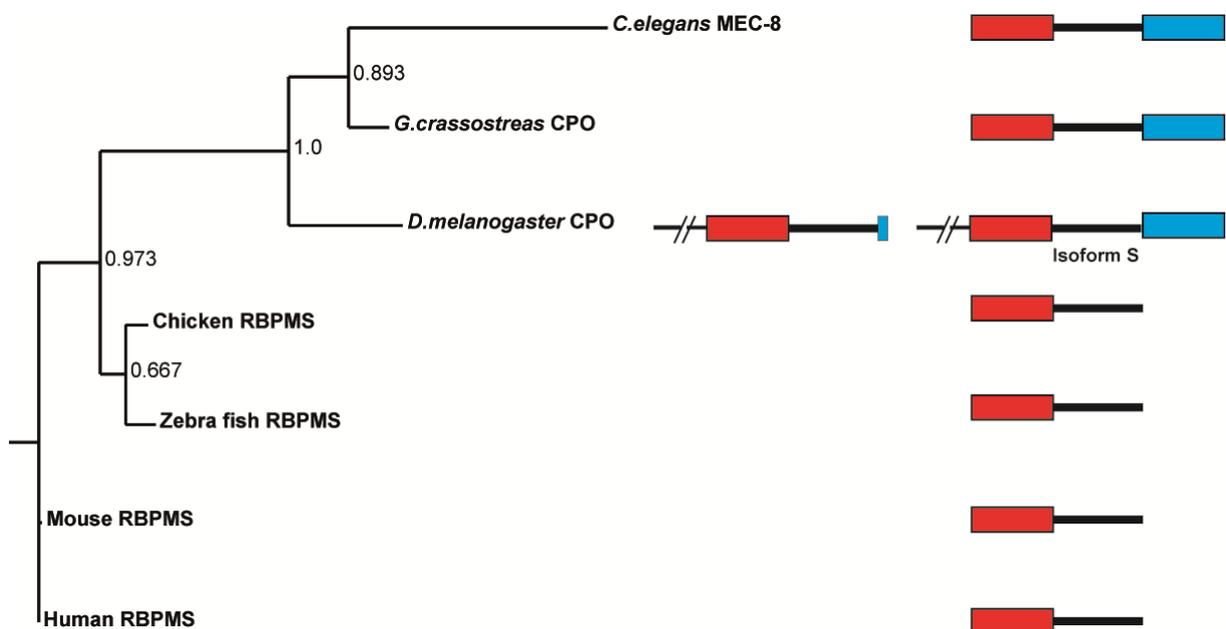


**Figure 80. CPO *D.mel* isoforme S**

(A) Les différentes isoformes de l'ARN messager du gène *dm-CPO*. (B) alignement des séquences de MEC-8 et de l'isoforme S de *dm-CPO*. (C) Phylogénie de la famille RBPMS en prenant en compte CPO-S et CPO.

Cependant une étude approfondie des 12 isoformes de CPO révèle que l'un d'entre eux (CPO-S) possède un second domaine RRM en position C-terminal et que celui-ci présente une séquence conservée avec le RRM2 de MEC-8 (*Figure 80.B*).

Etonnamment les protéines CPO chez les insectes étrangers à la famille des drosophiles sont beaucoup plus proches de la protéine MEC-8 présente chez les nématodes. Nous émettons donc l'hypothèse que la famille des drosophiles possède une place cruciale dans l'évolution des protéines de la famille RBPMS. CPO *D.mel* a la particularité selon son isoforme de présenter les caractéristiques des deux grands groupes de cette famille (*Figure 81*).

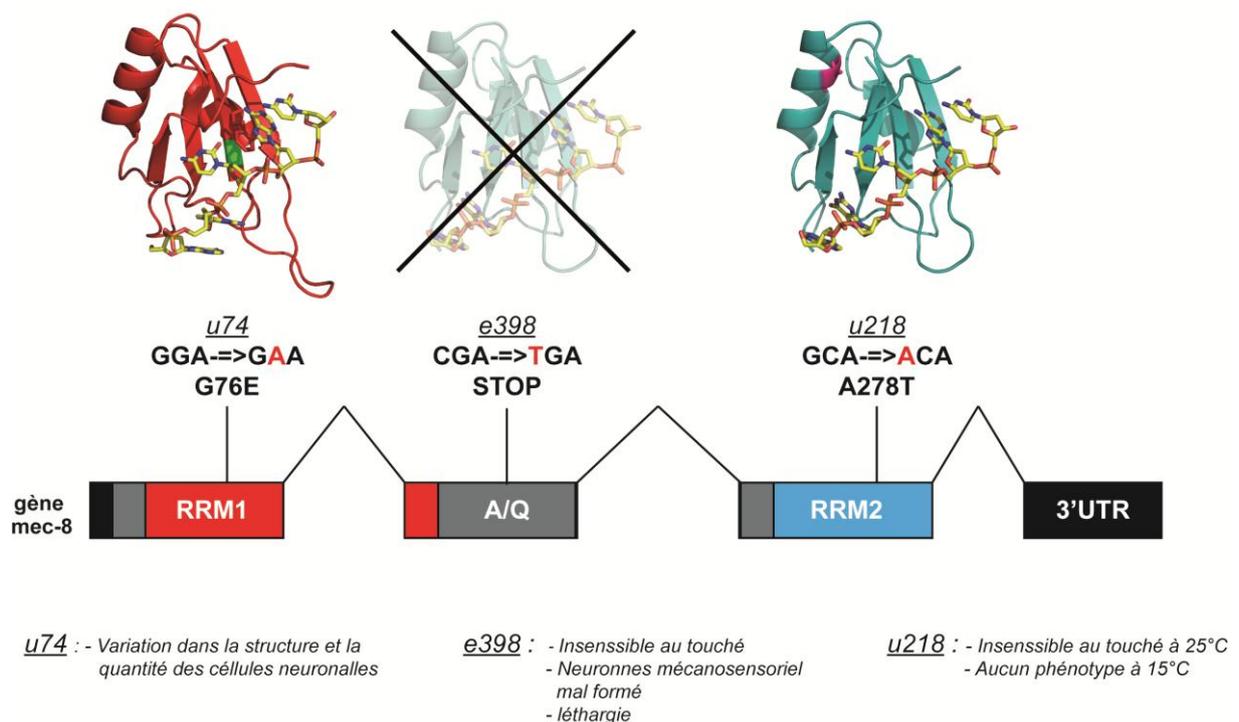


**Figure 81. Arbre phylogénétique de la famille de protéine RBPMS**

Arbre phylogénétique calculé à l'aide du programme PyML utilisant l'algorithme neighbour-joining avec paramètre par défauts. L'alignement des séquences a été réalisé avec le programme Muscle. Les rectangles rouges représentent les domaines RRM N-terminaux présents dans toute la famille de protéine RBPMS. Le rectangle bleu représente le domaine RRM C-terminal absent chez les vertébrés.

L'observation de la répartition du domaine RRM C-terminal dans le règne du vivant (nématodes, insectes, mollusque, crustacé..) nous permet de le lier directement à une spécificité de ce groupe d'organismes : la présence d'un stade larval dans leur cycle de vie.

L'analyse des mutants phénotypiques du gène *mec-8* ne permet cependant pas de lier le domaine RRM-2 à une fonction précise tel que le développement embryonnaire. Chacune de ces mutations qui concernent le RRM1 pour *u74* et le RRM2 pour *e398* et *u218* (**Figure 82**) affecte à la fois le développement des cellules neuronales et musculaires. A noter que ces mutations ne donnent pas de phénotype musculaire lorsque présentes seules, comme c'est souvent le cas chez *C.elegans* (Johnsen and Baillie 1997). Cependant combinés avec une mutation de *unc-52* non phénotypique, les trois mutants engendrent une létalité chez l'embryon avec des défauts de développement musculaire très importants (Davies, Spike et al. 1999). Le second domaine n'a donc vraisemblablement pas de fonction qui lui est propre mais agit sûrement de concert avec le RRM1 dans les fonctions que possède le facteur MEC-8.



**Figure 82. Les mutants phénotypiques de *mec-8***

Représentation des quatre exons du gène *mec-8* avec la localisation des trois mutations engendrant un phénotype viable chez *C.elegans*. La mutation *u74* est représentée en vert sur le RRM1 (rouge), la mutation *u218* est représentée en violet sur le domaine RRM2 (cyan).

## Les sites de liaison in-vivo potentiel de MEC-8

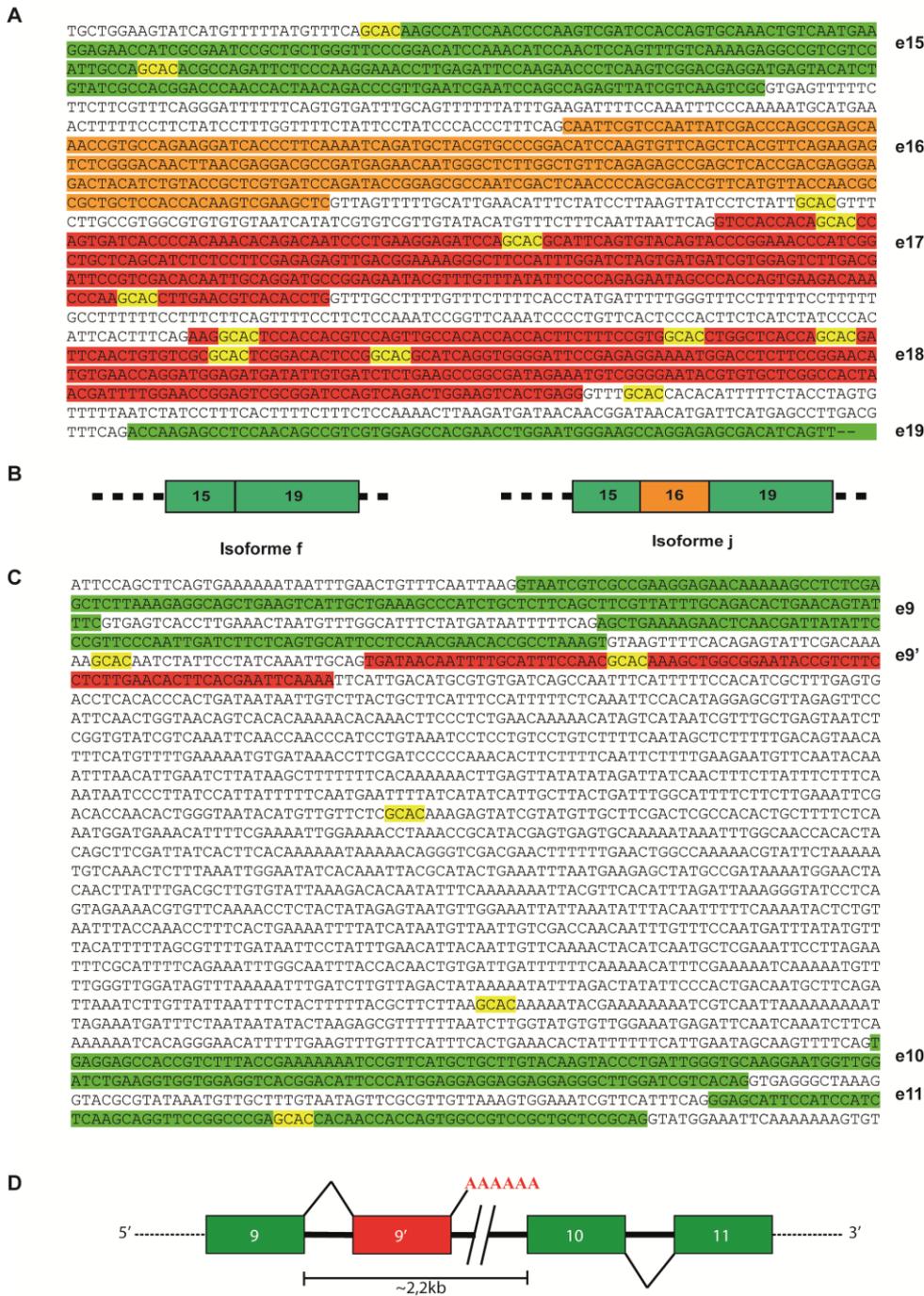
D'après nos études *in-vitro* la protéine MEC-8 a la possibilité de lier quatre motifs GCAC. Afin d'étayer cette hypothèse à un niveau génomique, nous avons analysé les séquences nucléotidiques des gènes dont les isoformes sont régulés par MEC-8.

Les deux gènes *unc-52* et *mec-2* sont des cibles de MEC-8 connues de longue date (Lundquist and Herman 1994, Calixto, Ma et al. 2010), cependant aucune séquence précise n'est connue comme étant une cible spécifique de MEC-8. Les séquences des régions de ces deux pré-ARNm dont l'isoforme varie avec une mutation de MEC-8 sont représentés **Figure 83**.

La cartographie des site GCAC montre une surreprésentation de ces motifs dans les régions directement affectées par un épissage alternatif dépendant de MEC-8. La position de ces motifs (**Figure 83**, **Figure 84**), qu'elle soit relative entre eux ou aux séquences conservées des jonctions exons/introns, ne nous permet pas d'extrapoler un modèle d'action précis du facteur MEC-8 de facons direct.

En effet les motifs GCAC semblent recouvrir aussi bien les points de branchement que les sites de reconnaissance des facteurs U1 et U2AFs. En outre, une étude détaillée du recouvrement de séquence de MEC-8 nécessiterai une information sur le repliement bidimensionnel de celle-ci au minimum. Cependant au vu de l'étendue de la surface recouverte par un quartet de GCAC (env 2kb chez *mec-2* **Figure 83.C**) cette information structurale sera peu fidèle si l'on utilise les moyens de prédiction bio-informatique classiques (RNAfold,mfold...).

Une information que l'on peut néanmoins tirer en analysant les gènes dont l'isoforme est affecté par un défaut de MEC-8 (Kelley, Yochem et al. 2015) est que cette protéine agit principalement en tant que régulateur négatif provoquant l'exclusion des exons comportant des motifs GCAC sur les introns qui les entours.



**Figure 83. Unc-52 et mec-2, deux pré-ARN messagers cibles de MEC-8**

(A) Séquence représentant les exons 15 à 19 de *unc-52*. (B) isoforme de *unc-52* induit par MEC-8. (C) Séquence représentant les exons 9 à 11 de *mec-2*. (D) Schéma représentant l'exon 9' et les 2.2kb suffisant à l'action de MEC-8 sur *mec-2*. En rouge et orange les séquences dont MEC-8 induit l'exclusion de l'ARNm, en jaune les sites GCAC.

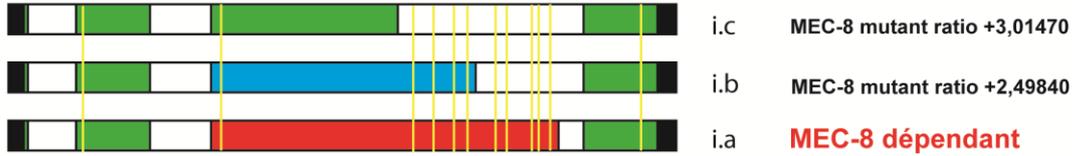
A

```

tttacggtactacacttacactccaacacatatattgtgtcgtcgtcATGGCTAGTGTGAGTTCATTTAAATTACTTTGGCT
TTCATATTTCTCATTATCTTTCTCACACTCACTTTCTCCTTGTTTCAGTTCTTTTTTGATCGCACTTCTCGCTACCCAGC
TCTCGGATGCATCGGAGGCTCGGTGGAGGAGGTGGTGGCTGCTGCCAACCATCTCAACCAGCTTGCGCTAACCCATGTG
GTGGAGCTTCTGCTGCTGDTAAGTTCCTTCTTCTGCTCTTTTTTATTATTTTGATTCAATATTTTCTAGCTATGCTGC
ACCAGCCGCTGGATACCAAGACCTCCAATCGGATATGGAGCACACACCCACCACCACCACCAGCTTTGCGAGGAGC
TCGGTGGAGGAGCCGCTACCCAGGAGCCGGAGCCGGAGGACAAATATGATACCGGAGTTCACGGAGGAGCTCAAGGAGGA
TAGCCCGGAGCCCAAGGAGCTCAAGGAGGATATGCCGGAGCCCAAGGAGCTCAAGGAGGATACGCCGGAGCCAGGGAGC
TCAAGGAGGATACGCCGGAGCCGCCAAGGAGCCCAAGGAGGATACGCTGGAGCCGTCCAAGGAGGACAAATCGCTTCCC
AAGGATACGCTGGAGCCGGAGCCCACTGAGCGCCGGAGGATCTTACAACCAAGGACCAGCCGCTATCAACGGAGGATAC
ACCTCTCAACAACAAGTCGGAGCTATCGCCCCAGCCCGGATACGGACAAGGACCAGCTGCCGTTTCATGTTGACGCCAC
CGCCGCCACATCCGAGGAGCCATCTCTCAACAAGGAGGATCCTCTGCCGCTGGAGGATACGGAGAGGAAGCCGGAGCCA
CTTCAGTCGTTCGGAGAGCAAACCAACACTGTTGTCCAGACCCGGAGAGACCGCCGAGGAAGGATACGGAGCCCAAGGAGGA
GTCGCCGGAGAGACCCGATCGTCGCTGTGCCACCTCTGCCCCAGCCCAACATACGAGAACGTTCAACCCAGC
TCCAGTTGCCCAACCACTGTCGAAGAGACCCAGCTCGAGGAGCCAGTCCAGCTGTCCAGAAGCCACCGTTGTCG
AGGAGTCCCACTTCGAGGAGCCAGCCCCAGCCCCAGCTGCCACTTCTGCTCCATTGTCGAGCCAGTTACTGAGGTGAAT
ACTGAAGAATATGAAGAGGAAGTTGTTGAGCCAGCTGAGCCACAAGTTGTTGAGCCACAAGTTGTCGAGCCAGTTCCACA
GCCACACGTTGACCAGTTCAACCAGTTCAACCAGCACAGAAAGTCACCCAGCCACAAGTTGTTGAGCCACAAGTCGTTG
TTGAGCCAAGCACACAACCAAGTTATTCACAAAGCACAGAGACCTTCGTCGAGACCTCTGCCCCAGTTGCACACAA
GTCAGTGTACCCAATTCCAGCAAGCAAACTGAGGCTCCAGAAGAGGTATACGAAGAGGAGACCGACGCACAGTTCGC
CGCACAGTCCAACCAACGGAATCCTCTTTGAAAGAGGTACCGCTGAGGAGACACCTGTACCAAGCACAGCCTTATG
TTCCACCAAGCAATTTTCACTGTATGTTCTGAAAACTAACACTTTTTTGTCTTCTTCCATTTCTTTCCCTAACATAA
GATCCTTGTTCAGGAGTCAAAGCCAGGATACGAGCAGACCCGCCAGCTGTTGAGACCAAGGAGGAATCTACGACGAG
CAGGAGCCGAGGAGACCCATGTTCGAGCCAGAGCAGCCAGCCGCTCGTTCGGTGGAGGATACTCGGCTACCCGAGCCCGC
CAAGCCAGATGGTGGATACCGCACCCACTAAatttcttttttagttccccataggtgccccattgctctgatctcgtttt
gattgttcgccaacaaccaccctcgacaaatctgctcaaaacatcaaaaattctcgtctttttttgcctcgttctttc
ctttgtctgtatatactgttgccttattttatttctcttatatataattttattatgtcattttcgtgagtagtagacaac
aattcactgactgaaattgtctcaatttcaataaaaatggtgatttttagaaa

```

B



**Figure 84. Zk180.5 un exemple de gène affecté par un dysfonctionnement de MEC-8**

(A) Séquence du gène *zk180.5*, les séquences surlignées en vert sont les exons de l'isoforme c, la séquence bleue représente la partie marquant la fin de l'exon 3 dans l'isoforme b et la séquence rouge la fin de l'exon 3 pour l'isoforme a. Les sites GCAC du gène sont surlignés en jaune. (B) Illustration des trois isoforme du gène *zk180.5*.

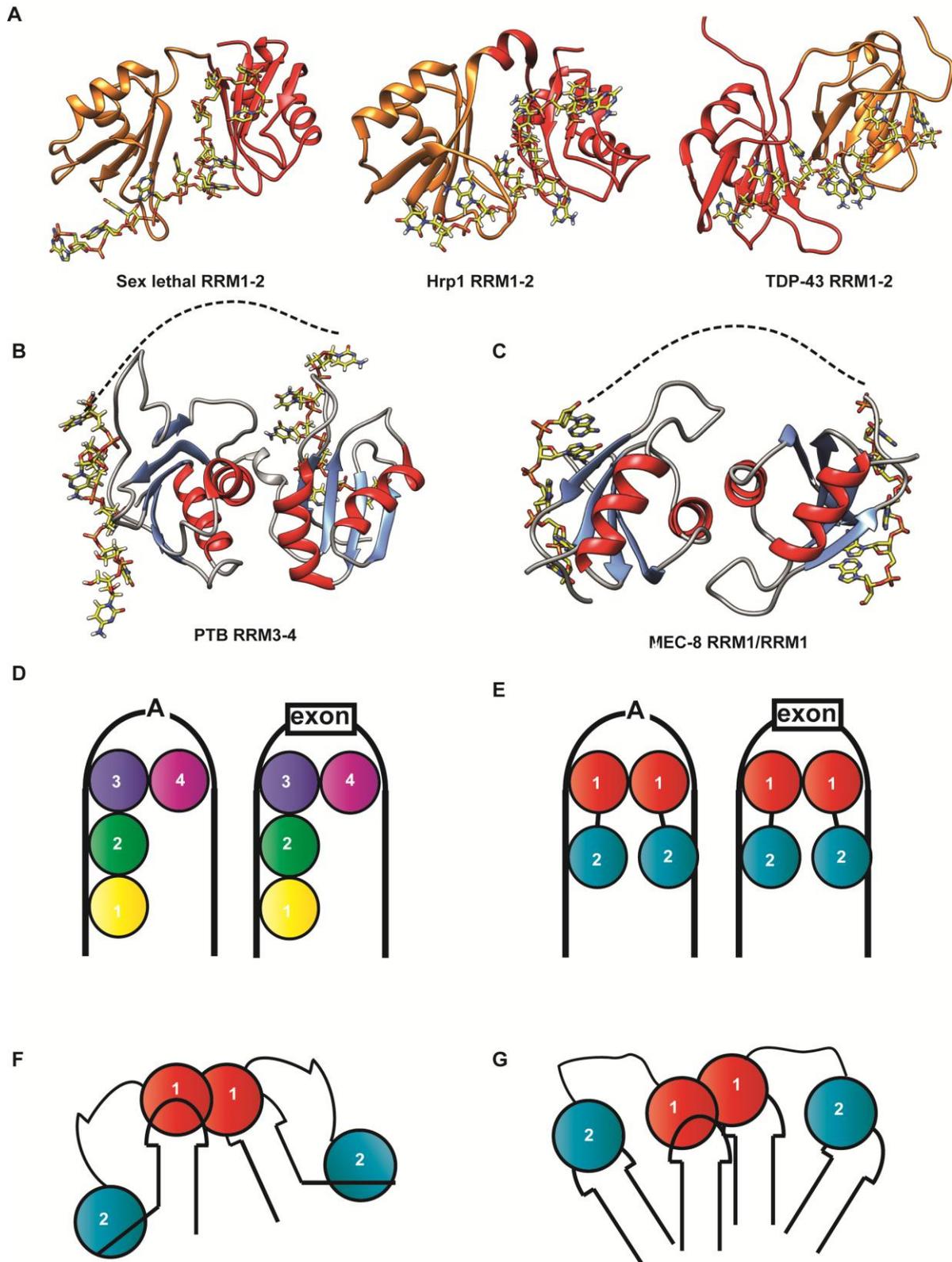
## Les modèles d'action hypothétique du facteur MEC-8

Très peu de données biologiques sont connues quant à l'action du facteur MEC-8 sur ses cibles. Cela rend l'élucidation du mécanisme par lequel MEC-8 agit sur les gènes qu'il régule compliqué. Néanmoins l'analyse de la structure du domaine RMM1 nous permet d'avancer des hypothèses.

En effet les domaines RRM s'associant en dimère étendent généralement leur interface de liaison à l'ARN de façons à reconnaître une séquence de taille supérieure continue, comme pour les protéines Sex lethal, Hrp1 ou TDP-43 (Handa, Nureki et al. 1999, Perez-Canadillas 2006, Lukavsky, Daujotyte et al. 2013). (*Figure 85.A*).

Les domaines RRM1 de MEC-8 ont la particularité de s'associer de façons à présenter deux régions de reconnaissances à l'ARN orientées de façons opposées (*Figure 85.C*). Cette conformation impose à l'ARN cible de former une boucle avec une séquence d'espacement d'une taille minimum de 8 nucléotides, cette structure particulière est très similaire à ce qui a été observé pour la protéine PTB (Oberstrass, Auweter et al. 2005, Auweter, Oberstrass et al. 2007)(*Figure 85.B*).

La protéine PTB utilise cette particularité pour réguler l'épissage de ses gènes cibles de différentes façons (Lamichhane, Daubner et al. 2010)(*Figure 85.D*) et on peut donc imaginer que MEC-8 inhibe la machinerie d'épissage en utilisant les mêmes mécanismes, soit en se liant aux séquences introniques aux extrémités d'un exon, soit en bloquant l'accessibilité au point de branchement du facteur U2AF (*Figure 85.E*). Une particularité de MEC-8 est de pouvoir aussi bien lier des séquences d'ARN simple brins qu'en tige-boucle. Cela permet d'envisager d'autres mécanismes d'action pour cette protéine à l'image de la protéine CsrB qui régule la structure de l'ARN RsmZ en se dimérisant et en liant de multiples motifs tige-boucle (Duss, Michel et al. 2014) (*Figure 85.F.G*).



**Figure 85. Les différents modèles de l'action de MEC**

(A) Représentation des tandem de RRM en complexe avec ARN des protéine Sexl-lethal, Hrp-1 et TDP-43. (B) Dimer RRM3-4 de PTB en complexe avec ARN. (C) Dimer de MEC-8 RRM1 en complexe avec ADN. (D) Modèle d'action d'écrit pour la protéine PTB. (E) Modèle d'action proposé pour MEC-8 déduit de la nature particulière du dimer RRM1. (F)(G) Modèle d'action proposé en rapport à la capacité des RRM1 et 2 de lier des tige-boucle.

## Conclusion

Ce travail nous a permis de répondre à de nombreuses questions biologiques sur la protéine MEC-8 et plus largement sur la famille de protéine RBPMS. Tout d'abord d'un point de vue structural, les structures libres et en complexe avec acide nucléique du domaine RRM1 ont été résolues.

Cela nous a permis d'identifier une interface de dimérisation de ce facteur, ainsi qu'un mécanisme de liaison à un motif ARN spécifique inconnue auparavant.

La seconde information structurale résultant de ce travail est la structure du RRM2 de MEC-8 en complexe avec un motif d'acide nucléique. Malgré la faible homologie entre les deux domaines RRM de MEC-8 ceux-ci lient la même séquence nucléotidique. Cette particularité a pour avantage d'ajouter de la complexité dans la séquence de reconnaissance du facteur MEC-8 (4 motif GCAC). Il nous a alors été permis d'identifier des zones riches en motif GCAC dans les séquences des gènes affectés par MEC-8.

De nombreuses interrogations subsistent néanmoins sur le comportement de la protéine MEC-8 et sur les façons dont cette protéine agit sur ses gènes cibles. Cependant les résultats obtenus lors de cette étude structurale permettront une investigation poussée de ces mécanismes dans un cadre *in-vivo* cette fois-ci.



## Références bibliographique

Afroz, T., L. Skrisovska, E. Belloc, J. Guillen-Boixet, R. Mendez and F. H. Allain (2014). "A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins." Genes Dev **28**(13): 1498-1514.

Allen, M. J. (2007). "What makes a fly enter diapause?" Fly (Austin) **1**(6): 307-310.

Auweter, S. D., F. C. Oberstrass and F. H. Allain (2006). "Sequence-specific binding of single-stranded RNA: is there a code for recognition?" Nucleic Acids Res **34**(17): 4943-4959.

Auweter, S. D., F. C. Oberstrass and F. H. Allain (2007). "Solving the structure of PTB in complex with pyrimidine tracts: an NMR study of protein-RNA complexes of weak affinities." J Mol Biol **367**(1): 174-186.

Bellen, H. J., S. Kooyer, D. D'Evelyn and J. Pearlman (1992). "The Drosophila couch potato protein is expressed in nuclei of peripheral neuronal precursors and shows homology to RNA-binding proteins." Genes Dev **6**(11): 2125-2136.

Bellen, H. J., H. Vaessin, E. Bier, A. Kolodkin, D. D'Evelyn, S. Kooyer and Y. N. Jan (1992). "The Drosophila couch potato gene: an essential gene required for normal adult behavior." Genetics **131**(2): 365-375.

Black, D. L. (2003). "Mechanisms of alternative pre-messenger RNA splicing." Annu Rev Biochem **72**: 291-336.

Blanco, F. J. and C. Bernabeu (2012). "The Splicing Factor SRSF1 as a Marker for Endothelial Senescence." Front Physiol **3**: 54.

Burd, C. G. and G. Dreyfuss (1994). "Conserved structures and diversity of functions of RNA-binding proteins." Science **265**(5172): 615-621.

Calixto, A., C. Ma and M. Chalfie (2010). "Conditional gene expression and RNAi using MEC-8-dependent splicing in *C. elegans*." Nat Methods **7**(5): 407-411.

Chabot, B., C. LeBel, S. Hutchison, F. H. Nasim and M. J. Simard (2003). "Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA." Prog Mol Subcell Biol **31**: 59-88.

Chen, Q., M. Jin, J. Zhu, Q. Xiao and L. Zhang (2013). "Functions of heterogeneous nuclear ribonucleoproteins in stem cell potency and differentiation." Biomed Res Int **2013**: 623978.

Chen, T., B. B. Damaj, C. Herrera, P. Lasko and S. Richard (1997). "Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain." Mol Cell Biol **17**(10): 5707-5718.

Cho, S., A. Hoang, R. Sinha, X. Y. Zhong, X. D. Fu, A. R. Krainer and G. Ghosh (2011). "Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly." Proc Natl Acad Sci U S A **108**(20): 8233-8238.

Christie, A. E., V. Roncalli, P. B. Lona, M. D. McCool, B. L. King, A. Bucklin, D. K. Hartline and P. H. Lenz (2013). "In silico characterization of the insect diapause-associated protein couch potato (CPO) in *Calanus finmarchicus* (Crustacea: Copepoda)." Comp Biochem Physiol Part D Genomics Proteomics **8**(1): 45-57.

Daubner, G. M., A. Brummer, C. Tocchini, S. Gerhardy, R. Ciosk, M. Zavolan and F. H. Allain (2014). "Structural and functional implications of the QUA2 domain on RNA recognition by GLD-1." Nucleic Acids Res **42**(12): 8092-8105.

Davies, A. G., C. A. Spike, J. E. Shaw and R. K. Herman (1999). "Functional overlap between the mec-8 gene and five sym genes in *Caenorhabditis elegans*." Genetics **153**(1): 117-134.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.

Delaney, K. J., S. G. Williams, M. Lawler and K. B. Hall (2014). "Climbing the vertebrate branch of U1A/U2B" protein evolution." RNA **20**(7): 1035-1045.

Dominguez, C., J. F. Fiset, B. Chabot and F. H. Allain (2010). "Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs." Nat Struct Mol Biol **17**(7): 853-861.

Dreyfuss, G., V. N. Kim and N. Kataoka (2002). "Messenger-RNA-binding proteins and the messages they carry." Nat Rev Mol Cell Biol **3**(3): 195-205.

Duss, O., E. Michel, M. Yulikov, M. Schubert, G. Jeschke and F. H. Allain (2014). "Structural basis of the non-coding RNA RsmZ acting as a protein sponge." Nature **509**(7502): 588-592.

Farazi, T. A., C. S. Leonhardt, N. Mukherjee, A. Mihailovic, S. Li, K. E. Max, C. Meyer, M. Yamaji, P. Cekan, N. C. Jacobs, S. Gerstberger, C. Bognanni, E. Larsson, U. Ohler and T. Tuschl (2014). "Identification of the RNA recognition element of the RBPMS family of RNA-binding proteins and their transcriptome-wide mRNA targets." RNA **20**(7): 1090-1102.

Foot, J. N., M. Feracci and C. Dominguez (2014). "Screening protein--single stranded RNA complexes by NMR spectroscopy for structure determination." Methods **65**(3): 288-301.

Fu, J., L. Cheng, Y. Wang, P. Yuan, X. Xu, L. Ding, H. Zhang, K. Jiang, H. Song, Z. Chen and Q. Ye (2015). "The RNA-binding protein RBPMS1 represses AP-1 signaling and regulates breast cancer cell proliferation and migration." Biochim Biophys Acta **1853**(1): 1-13.

Fu, X. D. and T. Maniatis (1992). "The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site." Proc Natl Acad Sci U S A **89**(5): 1725-1729.

Furukawa, M. T., H. Sakamoto and K. Inoue (2015). "Interaction and colocalization of HERMES/RBPMS with NonO, PSF, and G3BP1 in neuronal cytoplasmic RNP granules in mouse retinal line cells." Genes Cells **20**(4): 257-266.

Garcia-Blanco, M. A., S. F. Jamison and P. A. Sharp (1989). "Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns." Genes Dev **3**(12A): 1874-1886.

Gerber, W. V., S. A. Vokes, N. R. Zearfoss and P. A. Krieg (2002). "A role for the RNA-binding protein, hermes, in the regulation of heart development." Dev Biol **247**(1): 116-126.

Gerber, W. V., T. A. Yatskievych, P. B. Antin, K. M. Correia, R. A. Conlon and P. A. Krieg (1999). "The RNA-binding protein gene, hermes, is expressed at high levels in the developing heart." Mech Dev **80**(1): 77-86.

Glasscock, E. and M. A. Tanouye (2005). "Drosophila couch potato mutants exhibit complex neurological abnormalities including epilepsy phenotypes." Genetics **169**(4): 2137-2149.

Goodman, M. B., G. G. Ernstrom, D. S. Chelur, R. O'Hagan, C. A. Yao and M. Chalfie (2002). "MEC-2 regulates *C. elegans* DEG/ENaC channels needed for mechanosensation." Nature **415**(6875): 1039-1042.

Graveley, B. R. and T. Maniatis (1998). "Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing." Mol Cell **1**(5): 765-771.

Gredell, J. A., M. J. Dittmer, M. Wu, C. Chan and S. P. Walton (2010). "Recognition of siRNA asymmetry by TAR RNA binding protein." Biochemistry **49**(14): 3148-3155.

Handa, N., O. Nureki, K. Kurimoto, I. Kim, H. Sakamoto, Y. Shimura, Y. Muto and S. Yokoyama (1999). "Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein." Nature **398**(6728): 579-585.

Hapkova, I., J. Skarda, C. Rouleau, A. Thys, C. Notarnicola, M. Janikova, F. Bernex, M. Rypka, J. M. Vanderwinden, S. Faure, J. Vesely and P. de Santa Barbara (2013). "High expression of the RNA-binding protein RBPMS2 in gastrointestinal stromal tumors." Exp Mol Pathol **94**(2): 314-321.

Hastings, M. L., C. M. Wilson and S. H. Munroe (2001). "A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA." RNA **7**(6): 859-874.

Heim, A. E., O. Hartung, S. Rothamel, E. Ferreira, A. Jenny and F. L. Marlow (2014). "Oocyte polarity requires a Bucky ball-dependent feedback amplification loop." Development **141**(4): 842-854.

Hornberg, H., F. Wollerton-van Horck, D. Maurus, M. Zwart, H. Svoboda, W. A. Harris and C. E. Holt (2013). "RNA-binding protein Hermes/RBPMS inversely affects synapse density and axon arbor formation in retinal ganglion cells in vivo." J Neurosci **33**(25): 10384-10395.

Hui, J., L. H. Hung, M. Heiner, S. Schreiner, N. Neumuller, G. Reither, S. A. Haas and A. Bindereif (2005). "Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing." EMBO J **24**(11): 1988-1998.

Izquierdo, J. M., N. Majos, S. Bonnal, C. Martinez, R. Castelo, R. Guigo, D. Bilbao and J. Valcarcel (2005). "Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition." Mol Cell **19**(4): 475-484.

Johnsen, R. C. and D. L. Baillie (1997). Mutation. C. elegans II. D. L. Riddle, T. Blumenthal, B. J. Meyer and J. R. Priess. Cold Spring Harbor (NY).

Kankare, M., T. Salminen, A. Laiho, L. Vesala and A. Hoikkala (2010). "Changes in gene expression linked with adult reproductive diapause in a northern malt fly species: a candidate gene microarray study." BMC Ecol **10**: 3.

Kelley, M., J. Yochem, M. Krieg, A. Calixto, M. G. Heiman, A. Kuzmanov, V. Meli, M. Chalfie, M. B. Goodman, S. Shaham, A. Frand and D. S. Fay (2015). "FBN-1, a fibrillin-related protein, is required for resistance of the epidermis to mechanical deformation during *C. elegans* embryogenesis." Elife **4**.

Keren, H., G. Lev-Maor and G. Ast (2010). "Alternative splicing and evolution: diversification, exon definition and function." Nat Rev Genet **11**(5): 345-355.

Kwong, J. M., J. Caprioli and N. Piri (2010). "RNA binding protein with multiple splicing: a new marker for retinal ganglion cells." Invest Ophthalmol Vis Sci **51**(2): 1052-1058.

Lamichhane, R., G. M. Daubner, J. Thomas-Crusells, S. D. Auweter, C. Manatschal, K. S. Austin, O. Valniuk, F. H. Allain and D. Rueda (2010). "RNA looping by PTB: Evidence

using FRET and NMR spectroscopy for a role in splicing repression." Proc Natl Acad Sci U S A **107**(9): 4105-4110.

Lee, S. F., C. M. Sgro, J. Shirriffs, C. W. Wee, L. Rako, B. van Heerwaarden and A. A. Hoffmann (2011). "Polymorphism in the couch potato gene clines in eastern Australia but is not associated with ovarian dormancy in *Drosophila melanogaster*." Mol Ecol **20**(14): 2973-2984.

Leulliot, N., S. Quevillon-Cheruel, M. Graille, H. van Tilbeurgh, T. C. Leeper, K. S. Godin, T. E. Edwards, S. T. Sigurdsson, N. Rozenkrants, R. J. Nagel, M. Ares and G. Varani (2004). "A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNAse III." EMBO J **23**(13): 2468-2477.

Lim, L. P. and C. B. Burge (2001). "A computational analysis of sequence features involved in recognition of short introns." Proc Natl Acad Sci U S A **98**(20): 11193-11198.

Lin, C. H. and J. G. Patton (1995). "Regulation of alternative 3' splice site selection by constitutive splicing factors." RNA **1**(3): 234-245.

Lukavsky, P. J., D. Daujotyte, J. R. Tollervey, J. Ule, C. Stuani, E. Buratti, F. E. Baralle, F. F. Damberger and F. H. Allain (2013). "Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43." Nat Struct Mol Biol **20**(12): 1443-1449.

Lunde, B. M., C. Moore and G. Varani (2007). "RNA-binding proteins: modular design for efficient function." Nat Rev Mol Cell Biol **8**(6): 479-490.

Lundquist, E. A. and R. K. Herman (1994). "The *mec-8* gene of *Caenorhabditis elegans* affects muscle and sensory neuron function and interacts with three other genes: *unc-52*, *smu-1* and *smu-2*." Genetics **138**(1): 83-101.

Lundquist, E. A., R. K. Herman, T. M. Rogalski, G. P. Mullen, D. G. Moerman and J. E. Shaw (1996). "The mec-8 gene of *C. elegans* encodes a protein with two RNA recognition motifs and regulates alternative splicing of unc-52 transcripts." Development **122**(5): 1601-1610.

Mackereth, C. D., T. Madl, S. Bonnal, B. Simon, K. Zanier, A. Gasch, V. Rybin, J. Valcarcel and M. Sattler (2011). "Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF." Nature **475**(7356): 408-411.

Mancuso, V. P., J. M. Parry, L. Storer, C. Poggioli, K. C. Nguyen, D. H. Hall and M. V. Sundaram (2012). "Extracellular leucine-rich repeat proteins are required to organize the apical extracellular matrix and maintain epithelial junction integrity in *C. elegans*." Development **139**(5): 979-990.

Maniatis, T. and R. Reed (1987). "The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing." Nature **325**(6106): 673-678.

Manley, J. L. and A. R. Krainer (2010). "A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins)." Genes Dev **24**(11): 1073-1074.

Mannsfeldt, A. G., P. Carroll, C. L. Stucky and G. R. Lewin (1999). "Stomatin, a MEC-2 like protein, is expressed by mammalian sensory neurons." Mol Cell Neurosci **13**(6): 391-404.

Maris, C., C. Dominguez and F. H. Allain (2005). "The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression." FEBS J **272**(9): 2118-2131.

Maydanovych, O. and P. A. Beal (2006). "Breaking the central dogma by RNA editing." Chem Rev **106**(8): 3397-3411.

Mount, S. M. (1982). "A catalogue of splice junction sequences." Nucleic Acids Res **10**(2): 459-472.

Nagai, K., C. Oubridge, N. Ito, T. H. Jessen, J. Avis and P. Evans (1995). "Crystal structure of the U1A spliceosomal protein complexed with its cognate RNA hairpin." Nucleic Acids Symp Ser(34): 1-2.

Nagy, E. and L. E. Maquat (1998). "A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance." Trends Biochem Sci **23**(6): 198-199.

Notarnicola, C., C. Rouleau, L. Le Guen, A. Virsolvy, S. Richard, S. Faure and P. De Santa Barbara (2012). "The RNA-binding protein RBPMS2 regulates development of gastrointestinal smooth muscle." Gastroenterology **143**(3): 687-697 e681-689.

Oberstrass, F. C., S. D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D. L. Black and F. H. Allain (2005). "Structure of PTB bound to RNA: specific binding and implications for splicing regulation." Science **309**(5743): 2054-2057.

Oh, Y. L., B. Hahm, Y. K. Kim, H. K. Lee, J. W. Lee, O. Song, K. Tsukiyama-Kohara, M. Kohara, A. Nomoto and S. K. Jang (1998). "Determination of functional domains in polypyrimidine-tract-binding protein." Biochem J **331** ( Pt 1): 169-175.

Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler and P. A. Sharp (1986). "Splicing of messenger RNA precursors." Annu Rev Biochem **55**: 1119-1150.

Pan, Q., O. Shai, L. J. Lee, B. J. Frey and B. J. Blencowe (2008). "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing." Nat Genet **40**(12): 1413-1415.

Perez-Canadillas, J. M. (2006). "Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1." EMBO J **25**(13): 3167-3178.

Pozzoli, U. and M. Sironi (2005). "Silencers regulate both constitutive and alternative splicing events in mammals." Cell Mol Life Sci **62**(14): 1579-1604.

Price, S. R., P. R. Evans and K. Nagai (1998). "Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA." Nature **394**(6694): 645-650.

Ray, D., H. Kazan, E. T. Chan, L. Pena Castillo, S. Chaudhry, S. Talukder, B. J. Blencowe, Q. Morris and T. R. Hughes (2009). "Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins." Nat Biotechnol **27**(7): 667-670.

Ray, D., H. Kazan, K. B. Cook, M. T. Weirauch, H. S. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L. H. Matzat, R. K. Dale, S. A. Smith, C. A. Yarosh, S. M. Kelly, B. Nabet, D. Mecnas, W. Li, R. S. Laishram, M. Qiao, H. D. Lipshitz, F. Piano, A. H. Corbett, R. P. Carstens, B. J. Frey, R. A. Anderson, K. W. Lynch, L. O. Penalva, E. P. Lei, A. G. Fraser, B. J. Blencowe, Q. D. Morris and T. R. Hughes (2013). "A compendium of RNA-binding motifs for decoding gene regulation." Nature **499**(7457): 172-177.

Rogalski, T. M., E. J. Gilchrist, G. P. Mullen and D. G. Moerman (1995). "Mutations in the unc-52 gene responsible for body wall muscle defects in adult *Caenorhabditis elegans* are located in alternatively spliced exons." Genetics **139**(1): 159-169.

Rogalski, T. M., B. D. Williams, G. P. Mullen and D. G. Moerman (1993). "Products of the unc-52 gene in *Caenorhabditis elegans* are homologous to the core protein of the mammalian basement membrane heparan sulfate proteoglycan." Genes Dev **7**(8): 1471-1484.

Ruby, S. W. and J. Abelson (1991). "Pre-mRNA splicing in yeast." Trends Genet **7**(3): 79-85.

Sagnol, S., Y. Yang, Y. Bessin, F. Allemand, I. Hapkova, C. Notarnicola, J. F. Guichou, S. Faure, G. Labesse and P. de Santa Barbara (2014). "Homodimerization of RBPMS2 through a new RRM-interaction motif is necessary to control smooth muscle plasticity." Nucleic Acids Res **42**(15): 10173-10184.

Sattler, M., J. Schleucher and C. Griesinger (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients." Prog. Nucl. Magn. Reson. Spectrosc. **34**, **93–158**.

Schmidt, P. S., C. T. Zhu, J. Das, M. Batavia, L. Yang and W. F. Eanes (2008). "An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*." Proc Natl Acad Sci U S A **105**(42): 16207-16211.

Sharma, S., A. M. Falick and D. L. Black (2005). "Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex." Mol Cell **19**(4): 485-496.

Shimamoto, A., S. Kitao, K. Ichikawa, N. Suzuki, Y. Yamabe, O. Imamura, Y. Tokutake, M. Satoh, T. Matsumoto, J. Kuromitsu, H. Kataoka, K. Sugawara, M. Sugawara, M. Sugimoto,

M. Goto and Y. Furuichi (1996). "A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs." Proc Natl Acad Sci U S A **93**(20): 10913-10917.

Spartz, A. K., R. K. Herman and J. E. Shaw (2004). "SMU-2 and SMU-1, *Caenorhabditis elegans* homologs of mammalian spliceosome-associated proteins RED and fSAP57, work together to affect splice site choice." Mol Cell Biol **24**(15): 6811-6823.

Spike, C. A., A. G. Davies, J. E. Shaw and R. K. Herman (2002). "MEC-8 regulates alternative splicing of *unc-52* transcripts in *C. elegans* hypodermal cells." Development **129**(21): 4999-5008.

Spike, C. A., J. E. Shaw and R. K. Herman (2001). "Analysis of *smu-1*, a gene that regulates the alternative splicing of *unc-52* pre-mRNA in *Caenorhabditis elegans*." Mol Cell Biol **21**(15): 4985-4995.

Stamm, S., S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber, T. A. Thanaraj and H. Soreq (2005). "Function of alternative splicing." Gene **344**: 1-20.

Stark, H. and R. Luhrmann (2006). "Cryo-electron microscopy of spliceosomal components." Annu Rev Biophys Biomol Struct **35**: 435-457.

Stefl, R., F. C. Oberstrass, J. L. Hood, M. Jourdan, M. Zimmermann, L. Skrisovska, C. Maris, L. Peng, C. Hofr, R. B. Emeson and F. H. Allain (2010). "The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove." Cell **143**(2): 225-237.

Teplova, M., T. A. Farazi, T. Tuschl and D. J. Patel (2015). "Structural basis underlying CAC RNA recognition by the RRM domain of dimeric RNA-binding protein RBPMS." Q Rev Biophys: 1-11.

Teplova, M., M. Hafner, D. Teplov, K. Essig, T. Tuschl and D. J. Patel (2013). "Structure-function studies of STAR family Quaking proteins bound to their in vivo RNA target sites." Genes Dev **27**(8): 928-940.

Wang, Z. and C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." RNA **14**(5): 802-813.

Weinberg, D. E., K. Nakanishi, D. J. Patel and D. P. Bartel (2011). "The inside-out mechanism of Dicers from budding yeasts." Cell **146**(2): 262-276.

Williams, B. D. and R. H. Waterston (1994). "Genes critical for muscle development and function in *Caenorhabditis elegans* identified through lethal mutations." J Cell Biol **124**(4): 475-490.

Yochem, J., L. R. Bell and R. K. Herman (2004). "The identities of sym-2, sym-3 and sym-4, three genes that are synthetically lethal with mec-8 in *Caenorhabditis elegans*." Genetics **168**(3): 1293-1306.

Zeng, Q. and K. B. Hall (1997). "Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability." RNA **3**(3): 303-314.

Zhang, Q. and D. L. Denlinger (2011). "Elevated couch potato transcripts associated with adult diapause in the mosquito *Culex pipiens*." J Insect Physiol **57**(5): 620-627.

Zhang, S., J. Arnadottir, C. Keller, G. A. Caldwell, C. A. Yao and M. Chalfie (2004). "MEC-2 is recruited to the putative mechanosensory complex in *C. elegans* touch receptor neurons through its stomatin-like domain." Curr Biol **14**(21): 1888-1896.