

UNIVERSITÉ DE SHERBROOKE

**DÉTERMINATION DE LA STRUCTURE ET DE LA DYNAMIQUE DU  
DOMAINE DE LA PROTÉINE MIZ-1 FORMÉ DES DOIGTS DE ZINC 5 À 8  
PAR RÉSONANCE MAGNÉTIQUE NUCLÉAIRE**

Par

DAVID BERNARD

Programme de pharmacologie

Mémoire présenté à la Faculté de médecine et des sciences de la santé

En vue de l'obtention du grade de

Maître ès sciences (M.Sc.) en pharmacologie

Sherbrooke, Québec, Canada

Juillet 2013

Membres du jury d'évaluation

Pr. Pierre Lavigne, départements de biochimie et de pharmacologie

Pr. Guylain Boulay, département de pharmacologie

Pr. Martin Lepage, département de médecine nucléaire et radiobiologie

© David Bernard, 2013



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*ISBN: 978-0-499-00305-8*

*Our file Notre référence*

*ISBN: 978-0-499-00305-8*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

# Canada

## RÉSUMÉ

### DÉTERMINATION DE LA STRUCTURE DU DOMAINE DE LA PROTÉINE MIZ-1 FORMÉ DES DOIGTS DE ZINC 5 À 8 PAR RÉSONANCE MAGNÉTIQUE NUCLÉAIRE

Par

DAVID BERNARD

Programme de pharmacologie

Mémoire présenté à la Faculté de médecine et des sciences de la santé  
En vue de l'obtention du grade de Maître ès Sciences (M.Sc.) en pharmacologie, Faculté  
de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec,  
Canada, J1H 5N4

Miz-1 est un facteur de transcription qui active la transcription de gènes cytotostatiques tels que p15<sup>INK4B</sup> ou p21<sup>CIP1</sup>. Il s'agit d'une protéine de la famille BTB/POZ, qui possède un ensemble de 13 doigts de zinc de type Cys<sub>2</sub>His<sub>2</sub> dans sa portion C-terminale. L'activation de ces gènes par Miz-1 peut être régulée par les protéines SMAD. SMAD3 et 4 peuvent toutes deux se lier aux doigts de zinc 1 à 4 de Miz-1, et les autres doigts de zinc sont pressentis pour être responsables de la liaison à l'ADN. Les séquences reconnues par Miz-1 sur les promoteurs des deux gènes mentionnés ci-haut ont été identifiées, mais n'ont pas d'homologie entre elles. L'oncogène c-Myc a la possibilité de se lier à Miz-1, et cette interaction cause la répression des gènes normalement activés par Miz-1, favorisant ainsi la prolifération cellulaire. Cette interaction cruciale, de même que celle entre Miz-1 et l'ADN, est toutefois assez mal caractérisée. Le but du projet dont fait partie ce mémoire est d'éclaircir tout ce mécanisme de liaison.

Ce mémoire étudie la structure et les propriétés dynamiques des doigts de zinc 5 à 8 de Miz-1, qui, selon l'hypothèse initiale, seraient impliqués dans la liaison au promoteur des gènes-cibles de Miz-1. La résonance magnétique nucléaire est la technique qui a été utilisée dans le but d'obtenir ces résultats, et une partie importante de ce mémoire est dédiée à la théorie derrière cette puissante technique. La détermination de la structure de ces doigts de zinc est en fait une seule des nombreuses étapes du projet de l'élucidation du mécanisme de transrépression par c-Myc/Miz-1. Suite à la présentation des structures de ces doigts de zinc, ce mémoire s'intéresse à leurs caractéristiques dynamiques très particulières pour ce type de domaine protéique. Nous avons en effet découvert que les doigts de zinc 5 à 8 présentaient un niveau très élevé d'échange conformationnel, et qu'une portion du doigt de zinc 6 ne peut être caractérisée structurellement à cause de mouvements dans l'échelle de la micro/milliseconde, eux-mêmes dus à des répulsions électrostatiques. En conclusion, nous proposons que le ZF 6 ait un rôle de charnière entre deux ensembles de ZFs dans Miz-1. Ces résultats permettent l'actualisation du modèle de liaison de Miz-1 au promoteur de p15<sup>INK4B</sup> qui avait été élaboré comme hypothèse, et ils nous rapprochent de la compréhension de la transrépression par c-Myc.

Mots-clés : Doigts de zinc C<sub>2</sub>H<sub>2</sub>; résonance magnétique nucléaire; structure de protéines; dynamique de protéines; purification de protéines.

## Table des matières

RÉSUMÉ.....	II
LISTE DES TABLEAUX.....	V
LISTE DES ILLUSTRATIONS.....	VI
LISTE DES ABRÉVIATIONS.....	VII
<b>1. Introduction.....</b>	<b>1</b>
1.1 Protéines à doigts de zinc.....	1
1.1.1 Facteurs de transcription et types de doigts de zinc.....	1
1.1.2 Doigts de zinc de type Cys <sub>2</sub> His <sub>2</sub> .....	3
1.1.2.1 Caractéristiques structurales des doigts de zinc de type Cys <sub>2</sub> His <sub>2</sub> .....	3
1.1.2.2 Fonctions des doigts de zinc de type Cys <sub>2</sub> His <sub>2</sub> .....	4
1.2 Myc-interacting zinc finger protein-1 (Miz-1).....	6
1.2.1 La protéine c-Myc.....	6
1.2.1.1 Activation de la transcription par c-Myc.....	6
1.2.1.2 Répression de la transcription par c-Myc.....	7
1.2.1.3 c-Myc dans un contexte pathophysiologique.....	9
1.2.2 Gènes régulés par Miz-1.....	9
1.2.3 Caractéristiques de Miz-1.....	10
1.2.4 Activation de gènes par Miz-1.....	12
1.2.5 Répression de gènes par Miz-1.....	13
<b>2. Résonance magnétique nucléaire.....</b>	<b>15</b>
2.1 Concepts de base de la RMN.....	15
2.1.1 Moment cinétique intrinsèque (spin).....	15
2.1.2 Les niveaux d'énergie de spin.....	18
2.1.3 La détection de la magnétisation et le déplacement chimique.....	21
2.1.4 Relaxation.....	27
2.2 Formalisme de l'opérateur produit.....	30
2.3 Séquences de pulses de base.....	33
2.3.1 L'écho de spin (« spin echo ») sur un seul type de noyau.....	33
2.3.2 L'écho de spin sur deux types de noyaux.....	36
2.3.3 La séquence INEPT.....	37
2.3.4 La séquence HSQC.....	38
2.4 Spectres RMN.....	40
2.4.1 HSQC.....	40
2.4.2 HNCACB.....	41
2.4.3 CBCA(CO)NH.....	43
2.4.4 HNCO.....	45
2.4.5 C(CO)NH.....	45
2.4.6 H(CCO)NH.....	46
2.4.7 HCCH-TOCSY.....	46
2.4.8 <sup>15</sup> N- et <sup>13</sup> C-edited NOESY.....	48
2.4.9 Mesure des paramètres dynamiques d'un échantillon par RMN.....	49
2.4.10 CPMG-HSQC et dispersion de relaxation.....	51
<b>3. Présentation du projet.....</b>	<b>53</b>

3.1 Résumé du projet.....	53
3.2 Motivations.....	54
<b>4. Article.....</b>	<b>55</b>
4.1 Contribution des auteurs.....	55
4.2 Résumé de l'article.....	57
<b>Structural and Dynamical Characterization of the Miz-1 Zinc Fingers 5 to 8 by Solution-State NMR ....</b>	<b>58</b>
<b>Abstract.....</b>	<b>59</b>
<b>Introduction.....</b>	<b>60</b>
<b>Materials and methods.....</b>	<b>63</b>
Preparation of the Miz 5-8 construct.....	63
CD spectropolarimetry.....	64
NMR spectroscopy.....	64
<sup>15</sup> N Spin relaxation.....	65
Data processing and analysis.....	66
Structure calculations.....	66
<b>Results.....</b>	<b>67</b>
Folding of Miz 5-8.....	67
Resonance assignments and secondary structure of Miz 5-8.....	68
Evidence of conformational exchange on the $\mu$ s-ms timescale located in ZF 5 and 6.....	70
<sup>15</sup> N spin relaxation.....	72
Solution structures of Miz-1 ZFs 5 to 7.....	76
<b>Discussion and conclusion.....</b>	<b>78</b>
<b>Acknowledgements.....</b>	<b>82</b>
<b>References.....</b>	<b>82</b>
<b>6. Discussion.....</b>	<b>85</b>
<b>7. Conclusion.....</b>	<b>91</b>
<b>Remerciements.....</b>	<b>92</b>
<b>LISTE DES RÉFÉRENCES – LIVRES.....</b>	<b>93</b>
<b>LISTE DES RÉFÉRENCES – ARTICLES.....</b>	<b>93</b>
<b>Annexe 1 – Matériel supplémentaire à l'article.....</b>	<b>98</b>
<b>Annexe 2 – Autorisations de reproduction de figures.....</b>	<b>101</b>

**LISTE DES TABLEAUX**

Tableau 1 : Classes de doigts de zinc .....	2
Table 1: Structural statistics for ZFs 5 to 7 of Miz-1 .....	77
Tableau supplémentaire 1: Paramètres calculés de relaxation de dispersion .....	98

## LISTE DES ILLUSTRATIONS

Figure 1 : Structure générale d'un doigt de zinc canonique.....	3
Figure 2 : Contacts habituels entre doigts de zinc et ADN .....	5
Figure 3 : Activation de la transcription par c-Myc/Max (MONTAGNE, 2008, p.15).....	7
Figure 4 : Protéines impliquées dans la répression par c-Myc/Max et Miz-1 .....	8
Figure 5 : Structure du domaine BTB/POZ de Miz-1 .....	10
Figure 6 : Structure des ZFs 8 à 10 de Miz-1 .....	11
Figure 7 : Séquence complète de Miz-1 .....	12
Figure 8 : Protéines impliquées dans la transactivation par Miz-1 .....	13
Figure 9 : Protéines impliquées dans la transrépression par Miz-1 .....	14
Figure 10 : Schéma de l'expérience de Stern et Gerlach (COHEN-TANNOUJDI et al., 1998, p.386) .....	17
Figure 11 : Résultat de l'expérience de Stern et Gerlach (COHEN-TANNOUJDI et al., 1998, p.389).....	17
Figure 12 : Niveaux d'énergie de spin dans un champ magnétique.....	19
Figure 13 : Séparation des niveaux d'énergie dans un système de deux spins couplés.....	21
Figure 14 : Moments magnétiques, magnétisation et pulse 90° .....	22
Figure 15 : Précession des spins dans le référentiel tournant.....	23
Figure 16 : Déphasage des spins .....	24
Figure 17 : Exemple de FID (spectre 1D de Miz 5-8).....	24
Figure 18 : Exemple de spectre RMN 1D (Miz 5-8).....	26
Figure 19 : Écho de spin sur le noyau I.....	34
Figure 20 : Écho de spin sur deux types de noyaux .....	36
Figure 21 : Séquence de pulses INEPT .....	37
Figure 22 : Séquence de pulses d'un spectre HSQC (CAVANAGH et al., 1996, p.412).....	38
Figure 23 : Exemple de spectre HSQC (Miz 5-8) .....	39
Figure 24 : HSQC d'une protéine bien repliée ou dénaturée (Miz 5-8 à pH 7 ou 4) .....	41
Figure 25 : Exemple de HNCACB et d'attribution séquentielle d'une séquence SGNLKR de Miz 5-8 .....	42
Figure 26 : Exemple de CBCA(CO)NH d'une séquence SGNLKR de Miz 5-8 .....	44
Figure 27 : Exemple de HCCH-TOCSY d'une glutamine de Miz 5-8 .....	47
Figure 28 : Exemple de détermination de la valeur de T <sub>1</sub> de l'amide d'une tyrosine de Miz 5-8 .....	50
Figure 1: Characterization of Miz 5-8 .....	68
Figure 2: <sup>1</sup> H- <sup>15</sup> N HSQC spectrum of Miz 5-8 .....	69
Figure 3: Effect of pH on the intensities of Miz 5-8 backbone amide resonances. ....	71
Figure 4: <sup>15</sup> N relaxation measurements of the Miz 5-8 backbone amides on the ps-ns time-scale. ....	72
Figure 5: T <sub>1</sub> /T <sub>2</sub> and <sup>15</sup> N relaxation dispersion of Miz 5-8 backbone amides.....	73
Figure 6: <sup>15</sup> N relaxation dispersion of Miz 5-8 backbone amides. ....	75
Figure 7: NMR solution structures of (a) ZF 5, (b) ZF 6 and (c) ZF 7.....	76
Figure 8: Molecular model for the conformational exchange located in ZFs 5 and 6.....	81
Figure 29 : Lys19 dans le cœur hydrophobe du ZF 5 .....	87
Figure 30 : Modèle proposé de la liaison de Miz-1 à l'ADN avec les partenaires impliqués.....	89
Figure S1 : Distribution de NOEs inter-résidus en fonction de la structure primaire de Miz 5-8 .....	98

## LISTE DES ABRÉVIATIONS

<b>b-HLH-LZ :</b>	Domaine basique – hélice-boucle-hélice – glissière à leucines (Basic region – helix-loop-helix – leucine zipper)
<b>BTB/POZ :</b>	Domaine « BR-C, ttk and bab », ou domaine « Pox virus and zinc finger »
<b>CD :</b>	Spectroscopie par dichroïsme circulaire
<b>CSA :</b>	Anisotropie de déplacement chimique (chemical shift anisotropy)
<b>E-box :</b>	Séquence enhancer box (CANNTG)
<b>FID :</b>	Décroissance libre de l'induction (free induction decay)
<b>FPLC :</b>	Chromatographie rapide en phase liquide de protéines (Fast protein liquid chromatography)
<b>hNOE :</b>	Effet nucléaire Overhauser hétéronucléaire (Heteronuclear nuclear Overhauser effect)
<b>HPLC :</b>	Chromatographie en phase liquide à haute performance (High performance liquid chromatography)
<b>HSQC :</b>	Cohérence hétéronucléaire à simple quantum (Heteronuclear single-quantum coherence)
<b>Miz-1 :</b>	Protéine à doigts de zinc interagissant avec Myc (Myc-interacting zinc finger protein 1)
<b>NLS :</b>	Séquence signal de localisation nucléaire (Nuclear localization signal)
<b>NOE :</b>	Effet nucléaire Overhauser (Nuclear Overhauser Effect)
<b>NOESY :</b>	Spectroscopie d'effet nucléaire Overhauser (Nuclear Overhauser Effect Spectroscopy)
<b>RMN :</b>	Résonance magnétique nucléaire
<b>SMAD :</b>	Protéine analogue à la fois à la protéine MAD « mères contre la décapentaplégie » (« Mothers against decapentaplegic ») et à la protéine SMA « corps de petite taille » (« Small body size »); le nom de la protéine SMAD est la contraction des deux acronymes
<b>TGFβ :</b>	Facteur de croissance de transformation β (Transforming Growth Factor β)
<b>TOCSY :</b>	Spectroscopie de corrélation totale (Total Correlation Spectroscopy)
<b>ZF :</b>	Doigt de zinc (zinc finger)



## 1. Introduction

Ce mémoire présente les travaux que j'ai effectués lors de ma maîtrise au laboratoire du professeur Pierre Lavigne.

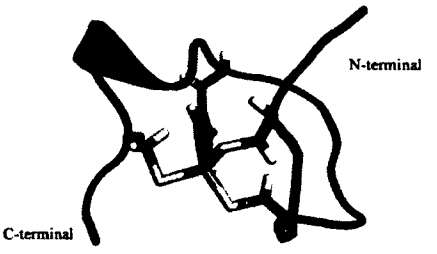
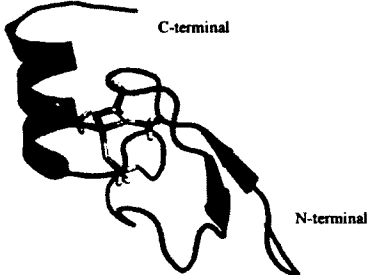
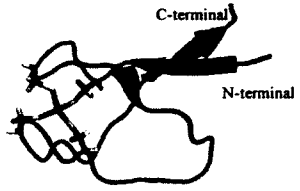
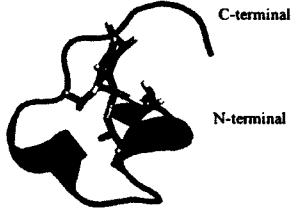
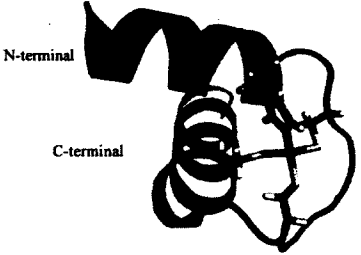
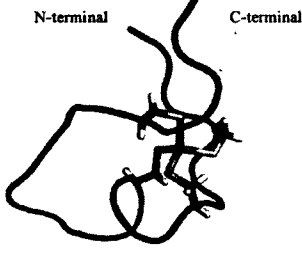

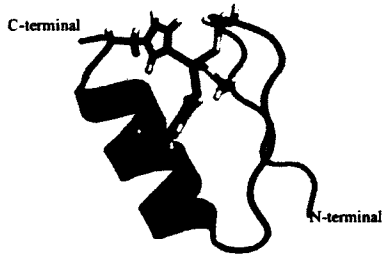
### 1.1 Protéines à doigts de zinc

#### 1.1.1 Facteurs de transcription et types de doigts de zinc

Il existe différentes superfamilles de facteurs de transcription, ces protéines qui peuvent se lier à l'ADN. On retrouve par exemple les protéines à b-HLH (« basic region – helix-loop-helix ») telle la protéine MyoD (MA et al., 1994), les protéines à LZ (« leucine zipper ») tel le dimère Fos-Jun (GLOVER et HARRISON, 1995), les b-HLH-LZ telle la protéine Max (SAUVÉ et al., 2004), de même que les protéines à doigts de zinc (ZF pour « zinc finger »). Le nom de chacune de ces familles réfère au domaine qui leur permet de se lier à l'ADN. Les protéines à ZFs peuvent être subdivisées en différentes classes, dépendamment du type de ZF qu'elles contiennent. Il en existe huit: les *gag knuckles*, les *treble clefs*, les *zinc ribbons*, les  $Zn_2/Cys_6$ , les *TAZ2 domain-like*, les *zinc-binding loops*, les *métallothionéines* et les *C2H2*, également notés  $Cys_2His_2$  (KRISHNA et al., 2003). Un exemple de chacun est présenté au tableau 1.

Il est à noter que les ZFs ont des fonctions assez diverses et ne sont aucunement limités à la liaison à l'ADN. En fait, le terme « ZF » désigne n'importe quel domaine protéique dont la structure est maintenue grâce à un ou plusieurs ions  $Zn^{2+}$ . En réalité, seuls les ZFs des classes treble clef,  $Zn_2/Cys_6$  et  $Cys_2His_2$  sont capables de se lier à l'ADN (KRISHNA et al., 2003). Parmi ces classes, celle des  $Cys_2His_2$  est de loin la plus étudiée, étant donné que près de la moitié des facteurs de transcription encodés dans le génome humain y appartiennent (EMERSON et THOMAS, 2009). C'est également à cette classe de ZFs qu'appartiennent les ZFs de la protéine étudiée dans ce mémoire, et par souci de concision, le terme ZF désignera implicitement des ZFs de type  $Cys_2His_2$  pour le reste de ce mémoire.

Tableau 1 : Classes de doigts de zinc

<p style="text-align: center;"><b>Gag knuckle</b></p>  <p style="text-align: center;">Identifiant PDB : 2L44 (QUINTAL et al., 2012)</p>	<p style="text-align: center;"><b>Treble clef</b></p>  <p style="text-align: center;">Identifiant PDB : 2W0T (LECHTENBERG et al., 2009)</p>
<p style="text-align: center;"><b>Zinc ribbon</b></p>  <p style="text-align: center;">Identifiant PDB : 2CON (DOI-KATAYAMA et al., 2005)</p>	<p style="text-align: center;"><b>Zn<sub>2</sub>/Cys<sub>6</sub></b></p>  <p style="text-align: center;">Identifiant PDB : 1CO4 (TURNER et al., 1998)</p>
<p style="text-align: center;"><b>TAZ2 domain-like</b></p>  <p style="text-align: center;">Identifiant PDB : 2KA6 (WOJCIAK et al., 2009)</p>	<p style="text-align: center;"><b>Zinc-binding loop</b></p>  <p style="text-align: center;">Identifiant PDB : 1HSO (NIEDERHUT et al., 2001)</p>
<p style="text-align: center;"><b>Métallothionéine</b></p>  <p style="text-align: center;">Identifiant PDB : 2KAK (PEROZA et al., 2009)</p>	<p style="text-align: center;"><b>Cys<sub>2</sub>His<sub>2</sub></b></p>  <p style="text-align: center;">Identifiant PDB : 2LVU (BÉDARD et al., 2012)</p>

## 1.1.2 Doigts de zinc de type Cys<sub>2</sub>His<sub>2</sub>

### 1.1.2.1 Caractéristiques structurales des doigts de zinc de type Cys<sub>2</sub>His<sub>2</sub>

En dépit de leur abondance, les ZFs ont un faible nombre de résidus conservés. Cette faible homologie est importante pour leur fonction de liaison à l'ADN, tel qu'il sera vu dans la section 1.1.2.2. La séquence consensus des ZFs est (Y/F)-X-C-X<sub>2.5</sub>-C-X<sub>3</sub>-(F/Y)-X<sub>5</sub>-Hph-X<sub>2</sub>-H-X<sub>3.5</sub>-H, où Hph représente un résidu hydrophobe, le plus fréquent étant une leucine (WOLFE et al., 2000). Une courte séquence désignée sous le nom d'espaceur (« linker ») comportant en général 5 acides aminés, le plus souvent TGEKP, relie deux ZFs consécutifs (CHOO et KLUG, 1993). Certaines mutations ponctuelles dans cet espaceur peuvent causer une perte d'affinité d'un facteur 10 à 100 pour l'ADN (WOLFE et al, 2000). Le motif adopte en général une structure  $\beta\beta\alpha$ , ou plus en détail un court feuillet  $\beta$  antiparallèle (formé des résidus (Y/F)-X-C-X<sub>2.5</sub>-C-X<sub>3</sub>-(F/Y)) suivi d'une hélice alpha (formée des résidus X<sub>3</sub>-Hph-X<sub>2</sub>-H-X<sub>3.5</sub>-H). Un exemple est présenté à la figure 1.

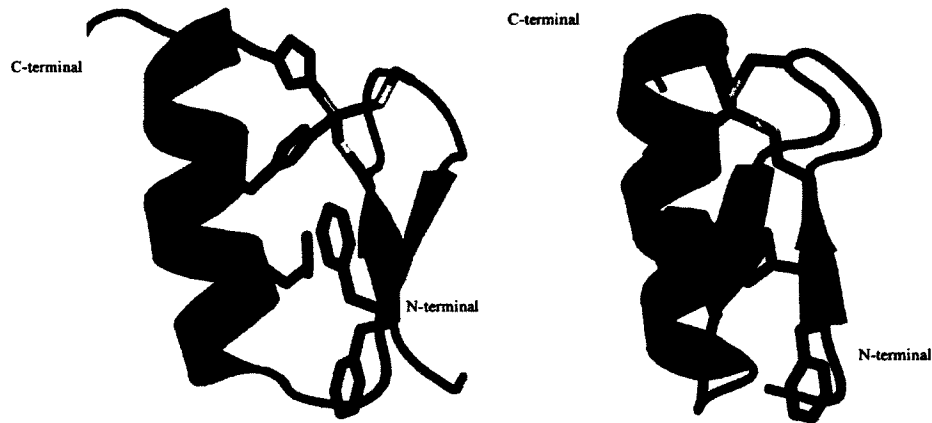


Figure 1 : Structure générale d'un doigt de zinc canonique

La structure présentée est celle du ZF 2 de EGR-1 (identifiant PDB 1ZAA; PAVLETICH et PABO, 1991). Les résidus représentés sont les résidus conservés, soit les deux cystéines et les deux histidines coordonnatrices de l'ion de zinc (représenté en gris), ainsi que les trois résidus du cœur hydrophobe. Les structures ont été réalisées à l'aide du logiciel PyMOL ([www.pymol.org](http://www.pymol.org)).

Les deux cystéines et les deux histidines sont les résidus les mieux conservés de la séquence consensus. Ce sont eux qui coordonnent l'ion  $Zn^{2+}$  maintenant la structure. La mutation de l'un de ces résidus empêche complètement le domaine de se replier, exception faite de la dernière histidine qui peut être remplacée par une cystéine (KRISHNA et al., 2003). Les trois autres résidus conservés forment un cœur hydrophobe qui apporte une stabilisation supplémentaire à la structure. Finalement, quatre résidus permettent la reconnaissance spécifique de bases d'ADN. Ces résidus sont situés aux positions -1, +2, +3 et +6 par rapport au premier résidu de l'hélice  $\alpha$  (il n'y a pas de position 0). D'autres résidus, situés à des positions variables, peuvent favoriser la liaison à l'ADN, mais de manière non spécifique, entre autres par des contacts avec les phosphates de la chaîne principale. Les résidus les plus communément impliqués dans ces interactions non spécifiques sont la première histidine coordonatrice de zinc (via son  $H_{N\delta}$ ), le premier résidu, souvent basique, du deuxième brin  $\beta$  (via les NH de la chaîne latérale), et la lysine qui se trouve en position 4 de l'espaceur (WOLFE et al., 2000).

### 1.1.2.2 Fonctions des doigts de zinc de type $Cys_2His_2$

Tel que mentionné précédemment, les ZFs de type  $Cys_2His_2$  forment le motif de liaison à l'ADN le plus courant. Certains ont également la capacité de se lier à l'ARN (BROWN, 2005) ou de médier des interactions protéine-protéine (BRAYER et SEGAL, 2008), mais ces fonctions ont été beaucoup moins étudiées que la liaison à l'ADN. Chaque ZF peut reconnaître de façon spécifique une séquence de 3 à 4 nucléotides consécutifs. En général, les ZFs se lient à l'ADN en trio, ce qui permet de reconnaître des séquences allant jusqu'à 10 nucléotides. Les ZFs en N-terminal et en C-terminal de la protéine interagissent respectivement avec les extrémités 3' et 5' de la séquence à laquelle ils se lient (DEL RIO et al., 1993). Lors de la liaison à l'ADN, les espaceurs, qui sont des séquences beaucoup plus flexibles que le reste du domaine ZF en l'absence d'ADN, adoptent une conformation stable et rigide (LAITY et al., 2000). Les résidus situés aux positions +6, +3 et -1 entrent en contact avec des bases consécutives d'un même brin d'ADN (qui peut être le brin codant ou non codant), et le résidu en position +2 entre en

contact avec la base complémentaire de la base située en 3' de celle contactée par le résidu en position -1 (ISALAN et al., 1997). Le tout est résumé par la figure 2.

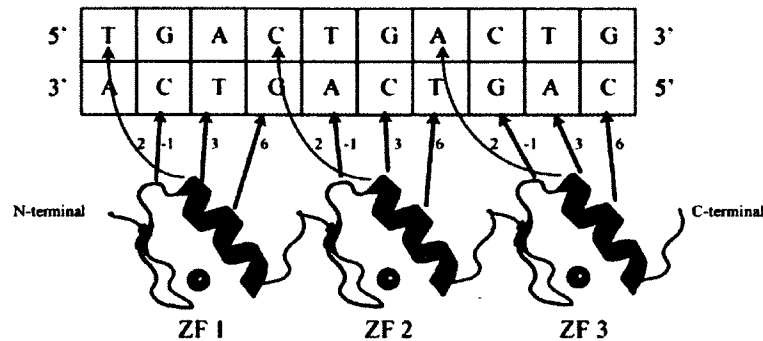


Figure 2 : Contacts habituels entre doigts de zinc et ADN

Ce schéma représente trois ZFs consécutifs (en vert) reliés à une séquence quelconque d'ADN. Les flèches correspondent au contact effectué entre le résidu situé à la position indiquée et le nucléotide. Les chiffres sont la position du résidu par rapport au premier résidu de l'hélice  $\alpha$  du ZF. Les structures ont été réalisées à l'aide du logiciel PyMOL ([www.pymol.org](http://www.pymol.org)).

L'homologie de séquence est faible d'un ZF à l'autre, particulièrement dans le cas des quatre résidus qui reconnaissent les bases de l'ADN (EMERSON et THOMAS, 2009), ce qui confère aux ZFs la possibilité de reconnaître quasiment n'importe quelle séquence de trois nucléotides. En effet, l'identité des nucléotides reconnus par les quatre résidus mentionnés ci-dessus est en partie dépendante de la nature du résidu et est spécifique à la position (-1, +2, +3 ou +6). Une arginine en position -1 ou +6 a une très forte affinité pour les guanines, alors qu'un aspartate en position +2 a une bonne affinité pour les cytosines, par exemple. Ceci a entraîné des chercheurs à établir un code de reconnaissance qui permettrait de prédire la séquence d'ADN reconnue par une protéine à ZFs quelconque (un exemple d'un tel code est présenté dans KAPLAN et al., 2005). Toutefois, d'autres facteurs encore inconnus entrent en ligne de compte, et jusqu'à présent il est difficile de prévoir avec certitude quelle séquence sera liée par un triplet de ZFs donné (WOLFE et al., 1999).

Finalement, bien que les ZFs se lient généralement par triplets à l'ADN, il existe de très nombreuses protéines possédant un grand nombre de ZFs, allant jusqu'à plus de 30 ZFs dans une seule protéine (rOAZ en possède 29, par exemple, voir TSAI et REED, 1998). Bien entendu, tous les ZFs ne se lient pas en même temps à l'ADN, en général

différents triplets seront utilisés, ce qui permet à une seule protéine de reconnaître plusieurs séquences (EMERSON et THOMAS, 2009). Il est toutefois possible que plusieurs triplets de ZFs d'une même protéine se lient au même gène en différents endroits. La protéine TFIIIA en est un excellent exemple (BROWN, 2005). Il est à noter que les protéines à poly-ZFs (plus de 3) sont apparues et se sont diversifiées relativement tard dans l'évolution : ces gènes sont les plus fréquents chez l'humain, et les mammifères en possèdent davantage que les autres animaux (EMERSON et THOMAS, 2009).

## **1.2 Myc-interacting zinc finger protein-1 (Miz-1)**

Ce mémoire s'intéresse au facteur de transcription à ZFs *Myc-interacting zinc finger protein-1* (Miz-1), également nommé ZBTB17. Afin de bien comprendre les enjeux autour de Miz-1, il est nécessaire de commencer par une courte description de la protéine de laquelle elle tire son nom, c'est-à-dire c-Myc.

### **1.2.1 La protéine c-Myc**

La protéine c-Myc est un facteur de transcription appartenant à la famille des b-HLH-LZ, tout comme son partenaire de dimérisation, la protéine Max. Cet hétérodimère reconnaît les séquences « enhancer box » (E-box) CANNTG, avec une préférence marquée pour les E-box canoniques CACGTG. Il est capable d'activer et de réprimer la transcription des gènes auxquels il se lie, qui se comptent par milliers (EILERS et EISENMAN, 2008).

#### **1.2.1.1 Activation de la transcription par c-Myc**

Le phénomène de transactivation par c-Myc est bien documenté (voir les revues par EILERS et EISENMAN, 2008; ou HERKERT et EILERS, 2010). L'hétérodimère c-Myc/Max se lie au promoteur d'un gène sur une E-box canonique, ce qui permet de recruter, via le long domaine TAD (« transactivation domain ») de c-Myc, différentes protéines audit promoteur, telles le cofacteur TRRAP ou l'histone acétyltransférase

hGCN5. Ceci permet l'ouverture de la chromatine et l'ARN polymérase peut venir se lier au promoteur (MCMAHON et al., 2000; MONTAGNE, 2008, p.11-17). La figure 3 résume ce processus.

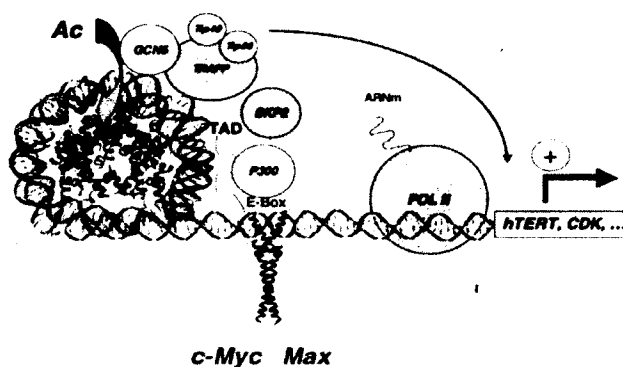


Figure 3 : Activation de la transcription par c-Myc/Max (MONTAGNE, 2008, p.15)

L'hétérodimère c-Myc/Max (en rouge et bleu, respectivement) se lie à une séquence E-box canonique CACGTG située dans le promoteur d'un gène transactivé par c-Myc. Le domaine de transactivation de c-Myc (TAD) recrute différentes protéines, telles TRRAP ou hGCN5, ce qui ouvre la chromatine et permet le recrutement de la polymérase. L'autorisation de reproduction de cette figure est en annexe.

Parmi les gènes régulés à la hausse par c-Myc, on retrouve les cyclines A2, B1, D2, D3 et E1, les kinases CDK1 et CDK4, de même que la protéine hTERT (ZELLER et al., 2003). En plus de ces gènes pro-prolifératifs, c-Myc régule à la hausse plusieurs gènes qui favorisent le métabolisme (MORRISH et al., 2008), de même que des ARNs ribosomiaux (GRANDORI et al., 2005). À noter que c-Myc n'active pas de gènes inactifs, il augmente simplement l'expression de gènes déjà actifs au niveau transcriptionnel, ce qui pourrait expliquer que les chercheurs ont eu tant de difficultés à tenter de définir un ensemble précis de gènes ciblés par Myc (LITTLEWOOD et al., 2012). Toutefois, malgré ce large éventail de gènes cibles, c-Myc est un mauvais activateur, ne causant une augmentation de l'expression de ces gènes que d'un facteur 2 au maximum (LOVÉN et al., 2012).

### 1.2.1.2 Répression de la transcription par c-Myc

En plus d'activer la transcription d'une multitude de gènes, c-Myc est capable de réprimer la transcription. Ce mécanisme est toutefois plus obscur, et plusieurs partenaires

peuvent être impliqués ou non pour la répression d'un gène en particulier (HERKERT et EILERS, 2010). Le mécanisme de transrépression par c-Myc le mieux décrit à ce jour est celui impliquant Miz-1 (PEUKERT et al., 1997). Cette protéine agit en général comme activateur de la transcription de gènes cyostatiques tels CDKN2B et CDKN1A (encodant respectivement p15<sup>INK4b</sup> et p21<sup>CIP1</sup>), mais lorsque c-Myc s'y associe, l'expression de ces gènes est au contraire réprimée (SEOANE et al., 2001). En se liant à Miz-1, c-Myc peut recruter différentes protéines qui inhibent l'expression d'un gène, telles l'histone désacétylase HDAC3, la méthyltransférase Dnmt3a ou alors la protéine Arf, qui causera la triméthylation des queues des histones (HERKERT et EILERS, 2010). De plus, la liaison de c-Myc à Miz-1 empêche Miz-1 de recruter p300 (STALLER et al., 2001) ainsi que la nucléophosmine (WANZEL et al., 2008), deux protéines nécessaires à la transactivation par Miz-1. Le tout est résumé par la figure 4.

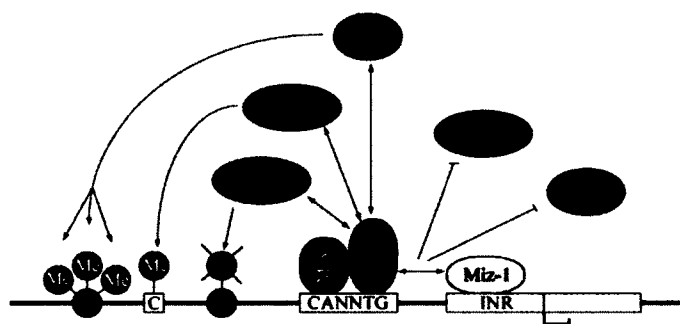


Figure 4 : Protéines impliquées dans la répression par c-Myc/Max et Miz-1

L'association entre Miz-1 (jaune) et l'hétérodimère c-Myc/Max (rouge et bleu) empêche le recrutement par Miz-1 de p300 et de la nucléophosmine (NPM) (orange). De plus, c-Myc peut recruter l'histone désacétylase (HDAC3), la méthyltransférase (Dnmt3a) ou la protéine Arf (toutes trois en cyan). Tout ceci inhibera les gènes reconnus par Miz-1.

La protéine c-Myc-S, une variante de c-Myc qui est incapable d'effectuer la transactivation mais qui conserve sa capacité à effectuer la transrépression, est capable d'induire un phénotype prolifératif chez des fibroblastes « myc-null », ce qui montre que la répression des gènes cyostatiques par c-Myc est un phénomène tout aussi important que l'activation des gènes pro-prolifératifs (COWLING et COLE, 2008). De plus, il est impossible de parler de la répression par l'interaction entre c-Myc et Miz-1 sans parler de MycV394D. Cette mutation du HLH de c-Myc empêche l'interaction avec Miz-1, mais conserve la capacité à hétérodimériser avec Max et à activer la transcription de gènes. La



seule inhibition de l'interaction entre c-Myc et Miz-1 freine énormément la prolifération cellulaire (HEROLD et al., 2002).

### 1.2.1.3 c-Myc dans un contexte pathophysiologique

L'activation et la répression de la transcription par c-Myc sont des phénomènes normaux, nécessaires à la croissance et la survie d'un organisme. Toutefois, il arrive que l'expression de Myc soit dérégulée, par exemple à cause de la substitution du gène encodant c-Myc vers un autre locus ayant un autre promoteur plus actif, tel celui de la chaîne lourde de l'immunoglobuline (TAUB et al, 1982; DALLA-FAVERA et al., 1982). Ceci rend Myc insensible à plusieurs formes de régulation, entre autre celle effectuée par le TGF $\beta$ . De plus, l'activation de gènes cytotostatiques par la voie des protéines SMAD n'a plus d'effet lors d'une surexpression de c-Myc (STALLER et al, 2001). Myc est alors surexprimé, ce qui cause une prolifération excessive des cellules dans lesquelles cette mutation est rencontrée, et c'est pourquoi c-Myc est un acteur important dans le cancer. Il est surexprimé dans environ 80% des cas de cancers du sein, 70% des cancers du côlon et du rectum, 90% des cancers gynécologiques, et 50% des carcinomes hépatocellulaires, pour ne nommer que ceux-là (ZELLER et al., 2003). Cette omniprésence à elle seule justifie les études servant à mieux comprendre ses mécanismes d'action.

### 1.2.2 Gènes régulés par Miz-1

Miz-1 est une protéine capable d'activer ou de réprimer la transcription de différents gènes, pour la plupart cytotostatiques. Ce sont les partenaires d'interaction de Miz-1 qui dicteront si un gène est activé ou réprimé. La littérature fait état de transactivation Miz-1-dépendante de gènes tels ceux encodant p15<sup>INK4b</sup> (STALLER et al., 2001), p21<sup>CIP1</sup> (SEOANE et al., 2002), WIF-1 (LICCHESI et al., 2010), ou Nramp1 (BOWEN et al., 2003). Lorsque c-Myc est associé à Miz-1, les gènes-cibles de Miz-1 sont réprimés. Il s'agit là d'un processus extrêmement important, au point où la délétion de Miz-1 est létale au stade embryonnaire à cause du manque de répression de p21<sup>CIP1</sup> (ADHIKARY et al., 2003). En plus des trois gènes mentionnés ci-haut, les gènes

encodant Mad4 (KIME et WRIGHT, 2003), p57<sup>KIP2</sup> (DAUPHINOT et al., 2001), C/EBP $\alpha$  (historiquement le premier gène pour lequel une régulation à la baisse par c-Myc a été découverte) (LI et al., 1994) et des protéines impliquées dans l'adhésion à la matrice extracellulaire telle l'intégrine  $\beta$ 1 (répression nécessaire à la différenciation des kératinocytes) (GEBHARDT et al., 2006) sont également réprimés par le complexe c-Myc/Miz-1, et il serait logique de penser que Miz-1 pourrait activer leur transcription en l'absence de c-Myc, bien que ce ne soit pas explicitement mentionné dans la littérature.

### 1.2.3 Caractéristiques de Miz-1

Miz-1 est un facteur de transcription de 803 acides aminés et de 93 kDa de la famille des facteurs de transcription à domaine BTB/POZ. Son N-terminal est constitué d'un domaine « BR-C, ttk and bab », également appelé « Pox virus and zinc finger », ou plus simplement BTB/POZ. La structure de ce domaine a été déterminée par cristallographie par le groupe de Stogios. Il s'agit d'un domaine qui permet des

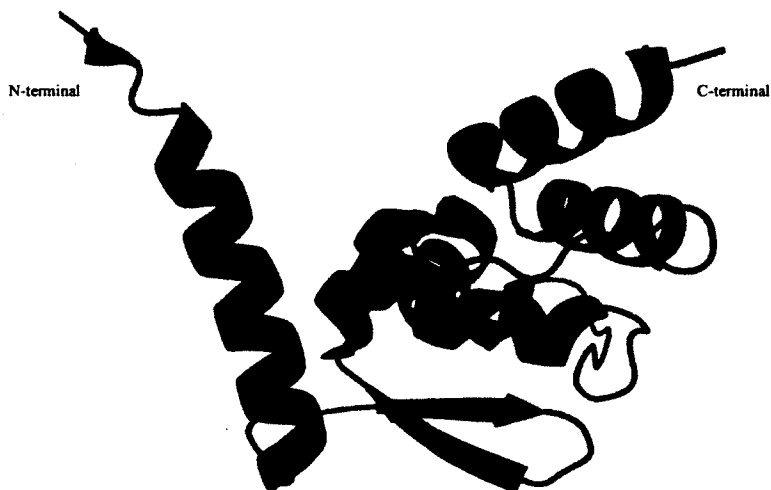


Figure 5 : Structure du domaine BTB/POZ de Miz-1

Structure obtenue par cristallographie d'un monomère du domaine BTB/POZ de Miz-1 (identifiant PDB 3M52; STOGIOS et al., 2010). Ce domaine se retrouve sous forme de dimère dans la structure cristalline. Il forme la portion N-terminale de la protéine Miz-1. La structure a été réalisée à l'aide du logiciel PyMOL ([www.pymol.org](http://www.pymol.org)).

interactions protéine-protéine, par exemple l'homodimérisation (STOGIOS et al., 2010) ou l'interaction avec la nucléophosmine (WANZEL et al., 2008). La structure de ce domaine est présentée à la figure 5.

Le domaine C-terminal de Miz-1 est constitué de 13 ZFs, dont 12 consécutifs. La structure des ZFs 8 à 10 a été déterminée par notre laboratoire récemment (BÉDARD et al., 2012) et est présentée à la figure 6, mais la structure des autres ZFs est toujours inconnue.

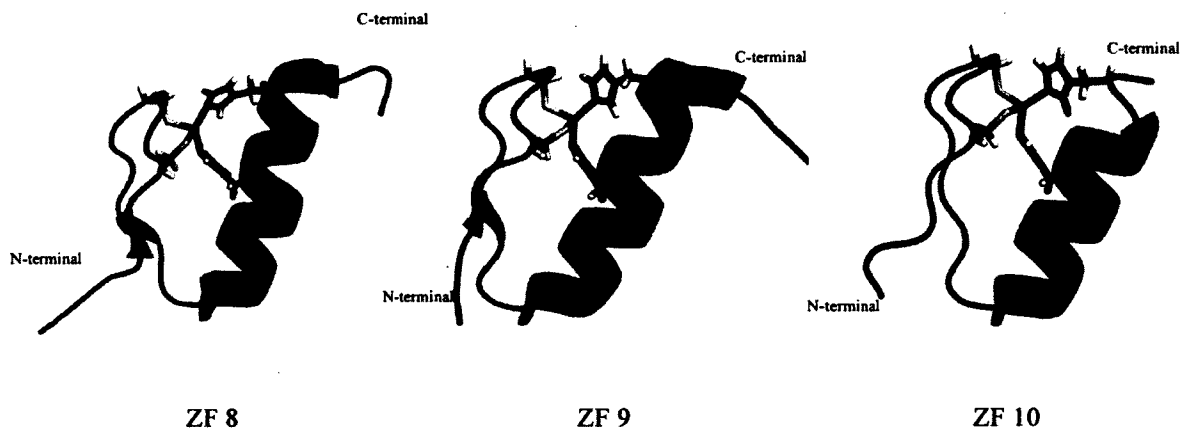


Figure 6 : Structure des ZFs 8 à 10 de Miz-1

Structures des ZFs individuels 8 (identifiant PDB 2LVR), 9 (identifiant PDB 2LVT) et 10 (identifiant PDB 2LVU) (BÉDARD et al., 2012). Dans les trois cas, les seuls résidus représentés sont les deux cystéines et deux histidines coordonatrices de l'ion de zinc (en gris). Les structures ont été réalisées à l'aide du logiciel PyMOL ([www.pymol.org](http://www.pymol.org)).

Le 13<sup>e</sup> ZF est séparé du 12<sup>e</sup> par un domaine de structure indéterminée, et un autre domaine de structure indéterminée sépare le domaine BTB/POZ du premier ZF. Ces deux domaines sans structure déterminée ont été nommés *Myc-interacting domains* (MIDs), le groupe de Peukert ayant déterminé ces séquences comme étant celles responsables de l'interaction entre c-Myc et Miz-1, et qui permettent de se lier au domaine HLH de c-Myc (PEUKERT et al., 2001). Toutefois, des résultats obtenus au laboratoire par spectroscopie de dichroïsme circulaire (CD) et de résonance magnétique nucléaire (RMN) pointent vers une interaction impliquant les quatre premiers ZFs (Mikaël Bédard, communication personnelle). La figure 7 présente la séquence complète de Miz-1 (banque de données UNIPROT).

10            20            30            40            50            60            70            80  
**MDFPQHSQEV LEQLMQQRQL GLLCDCTFVV DGVHFKAKKA VLAACSEYFK MLFVDQKDVV HLDISHAAGL GQVLEPFMYA**  
90            100            110            120            130            140            150            160  
**KLSLSPENVD DVLAVATFLQ MQDIITACHA LKSLAEPATS PGGNAEALAT EGGDKRAKEE KVATSTLSRL EQAGRSTPIG**  
170            180            190            200            210            220            230            240  
**PSRDLKEERG GOAQAASGA EQTEKADAPR EPPPVELKPD PTSGMAAAEA EAALSESSEQ EMEVEPARKG EEEQKEQEEQ**  
250            260            270            280            290            300            310            320  
**EEEGAGPAEV KEEGSQLENG EAPEENENEE SAGTDSGQEL GSEARGLRSG TYGDRTESKA YGSVI**  
330            340            350            360            370            380            390            400  
**SGEKP**  
410            420            430            440            450            460            470            480  
**SGEKP**  
490            500            510            520            530            540            550            560  
**SGEK P**  
570            580            590            600            610            620            630            640  
**DNIRP**  
650            660            670            680            690            700            710            720  
**AGIKILEPEE GSEVSVTVTD DMVTLATEAL AATAVTQLTV VPVGAAVTAD ETEVLKAEIS KAVKQVQEED PNTHIL**  
730            740            750            760            770            780            790            800  
**T AQALVMFQTD ADFYQQYGPG GTWPAGQVLQ AGELVFRPRD GAEGQPALAE TSPTAPECPP**  
803  
**PAE**

Figure 7 : Séquence complète de Miz-1

Séquence primaire complète de la protéine Miz-1 tirée de la banque de données UNIPROT. Chacun des acides aminés est représenté par le code à une lettre classique. Les nombres indiquent la position dans la séquence de l'acide aminé situé sous le chiffre souligné. Le domaine BTB/POZ est surligné en jaune, et chacun des ZFs est surligné en cyan.

#### 1.2.4 Activation de gènes par Miz-1

À la base, Miz-1 est un activateur de la transcription. Miz-1 ne possède pas de « nuclear localization sequence » (séquence NLS), et peut être co-transportée au noyau grâce à c-Myc (PEUKERT et al., 1997). La voie du TGF $\beta$  est l'une des voies permettant l'activation de Miz-1, ce qu'elle fait via l'activation des protéines SMAD. En effet, SMAD3 et SMAD4 peuvent se lier directement aux ZFs 1 à 4 de Miz-1 (SEOANE et al., 2001). Parallèlement, la nucléophosmine (NPM1), une protéine généralement séquestrée dans les nucléoles par la protéine ribosomale L23, sera relâchée dans le nucléoplasme lors d'une déplétion de L23. La nucléophosmine peut alors se lier au domaine BTB/POZ de Miz-1, ce qui est une condition d'activation de cette dernière (WANZEL et al., 2008). Miz-1 est incapable d'activer la transcription en l'absence de Sp1, un autre facteur de transcription à ZFs. Miz-1 et Sp1 n'interagissent pas ensemble, mais doivent tous deux se

lier au même promoteur afin de recruter le complexe coactivateur p300/CBP, ce qui permet l'initiation de la transcription (BOWEN et al., 2003). Les domaines formés des acides aminés 190-284 et 683-715 de Miz-1 se lient au domaine TAZ2 de p300 (STALLER et al., 2001). Tout ceci est résumé par la figure 8. Il est à noter qu'alors que la séquence reconnue par Sp1 est connue (les « GC boxes », BOWEN et al., 2003), la séquence reconnue par Miz-1, elle, ne l'est pas. Miz-1 reconnaît les séquences -105/-81 et -46/-32 sur le promoteur de p21<sup>CIP1</sup> ainsi que -155/-140 et -1/+13 sur le promoteur de p15<sup>INK4b</sup>, mais ces séquences n'ont aucune homologie entre elles, ce qui laisse supposer l'implication de différents sous-ensembles de ZFs de Miz-1 pour la liaison de ces promoteurs (SEOANE et al., 2002). Un autre article fait plutôt état que les sites de liaison de Miz-1 au promoteur de p21<sup>CIP1</sup> seraient plutôt les séquences -2267/-2237, -1384/-1354, -90/-64, -55/-63 et -27/-17 (LEE et al., 2011).

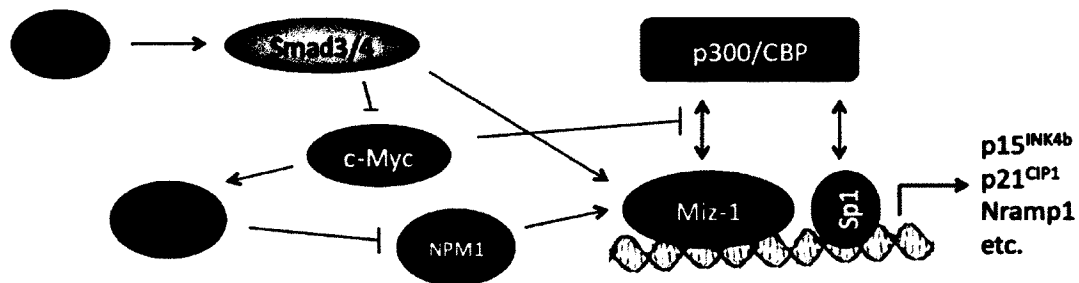


Figure 8 : Protéines impliquées dans la transactivation par Miz-1

Le TGFβ (gris-bleu) active la voie des SMADs (jaune). Celles-ci, en plus d'activer directement Miz-1 (vert) en s'y liant, inhibent c-Myc (rouge), ce qui empêche ce dernier d'empêcher l'interaction Miz-1/p300 (violet) et d'activer la protéine ribosomale L23 (brun), un inhibiteur de la nucléophosmine (NPM1, bleu), un activateur nécessaire à l'activité de Miz-1. Miz-1 et Sp1 (gris) co-recrutent p300, ce qui permet l'initiation de la transcription.

### 1.2.5 Répression de gènes par Miz-1

Le mécanisme de transrépression de Miz-1 le mieux décrit implique l'association de c-Myc à Miz-1. Les régions comprenant les acides aminés 269-308 et 637-718 de Miz-1 sont responsables de l'interaction avec c-Myc selon la banque de données UNIPROT (bien que, tel que mentionné plus tôt, des résultats obtenus au laboratoire laissent croire que les ZFs 1 à 4 soient impliqués). Ceci signifie que c-Myc inhibe compétitivement l'interaction de Miz-1 avec p300, puisque cette dernière se lie aux mêmes régions

(STALLER et al., 2001). Suite à cette interaction, c-Myc recrute les protéines mentionnées dans la section 1.2.1.2, entraînant la répression des gènes-cibles de Miz-1. Il est en outre à noter que les promoteurs des gènes réprimés par l'association c-Myc/Miz-1 présentent généralement une ou plusieurs séquences E-box non-canoniques (CAGCTG), ce qui laisserait croire que l'hétérodimère c-Myc/Max pourrait s'y lier directement. Il n'a toutefois pas été rapporté dans la littérature si cette liaison se produit, ou si elle est nécessaire au phénomène de transrépression par c-Myc/Miz, mais cette liaison est probable, puisque l'hétérodimère c-Myc/Max ne favorise les E-box canoniques par rapport aux non-canoniques que par un facteur 10 dans sa constante de dissociation (BEAULIEU, 2011, p.90-114).

Néanmoins, c-Myc n'a pas l'exclusivité de la transrépression effectuée par Miz-1. La protéine BCL6, un autre facteur de transcription à ZFs possédant un domaine BTB/POZ en N-terminal, peut s'associer à Miz-1 via leur domaine BTB/POZ pour réprimer le gène de  $p21^{CIP1}$  de façon c-Myc-indépendante (PHAN et al., 2005). Une troisième protéine à ZFs possédant un domaine BTB/POZ en N-terminal, KR-POK, qui est également appelée ZBTB36, peut causer l'inhibition du gène de  $p21^{CIP1}$  en s'associant à Miz-1, mais KR-POK peut également causer cette répression en l'absence de Miz-1 (LEE et al., 2011). Gfi-1, un autre facteur de transcription à ZFs, peut interagir avec Miz-

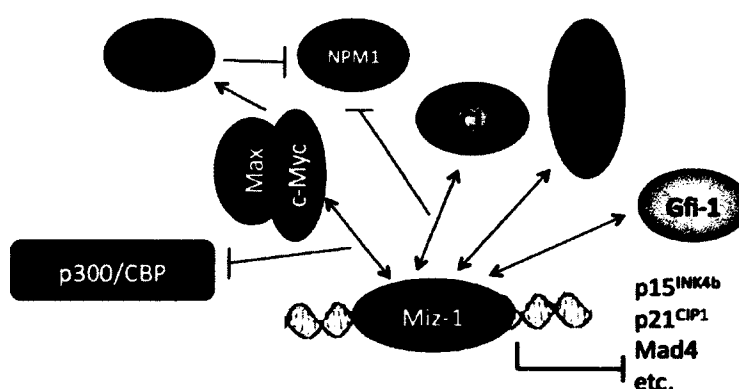


Figure 9 : Protéines impliquées dans la transrépression par Miz-1

Plusieurs interactions entre Miz-1 et un autre partenaire peuvent causer une répression de la transcription. L'interaction entre Miz-1 (vert) et Gfi-1 (jaune), KR-POK (rose) ou BCL6 (orange), trois protéines à ZFs, cause l'inhibition de la transcription de certains gènes-cibles de Miz-1. BCL6 agit en inhibant la nucléophosmine (NPM1, bleu). L'interaction entre Miz-1 et c-Myc/Max (rouge et bleu) inhibe le recrutement de p300 (violet) et active l'inhibiteur de la nucléophosmine, L23 (brun).

1 pour réprimer les gènes encodant p15<sup>INK4b</sup> et p27<sup>KIP1</sup>, répression qui est possible en présence ou en absence de c-Myc. Gfi-1 et c-Myc n'interagissent pas directement ensemble, mais plutôt se lient à des régions différentes de Miz-1. La répression d'un gène par Gfi-1/Miz-1/c-Myc est plus importante que celle effectuée par Miz-1 avec un seul de ces deux partenaires d'interaction (BASU et al., 2009). Ce ne sont là que des exemples, et le tout est résumé par la figure 9.

## **2. Résonance magnétique nucléaire**

Il existe trois méthodes pour déterminer la structure tertiaire d'une protéine : la diffractométrie de rayons X, également appelée cristallographie, la diffraction des électrons, qui est très semblable à la cristallographie, et la résonance magnétique nucléaire (RMN). Dans le cadre de ma maîtrise, c'est la RMN qui a été l'outil de prédilection pour déterminer la structure de Miz 5-8, puisque la RMN permet d'étudier des échantillons en solution dans l'eau plutôt que figés dans un cristal. Elle permet aussi d'étudier la dynamique des échantillons étudiés, dynamique qui s'est avérée être un sujet très central dans l'article étant donné l'omniprésence de mouvements importants dans la protéine étudiée.

### **2.1 Concepts de base de la RMN**

Ce chapitre a pour but d'expliquer la théorie et les concepts de base de cette puissante technique qu'est la RMN. Ce chapitre vise à démystifier la physique derrière cet appareil. Il est à noter que tous les symboles en gras correspondent à des vecteurs.

#### **2.1.1 Moment cinétique intrinsèque (spin)**

(GRIFFITHS, 2005, p.171-189; COHEN-TANNOUDJI et al., 1998, p.386-403; COHEN-TANNOUDJI et al., 2007, p.957-970)

Afin de bien comprendre comment fonctionne la résonance magnétique nucléaire, il est important de comprendre ce qu'elle mesure exactement, soit le moment cinétique

intrinsèque du noyau, également appelé son *spin* (c'est également la même chose, à un facteur constant près, que son moment magnétique intrinsèque  $\mu$ ). Plusieurs manuels de référence sur la RMN (SILVERSTEIN *et al.*, 2005, p.127; LAMBERT et MAZZOLA, 2004, p.1-2) décrivent ce phénomène selon l'explication donnée par la mécanique classique, c'est-à-dire qu'en tournant sur lui-même, l'électron (ou le proton, ou le noyau), comme toute charge en rotation, produit un moment magnétique de par la sommation du moment cinétique orbital de chaque point le constituant (comme une boule tournant sur elle-même). Cette explication est toutefois fautive. Si elle était vraie, cela impliquerait que la vitesse de « l'équateur » de l'électron soit supérieure à la vitesse de la lumière. Ce calcul est assez simple à effectuer, avec un moment cinétique  $I\omega = \frac{2}{5} m_e r_c^2 \cdot \frac{v_{\text{eq}}}{r_c}$  égal à  $\frac{\hbar}{2}$  et un rayon classique de l'électron égal à  $r_c = \frac{e^2}{4\pi\epsilon_0 m_e c^2}$ , une vitesse tangentielle d'environ  $171c$  est obtenue;  $\hbar$  est la constante de Planck divisée par  $2\pi$ ,  $e$  la charge électronique,  $\epsilon_0$  la permittivité électrique du vide,  $m_e$  la masse électronique,  $c$  la vitesse de la lumière dans le vide,  $I$  le moment d'inertie et  $\omega$  la fréquence angulaire.

Quantiquement, l'électron est une particule ponctuelle (donc sans volume) et ne peut donc pas tourner sur lui-même. L'autre explication possible est que le spin est tout simplement un moment cinétique *intrinsèque*, c'est-à-dire une propriété caractéristique de l'électron, tout comme sa charge ou sa masse, et n'a aucun équivalent en physique classique. Bien que cette contre-justification ne s'applique pas au proton (le même calcul donne environ  $0.3c$ , et le proton possède bel et bien un volume) ou au noyau atomique, la nature du spin du proton (et du noyau) est la même que celle de l'électron, et ce n'est donc pas simplement la rotation de la particule sur elle-même qui en est à l'origine.

Le spin (à tout le moins l'un de ses effets) a été découvert par Otto Stern et Walther Gerlach en 1922. À l'aide du montage illustré à la figure 10, ils ont fait passer un faisceau d'atomes d'argent neutres à travers un champ magnétique fortement inhomogène. Les atomes étant neutres, la déviation de ce faisceau n'est pas due au



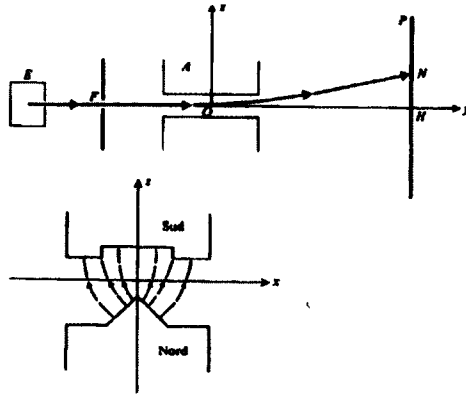


Figure 10 : Schéma de l'expérience de Stern et Gerlach (COHEN-TANNOUDJI et al., 1998, p.386)

La source E émet un faisceau d'atomes d'argent neutres qui passe à travers le focaliseur F. Le faisceau passe ensuite entre les enceintes de l'aimant A, qui produisent un champ magnétique fortement inhomogène, représenté au bas de la figure. Le faisceau est alors dévié et poursuit sa trajectoire avant de frapper l'écran P. L'autorisation de reproduction de cette figure est en annexe.

déplacement de charges dans un champ magnétique, mais bien aux propriétés magnétiques des atomes uniquement. En fait, le champ magnétique dans cette situation particulière n'affectera qu'un seul élément : l'électron non apparié situé dans l'orbitale 5s de l'atome. En étant dans une orbitale s, cet électron n'a aucun moment magnétique orbital (nombre quantique  $l = 0$ ). Il ne reste donc que le moment magnétique intrinsèque de cet électron qui peut jouer un rôle. Selon la physique classique, la force exercée par le champ magnétique sera égale à  $\mu_z \nabla B_z$ , où  $\mu$  est le moment magnétique et  $B$  le champ magnétique ambiant. Cette force est responsable de la déviation du faisceau, et donc la grandeur de la déviation  $y$  est proportionnelle. Comme  $\mu_z$  devrait pouvoir prendre



Figure 11 : Résultat de l'expérience de Stern et Gerlach (COHEN-TANNOUDJI et al., 1998, p.389)

La ligne verticale correspond à l'écran P de la figure 10. Les autres lignes représentent la probabilité qu'un atome d'argent ait frappé l'écran à cette coordonnée. La ligne pointillée est le résultat prédit par la mécanique classique, et la ligne pleine formant deux cloches est le résultat obtenu. L'autorisation de reproduction de cette figure est en annexe.

n'importe quelle valeur entre  $-\mu$  et  $+\mu$ , le faisceau devrait pouvoir être dévié de façon à frapper l'écran n'importe où entre les points  $N_1$  et  $N_2$  avec une probabilité égale et ainsi former une seule tache (ligne pointillée sur la figure 11).

Ce n'est pas ce qui est observé : en fait, on retrouve deux taches, l'une à  $N_1$  et l'autre à  $N_2$  (ligne pleine sur la figure 11). La physique quantique permet d'expliquer ceci : le moment magnétique intrinsèque de l'électron a uniquement deux valeurs possibles ( $\frac{\hbar}{2}$  et  $-\frac{\hbar}{2}$ ), au lieu de la vision classique selon laquelle il s'agit d'un vecteur pouvant adopter n'importe quelle orientation et donc n'importe quelle valeur de composante en  $z$ . Le spin est une valeur quantifiée.

### 2.1.2 Les niveaux d'énergie de spin

(Griffiths, 2005, p.171-189)

Tel que mentionné plus tôt, la différence entre le moment cinétique intrinsèque  $S$  et le moment magnétique intrinsèque  $\mu$  n'est qu'un facteur constant. Ce facteur se nomme le rapport gyromagnétique (ou magnétogyrique) et est défini ainsi :  $\mu = \gamma S = \frac{gq}{2m} S$  où  $g$  est le facteur de Landé (propre à chaque particule),  $q$  est la charge et  $m$  est la masse. Chaque particule a donc un rapport gyromagnétique qui lui est propre. Celui de l'électron, par exemple, est d'environ  $176 \cdot 10^9 \text{ rad} \cdot \text{s}^{-1} \cdot \text{T}^{-1}$  et celui du proton est d'environ  $270 \cdot 10^6 \text{ rad} \cdot \text{s}^{-1} \cdot \text{T}^{-1}$ .

Le spin est à la source de différents niveaux d'énergie. En temps normal, ces niveaux sont dégénérés (c'est-à-dire impossibles à discriminer, ou tout simplement d'énergie égale). Ces niveaux d'énergie se séparent lorsque les spins sont plongés dans un champ magnétique externe. Cette séparation provient du fait que les moments magnétiques ont tendance à s'aligner avec le champ magnétique – de façon parallèle pour les spins positifs, et de façon antiparallèle pour les spins négatifs. La valeur de ces niveaux d'énergie sera de  $E = -\gamma B_0 \hbar m_s$ , où  $m_s$  est un nombre variant entre  $-S$  et  $+S$  par intervalles de 1. Si  $S$  est entier,  $m_s$  devra être entier, et si  $S$  est demi-entier,  $m_s$  ne pourra

prendre que des valeurs demi-entières ( $\pm 1/2, \pm 3/2 \dots$ ). Une particule de spin nul n'a donc qu'un seul niveau d'énergie de spin (voir la droite verte sur la figure 12). Une particule de spin  $1/2$  aura deux niveaux d'énergie (droites bleues), une particule de spin 1 en aura 3 (droites rouges) et une particule de spin  $S$  aura  $2S+1$  niveaux d'énergie de spin.

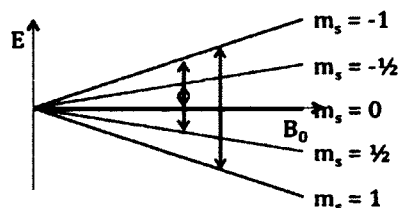


Figure 12 : Niveaux d'énergie de spin dans un champ magnétique

Ce graphique montre la séparation des niveaux d'énergie  $E$  de spin en fonction de l'intensité  $B_0$  du champ magnétique externe pour une particule de spin 0 (vert), de spin  $1/2$  (bleu) ou de spin 1 (rouge). Les flèches représentent des transitions possibles entre les niveaux. Le nombre quantique  $m_s$  représente un niveau d'énergie de spin particulier.

Ces niveaux d'énergie ne sont pas détectables en tant que tels, ce sont en fait les transitions des particules entre ces niveaux d'énergie qui sont visibles. Ceci signifie que les particules de spin nul, comme le  $^{12}\text{C}$ , sont complètement invisibles en RMN. Les particules de spin 1 (comme le  $^{14}\text{N}$ ) ou supérieur à 1 présentent plus d'une transition possible, mais elles présentent aussi des effets quadripolaires électriques, qui vont causer un élargissement des résonances et un raccourcissement important des périodes de relaxation, ce qui diminue grandement l'intérêt de les étudier par RMN (LAMBERT et MAZZOLA, 2004, p.22). Il ne reste donc que les particules de spin  $1/2$ , par exemple le proton ( $^1\text{H}$ ), le  $^{13}\text{C}$ , le  $^{15}\text{N}$ , le  $^{19}\text{F}$  ou le  $^{23}\text{Na}$ , qui ont la possibilité de donner des informations intéressantes et intelligibles par RMN. Il est à noter que tous les noyaux étudiés en RMN sont des isotopes stables, non radioactifs, contrairement à la croyance populaire.

Il est facile de calculer la valeur moyenne de  $S$  d'une particule plongée dans un champ magnétique  $B_0 \hat{z}$  à l'aide de la mécanique quantique. Néanmoins, détailler ce calcul requerrait la définition de plusieurs notions telles que le Hamiltonien, les matrices de Pauli ou les opérateurs up et down. C'est pourquoi seul le résultat est présenté ici :

$$\langle \mathbf{S} \rangle = \begin{pmatrix} \langle S_x \rangle \\ \langle S_y \rangle \\ \langle S_z \rangle \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} \sin \theta \cos \gamma B_0 t \\ \sin \theta \sin \gamma B_0 t \\ \cos \theta \end{pmatrix}$$

La constante  $\theta$  est simplement liée à la probabilité du spin d'être dans l'état  $+1/2$  ou  $-1/2$ , et  $t$  est le temps. Donc, le résultat obtenu est que  $\mathbf{S}$  (et  $\boldsymbol{\mu}$ ) est un vecteur qui a une trajectoire circulaire appelée *précession* dans le plan  $xy$ , soit autour du champ magnétique. En effet, la composante temporelle en  $x$  est cosinusoidale, celle en  $y$  est sinusoidale, et il n'y a aucune dépendance temporelle en  $z$ . Ceci montre également que cette précession s'effectue à une fréquence angulaire  $\omega_0 = \gamma B_0$ . Ce phénomène est appelé la *précession de Larmor*, et  $\omega_0$  est nommée *fréquence de Larmor*.

Tel que mentionné ci haut, ce sont les transitions entre les niveaux d'énergie qui sont sensibles aux irradiations utilisées en RMN. La différence entre deux niveaux d'énergie consécutifs est égale à  $\gamma B_0 \hbar$ . Il est donc possible d'exciter un noyau (le faire passer de l'état moins énergétique  $+1/2$  à l'état plus énergétique  $-1/2$ , également respectivement dénommés l'état  $\alpha$  et  $\beta$  dans certains livres) en l'irradiant avec une radiofréquence ayant la même énergie que cette transition. Comme l'énergie peut s'écrire sous la forme  $\hbar\omega$ , ceci signifie que la radiofréquence utilisée doit avoir une fréquence angulaire  $\omega = \gamma B_0$ , soit la même fréquence que celle à laquelle précessent les spins.

Avant de terminer cette section, effectuons une petite parenthèse pour introduire la séparation des niveaux d'énergie dans un système aux spins *couplés*, c'est-à-dire des spins voisins qui interagissent entre eux. Un système à plusieurs spins  $1/2$  possède bien deux niveaux d'énergie par spin, mais ces niveaux ne sont pas situés à des valeurs aléatoires l'un par rapport à l'autre à cause du couplage. Prenons l'exemple le plus simple, soit un système à deux spins  $S$  et  $I$ . Dans un tel système couplé, il y aura quatre niveaux d'énergie,  $\alpha_S\alpha_I$ ,  $\alpha_S\beta_I$ ,  $\beta_S\alpha_I$  et  $\beta_S\beta_I$ . Il existe dans ce système un total de six transitions énergétiques possibles. La transition  $\alpha \leftrightarrow \beta$  du spin  $S$ , qui a une énergie égale à  $\hbar\omega_S$ , peut s'effectuer entre les états  $\alpha_S\alpha_I$  et  $\beta_S\alpha_I$ , ou entre les états  $\alpha_S\beta_I$  et  $\beta_S\beta_I$ . Il en est de même pour la transition  $\alpha \leftrightarrow \beta$  du spin  $I$ , qui a une énergie égale à  $\hbar\omega_I$ . Dans ces quatre cas, un

seul spin effectue une transition énergétique, et on appelle ce type de transition une *transition à un quantum* (« single quantum »). Il est également possible que les deux spins changent d'état simultanément. La transition  $\alpha_S\alpha_I \leftrightarrow \beta_S\beta_I$  est appelée *transition double quantum* et a une valeur énergétique égale à la somme de l'énergie des deux transitions à un quantum, soit  $\hbar(\omega_S + \omega_I)$ . La transition  $\alpha_S\beta_I \leftrightarrow \beta_S\alpha_I$ , elle, est appelée *transition zéro quantum*, et sa valeur énergétique est égale à la différence de l'énergie des deux transitions à un quantum, soit  $|\hbar(\omega_S - \omega_I)|$ . Le tout est résumé à la figure 13.

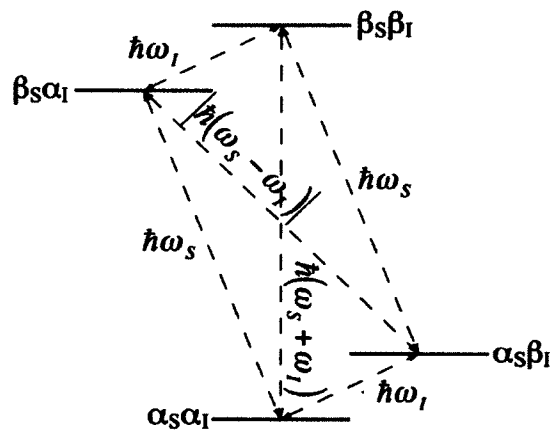


Figure 13 : Séparation des niveaux d'énergie dans un système de deux spins couplés

Les niveaux d'énergie sont représentés par les lignes horizontales. Leur valeur énergétique est plus élevée en montant dans la figure. L'état des deux spins couplés S et I est indiqué à côté du niveau. Les flèches pointillées indiquent les transitions possibles, et la différence énergétique entre les deux niveaux impliqués est indiquée à côté de celles-ci.

### 2.1.3 La détection de la magnétisation et le déplacement chimique

(LAMBERT et MAZZOLA, 2004, p.5-13)

Pour un échantillon contenant une grande quantité de spins dans un champ magnétique, il y a à la base un léger excès de spins dans l'état  $+1/2$  par rapport au nombre de spins dans l'état  $-1/2$  (puisque ce dernier est plus énergétique que le premier), ce qui donne lieu à une *magnétisation* nette  $M$ , parallèle au champ magnétique externe, à l'état d'équilibre (voir figure 14, à gauche). La magnétisation est définie comme la sommation de l'ensemble des moments magnétiques intrinsèques. Puisqu'il n'y a qu'un léger excès de spins dans l'état  $+1/2$ , cette magnétisation est très faible par rapport au champ magnétique externe, elle n'est pas détectable telle quelle. En irradiant l'échantillon avec la

radiofréquence  $\omega_0$ , des spins  $+1/2$  sont excités dans l'état  $-1/2$ , et en irradiant pendant un temps approprié, assez de spins (soit la moitié de l'excès de spin dans l'état  $+1/2$ ) changeront de niveau pour faire basculer la magnétisation jusqu'à ce qu'elle se retrouve dans le plan xy (voir figure 14, à droite). Cette irradiation est appelée un *pulse 90°*, puisque la magnétisation bascule de  $90^\circ$ . C'est l'intensité de cette magnétisation transverse qui est détectée par un spectromètre RMN. Il est à noter qu'immédiatement après un pulse  $90^\circ$ , tous les spins se retrouvent *en phase*, c'est-à-dire que tous les moments magnétiques sont au même angle dans le plan xy, représentés par des flèches plus épaisses à la figure 14. De plus, il est à noter qu'il est possible de faire basculer la magnétisation selon n'importe quel angle en excitant la fraction appropriée de spins.

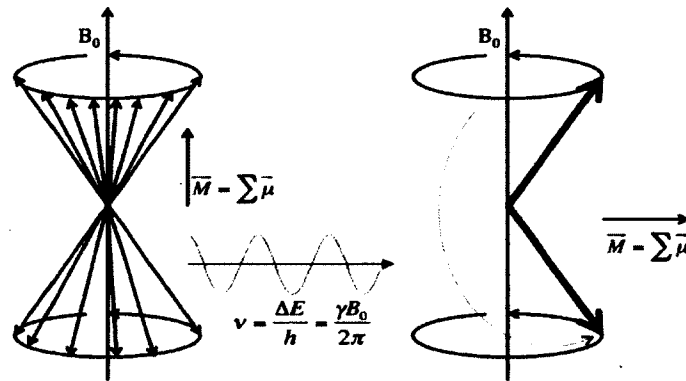


Figure 14 : Moments magnétiques, magnétisation et pulse  $90^\circ$

L'état d'équilibre d'un échantillon à plusieurs spins plongé dans un champ magnétique externe  $B_0$  est représenté à gauche. Chacune des flèches droites originaires du centre représente un spin individuel, et les flèches circulaires montrent le sens de précession des spins. La somme des spins  $\mu$  donne la magnétisation  $M$ , parallèle au champ magnétique externe. L'application d'un pulse  $90^\circ$  (onde sinusoïdale) excite les spins, ceux-ci se mettent en phase, et la magnétisation se retrouve perpendiculaire au champ magnétique externe (à droite).

Une fois dans le plan xy, chaque moment magnétique poursuivra sa précession autour du champ magnétique externe, et la magnétisation nette, composée de ces moments magnétiques, se mettra elle aussi à décrire un mouvement de précession tel que mentionné plus tôt. Dépendamment de la force du champ magnétique, cette précession se fait à plusieurs centaines de mégahertz, et il est difficile d'imaginer le comportement de la magnétisation dans ces conditions. C'est pourquoi, au lieu d'utiliser un système d'axes (également appelé *référentiel*) xyz fixes, on observe le comportement de la magnétisation dans un référentiel dont les axes x et y tournent à la fréquence de référence de l'appareil

utilisé. Dans ce *référentiel tournant*, la magnétisation précessionnant à la fréquence de référence de l'appareil semble immobile. Néanmoins, la plupart des moments magnétiques n'effectue pas sa précession exactement à la fréquence de référence de l'appareil, car la présence dans leur environnement d'autres particules ayant un spin non nul (les électrons étant ceux qui produisent l'effet le plus grand) va causer une très légère perturbation des niveaux d'énergie de spin, entraînant une très légère différence de fréquence de précession. Ces différences font que dans le référentiel tournant, la plupart des moments magnétiques ne seront pas immobiles, mais précessionneront eux aussi (voir figure 15), et la fréquence apparente de précession sera égale à la différence entre la fréquence réelle et la fréquence de référence :

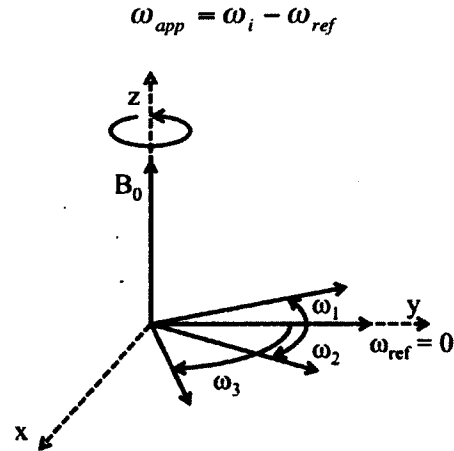


Figure 15 : Précession des spins dans le référentiel tournant

Les axes de coordonnées  $x$ ,  $y$  et  $z$  sont représentés en pointillés, et le champ magnétique externe  $B_0$  est parallèle à l'axe des  $z$ . La flèche circulaire en haut représente le fait que le référentiel est tournant. Trois spins sont représentés, et leur fréquence de précession dans le référentiel tournant est indiquée par  $\omega_1$ ,  $2$  et  $3$ .

Ainsi, suite à un pulse  $90^\circ$  (au temps 0), la magnétisation sera en phase, tel que mentionné plus haut. Après ce pulse, les différents moments magnétiques précessionneront chacun à leur propre fréquence, résultant en une perte (réversible) de l'intensité de la magnétisation, également appelé *déphasage* (voir figure 16). Il est à noter que d'autres processus, notamment différents types de *relaxation*, vont causer une diminution (irréversible cette fois) de l'intensité de la magnétisation.

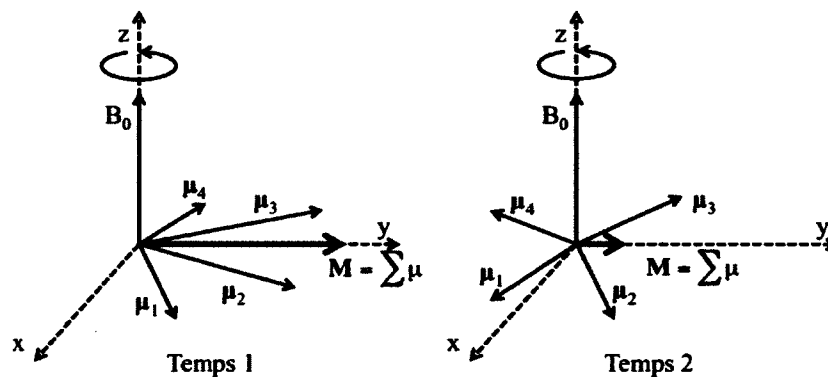


Figure 16 : Déphasage des spins

Les axes de coordonnées  $x$ ,  $y$  et  $z$  sont représentés en pointillés, et le champ magnétique externe  $B_0$  est parallèle à l'axe des  $z$ . La flèche circulaire en haut représente le fait que le référentiel est tournant. Quatre spins sont représentés, représentés par les différents  $\mu_i$ . Leur somme donne la magnétisation  $M$ . Au temps 1, alors que les spins sont encore presque en phase,  $M$  est encore grand, mais au temps 2, après que les spins aient évolué et se soient déphasés,  $M$  est beaucoup plus faible.

Le détecteur mesure l'intensité de la magnétisation en fonction du temps. Cette détection peut être effectuée sur l'axe des  $x$  ou l'axe des  $y$ , positif ou négatif dans les deux cas. Ce signal en fonction du temps s'appelle le *free induction decay* (la décroissance libre de l'induction) ou *FID*, dont voici un exemple ci-dessous.

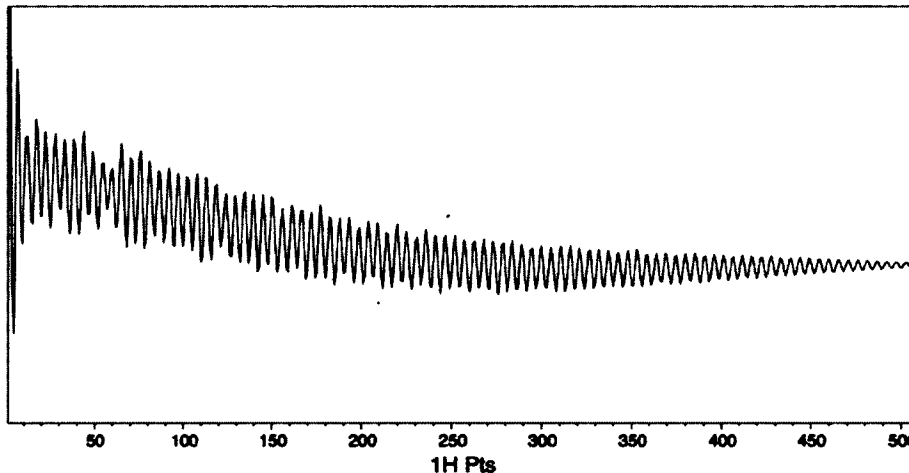


Figure 17 : Exemple de FID (spectre 1D de Miz 5-8)

L'axe horizontal correspond au temps et l'axe vertical à l'intensité de la magnétisation. Les deux axes sont en unités arbitraires.

La FID ne permet pas d'obtenir directement beaucoup d'informations sur les différents atomes composant l'échantillon étudié. Ce qui intéresse le spectroscopiste RMN, c'est la composition en fréquences de la magnétisation, alors que la FID est un



graphique en fonction du temps. C'est pourquoi le spectroscopiste effectuera une *transformée de Fourier*, une opération mathématique permettant d'obtenir le spectre des fréquences qui composent la FID. Il existe plusieurs définition d'une transformée de Fourier, voici celle présentée dans le livre de Griffiths (GRIFFITHS, 2005, p.61) :

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

Cette transformation permet donc d'obtenir un spectre montrant la fréquence de précession de chacun des spins d'un échantillon. Néanmoins, deux spectres d'un même échantillon seront différents s'ils sont mesurés sur deux spectromètres différents, étant donné que la séparation des niveaux d'énergie de spin est proportionnelle à  $B_0$ . Pour remédier à ce problème, les spectroscopistes RMN utilisent comme unité de mesure le *déplacement chimique* au lieu de la fréquence ou de la fréquence relative.

$$\delta = 10^6 \frac{\omega_i - \omega_{ref}}{\omega_{ref}}$$

Le facteur  $10^6$  est ajouté pour obtenir des valeurs plus faciles à utiliser, et ces valeurs ont donc comme unité de mesure le ppm (partie par million). En effet,  $\omega_i - \omega_{ref}$  représente typiquement quelques centaines de hertz, alors que  $\omega_{ref}$  est généralement de l'ordre de quelques centaines de mégahertz.

Différents facteurs influencent le déplacement chimique d'un atome, mais le plus important d'entre eux est le *blindage* du noyau. Ce concept décrit qualitativement l'atténuation du champ magnétique par le nuage électronique orbitant autour du noyau. Plus un noyau est blindé, plus son déplacement chimique sera faible, et vice versa. C'est en fait le niveau de blindage des noyaux qui est affecté par son environnement chimique, et ceci affecte à son tour les niveaux d'énergie de spin.

Pour résumer tout ce qui a été dit jusqu'à présent, les niveaux d'énergie de spin se séparent dans un champ magnétique. Les spins s'alignent selon (ou contre) ce champ magnétique et précessent autour de celui-ci. La magnétisation, un vecteur formé de la sommation des spins, est donc parallèle au champ magnétique externe. En irradiant l'échantillon à la fréquence correspondant à la différence entre les niveaux d'énergie de spin pendant une période de temps appropriée, il est possible de faire basculer la magnétisation dans le plan xy à l'aide d'un pulse  $90^\circ$ , où son intensité pourra être mesurée en fonction du temps (FID). Les spins précessent à des fréquences très légèrement différentes à cause de perturbations des niveaux d'énergie dues à leur environnement chimique qui affecte leur blindage, ce qui résulte en la création d'un spectre de fréquences (après l'application de la transformée de Fourier). Donc, un pulse  $90^\circ$  permet l'obtention d'un spectre RMN à une dimension, dont un exemple est présenté ci-dessous à la figure 18 (ce spectre 1D est la transformée de Fourier de la FID présentée à la figure 17). Noter qu'en RMN, une *dimension* désigne un type de noyau.

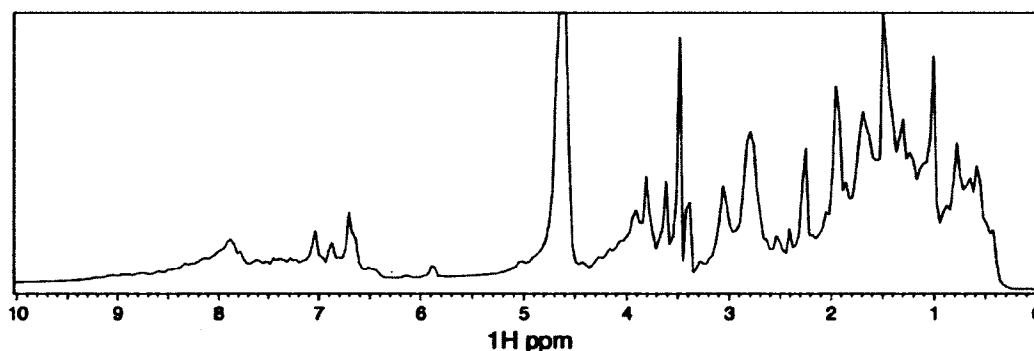


Figure 18 : Exemple de spectre RMN 1D (Miz 5-8)

L'axe horizontal représente le déplacement chimique des protons en ppm, et l'axe vertical représente l'intensité du signal en unités arbitraires. Les résonances entre 5.5 et 10 ppm appartiennent aux protons amide et aromatiques de la protéine, alors que celles situées entre 0 et 4.5 ppm appartiennent aux protons aliphatiques. La grande résonance située à 4.63 ppm est celle des protons de l'eau.

Il est possible de jouer avec la magnétisation à l'aide de pulses d'intensité et de direction variables ou de gradients de champs magnétiques effectués dans un ordre précis et à des moments précis. Ceci peut permettre d'obtenir des informations différentes ou de meilleurs rapports signal sur bruit, par exemple. Ces séries de manipulations sont appelées *séquences de pulses*.

### 2.1.4 Relaxation

(JACOBSEN, 2007, p.162-175; JARYMOWYCZ et STONE, 2006)

Les phénomènes de relaxation sont extrêmement importants en RMN. L'effet nucléaire Overhauser, qui sera mentionné plus tard, résulte en effet de la relaxation. Toutefois, en général, lorsqu'on parle de relaxation en RMN, on fait référence à deux phénomènes en particulier : la relaxation spin-spin et la relaxation spin-réseau.

La diminution irréversible de l'intensité de la magnétisation due à la perte de la cohérence entre les spins mentionnée à la section précédente est appelée *relaxation spin-spin*, ou *relaxation*  $T_2$ . En effet, l'intensité maximale de la magnétisation cohérente décroît exponentiellement au fil du temps selon une équation telle celle présentée ci-dessous :

$$\mathbf{M}_{\text{transverse}} = \sqrt{\mathbf{M}_x^2 + \mathbf{M}_y^2} = \mathbf{M}_0 e^{-R_2 t}$$

$R_2$  est une constante, qui a été définie comme étant le *taux de relaxation spin-spin*, et qui est également définie comme étant égale à  $\frac{1}{T_2}$ . La constante  $T_2$  est appelée la *période de relaxation spin-spin* ou *transverse*.

L'autre processus de relaxation important implique le retour de la magnétisation à l'équilibre. Tel que mentionné dans la section 2.1.3, à l'équilibre, il y a un léger excès de spins dans l'état  $+\frac{1}{2}$ . Or, pour que la magnétisation se trouve dans le plan transverse, la moitié de ces spins en excès a été excitée à l'état  $-\frac{1}{2}$ . Ces spins reviendront éventuellement à leur état initial, ramenant graduellement la magnétisation en  $z$ , et la magnétisation en  $z$  croîtra jusqu'à ce que l'état d'équilibre soit atteint à nouveau. Ce processus s'appelle la *relaxation spin-réseau*, ou *relaxation*  $T_1$ . La dépendance temporelle de cette magnétisation sera elle aussi exponentielle et obéira à l'équation suivante :

$$\mathbf{M}_z = \mathbf{M}_0 (1 - e^{-R_1 t})$$

$R_1$  est encore une fois une constante, appelée le *taux de relaxation spin-réseau*, qui est égale à  $\frac{1}{T_1}$ , où  $T_1$  est la *période de relaxation spin-réseau* ou *longitudinale*. Les relaxations  $T_1$  et  $T_2$  sont des processus qui s'effectuent simultanément, et  $T_2$  est toujours plus court que  $T_1$ . Il est à noter que  $T_1$  dicte le temps minimal qu'il est requis d'attendre avant de pouvoir prendre une nouvelle mesure sur l'échantillon, c'est-à-dire entre  $3T_1$  (CAVANAGH et al., 1996, p.172) et  $5T_1$  (JACOBSEN et al., 2007, p.153).

Les deux types de relaxation sont causés par des transferts d'énergie via des interactions dipôle-dipôle et par l'anisotropie du déplacement chimique (CSA).

La CSA est causée par une variation du blindage d'un noyau en fonction de l'orientation de ce dernier dans le champ magnétique externe. Le nuage électronique n'est pas quelque chose de statique, la distribution de charge varie énormément dans le temps, suite à la fluctuation du lien  $^{15}\text{N-H}$  par rapport au champ magnétique externe à cause des mouvements internes ou de la rotation de la molécule sur elle-même. Ces fluctuations du nuage électronique causent des fluctuations du blindage du noyau, et par le fait même des fluctuations du champ magnétique ressenti par celui-ci. Lorsque les fluctuations s'effectuent à la fréquence de résonance des spins, la relaxation se produira. La CSA affecte à la fois  $T_1$  et  $T_2$ .

Les transferts d'énergie par interactions dipôle-dipôle dépendent de la fréquence à laquelle une molécule se réoriente sur elle-même dans l'échantillon étudié, ce qui cause la réorientation d'un dipôle (le spin d'un proton) par rapport à l'autre (le spin du  $^{15}\text{N}$  qui y est lié) car les spins restent alignés avec ou contre le champ magnétique externe. Le *temps de corrélation rotationnelle*  $\tau_c$  est défini comme le temps que met une molécule à tourner sur elle-même d'un radian (LAMBERT et MAZZOLA, 2004, p.132).  $\tau_c$  dépend de la masse de la molécule, de sa forme, ainsi que de la viscosité du solvant, et sera différent pour chaque molécule. Une molécule de forme globulaire aura un  $\tau_c$  isotrope, c'est-à-dire le même dans toutes les directions, alors qu'une molécule allongée présentera de l'*anisotropie de diffusion rotationnelle*, c'est-à-dire que  $\tau_c$  aura des valeurs différentes

dépendamment de l'axe autour duquel la rotation se produira.  $\tau_c$  sera ainsi plus court lorsque la molécule tourne autour de son axe long, et plus long lorsqu'elle tourne autour d'un axe perpendiculaire à celui-ci. Ainsi,  $T_1$  est affecté par l'anisotropie de diffusion rotationnelle, donc par la réorientation des atomes individuels de la protéine. Dans le cas par exemple d'une molécule de forme allongée, si un vecteur N-H est aligné le long de l'axe de la molécule, il aura un  $T_1$  plus long qu'un vecteur N-H perpendiculaire à cet axe, puisqu'un solide de forme allongée peut tourner plus facilement autour de son axe que perpendiculairement à celui-ci.

C'est lorsque les fluctuations dipôle-dipôle et de CSA se produisent à une fréquence de transition que la relaxation se produit. Dans le cas des protéines et autres macromolécules, les interactions dipôle-dipôle qui causent la relaxation se produisent rarement, et  $T_1$  est long. Plus une protéine est massive, plus  $\tau_c$  sera long, et plus les valeurs de  $T_1$  des atomes la composant seront longues. La dépendance de  $T_1$  en fonction de la masse est toutefois complètement modifiée lorsque  $\omega_0\tau_c$  s'approche de 1 ou devient plus faible que 1 (ce qui arrive avec des molécules de moins de 1kDa en général).

Le phénomène  $T_2$  est aussi régi par les interactions dipôle-dipôle via  $\tau_c$  : plus celui-ci est long, plus la protéine tournera lentement sur elle-même, et plus l'environnement externe de la protéine aura un effet sur les atomes individuels de la protéine. À l'inverse, une protéine ayant un  $\tau_c$  court tournera rapidement sur elle-même et l'effet de l'environnement externe à la protéine sera moyenné, ce qui diminue les variations de champ magnétique subies par les spins individuels, d'où une plus grande cohérence et un  $T_2$  plus long. Toutefois, contrairement à  $T_1$ ,  $T_2$  peut être affecté par des variations de l'environnement chimique du noyau. Si une section de la protéine subit un changement conformationnel, les spins situés dans cette section n'auront pas le même environnement d'une protéine à l'autre, résultant en une perte supplémentaire de la cohérence et donc un raccourcissement de  $T_2$  par un facteur dénommé  $R_{ex}$  (qui sera détaillé à la section 2.4.10. Finalement,  $T_2$  dicte la durée de la FID, de même que la largeur des résonances sur le spectre. En effet, par la transformée de Fourier, plus la FID est longue, et plus les résonances seront fines et précises.

## 2.2 Formalisme de l'opérateur produit

(JACOBSEN, 2007, p.242-253)

Le but premier de la section « résonance magnétique nucléaire » est d'expliquer la théorie derrière le *spectre HSQC* (« heteronuclear single-quantum coherence »), un spectre à deux dimensions qui est utilisé en abondance dans l'étude de structures de protéines. Ceci servira à faciliter la compréhension de n'importe quelle séquence de pulses relativement simple. Afin d'expliquer correctement la séquence de pulses nécessaire pour obtenir un spectre HSQC, le modèle des vecteurs couramment utilisé pour vulgariser le comportement des spins nucléaires dans un échantillon n'est pas suffisant car trop simpliste. Pour les spectres à deux dimensions, c'est le modèle de *l'opérateur produit* (« product operator ») qui est utilisé. Ce modèle est une simplification des représentations matricielles formelles, qui sont plus proches de la réalité, mais qui sont plus laborieuses à utiliser.

Dans le formalisme de l'opérateur produit, le spin d'un proton ( $^1\text{H}$ ) sera généralement représenté par la lettre  $I$ , alors que le spin d'un atome lourd ( $^{13}\text{C}$  ou  $^{15}\text{N}$  en général) sera représenté par la lettre  $S$  (en gras dans les deux cas puisqu'il s'agit de vecteurs). Chaque vecteur décrit l'état dans lequel un spin se trouve. Un spin décrit par  $-I_z$  correspondra donc au spin d'un atome d'hydrogène pointant vers l'axe des  $z$  négatif, ou plus concrètement à une magnétisation pointant dans le sens inverse au champ magnétique externe appliqué sur l'échantillon (qui est par définition dans la direction  $+z$ ). Si le spin est orienté dans une direction plus quelconque, on utilisera la notation  $c_x I_x + c_y I_y + c_z I_z$ , où les  $c_i$  sont tout simplement les composantes du vecteur. Il est donc aussi simple de décrire la magnétisation d'une seule population de noyaux identiques non couplés à d'autres noyaux qu'avec les représentations vectorielles. Lorsqu'un proton est couplé à un autre noyau de spin non nul, toutefois, un deuxième vecteur doit être utilisé, et c'est là que le modèle des vecteurs devient peu pratique. En effet, lorsque ces deux vecteurs sont en antiphase, la somme de leurs composantes donne zéro. Or, pour suivre ce qui arrive à la magnétisation lors d'une séquence de pulses, les périodes où la magnétisation est en antiphase sont très importantes. Dans le formalisme de l'opérateur

produit, une magnétisation en antiphase est représentée par la notation  $2I_iS_j$  ou  $2S_iI_j$ . Ces deux notations ne sont pas équivalentes, car elles représentent une multiplication matricielle, qui n'est pas commutative, sauf dans certains cas particuliers. On écrit toujours l'opérateur *observable* (celui qui se trouve dans le plan xy puisque c'est dans ce plan que les mesures sont effectuées) en premier, donc si, suite à un pulse, on obtient un opérateur produit de la forme  $2I_zS_{x-y}$ , on le réécrira  $2S_{x-y}I_z$  (dans ce cas particulier, la multiplication est commutative). Le facteur 2 est un simple facteur de normalisation et peut donc être ignoré pour l'explication présentée ici. Prenons par exemple le cas d'un amide. Le spin du  $^1\text{H}$  est couplé au spin du  $^{15}\text{N}$ , couplage qui est à l'origine de l'antiphase. Dans le cas où la magnétisation du  $^{15}\text{N}$  serait en z et celle des protons dans l'état  $\alpha$  en +x (et celle des protons dans l'état  $\beta$  en -x), l'opérateur produit s'écrirait  $+2I_xS_z$ , ce qui se lit « magnétisation du  $^1\text{H}$  sur l'axe des x, en antiphase par rapport à son partenaire de couplage  $^{15}\text{N}$  en z ». Ceci signifie en outre que  $-2I_xS_z$  n'est pas équivalent à  $+2I_xS_z$ , car les protons dans l'état  $\alpha$  sont alors en -x plutôt qu'en +x. Finalement, puisqu'ils ont été mentionnés ci-haut sans préciser davantage leur nature, les seuls opérateurs observables sont  $S_x, S_y, I_x, I_y, S_xI_z, S_yI_z, I_xS_z$  et  $I_yS_z$ . Ces opérateurs correspondent à des transitions à un quantum, alors que toutes les autres combinaisons correspondent à des combinaisons de transitions zéro quantum ou double quantum, qui ne sont pas directement détectables.

Le référentiel xyz utilisé ici est un référentiel tournant à la fréquence de résonance de l'atome étudié. Sous l'effet du déplacement chimique, la magnétisation effectuera une précession à une fréquence  $\nu_r$  différente de la fréquence de référence  $\nu_o$  autour du champ magnétique externe, conventionnellement dans le sens antihoraire. La fréquence relative  $\Delta\nu = \nu_o - \nu_r$ , qui est reliée au déplacement chimique, peut être exprimée en terme de fréquence angulaire  $\Omega = 2\pi\Delta\nu$  en rad/s. Après un délai d'évolution  $\tau$ , les différentes composantes évoluent ainsi :

$$I_x \xrightarrow{\tau} I_x \cos\Omega\tau + I_y \sin\Omega\tau$$

$$I_y \xrightarrow{\tau} I_y \cos\Omega\tau - I_x \sin\Omega\tau$$

$$I_z \xrightarrow{\tau} I_z$$

Pour une combinaison de différents états, il suffit de faire évoluer chacune des composantes indépendamment. Dans le cas d'un état en antiphase évoluant sous l'effet du déplacement chimique, il en va de même, par exemple :

$$2S_x I_z \xrightarrow{\Omega_s \tau} 2(S_x \cos \Omega_s \tau + S_y \sin \Omega_s \tau) I_z = 2S_x I_z \cos \Omega_s \tau + 2S_y I_z \sin \Omega_s \tau$$

Dans tout cas d'évolution, l'état initial est multiplié par un cosinus, alors que l'état vers lequel la magnétisation se dirige est multiplié par un sinus.

Le couplage scalaire (« J-coupling »), c'est-à-dire l'effet des spins adjacents au spin étudié, est un autre phénomène causant une évolution. Cette évolution peut être vue comme un cycle qui tourne dans le sens antihoraire, tout comme l'évolution sous l'effet du déplacement chimique. Une magnétisation en phase évoluera vers une magnétisation antiphase, qui évoluera vers une magnétisation en phase, et ainsi de suite, de cette façon :

$$\underset{\text{phase}}{I_x} \rightarrow \underset{\text{antiphase}}{2I_y S_z} \rightarrow \underset{\text{phase}}{-I_x} \rightarrow \underset{\text{antiphase}}{-2I_y S_z} \rightarrow \underset{\text{phase}}{I_x}$$

La différence de fréquence entre les deux composantes évoluant sous l'effet du couplage scalaire est égale à J, donc chaque composante a une fréquence de  $\pm J/2$  dans le référentiel tournant, donc chacune des composantes effectuera sa précession à une fréquence angulaire égale à  $\pi J$ . Voici comment la magnétisation évolue sous l'effet du couplage scalaire après un délai  $\tau$  :

$$\begin{aligned} I_x &\xrightarrow{\pi J \tau} I_x \cos \pi J \tau + 2I_y S_z \sin \pi J \tau \\ I_y &\xrightarrow{\pi J \tau} I_y \cos \pi J \tau - 2I_x S_z \sin \pi J \tau \\ 2I_x S_z &\xrightarrow{\pi J \tau} 2I_x S_z \cos \pi J \tau + I_y \sin \pi J \tau \\ 2I_y S_z &\xrightarrow{\pi J \tau} 2I_y S_z \cos \pi J \tau - I_x \sin \pi J \tau \end{aligned}$$



Lorsque la magnétisation évolue à la fois sous l'effet du déplacement chimique et du couplage scalaire, il suffit d'appliquer les deux effets indépendamment (l'ordre n'a pas d'importance).

Les pulses agissent sur l'opérateur produit de la même façon que sur les vecteurs. Ils sont toujours spécifiques au type de noyau irradié (donc un pulse sur le  $^{13}\text{C}$  n'affecte pas la magnétisation des protons par exemple, la fréquence des pulses n'étant pas la même) et font basculer la magnétisation dans le sens antihoraire pour un angle positif. L'effet d'un pulse sur une magnétisation en phase est plutôt intuitif, donc voici en exemple l'effet d'un pulse de  $90^\circ$  autour de l'axe des y à la fréquence du proton :

$$2I_x S_z \rightarrow 2(-I_z) S_z = -2I_z S_z; \quad 2S_y I_z \rightarrow 2S_y I_x; \quad 2I_y S_z \rightarrow 2I_y S_x$$

Suite à certains pulses, il est possible d'obtenir des opérateurs non observables, tels  $2I_z S_z$  ou  $2S_y I_x$ . Ces opérateurs sont néanmoins nécessaires pour suivre ce qui arrive à la magnétisation lors d'une séquence de pulses.

## 2.3 Séquences de pulses de base

Afin de simplifier l'explication de la séquence de pulses HSQC, voici quelques séquences de pulses de base ainsi que leur effet sur la magnétisation et son évolution.

### 2.3.1 L'écho de spin (« spin echo ») sur un seul type de noyau

(MANDAL et MAJUMDAR 2003, p.2-4)

Immédiatement après le pulse initial de  $90^\circ$  transférant la magnétisation dans le plan xy, la magnétisation sera en phase. Sous l'effet du déplacement chimique, du couplage scalaire et des différents types de relaxation, les différentes populations de spins précéderont à des fréquences différentes et la phase sera très rapidement perdue. La perte de cohérence due au couplage scalaire ou au déplacement chimique est toutefois réversible (contrairement à celle due à la relaxation, qui ne sera pas traitée davantage dans

ce chapitre). Il est en effet possible de ramener la magnétisation en phase à l'aide d'une séquence d'écho de spin. Il existe deux types d'échos de spin pour les expériences à deux dimensions : l'écho sur un seul type de noyau et l'écho sur les deux types.

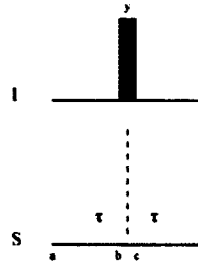


Figure 19 : Écho de spin sur le noyau I

Le temps évolue de gauche à droite, le rectangle noir représente un pulse  $180^\circ$  sur le noyau I appliqué selon la direction de l'axe des y.  $\tau$  est le délai entre les points a et b, ainsi qu'entre les points c et d.

Voici ce qui se passe lorsqu'on applique un pulse de  $180^\circ$  sur un seul type d'atome exactement au centre d'un délai  $2\tau$  (en considérant que l'état initial est tout simplement la magnétisation suivant immédiatement le pulse de  $90^\circ$  qui amène la magnétisation dans le plan xy). Le premier aspect pris en compte est l'effet du déplacement chimique sur le noyau I (voir figure 19 pour les points de repère).

$$\begin{aligned}
 a \rightarrow b & \quad \mathbf{I}_y \xrightarrow{\Omega_I \tau} \mathbf{I}_y \cos \Omega_I \tau - \mathbf{I}_x \sin \Omega_I \tau \\
 b \rightarrow c & \quad \xrightarrow{(180^\circ) \mathbf{I}_y} \mathbf{I}_y \cos \Omega_I \tau + \mathbf{I}_x \sin \Omega_I \tau \\
 c \rightarrow d & \quad \xrightarrow{\Omega_I \tau} (\mathbf{I}_y \cos \Omega_I \tau - \mathbf{I}_x \sin \Omega_I \tau) \cos \Omega_I \tau + (\mathbf{I}_x \cos \Omega_I \tau + \mathbf{I}_y \sin \Omega_I \tau) \sin \Omega_I \tau \\
 & \quad = \mathbf{I}_y (\cos^2 \Omega_I \tau + \sin^2 \Omega_I \tau) + \mathbf{I}_x (\cos \Omega_I \tau \sin \Omega_I \tau - \sin \Omega_I \tau \cos \Omega_I \tau) \\
 & \quad = \mathbf{I}_y
 \end{aligned}$$

Le couplage scalaire aura l'effet suivant sur le noyau I dans cette séquence :

$$\begin{aligned}
 a \rightarrow b & \quad \mathbf{I}_y \xrightarrow{\pi J \tau} \mathbf{I}_y \cos \pi J \tau - 2 \mathbf{I}_x \mathbf{S}_z \sin \pi J \tau \\
 b \rightarrow c & \quad \xrightarrow{(180^\circ) \mathbf{I}_y} \mathbf{I}_y \cos \pi J \tau + 2 \mathbf{I}_x \mathbf{S}_z \sin \pi J \tau \\
 c \rightarrow d & \quad \xrightarrow{\pi J \tau} (\mathbf{I}_y \cos \pi J \tau - 2 \mathbf{I}_x \mathbf{S}_z \sin \pi J \tau) \cos \pi J \tau + (2 \mathbf{I}_x \mathbf{S}_z \cos \pi J \tau + \mathbf{I}_y \sin \pi J \tau) \sin \pi J \tau \\
 & \quad = \mathbf{I}_y (\cos^2 \pi J \tau + \sin^2 \pi J \tau) + 2 \mathbf{I}_x \mathbf{S}_z (\cos \pi J \tau \sin \pi J \tau - \sin \pi J \tau \cos \pi J \tau) \\
 & \quad = \mathbf{I}_y
 \end{aligned}$$

Un pulse de  $180^\circ$  sur le noyau **I** ramène donc la magnétisation de ce noyau à l'état initial à la fois au niveau du déplacement chimique et du couplage scalaire. Le pulse de  $180^\circ$  sur le noyau **I** aura l'effet suivant sur le noyau **S**, d'abord au niveau du déplacement chimique :

$$\begin{aligned}
 a \rightarrow b & \quad S_y \xrightarrow{\Omega_S \tau} S_y \cos \Omega_S \tau - S_x \sin \Omega_S \tau \\
 b \rightarrow c & \quad \xrightarrow{(180^\circ)I_y} S_y \cos \Omega_S \tau - S_x \sin \Omega_S \tau \\
 c \rightarrow d & \quad \xrightarrow{\Omega_S \tau} (S_y \cos \Omega_S \tau - S_x \sin \Omega_S \tau) \cos \Omega_S \tau - (S_x \cos \Omega_S \tau + S_y \sin \Omega_S \tau) \sin \Omega_S \tau \\
 & \quad = S_y (\cos^2 \Omega_S \tau - \sin^2 \Omega_S \tau) - 2S_x \sin \Omega_S \tau \cos \Omega_S \tau \\
 & \quad = S_y \cos 2\Omega_S \tau - S_x \sin 2\Omega_S \tau
 \end{aligned}$$

Voici ce qui arrive à l'évolution du noyau **S** sous l'effet du couplage scalaire :

$$\begin{aligned}
 a \rightarrow b & \quad S_y \xrightarrow{\pi J \tau} S_y \cos \pi J \tau - 2S_x I_z \sin \pi J \tau \\
 b \rightarrow c & \quad \xrightarrow{(180^\circ)I_y} S_y \cos \pi J \tau + 2S_x I_z \sin \pi J \tau \\
 c \rightarrow d & \quad \xrightarrow{\pi J \tau} (S_y \cos \pi J \tau - 2S_x I_z \sin \pi J \tau) \cos \pi J \tau + (2S_x I_z \cos \pi J \tau + S_y \sin \pi J \tau) \sin \pi J \tau \\
 & \quad = S_y (\cos^2 \pi J \tau + \sin^2 \pi J \tau) + 2S_x I_z (\cos \pi J \tau \sin \pi J \tau - \sin \pi J \tau \cos \pi J \tau) \\
 & \quad = S_y
 \end{aligned}$$

Un pulse de  $180^\circ$  sur l'atome **I** n'affecte donc pas l'évolution sous l'effet du déplacement chimique du noyau **S**, il évolue pendant un délai  $2\tau$  comme si le pulse n'existait pas. Toutefois, ce pulse annule l'effet du couplage scalaire sur le noyau **S** pendant ce délai. En effet, le pulse effectué sur le noyau **I** affecte le noyau **S** à cause du couplage entre les deux, couplage qui est visible par le terme  $2S_x I_z$  (et qui est le terme affecté par le pulse).

### 2.3.2 L'écho de spin sur deux types de noyaux

(MANDAL et MAJUMDAR, 2003, p.4)

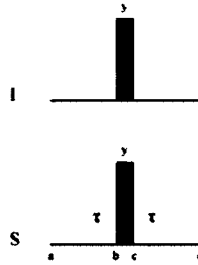


Figure 20 : Écho de spin sur deux types de noyaux

Le temps évolue de gauche à droite, chacun des rectangles représente un pulse  $180^\circ$ .  $\tau$  est le délai entre les points a et b, ainsi qu'entre les points c et d.

Le pulse de  $180^\circ$  peut être appliqué à la fois sur le noyau S et le noyau I. Individuellement, chacun de ces pulses agira tel que démontré ci-haut sur l'évolution de la magnétisation sous l'effet du déplacement chimique, c'est-à-dire qu'il ramènera la magnétisation à son état initial. Par souci de concision, cette démonstration ne sera pas répétée. Voici par contre ce qui arrive à l'évolution sous l'effet du couplage scalaire (voir figure 20 pour les points de repère) :

$$\begin{aligned}
 a \rightarrow b & \quad \mathbf{I}_y \xrightarrow{\pi J \tau} \mathbf{I}_y \cos \pi J \tau - 2\mathbf{I}_x \mathbf{S}_z \sin \pi J \tau \\
 b \rightarrow c & \quad \xrightarrow{(180^\circ)(\mathbf{I}_y + \mathbf{S}_y)} \mathbf{I}_y \cos \pi J \tau - 2(-\mathbf{I}_x)(-\mathbf{S}_z) \sin \pi J \tau \\
 c \rightarrow d & \quad \xrightarrow{\pi J \tau} (\mathbf{I}_y \cos \pi J \tau - 2\mathbf{I}_x \mathbf{S}_z \sin \pi J \tau) \cos \pi J \tau - (2\mathbf{I}_x \mathbf{S}_z \cos \pi J \tau + \mathbf{I}_y \sin \pi J \tau) \sin \pi J \tau \\
 & \quad = \mathbf{I}_y (\cos^2 \pi J \tau - \sin^2 \pi J \tau) - 2\mathbf{I}_x \mathbf{S}_z (\cos \pi J \tau \sin \pi J \tau + \sin \pi J \tau \cos \pi J \tau) \\
 & \quad = \mathbf{I}_y \cos 2\pi J \tau - 2\mathbf{I}_x \mathbf{S}_z \sin 2\pi J \tau
 \end{aligned}$$

Et il en va de même pour ce qui arrive au noyau S. Donc, un écho de spin sur les deux types de noyaux élimine l'évolution sous l'effet du déplacement chimique au cours du délai  $2\tau$  pour les deux noyaux, mais permet à la magnétisation de continuer d'évoluer sous l'effet du couplage scalaire.

### 2.3.3 La séquence INEPT

(CAVANAGH et al., 1996, p.91-93, 411)

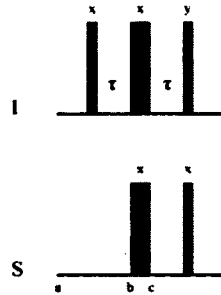


Figure 21 : Séquence de pulses INEPT

Le temps évolue de gauche à droite. Les rectangles noirs représentent chacun un pulse différent, les minces représentent des pulses  $90^\circ$  alors que les larges au centre correspondent à des pulses  $180^\circ$ .  $\tau$  est le délai entre le pulse  $90^\circ$  initial et les pulses  $180^\circ$  centraux, ainsi qu'entre les pulses  $180^\circ$  et les pulses  $90^\circ$  finaux.

Il est possible d'exciter et de détecter le signal de noyaux autres que l'hydrogène. En RMN de protéines, les noyaux de  $^{13}\text{C}$  et de  $^{15}\text{N}$  sont d'un grand intérêt. Toutefois, l'excitation et la détection du signal de ces noyaux présente un désavantage majeur par rapport à l'hydrogène : le rapport signal sur bruit. En effet, ce rapport est proportionnel à  $\gamma_{ex}\gamma_{det}^{3/2}$ , où  $\gamma_{ex}$  est le rapport gyromagnétique du noyau excité et  $\gamma_{det}$  est celui du noyau détecté. Or, le rapport gyromagnétique du  $^{13}\text{C}$  est quatre fois plus faible que celui du proton, et celui du  $^{15}\text{N}$  est dix fois plus faible. Néanmoins, il existe une façon très simple de compenser cette lacune : le transfert de la magnétisation, c'est-à-dire de faire passer la magnétisation d'un type de noyau à un autre. La séquence INEPT (« insensitive nucleus enhanced by polarization transfer ») est la méthode la plus simple pour atteindre ce but. Voici l'évolution de la magnétisation lors de cette séquence :

$$\begin{aligned} \mathbf{I}_z &\xrightarrow{(90^\circ)\mathbf{I}_x} -\mathbf{I}_y \xrightarrow{\frac{\pi J \tau}{2}, (180^\circ)(\mathbf{I}_x + \mathbf{S}_x), \frac{\pi J \tau}{2}} \mathbf{I}_y \cos \pi J \tau - 2\mathbf{I}_x \mathbf{S}_z \sin \pi J \tau \\ &\xrightarrow{\tau = \frac{1}{2J}} -2\mathbf{I}_x \mathbf{S}_z \xrightarrow{(90^\circ)(\mathbf{I}_y + \mathbf{S}_y)} -2\mathbf{I}_z \mathbf{S}_y = -2\mathbf{S}_y \mathbf{I}_z \end{aligned}$$

Donc, lors d'une séquence INEPT, la magnétisation passe du noyau I au noyau S. En effet, dans l'état  $-2\mathbf{S}_y \mathbf{I}_z$ , c'est la magnétisation du noyau S qui se retrouve (et évolue) dans le plan xy. Or, comme c'est le noyau I qui a été excité (pulse initial de  $90^\circ$ ),

l'intensité de la magnétisation demeure la même, c'est-à-dire dépendante de  $\gamma_I$  et non de  $\gamma_S$ . Les spectres RMN de noyaux lourds ainsi obtenus ont donc un rapport signal sur bruit bien meilleur.

### 2.3.4 La séquence HSQC

(CAVANAGH et al., 1996, p.411-414, 416-418)

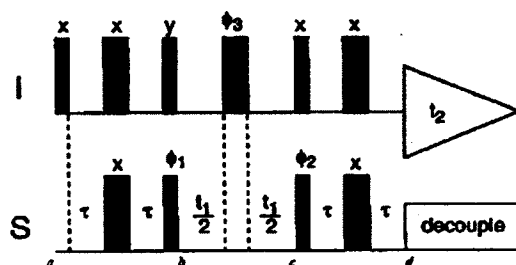


Figure 22 : Séquence de pulses d'un spectre HSQC (CAVANAGH et al., 1996, p.412)

Le temps évolue de gauche à droite. Chacun des rectangles noirs représente un pulse différent, les minces représentent des pulses  $90^\circ$  alors que les larges au centre correspondent à des pulses  $180^\circ$ . Tous les délais  $\tau$  ont la même durée.  $t_1$  et  $t_2$  sont les périodes de temps qui sont variés d'une répétition à l'autre de l'expérience et en fonction desquelles sont effectuées les transformées de Fourier. Le rectangle « decouple » correspond à l'irradiation constante du noyau S pour que la moyenne de sa magnétisation donne zéro. L'autorisation de reproduction de cette image est en annexe.

Le HSQC est le spectre de base pour le biologiste structural, et donc sa séquence de pulses est importante à comprendre. Il est bien entendu possible de modifier cette séquence de pulses afin d'obtenir un meilleur rapport signal sur bruit, mais ces modifications ne seront pas détaillées ici. La séquence de pulses est montrée à la figure 22.

Les symboles  $\phi_1$ ,  $\phi_2$  et  $\phi_3$  représentent la direction selon laquelle le pulse est effectué. Ces valeurs varient au fil des répétitions, afin d'effectuer ce qu'on appelle des *cycles de phases*, qui permettent de retirer des artéfacts tout en augmentant le rapport signal sur bruit, mais ceci ne sera pas détaillé ici. Pour le cas présent, considérons que  $\phi_1$  et  $\phi_2$  sont selon la direction  $+x$ , et  $\phi_3$  dans la direction  $+y$ . Premièrement, on remarque que de  $a$  à  $b$ , la séquence de pulses est celle d'un INEPT. Par la suite, de  $b$  à  $c$ , on a un pulse  $180^\circ$  sur un seul type de noyau. Finalement, de  $c$  à  $d$ , on a ce qu'on appelle un *reverse*

*INEPT*, qui sert à ramener la magnétisation sur le noyau initial suite à un *INEPT*. Voici le comportement de la magnétisation en détails :

$$\begin{aligned}
 a \rightarrow b & \quad \mathbf{I}_z \xrightarrow{\text{INEPT}} -2\mathbf{S}_y\mathbf{I}_z \\
 b \rightarrow c & \quad \xrightarrow{\frac{\Omega_S t_1}{2}, (180^\circ)\mathbf{I}_y, \frac{\Omega_S t_1}{2}} -2(\mathbf{S}_y \cos \Omega_S t_1 - \mathbf{S}_x \sin \Omega_S t_1)(-\mathbf{I}_z) \\
 & \quad = 2\mathbf{S}_y\mathbf{I}_z \cos \Omega_S t_1 - 2\mathbf{S}_x\mathbf{I}_z \sin \Omega_S t_1 \\
 c \rightarrow d & \quad \xrightarrow{(90^\circ)(\mathbf{I}_x + \mathbf{S}_x)} -2\mathbf{I}_y\mathbf{S}_z \cos \Omega_S t_1 + 2\mathbf{S}_x\mathbf{I}_y \sin \Omega_S t_1 \\
 & \quad \xrightarrow{\pi J \tau, (180^\circ)(\mathbf{I}_x + \mathbf{S}_x), \pi J \tau} -(2\mathbf{I}_y\mathbf{S}_z \cos 2\pi J \tau - \mathbf{I}_x \sin 2\pi J \tau) \cos \Omega_S t_1 + 2\mathbf{S}_x(-\mathbf{I}_y) \sin \Omega_S t_1 \\
 & \quad \xrightarrow{\tau = \frac{1}{4J}} \mathbf{I}_x \cos \Omega_S t_1 - 2\mathbf{S}_x\mathbf{I}_y \sin \Omega_S t_1
 \end{aligned}$$

Le terme  $-2\mathbf{S}_x\mathbf{I}_y \sin \Omega_S t_1$  n'est pas une observable, tel que mentionné plus haut. Le terme  $\mathbf{I}_x \cos \Omega_S t_1$  montre quant à lui que la magnétisation est en partie revenue sur le noyau I (le proton). En laissant évoluer ce terme pendant l'acquisition, on obtient un spectre 1D de proton, comme précédemment, mais le  $\cos \Omega_S t_1$  sera non nul uniquement pour les fréquences où la magnétisation du noyau S est en phase elle aussi, ce qu'on appelle une *cohérence*. En prenant plusieurs spectres à des délais  $t_1$  différents, on obtient

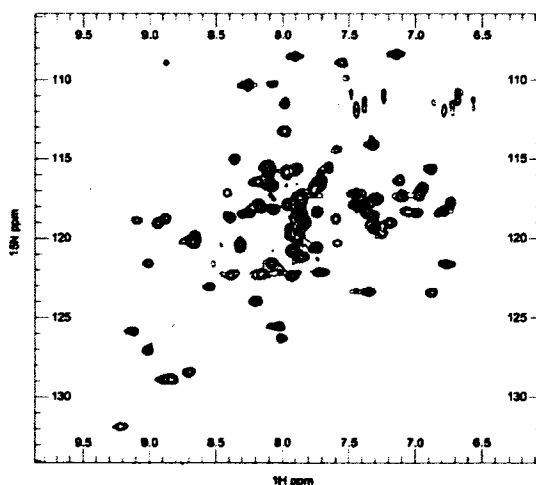


Figure 23 : Exemple de spectre HSQC (Miz 5-8)

L'axe des abscisses représente le déplacement chimique des protons et l'axe des ordonnées, à celui des noyaux d'azote. Tous deux sont en ppm. Chacune des résonances visibles correspond à une paire  $^{15}\text{N}-^1\text{H}$ .

d'autres résonances en cohérence, et en prenant la transformée de Fourier dans la dimension  $t_1$  en plus de celle qu'on prend habituellement (qui donne un spectre  $^1\text{H}$ ), on obtient un spectre à deux dimensions tel que celui présenté ci-dessous à la figure 23. Les informations contenues dans un spectre 2D varient selon la séquence de pulses utilisée, et la section 2.4 décrit quelles informations peuvent être tirées de quel spectre, dont le HSQC.

## 2.4 Spectres RMN

Cette section énumère différents spectres RMN utilisés en biologie structurale, ainsi que les informations qui peuvent en être extraites.

### 2.4.1 HSQC

(JACOBSEN, 2007, p.598-599)

Le HSQC est le spectre utilisé le plus fréquemment par les biologistes structuraux dans le cas d'un échantillon ayant une concentration de l'ordre du millimolaire et *uniformément marqué* (c'est-à-dire dont un ou plusieurs types d'atomes ont été entièrement remplacés par des isotopes magnétiquement actifs; un échantillon uniformément marqué  $^{15}\text{N}$  par exemple correspond à un échantillon dont tous les atomes d'azote sont des  $^{15}\text{N}$ , aucun  $^{14}\text{N}$  n'y est présent). Il s'agit d'un spectre à deux dimensions (en général  $^1\text{H}$  avec  $^{15}\text{N}$  ou  $^{13}\text{C}$ ). Il s'acquiert rapidement : une quinzaine de minutes est en général suffisant pour obtenir un spectre clair, mais il est possible d'en faire de plus longs dans le but d'avoir un meilleur rapport signal sur bruit (qui est proportionnel à  $n^{1/2}$ , où  $n$  est le nombre de répétitions). Un [ $^{15}\text{N}$ - $^1\text{H}$ ]-HSQC permet de voir une seule résonance par acide aminé (à l'exception des prolines qui n'en présentent aucune, des asparagines et des glutamines qui en ont deux supplémentaires, des tryptophanes qui en possèdent une supplémentaire et possiblement des arginines dépendamment du pH à cause de leur chaîne latérale) et donc son examen permet de détecter rapidement des changements de structure, ainsi que l'identité des résidus les subissant. Il est également facile de voir si une protéine est dénaturée par la présence d'une importante concentration de résonances, également



appelées pics, situées au centre, alors qu'au contraire, une protéine dans son état natif présentera des pics bien dispersés. La figure 24 ci-dessous compare un HSQC de protéine dans son état natif (à gauche) à celui d'une protéine dénaturée (à droite) :

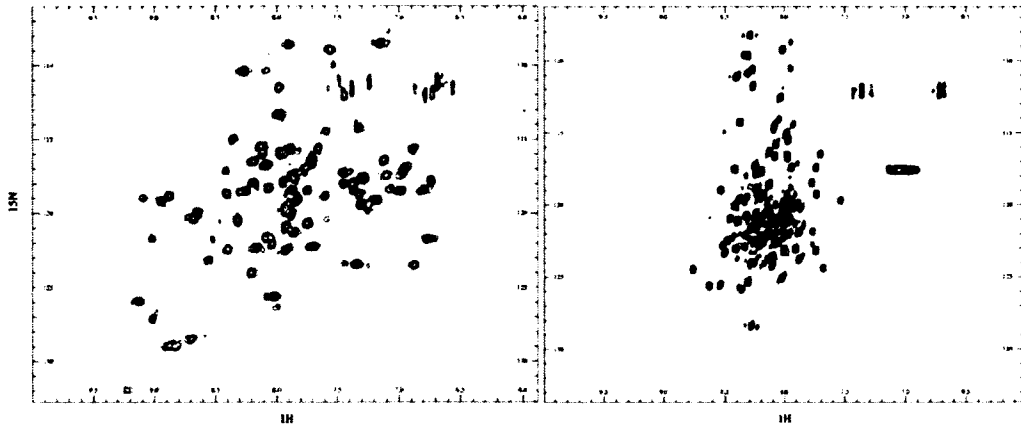


Figure 24 : HSQC d'une protéine bien repliée ou dénaturée (Miz 5-8 à pH 7 ou 4)

Dans les deux spectres, l'axe des abscisses représente le déplacement chimique des protons et l'axe des ordonnées, à celui des noyaux d'azote. Tous deux sont en ppm. Les deux spectres ont la même échelle dans les deux dimensions. Le spectre de gauche est le HSQC de la protéine Miz 5-8 correctement repliée, alors que le spectre de droite est le HSQC de la protéine Miz 5-8 dénaturée.

En plus du [ $^{15}\text{N}$ - $^1\text{H}$ ]-HSQC, il est possible d'effectuer un [ $^{13}\text{C}$ - $^1\text{H}$ ]-HSQC, qui permet de voir chacune des cohérences de résonances C-H de l'échantillon, c'est-à-dire quel proton est relié à quel carbone. En biologie structurale, les échantillons étudiés sont très souvent des protéines, et donc possèdent une très grande quantité de paires C-H, résultant en un spectre généralement encombré.

#### 2.4.2 HNCACB

(JACOBSEN, 2007, p.617-621)

Le [ $^{15}\text{N}$ - $^1\text{H}$ ]-HSQC est très utile pour voir les changements qu'un échantillon subit, mais pour en tirer des informations intéressantes, il est nécessaire d'attribuer chacune des résonances le composant, ce qui n'est pas possible à l'aide d'un spectre 2D seul. C'est pour cette raison que les biologistes structuraux utilisent le HNCACB, un spectre 3D (il existe également des spectres avec davantage de dimensions encore). Dans la présentation habituelle de ce spectre, une tranche spécifique (en z) correspond à un

déplacement chimique en azote, une colonne spécifique correspond à l'hydrogène, et une rangée correspond au carbone. À une tranche et une colonne correspondant à une paire N-H du squelette polypeptidique, en temps normal, quatre résonances seront visibles : le  $C_\alpha$  et le  $C_\beta$  appartenant au même résidu que l'amide, ainsi que ceux appartenant au résidu précédent, avec une intensité légèrement moindre. Dans le jargon des spectroscopistes, on parle de *résidu en i* et de *résidu en i-1* respectivement pour désigner les résidus en question. Les  $C_\alpha$  seront identifiables par des pics positifs, alors que les  $C_\beta$  seront associés aux pics négatifs. Il existe des statistiques sur le déplacement chimique des  $C_\alpha$  et  $C_\beta$  en fonction des résidus (disponibles publiquement sur la BioMagnetic Resonance Databank, BMRB), et ces statistiques permettent d'identifier à quel type de résidu appartiennent cet amide et ces carbones.

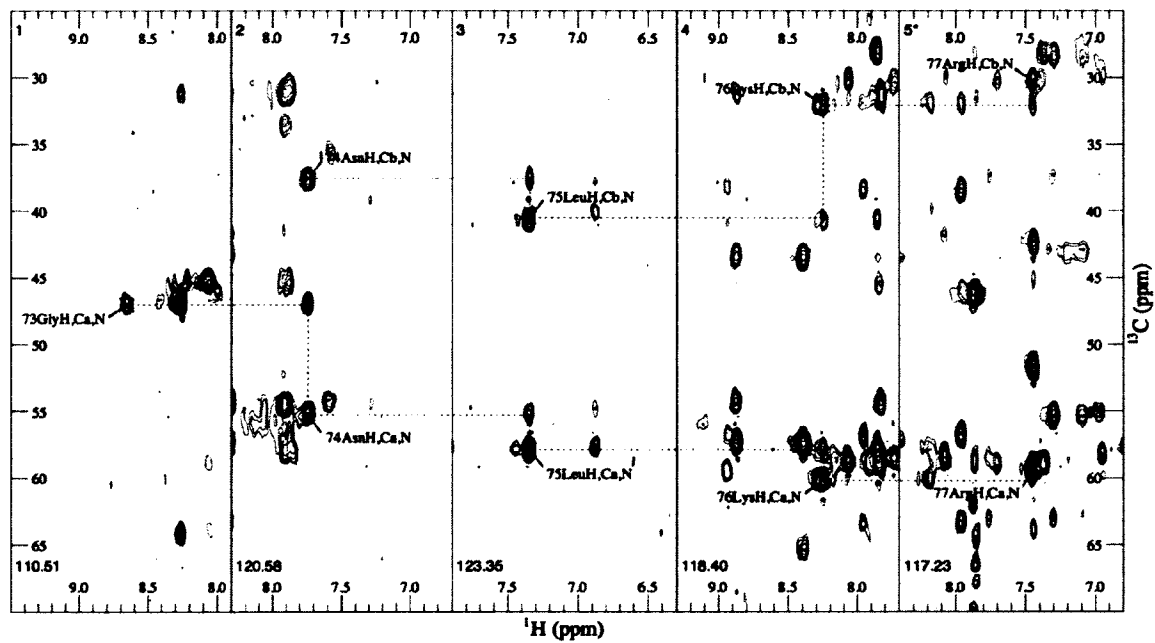


Figure 25 : Exemple de HNCACB et d'attribution séquentielle d'une séquence SGNLKR de Miz 5-8

Cette figure juxtapose cinq tranches d'un spectre 3D HNCACB. En abscisse se retrouve le déplacement chimique des protons et en ordonnée celui du  $^{13}\text{C}$ . En z est le déplacement chimique en  $^{15}\text{N}$ , qui est indiqué dans le coin inférieur gauche de chacune des tranches. Dans chacune des tranches sont identifiées les résonances appartenant aux résidus 73 à 77 de la protéine Miz 5-8. Les résonances en vert sont d'intensité positive et appartiennent à des  $^{13}\text{C}_\alpha$  alors que les résonances rouges sont d'intensité négative et appartiennent à des  $^{13}\text{C}_\beta$ . Les lignes pointillées servent à indiquer l'itinéraire effectué pour identifier les résonances appartenant aux résidus consécutifs.

En partant des déplacements chimiques du  $C_\alpha$  et du  $C_\beta$  en  $i-1$  d'un amide particulier et en cherchant quel amide possède en  $i$  un  $C_\alpha$  et un  $C_\beta$  ayant les mêmes déplacements chimiques, on peut identifier des résidus consécutifs. Si la séquence primaire de la protéine est connue, il sera ainsi possible d'identifier à quel atome précisément appartiennent chacune des résonances visibles sur le spectre. Il est également possible d'effectuer cette progression dans l'autre sens. Un exemple d'une telle identification consécutive est présenté à la figure 25, où les  $C_\alpha$  sont en vert et les  $C_\beta$  en rouge. Seules les résonances des résidus d'intérêt situés en  $i$  ont été identifiées. La tranche appartenant à la Leu75 montre clairement les quatre résonances mentionnées ci-dessus. Les tranches appartenant à la Lys76 et à l'Arg77 en contiennent également quatre chacune, mais celles-ci sont plus difficiles à distinguer à cause du chevauchement. La tranche de l'Asn74 en contient seulement trois, puisque l'acide aminé en  $i-1$  de cette asparagine est une glycine, qui ne possède pas de  $C_\beta$ . Quant à la tranche de la Gly73, une seule résonance  $\gamma$  est visible, la raison en sera détaillée dans la prochaine section.

### 2.4.3 CBCA(CO)NH

(JACOBSEN, 2007, p.617-621)

Le HNCACB est le spectre le plus utile afin d'attribuer séquentiellement les résidus d'une protéine, mais il a quelques inconvénients. Premièrement, les  $C_\alpha$  et  $C_\beta$  des sérines et des thréonines peuvent être au même déplacement chimique. Or, comme ces deux pics sont d'intensité inverse, ils s'annuleront s'ils sont superposés (comme c'est le cas dans le cadre le plus à gauche de la figure 25, où le résidu en  $i-1$  est une sérine). Deuxièmement, il arrive que certaines résonances soient tout simplement faibles pour le résidu en  $i$ , alors le résidu en  $i-1$  sera tout simplement invisible. Troisièmement, il arrive que les résonances du résidu en  $i-1$  se superposent à celles du résidu en  $i$ , en particulier pour les  $C_\alpha$ , ce qui rend difficile de continuer la progression de l'attribution séquentielle (comme c'est le cas dans le cadre le plus à droite de la figure 25). Pour circonvier à ces inconvénients, les biologistes structuraux utilisent le CBCA(CO)NH, qui est plus sensible que le HNCACB. Ce spectre corrèle une paire N-H amide au  $C_\alpha$  et au  $C_\beta$  du résidu situé en  $i-1$  uniquement, puisque la magnétisation doit passer par le carbonyle, contrairement au

HNCACB. Il est à noter que le carbonyle lui-même n'est pas visible, la séquence de pulses ne laissant pas évoluer la magnétisation sur cet atome, et c'est pourquoi il est entre parenthèses dans le nom du spectre. Et contrairement au HNCACB, les résonances des

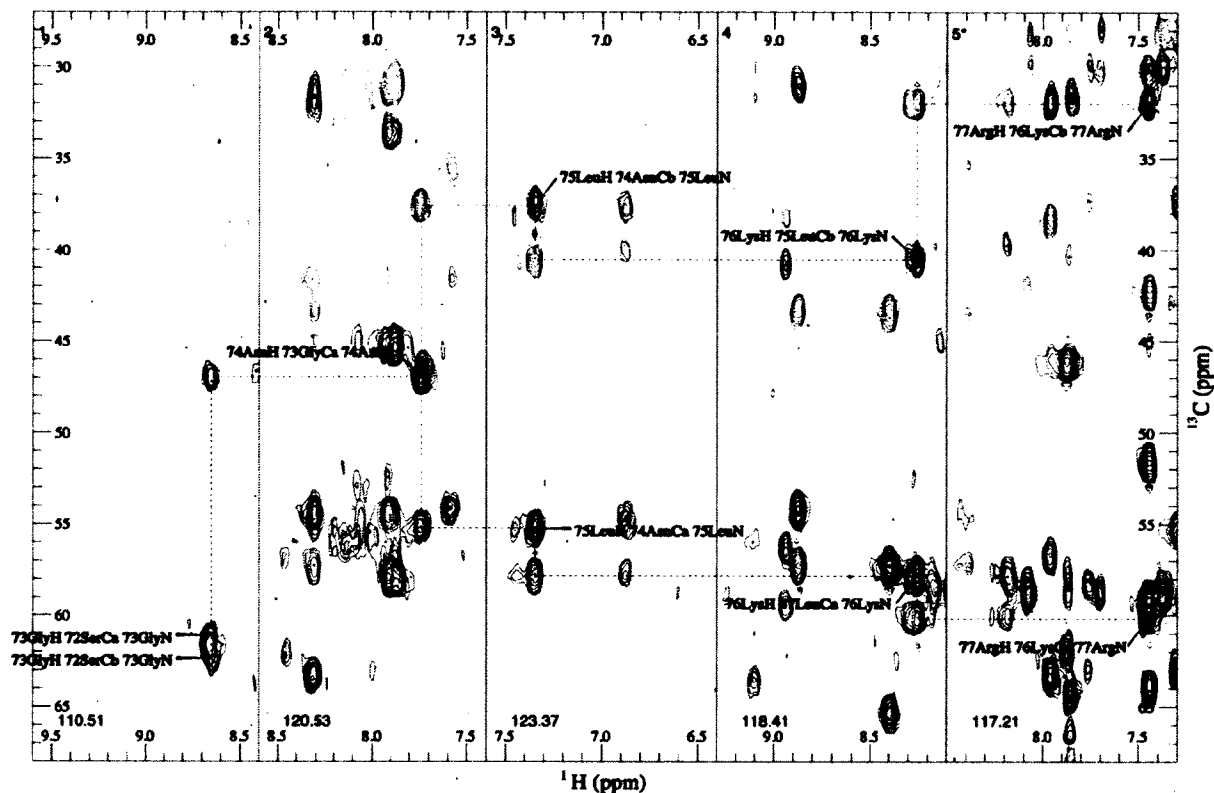


Figure 26 : Exemple de CBCA(CO)NH d'une séquence SGNLKR de Miz 5-8

Superposition d'un spectre CBCA(CO)NH et du spectre HNCACB de la figure 25. Les résonances en vert et celles en rouges appartiennent au HNCACB, celles en bleu appartiennent au CBCA(CO)NH. En abscisse se retrouve le déplacement chimique des protons et en ordonnée celui du  $^{13}\text{C}$ . En z est le déplacement chimique en  $^{15}\text{N}$ , qui est indiqué dans le coin inférieur gauche de chacune des tranches. Dans chacune des tranches sont identifiées les résonances du spectre CBCA(CO)NH appartenant aux résidus 73 à 77 de la protéine Miz 5-8. Les lignes pointillées servent à indiquer l'itinéraire effectué pour identifier les résonances appartenant aux résidus consécutifs.

deux types de carbone sont positives. Donc, en superposant le CBCA(CO)NH au HNCACB, il est possible de dire quelle résonance appartient au résidu en i-1, ce qui permet une attribution plus aisée (voir un exemple sur la figure 26, le CBCA(CO)NH est en bleu et le HNCACB est en vert et rouge, et seules les résonances d'intérêt sur le CBCA(CO)NH ont été identifiées).

#### 2.4.4 HNCO

(JACOBSEN, 2007, p.611-614)

Le déplacement chimique des atomes du squelette polypeptidique permet d'obtenir plusieurs informations intéressantes au niveau structural, entre autres les angles dièdres  $\phi$  et  $\psi$ . Ces angles sont en général déterminés automatiquement à l'aide d'un programme informatique. La connaissance de ces angles permet également de déterminer la structure secondaire de la protéine. Le HNCACB et le CBCA(CO)NH permettent de déterminer l'ensemble des atomes du squelette polypeptidique, à l'exception du carbone du groupement carbonyle (le déplacement chimique de l'oxygène n'est pas requis pour les calculs, et de toute façon, le  $^{16}\text{O}$  et le  $^{18}\text{O}$  sont de spin nul, et le  $^{17}\text{O}$  est de spin 5/2, tous les autres isotopes de l'oxygène sont instables). C'est le spectre HNCO qui permet de combler cette information. Ce spectre corrèle tout simplement les paires N-H au carbonyle du résidu en  $i-1$  et est très simple à analyser.

#### 2.4.5 C(CO)NH

(MOY et al., 1995)

Suite à l'analyse des spectres précédents, il ne reste que les atomes lourds des chaînes latérales à identifier avant de passer aux hydrogènes. Le C(CO)NH permet d'attribuer les carbones de ces chaînes latérales, et en général les biologistes structuraux omettent les hétéroatomes. Ce spectre corrèle les paires N-H amides avec l'ensemble des carbones du résidu situé en  $i-1$  de ces amides, à la condition que ces carbones soient adjacents et qu'un hydrogène y soit lié. Ainsi, les carbones quaternaires (tel le  $\text{C}_\gamma$  d'une phénylalanine) ou séparés du groupement amide par un hétéroatome (tel le  $\text{C}_\epsilon$  d'une méthionine) ne sont pas détectables sur ce spectre, et il en va de même que tous les carbones situés au-delà de ces atomes (l'ensemble des carbones aromatiques de la phénylalanine mentionnée précédemment, par exemple). Ce spectre a en outre l'avantage de permettre la confirmation de l'attribution effectuée à l'aide des spectres précédents, une lysine par exemple aura cinq résonances sur ce spectre.

### 2.4.6 H(CCO)NH

(MOY et al., 1995)

Après avoir déterminé le déplacement chimique de la majorité des atomes lourds à l'aide des spectres précédents, les protons doivent être attribués également, car ce sont eux qui donneront l'information nécessaire à la détermination de la structure tertiaire de la protéine. Le H(CCO)NH est l'un des spectres permettant ceci. Il s'agit d'un spectre 3D qui corrèle les paires N-H à l'ensemble des hydrogènes liés à des carbones du résidu situé en  $i-1$ . Tout comme dans le cas du C(CO)NH, les atomes situés au-delà d'un carbone quaternaire ou d'un hétéroatome ne seront pas visibles sur ce spectre. Tout comme pour les atomes de carbones, il existe des statistiques sur la BMRB qui facilitent l'identification du déplacement chimique des atomes d'hydrogène en fonction de l'acide aminé et de leur position sur l'acide aminé. Ce spectre présente quelques inconvénients. Tout d'abord, plusieurs résonances, dépendantes de l'acide aminé et de la position, sont souvent difficiles à voir ou invisibles. C'est le cas du  $H_\alpha$  des cystéines, des alanines, de l'histidine et des résidus aromatiques; des  $H_\beta$  des glutamates et glutamines, des  $H_\gamma$  des méthionines, ainsi que la quasi-totalité des protons de l'isoleucine. Il est à noter que cette dernière affirmation repose sur ce que j'ai pu observer sur mes propres spectres. Ensuite, il arrive que deux protons liés au même carbone aient un déplacement chimique qui leur est propre, ce qui arrive très souvent avec les prolines. Il est alors difficile de déterminer s'il s'agit, par exemple, des deux  $H_\beta$  ou des deux  $H_\delta$  qui sont distincts, et les deux autres qui sont confondus.

### 2.4.7 HCCH-TOCSY

(JACOBSEN, 2007, p.609-610; MOY et al., 1995)

Le HCCH-TOCSY (« Total Correlation Spectroscopy ») est l'autre spectre permettant de déterminer le déplacement chimique des atomes d'hydrogène. Il s'agit encore une fois d'un spectre 3D, qui présente le  $^1\text{H}$  à la fois en x et en y, et le  $^{13}\text{C}$  en z, dans la disposition habituelle. Tout comme le H(CCO)NH, il montre l'ensemble des protons liés à des carbones sur un même résidu, mais dans ce cas-ci ils sont corrélés à une

paire C-H plutôt qu'à une paire N-H. Encore une fois, le transfert de la magnétisation est arrêté par la présence d'un carbone quaternaire ou d'un hétéroatome. Ce spectre est légèrement plus complexe à analyser qu'un H(CCO)NH, étant donnée la présence de plus d'un carbone par résidu (glycine exceptée). Tout d'abord, sur la diagonale se retrouvent les résonances correspondant aux paires C-H. Sur la même ligne horizontale se retrouvent les résonances des hydrogènes du même résidu qui ne sont séparés de la paire C-H que par des carbones non quaternaires. Prenons par exemple la glutamine présentée sur le spectre HCCH-TOCSY visible à la figure 27 (seules les résonances appartenant à la glutamine ont été identifiées). La rangée du bas présente la tranche correspondant au déplacement chimique du  $C_\alpha$  de la glutamine; la résonance sur la diagonale appartient donc à la paire  $C_\alpha$ - $H_\alpha$ . Les protons  $H_\beta$  et  $H_\gamma$  de la même glutamine lui sont corrélées et sont donc situés sur la même ligne horizontale. En se rendant à la tranche correspondant

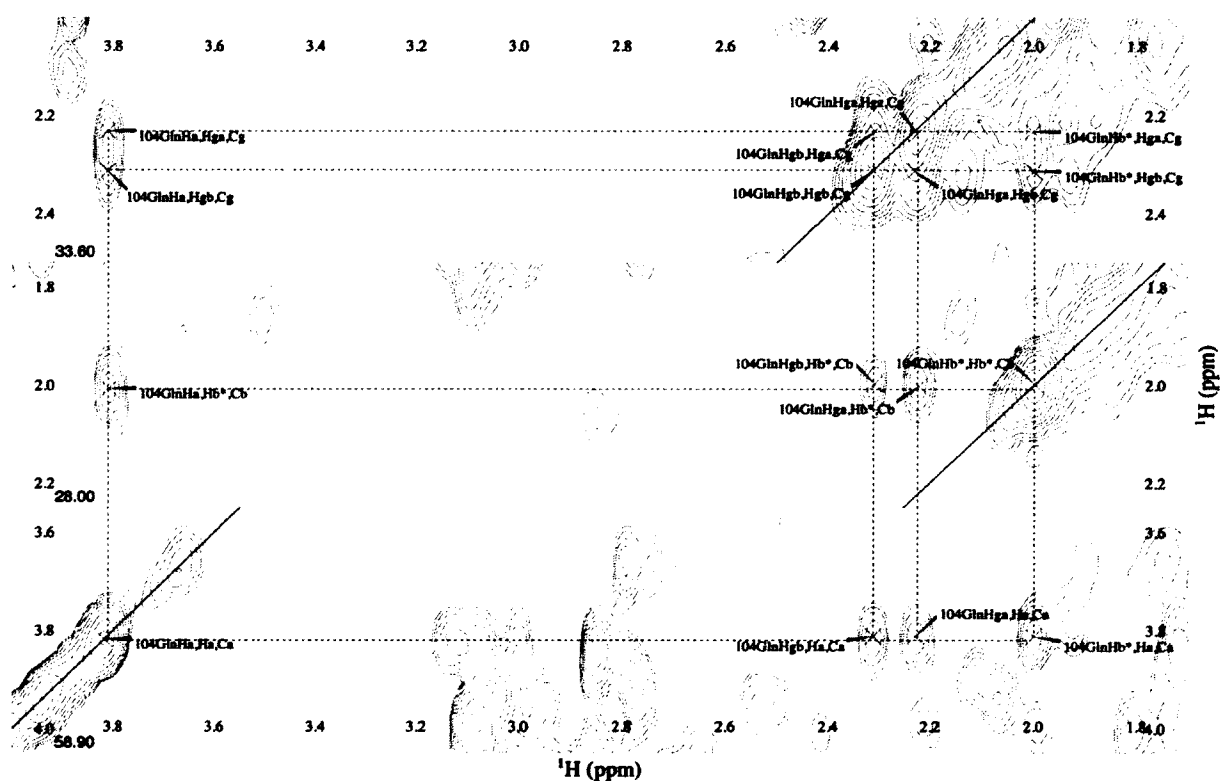


Figure 27 : Exemple de HCCH-TOCSY d'une glutamine de Miz 5-8

Cette figure juxtapose trois tranches d'un spectre 3D HCCH-TOCSY. L'axe des abscisses et l'axe des ordonnées représentent tous deux le déplacement chimique des protons, et en z se retrouve celui du  $^{13}\text{C}$ , qui est indiqué dans le coin inférieur gauche de chaque tranche. Toutes les résonances appartenant à la Gln104 de la protéine Miz 5-8 sont identifiées. Les pointillés ont été ajoutés pour permettre de voir la correspondance entre les résonances.

au  $C_\beta$ , la résonance sur la diagonale correspond à la paire  $C_\beta$ - $H_\beta$ , et on retrouve les résonances appartenant aux protons  $H_\alpha$  et  $H_\gamma$  sur la même ligne. Les deux protons  $H_\beta$  étant équivalents, on retrouve une seule résonance sur la diagonale appartenant à ce résidu. Il en est tout autrement des protons  $H_\gamma$ , qui sont distincts. Ceci peut être observé par la présence de deux lignes horizontales, au même déplacement chimique en carbone, qui appartiennent au même résidu, comme on peut le voir dans la rangée du haut.

Cet exemple montre le grand avantage du HCCH-TOCSY sur le H(CCO)NH : il permet d'identifier sans ambiguïté les protons non équivalents liés à un même carbone. Il a toutefois plusieurs désavantages. D'abord, les  $H_\alpha$  sont souvent difficiles à voir, soit à cause de la faiblesse de leurs résonances (celles-ci sont atténuées par la suppression des protons de l'eau, qui ont un déplacement chimique similaire; il s'agit d'une étape qui est ajoutée à la plupart des séquences de pulses afin de rendre le solvant invisible sur les spectres) ou du fait qu'elles soient perdues dans une ligne de base souvent non nulle dans cette région du spectre. En outre, les résonances sont très nombreuses et il devient ainsi difficile de les discriminer.

#### 2.4.8 $^{15}\text{N}$ - et $^{13}\text{C}$ -edited NOESY

(JACOBSEN, 2007, p.428-429, 436-437, 586-590)

Tous les spectres décrits précédemment ont une caractéristique commune : ils mettent en corrélation des atomes qui sont reliés les uns aux autres via un couplage scalaire (J), et donc qui sont situés près les uns des autres dans la séquence de la molécule. Ceci ne permettrait que de déterminer la structure primaire de la protéine, qui est généralement déjà connue. Pour déterminer la structure tertiaire, il faut connaître quels atomes sont proches les uns des autres dans l'espace. C'est cette information que donnent les différents spectres NOESY (« Nuclear Overhauser Effect Spectroscopy »). L'effet Overhauser (appelé simplement un NOE dans le jargon des spectroscopistes) est un transfert de magnétisation dû à une relaxation simultanée de deux atomes (via des transitions zéro quantum ou double quantum), et se fait à travers l'espace plutôt qu'à travers des liaisons chimiques. L'intensité des résonances sera proportionnelle à  $\frac{1}{r^6}$ , où r



est la distance entre les atomes. En général, on considère que la distance maximale à laquelle un NOE est visible sera d'environ 6Å ( $1\text{Å} = 10^{-10}\text{ m}$ ). Il est possible d'ajuster un paramètre appelé le temps de mélange (« mixing time ») dans la séquence de pulses afin de voir uniquement des NOEs sur des distances plus courtes.

Un NOESY est à la base un spectre 2D, qui montre le déplacement chimique en  $^1\text{H}$  en x et en y. Sur la diagonale se retrouvera la résonance du proton étudié, et chacune des résonances situées sur la même ligne verticale correspondra à un proton situé proche du proton étudié. Il suffit de les identifier à l'aide de leur déplacement chimique, attribué grâce au H(CCO)NH et au HCCH-TOCSY. Toutefois, dans le cas d'une protéine, analyser un NOESY 2D relève de l'utopie, étant donné la très grande quantité de protons contenue dans une protéine typique. C'est pourquoi la séquence de pulses est altérée de façon à créer une troisième dimension. Ainsi, les résonances sur la diagonale ne correspondront non pas à un proton, mais à une paire C-H ou N-H, et le déplacement chimique du noyau lourd sera la coordonnée en z. Ces spectres sont donc respectivement appelés «  $^{13}\text{C}$ -edited NOESY » et «  $^{15}\text{N}$ -edited NOESY ». Le  $^{13}\text{C}$ -edited NOESY est généralement scindé en deux spectres, dans l'un des deux, seuls les carbones aliphatiques ( $\delta < 80\text{ppm}$ ) seront visibles, alors que l'autre montrera uniquement les carbones aromatiques ( $\delta > 100\text{ppm}$ ).

#### **2.4.9 Mesure des paramètres dynamiques d'un échantillon par RMN**

(JACOBSEN, 2007, p.162-175; HORE, 2001, p.61-64)

Beaucoup d'information sur le comportement dynamique d'une protéine peut être obtenue à partir de la relaxation des spins  $^{15}\text{N}$  du groupement amide de chacun des acides aminés. Le premier de ces paramètres est relié à un autre type de NOE que ceux décrits dans la section 2.4.8. Ce paramètre est appelé le NOE hétéronucléaire (hNOE) car les NOEs qui sont étudiés dans les spectres NOESY décrits à la section 2.4.8 se produisent entre protons, alors que celui-ci se produit par relaxation simultanée d'un proton et d'un atome lourd ( $^{15}\text{N}$ ). Pour mesurer le hNOE, deux spectres semblables au HSQC sont enregistrés, le premier sans présaturation des protons amide, et le second avec

présaturation. Le hNOE est le rapport de l'intensité des résonances avec/sans présaturation. Cette étape, en l'absence de mouvements internes, crée un état stationnaire qui annule la magnétisation nette du proton. Pour des molécules de la taille des protéines, si le vecteur N-H était parfaitement rigide, la valeur du hNOE serait élevée (à peu près égale à 0.87). La présence de mouvements internes diminuera la valeur du hNOE (qui peut alors descendre jusqu'à -3.5 dans le cas d'une protéine complètement désordonnée). En général, un hNOE faible ( $< 0.6$ ) signifie que l'amide est situé dans une région flexible de la protéine, et inversement un hNOE fort ( $> 0.6$ ) signifie qu'il est plutôt situé dans une région plus rigide.

Les deux autres paramètres dynamiques les plus fréquemment étudiés sont tout simplement les valeurs de  $T_1$  et de  $T_2$ , constantes décrites dans la section 2.1.4, des noyaux  $^{15}\text{N}$  appartenant au squelette polypeptidique. Dans les deux cas, ces valeurs sont obtenues à partir de deux séries de spectres similaires à des HSQC. La séquence de pulses de ces spectres est modifiée de façon à contenir des délais pendant lesquels les spins relaxeront uniquement selon  $T_1$  ou selon  $T_2$  et  $T_1$  (mais comme  $T_1$  est plus long que  $T_2$ , en

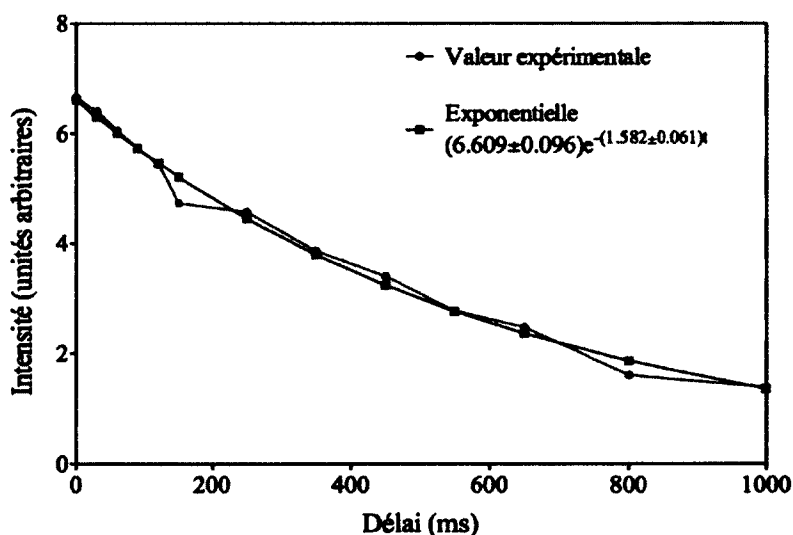


Figure 28 : Exemple de détermination de la valeur de  $T_1$  de l'amide d'une tyrosine de Miz 5-8

Graphique de l'intensité de la résonance appartenant à la Tyr4 de la protéine Miz 5-8 en fonction du délai appliqué dans la séquence de pulses servant à mesurer  $T_1$ . La courbe en rouge correspond aux données mesurées sur lesdits spectres, et la courbe en bleu est l'exponentielle décroissante dont les paramètres correspondent le mieux aux données expérimentales. Le paramètre multipliant  $t$  dans l'exponentielle est  $R_1$ .

utilisant des délais assez courts, il est possible de négliger l'évolution selon  $T_1$  dans ce dernier cas). Le délai de chaque spectre enregistré sera différent, et en traçant le graphique de l'intensité de chaque résonance en fonction de la durée du délai, il sera possible de l'approximer à une courbe exponentielle et de connaître le  $T_1$  ou le  $T_2$  de ce résidu à partir de l'équation de cette exponentielle. Un exemple est présenté à la figure 28.

#### 2.4.10 CPMG-HSQC et dispersion de relaxation

(JACOBSEN, 2007, p.229-231; WIKSTRÖM et al., 1999, LUZ et MEIBOOM, 1963; MULDER et al., 2001)

Il arrive qu'un même atome puisse avoir plus d'une valeur de déplacement chimique. Ceci peut être dû à un échange conformationnel entre deux possibilités de structures, ou alors à l'échange d'un proton labile entre la molécule étudiée et le solvant. Dépendamment de la vitesse à laquelle s'effectue cet échange, cela aura des conséquences différentes. Si l'échange est lent par rapport à la fréquence à laquelle les spectres sont mesurés dans l'appareil, une résonance sera visible à chacun des déplacements chimiques adoptés par l'atome, et l'intensité de ces résonances sera proportionnelle à la population occupant l'état qui lui est associé. Si le taux d'échange est très rapide, une seule résonance sera observée, au déplacement chimique correspondant à la valeur moyenne des déplacements chimiques limites, pondérée par la population de chacun des états. Finalement, si le taux d'échange est intermédiaire, les résonances sembleront disparaître, car en fait elles s'étaleront entre les valeurs possibles de déplacement chimique, résultant en une diminution importante de l'intensité maximale. Cette disparition est appelée *coalescence* (JACOBSEN et al., 2007).

Dans plusieurs cas, les résonances peuvent redevenir visibles à l'aide d'un spectre appelé CPMG-HSQC (Carr-Purcell-Meiboom-Gill, du nom des scientifiques qui ont développé cette technique). Ce spectre introduit plusieurs pulses  $180^\circ$  afin de constamment ramener la magnétisation à son état originel, ce qui permet d'observer les résonances des protons qui sont en coalescence dans un spectre normal. Ceci sert simplement à confirmer la présence d'échange conformationnel. Il est possible de mesurer

$R_{ex}$  (voir section 2.1.4) en effectuant une série de spectres CPMG ayant différentes fréquences de répétition desdits pulses  $180^\circ$  ( $\nu_{CPMG}$ ) et en traçant le graphique de  $R_{2,eff}$ , le *taux de relaxation transverse effectif*, en fonction de  $\nu_{CPMG}$ . Cette expérience s'appelle la *dispersion de relaxation*.

$$R_{2,eff}(\nu_{CPMG}) = \frac{1}{T} \ln \frac{I(\nu_{CPMG})}{I_0}$$

Ici,  $T$  est la période d'évolution CPMG (qui doit être la même pour chacun des spectres CPMG enregistrés),  $I_0$  est l'intensité des résonances en l'absence de pulses CPMG, et  $I(\nu_{CPMG})$  est l'intensité des résonances en présence de ces pulses. Si on voit  $R_{2,eff}$  comme l'inverse d'un hypothétique  $T_2$  effectif, la valeur de  $R_{2,eff}$  sera grande, et donc le  $T_{2,eff}$  sera court, pour les valeurs faibles de  $\nu_{CPMG}$ , puisqu'il y aura peu de pulses qui ramèneront la magnétisation en phase pendant la période  $T$  si de l'échange conformationnel se produit. Inversement, lorsque  $\nu_{CPMG}$  est grand, la valeur de  $R_{2,eff}$  sera plus faible, et donc  $T_{2,eff}$  plus long, puisque la perte de cohérence due à l'échange conformationnel est éliminée des causes de raccourcissement de  $T_2$ .

La valeur de  $R_{ex}$  peut être déterminée en calculant les paramètres d'une courbe théorique qui passerait par les points du graphique de  $R_{2,eff}$  en fonction de  $\nu_{CPMG}$ . L'équation de cette courbe est la suivante dans le cas dans le cas d'un échange entre deux conformations, où la population  $P_B$  de l'une des deux conformations est faible par rapport à la population  $P_A$  de l'autre état :

$$R_{2,eff}(\nu_{CPMG}) = R_{2,eff}(\nu_{CPMG} \rightarrow \infty) + R_{ex} \left( 1 - \frac{4\nu_{CPMG}}{k_{ex}} \tanh \frac{k_{ex}}{4\nu_{CPMG}} \right)$$

$$R_{ex} = \frac{P_A P_B \Delta\omega^2}{k_{ex}}$$

$R_{2,eff}(\nu_{CPMG} \rightarrow \infty)$  est la valeur de  $R_{2,eff}$  lorsque la fréquence de répétitions des pulses  $180^\circ$  devient infinie,  $R_{ex}$  est la valeur recherchée,  $k_{ex}$  est la constante d'échange et

$\Delta\omega$  est la différence du déplacement chimique du  $^{15}\text{N}$  entre les deux états. Ces paramètres peuvent être calculés à l'aide d'un logiciel permettant de tracer des graphiques. Des exemples de graphiques des valeurs obtenues de  $R_{2,\text{eff}}$  et de la courbe théorique qui satisfait ces valeurs en fonction de  $\nu_{\text{CPMG}}$  est montré à la figure 5c de l'article.

### 3. Présentation du projet

#### 3.1 Résumé du projet

Le but ultime du projet est la compréhension structurale du mécanisme de transrépression par le complexe c-Myc/Miz-1, et de savoir comment celui-ci se lie à l'ADN. Toutefois, il s'agit là d'un projet beaucoup plus vaste que ce qui peut être déterminé au cours d'une seule maîtrise. C'est pourquoi ce mémoire s'intéresse uniquement à l'une des étapes de ce projet, c'est-à-dire la détermination de la structure des ZFs 5 à 8 de Miz-1 par RMN en l'absence d'ADN. Ce domaine de Miz-1 sera dénommé Miz 5-8 tout au long de ce mémoire. La raison pour laquelle ce domaine en particulier a été choisi est que des prédictions bioinformatiques effectuées précédemment au laboratoire ont déterminé que les triplets de ZFs 5 à 7 et 10 à 12 étaient ceux qui auraient le plus de chances de se lier au promoteur de  $p15^{\text{INK4B}}$ . Ces prédictions sont basées sur le code de reconnaissance de chaque acide aminé en fonction de sa position dans l'hélice  $\alpha$  établi par le groupe de Kaplan (KAPLAN et al., 2005).

Suite aux expériences RMN menées sur Miz 5-8, nous avons ajouté la caractérisation de la dynamique de la protéine parmi les objectifs à atteindre. En effet, l'analyse de ces expériences a permis de découvrir dans cette protéine une dynamique présente à un niveau surprenant pour des ZFs, et nous avons décidé d'étudier davantage cette particularité.

### 3.2 Motivations

Le but premier du projet dans son ensemble est de comprendre le mécanisme d'action de l'oncogène c-Myc, puisqu'il est omniprésent dans le cancer. De cette connaissance découlerait la possibilité de développer des moyens d'action sur ce mécanisme dans le but de l'inhiber, et ainsi développer une thérapie contre les cancers c-Myc-dépendants. Il a été démontré, par Soucek entre autres (SOUCEK et al., 2013), que cibler c-Myc est un moyen particulièrement efficace de combattre les cellules cancéreuses. Grâce à une thérapie génique, elle a montré que les cellules cancéreuses sont dépendantes à l'oncogène et qu'elles ne survivent pas lorsqu'il est suffisamment inhibé, contrairement aux cellules saines. En effet, c-Myc a la particularité d'avoir une fonction non-redondante, ce qui explique l'intensité de la réponse que Soucek a observée. L'inhibiteur qu'elle a utilisé était la protéine Omomyc, une protéine agissant comme dominant-négatif de la transactivation par c-Myc. Omomyc conserve toutefois la possibilité de causer la transrépression c-Myc-dépendante. Comprendre ce dernier mécanisme dans l'optique de le cibler dans un traitement anticancéreux est donc un projet particulièrement intéressant, et c'est la raison pour laquelle Miz-1 est le sujet de ce projet. Comme la liaison à l'ADN devrait être médiée par les ZFs 5 à 7 selon les prédictions, la structure de Miz 5-8 est une étape cruciale du projet.

#### **4. Article**

Structural and Dynamical Characterization of the Miz-1 Zinc Fingers 5 to 8 by Solution-State NMR

David Bernard, Mikaël Bédard, Josée Bilodeau and Pierre Lavigne

Statut du manuscrit : accepté le 6 août 2013

Publié sous le doi : 10.1007/s10858-013-9770-6

Titre du journal : Journal of Biomolecular NMR

##### **4.1 Contribution des auteurs**

Dans le cadre de cette étude, Josée Bilodeau et moi avons commencé par mettre au point un protocole d'expression dans des bactéries *E. coli* d'un gène de Miz 5-8 cloné par Jean-François Naud, et j'ai commencé la mise au point du protocole de purification sur FPLC. Suite à la réalisation qu'il manquait quelques acides aminés à la protéine, Mikaël Bédard et moi avons créé un nouveau clone de Miz 5-8, et Josée Bilodeau, Mikaël Bédard et moi avons poursuivi la mise au point du protocole d'expression. J'ai raffiné le protocole de purification par l'ajout d'une étape sur HPLC (ce pour quoi je remercie le laboratoire du professeur Yves Dory pour m'avoir permis d'utiliser leur appareil). J'ai effectué l'analyse de l'ensemble des spectres RMN sur le logiciel CcpNmr avec les conseils de Mikaël Bédard et de mon directeur de maîtrise, le Pr. Pierre Lavigne, et il en a été de même pour les calculs de structure sur les logiciels CNS et ARIA. Pierre Lavigne a interprété les résultats de dynamique. Le manuscrit a été rédigé à parts égales par Pierre Lavigne et moi. J'ai également fait l'ensemble des figures.

## 4.2 Résumé de l'article

Miz-1 est un facteur de transcription de la famille BTB/POZ qui active la transcription de gènes cytotostatiques, tels p15<sup>INK4b</sup> ou p21<sup>CIP1</sup>. L'extrémité C-terminale de Miz-1 contient la séquence consensus de 13 domaines de type doigt de zinc (ZF) Cys<sub>2</sub>His<sub>2</sub>. Les ZFs 1 à 4 interagissent avec SMAD3/4, et il est attendu que le reste serve à reconnaître les promoteurs des gènes-cibles. Nous avons noté des caractéristiques inhabituelles dans le ZF 5 et dans l'espaceur entre les ZFs 5 et 6. Effectivement, on retrouve un glutamate à la place du résidu basique conservé situé deux positions avant la dernière histidine de l'hélice du ZF 5, et la séquence de l'espaceur est DTDKE au lieu du traditionnel TGEKP. Dans une structure  $\beta\beta\alpha$  canonique, des éléments de structure primaire si inhabituels causeront de sévères répulsions électrostatiques. Dans ce contexte, nous avons caractérisé la structure et la dynamique d'une construction de Miz-1 comprenant les ZFs 5 à 8 (Miz 5-8) par RMN en solution. Les ZFs 5, 7 et 8 adoptent la conformation  $\beta\beta\alpha$  classique aux ZFs Cys<sub>2</sub>His<sub>2</sub>, mais le nombre de NOEs à longue distance était insuffisant pour définir un tel repliement pour le ZF 6. Nous montrons grâce à des expériences de dispersion de relaxation <sup>15</sup>N que cette lacune en NOEs est due à la présence importante de mouvements à l'échelle micro-/milliseconde. Puisque cette région chargée négativement devrait être située près des phosphodiesteres du squelette lorsque liée à l'ADN, nous suggérons qu'en plus de favoriser les recherches conformationnelles, cette région sert à éloigner les ZFs 1 à 4 de l'ADN.



## **Structural and Dynamical Characterization of the Miz-1 Zinc Fingers 5 to 8 by Solution-State NMR**

David Bernard, Mikaël Bédard, Josée Bilodeau and Pierre Lavigne\*

Département de Biochimie, Institut de Pharmacologie de Sherbrooke, Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, 3001, 12e avenue Nord, Sherbrooke, QC, Canada, J1H 5N4, Regroupement Stratégique sur la Fonction, la Structure et l'Ingénierie des Protéines (PROTEO) and Groupe de Recherche Axé sur la Structure des Protéines (GRASP).

Phone number: (819) 820-6868 ext. 15462

Fax number: (819) 564-5400

E-mail address of corresponding author: [pierre.lavigne@usherbrooke.ca](mailto:pierre.lavigne@usherbrooke.ca)

*Keywords: Miz-1, solution structure, zinc fingers, protein dynamics, conformational exchange*

Abbreviations: CD – circular dichroism; Miz 5-8 – Miz-1 zinc fingers 5 to 8; NMR – nuclear magnetic resonance spectroscopy; water-TFA – 0.1% trifluoroacetic acid in water; ZF – zinc finger

## Abstract

Miz-1 is a BTB/POZ transcription factor that activates the transcription of cytostatic genes, such as p15<sup>INK4B</sup> or p21<sup>CIP1</sup>. The C-terminus of Miz-1 contains 13 consensus C<sub>2</sub>H<sub>2</sub> zinc finger domains (ZF). ZFs 1 to 4 have been shown to interact with SMAD3/4, while the remaining ZFs are expected to bind the promoters of target genes. We have noted unusual features in ZF 5 and the linker between ZFs 5 and 6. Indeed, a glutamate is found instead of the conserved basic residue two positions before the second zinc-coordinating histidine on the ZF 5 helix, and the linker sequence is DTDKE in place of the classical TGEKP sequence. In a canonical  $\beta\beta\alpha$  fold, such unusual primary structure elements should cause severe electrostatic repulsions. In this context, we have characterized the structure and the dynamics of a Miz-1 construct comprising ZFs 5 to 8 (Miz 5-8) by solution-state NMR. Whilst ZFs 5, 7 and 8 were shown to adopt the classical  $\beta\beta\alpha$  fold for C<sub>2</sub>H<sub>2</sub> ZFs, the number of long-range NOEs was insufficient to define a classical fold for ZF 6. We show by using <sup>15</sup>N-relaxation dispersion experiments that this lack of NOEs is due to the presence of extensive motions on the  $\mu$ s-ms timescale. Since this negatively charged region would have to be located near the phosphodiester backbone in a DNA complex, we propose that in addition to promoting conformational searches, it could serve as a hinge region to keep ZFs 1-4 away from DNA.

## Introduction

Myc-interacting Zinc Finger Protein-1 (Miz-1) is a member of the BTB/POZ transcription factor family. It is 803-amino-acid long and weighs 93 kDa. Like all members of the family, it possesses a BTB/POZ protein interaction domain at its N-terminus and a number (thirteen) of putative zinc finger domains (ZF) in its C-terminus. The crystal structure of the Miz-1 BTB/POZ has been solved (Stead et al. 2007; Stogios et al. 2010). Twelve of the Miz-1 ZFs are consecutive, and the thirteenth is remote from the twelfth by a sequence that has been identified as a Myc-interacting domain. Another such domain exists between the BTB/POZ domain and the first ZF (Peukert et al. 1997). Miz-1, which can be activated through the TGF- $\beta$  pathway, activates the transcription of various cytostatic genes, such as p15<sup>INK4B</sup> and p21<sup>CIP1</sup>. Miz-1 and Sp1 both bind the core promoter of those genes (Bowen 2003) and nucleophosmin binds to the BTB/POZ domain of Miz-1 (Wanzel et al. 2008). This allows histone acetyltransferase p300 to be recruited, which activates transcription. When c-Myc binds Miz-1 (as a c-Myc/Max heterodimer), the recruitment of p300 becomes impossible and the complex represses transcription (Seoane et al. 2001; Staller et al. 2001).

It is thought that a subset of the thirteen putative C<sub>2</sub>H<sub>2</sub> ZFs is responsible for the DNA-binding activity associated with the Miz-1 protein. Indeed, ZFs are very common DNA-binding motifs. In fact, it is the most common motif found in the human genome (Razin et al. 2012; Wolfe et al. 2000). C<sub>2</sub>H<sub>2</sub> ZFs are ~30-amino-acid long with the consensus sequence F/Y-X-C-X<sub>2,4</sub>-C-X<sub>3</sub>-F-X<sub>5</sub>-L-X<sub>2</sub>-H-X<sub>3</sub>-H-X<sub>5</sub>, and they usually adopt a  $\beta\beta\alpha$  conformation (Berg 1988). This conformation is held in place by the two cysteines and the two histidines, which coordinate a Zn<sup>2+</sup> ion, as well as three conserved large hydrophobic amino acids. This makes the N-terminus of the  $\alpha$  helix protrude, and allows it to enter the major groove of the DNA. In a canonical ZF, four residues will contact the DNA bases, allowing recognition of a particular sequence. Residues situated at positions -1, 3 and 6 relative to the first residue of the helix usually contact consecutive bases on one strand, and the residue in position 2 will contact one base on the other strand. This last base is in fact the complementary base to the one contacted by the residue in position 6 of

the preceding ZF. All four of these contacts can be base-specific. ZFs usually bind DNA by groups of three, allowing specific recognition of sequences 10bp long (reviewed in Wolfe et al. (Wolfe et al. 2000)).

ZFs are typically separated by a conserved linker with the consensus sequence TGEKP. Interestingly, the lysine in the linker is often involved in non-specific salt bridges with phosphate groups when ZFs are bound to DNA. Moreover, a salt bridge between the conserved glutamate and a conserved basic side chain (arginine or lysine) in the preceding helix has been demonstrated to be important in the modulation of inter-ZF movements. Indeed, when a salt bridge is not possible, the inter-ZF movements are amplified and this is thought to be important for the mechanism of recognition of specific DNA sequences by polydactyl ZF proteins (Zandarashvili et al. 2012).

Recent bioinformatics studies have demonstrated that the residues at positions -1, 2, 3, 5 and 6 in the  $\alpha$ -helix are the only residues in ZFs displaying selective pressure during evolution (Emerson and Thomas 2009). While this could be interpreted as an evolutive pressure to bind diverse DNA sequences, it could also mean that they have evolved for other functions, such as protein binding. In fact, it is becoming increasingly recognized that ZFs are also important actors in protein-protein interactions (Brayer and Segal 2008). In accordance with this growing concept, the first 4 ZFs of Miz-1 have been shown to interact with the MH1 domain of SMAD3 and 4 (Seoane et al. 2001).

While the binding of Miz-1 has been narrowed down to two distant sites at the promoters of p15<sup>INK4B</sup> (-155/-140, -1/13) and p21<sup>CIP1</sup> (-105/-81, -46/-32) (Seoane et al. 2002), the exact ZFs involved have yet to be identified. Interestingly, application of the most recent bioinformatics programs fails to generate a consistent prediction as to which ZF subset (equal to or larger than 3) have the highest probability of binding the identified sequences on the p15<sup>INK4B</sup> and p21<sup>CIP1</sup> promoters.

In this context, we have undertaken a structural and dynamical characterization of the Miz-1 ZFs in absence of DNA with determination of the NMR structure of ZF

domains 8 to 10 (Bédard et al. 2012). The three ZFs were demonstrated to adopt the typical  $\beta\beta\alpha$  fold. In accordance with previous reports (Hyre and Klevit 1998; Potter 2005; Zandarashvili et al. 2012), the two canonical linkers (TGEKP) were shown to be the most flexible regions on the ps-ns timescale without, however, being totally unfolded. In the present study, we extend our structural and dynamical characterization of Miz-1 ZFs with a construct composed of ZFs 5 to 8 (henceforth referred to as Miz 5-8).

Whereas ZF 5, ZF 7 and ZF 8 are shown to adopt a stable  $\beta\beta\alpha$  fold, the structure of ZF 6 appears to undergo a slow ( $\mu\text{s}$ -ms) conformational change. Moreover, we also note the presence of extensive conformational exchange in the  $\mu\text{s}$ -ms timescale at the interface between the helices of ZF 5, ZF 6 and ZF 7 and the contiguous linkers. Although such movements have been noted between ZF 1 and ZF 2 of Egr-1 (Zif268) (Zandarashvili et al. 2012), those observed in the Miz 5-8 construct appear exacerbated. In fact, the movements in the Egr-1 construct have been attributed to the absence of the conserved E in the linker (TGQKP) between ZF 1 and ZF 2. Interestingly, two out of three linkers in the Miz 5-8 construct have more peculiar sequences; DTDKE and IADGP. As described in detail, electrostatic repulsions located at the interface between the helix of ZF 5 and the contiguous linker are most likely responsible for the local unfolding in ZF 6. While a role for such unusual slow conformational exchange in the DNA and/or protein binding remains to be determined, we show by homology modeling that DNA binding by ZF 5 and 6 is highly unlikely. Indeed, because of the sequence of the linker between ZF 5 and 6, intramolecular and intermolecular electrostatic repulsion in a canonical DNA-bound configuration would prevent DNA binding. We propose that this peculiar sequence and resultant dynamical behavior might be important for Miz-1 function by promoting the availability of ZF 1-4 to engage in protein-protein interaction with SMAD3/4.

## Materials and methods

### Preparation of the Miz 5-8 construct

The cDNA fragment of ZFs 5 to 8 of Miz-1 (residues 416-526, Miz 5-8, see primary structure in Fig. 1a) was generated by PCR from the complete cDNA of Miz-1 using primers F (5'-CAT ATG AAG CCC TAC CAG TGC GAC TAC TGC GGC CG-3') and R (5'-GGA TCC CTA ACC TGT GTG AAT GCG GAC GTG CCG CTG-3'). The PCR product was inserted into a pDrive (Qiagen) vector prior to a digestion by NdeI and BamHI. The digested product was inserted into a pET-3a (Novagen) plasmid by the same restriction sites and the construct transformed in BL21 star (DE3\*) one shot competent cells (Invitrogen). Bacteria were grown in LB or in M9 broth (supplemented with 10  $\mu$ M FeSO<sub>4</sub> and 100  $\mu$ M CaCl<sub>2</sub>) containing <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C-glucose to an O.D. of 0.6 at 600 nm; they were induced with 1 mM IPTG and 100  $\mu$ M ZnCl<sub>2</sub> for 21 hours at 30°C and they were harvested by centrifugation. The cell pellet was resuspended in lysis buffer (800 mM NaCl, 50 mM KH<sub>2</sub>PO<sub>4</sub>, 100 mM DTT) and frozen at -80°C for at least 2 hours prior to thawing in hot water. The lysate was treated with 50 mg/mL of DNaseI for 1 hour at room temperature and centrifugated at 30 000 RCF for 30 minutes. The supernatant was diluted by a factor 2 with Tris-HCl pH 7.6 and purified by FPLC with a HiTrap SP HP column (GE Healthcare) using buffer A (50 mM sodium acetate, pH 5.00). After eluting impurities using a mix of buffer A containing 14% of buffer B (50 mM sodium acetate, 3 M NaCl, pH 5.00), Miz 5-8 was eluted using a gradient of buffer B. Fractions containing Miz 5-8 were pooled and desalted on FPLC using HiTrap Desalting columns (GE Healthcare) with water containing 0.1% trifluoroacetic acid (water-TFA) as elution buffer. The sample was then concentrated using ultracentrifugation (Amicon Ultra 3K, Millipore). Finally, the protein was purified by reverse-phase HPLC with a Discovery BIO Wide Pore C18 column (Supelco) preconditioned with water-TFA and eluted using a gradient of acetonitrile. Fractions containing Miz 5-8 were lyophilized and resuspended in water-TFA so that cysteines remain in reduced form.

## CD spectropolarimetry

CD measurements were performed with a Jasco J-810 spectropolarimeter equipped with a Jasco Peltier-type thermostat. The CD spectra were recorded at 20°C with 40 μM of Miz 5-8 at different pH and Zn<sup>2+</sup> concentrations in a 1 mm quartz cell. CD spectra were averaged over nine scans in order to enhance the signal-to-noise ratio by a factor 3. They were recorded between 190 and 250 nm with a wavelength step of 0.2 nm and smoothed using Spectra Manager (JASCO Corporation). Data was converted from CD signal to mean residue ellipticity using equation 1, where  $\delta$  is the ellipticity in degrees, MRW is the mean residue weight,  $c$  is the concentration of the sample ( $\text{g} \cdot \text{ml}^{-1}$ ) and  $l$  is the path length (cm).

$$[\Theta] = \frac{\delta \cdot MRW}{10 \cdot c \cdot l} \text{ (equation 1)}$$

## NMR spectroscopy

All NMR experiments were performed at 25°C on a Varian (Agilent) Unity INOVA operating at <sup>1</sup>H frequency of 600 MHz and equipped with an indirect detection H/C/N room temperature probe and Z-axis pulsed-field gradient capacity. After refolding the uniformly <sup>13</sup>C, <sup>15</sup>N-labeled Miz 5-8 construct, the buffer was changed to 10 mM sodium cacodylate, pH 7.0, 50 mM KCl, 2 mM TCEP, 4.4 molar equivalents ZnCl<sub>2</sub>, 10% D<sub>2</sub>O using ultracentrifugation (Amicon Ultra 3K, Millipore). The final concentration of the Miz 5-8 construct was 1.16 mM.

The chemical shifts of Miz 5-8 <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N nuclei were assigned using standard multi-dimensional and heteronuclear NMR experiments: <sup>1</sup>H-<sup>15</sup>N HSQC, <sup>1</sup>H-<sup>13</sup>C HSQC, 3D- HNCACB, CBCA(CO)NH, HNCO, C(CO)NH, H(CCO)NH, HCCH-TOCSY and <sup>13</sup>C and <sup>15</sup>N-edited-NOESY-HSQC with mixing times of 150 ms. All pulse sequences used were taken from the Biopack repertoire and based on the work of L. E. Kay. The spectra were referenced as described in Wishart et al. (Wishart et al. 1995).

## <sup>15</sup>N Spin relaxation

<sup>15</sup>N-T<sub>1</sub>, T<sub>2</sub> and {<sup>1</sup>H}-<sup>15</sup>N NOE experiments were recorded using previously described pulse sequences available in the Biopack repertoire (Farrow et al. 1994). Delays of 0, 30, 60, 90, 120, 150, 250, 350, 450, 550, 650, 800 and 1000 ms and of 0, 10, 30, 50, 70, 90, 110, 130, 150, 170, 250, 350 and 490 ms were used to obtain <sup>15</sup>N-T<sub>1</sub> and T<sub>2</sub> respectively. {<sup>1</sup>H}-<sup>15</sup>N NOE measurements were done by comparing HSQC spectra with and without a 6-second proton saturation.

<sup>15</sup>N backbone CPMG relaxation dispersion profiles were acquired in a constant time (60 ms) and interleaved manner, using a modified version of the pulse sequence from the Biopack repertoire and based on the work of Palmer et al. (Palmer et al. 2001; Wang et al. 2001) and Kay et al. (Mulder et al. 2001b). The field strengths ( $\nu_{\text{CPMG}}=1/4\tau_{\text{CPMG}}$ ) used were 33.3, 66.7, 100, 133, 167, 200, 267, 333, 400, 467, 533, 600, 667, 733, 800, 867, 933 and 1000 Hz, where  $2\tau_{\text{CPMG}}$  is the delay between the centers of two consecutive refocusing pulses (Mulder et al. 2001b). Experiments with  $\nu_{\text{CPMG}} = 100, 333, 867$  Hz were repeated twice to estimate the extent of the experimental error. Given that only one static field was used and owing to the complexity of the system, we have limited our analysis of the relaxation profiles to the estimation of  $R_{\text{ex}}$  by fitting the graphs of  $R_{2,\text{eff}}$  with the  $R_{2,\text{eff}}(\nu_{\text{CPMG}})$  simplified expression for a two-site exchange in the fast exchange limit (Luz and Meiboom 1963; Mulder et al. 2001a):

$$R_{2,\text{eff}}(\nu_{\text{CPMG}}) = R_{2,\text{eff}}(\nu_{\text{CPMG}} \rightarrow \infty) + R_{\text{ex}} \left( 1 - \frac{4\nu_{\text{CPMG}}}{k_{\text{ex}}} \tanh\left(\frac{k_{\text{ex}}}{4\nu_{\text{CPMG}}}\right) \right) \text{ equation 2}$$

where

$$R_{\text{ex}} = \frac{P_A P_B \Delta\omega^2}{k_{\text{ex}}} \text{ equation 3}$$

where the data were plotted using:



$$R_{2,eff}(v_{CPMG}) = \frac{1}{T} \ln \frac{I(v_{CPMG})}{I_0} \text{ equation 4}$$

$R_{2,eff}(v_{CPMG} \rightarrow \infty)$  is the effective  $R_2$  when the frequency of refocusing reaches an infinite value,  $I(v_{CPMG})$  and  $I_0$  are the resonance intensities of the correlations between backbone  $^{15}\text{N}$  and  $^1\text{H}_\text{N}$  in presence and absence of CPMG refocusing, respectively.  $T$  is the total and constant duration of the CPMG period (60 ms).  $P_A$  and  $P_B$  are the populations of the two interchanging states,  $\Delta\omega$  is the difference in  $^{15}\text{N}$  chemical shifts in both states and  $k_{ex}$ , the exchange rate, is given by  $k_A+k_B$ , the two rate constants of the exchange equilibrium. In the fast exchange limit ( $k_{ex} \gg \Delta\omega$ ), the populations and  $\Delta\omega$  are inseparable. Nonetheless, estimation of  $R_{ex}$  values at one field is valid using this approach (Lipchock and Loria 2009).

### Data processing and analysis

The NMR spectra were processed using NMRPipe (Delaglio et al. 1995) and analyzed with CcpNmr Analysis (Vranken et al. 2005). The  $^{15}\text{N}$ - $T_1$  and  $T_2$  time constants and their uncertainties were obtained by fitting the decays of the peak heights to the classical two-parameter single exponential using the least squares fitting program in CcpNmr Analysis. The relaxation dispersion profiles were fit to equation 2 with the program GraphPad Prism 5 in order to get  $R_{ex}$  values and their uncertainties.

### Structure calculations

All NOEs were assigned manually and converted into distance restraints using CcpNmr Analysis. The program DANGLE was used to obtain  $\phi$  and  $\psi$  dihedral angles on the basis of backbone and  $^{13}\text{C}_\beta$  chemical shift value. Dihedral  $\phi$  and  $\psi$  angles were calculated for 99 residues and they confirmed the typical  $\beta\beta\alpha$  fold for all four ZFs, except for the first  $\beta$  strand of both ZFs 7 and 8. NOESY spectra revealed sequential distance restraints typical of  $\alpha$  helices, which appear in the same regions predicted by DANGLE to

fold into helical structure. Structures were calculated using the program ARIA2.2 (Rieping et al. 2007) in conjunction with CNS (Brünger et al. 1998). Calculations were first carried out without zinc atom and zinc coordination restraints. In ZF 5 and ZF 7, the conserved cysteines and histidines were positioned correctly to allow coordination of zinc. By contrast, in ZF 6, due to the lack of NOEs around the two cysteines, only the two histidines were well positioned. However, NOEs between His50  $^1\text{H}_{\delta 2}$  and Cys34  $^1\text{H}_{\beta a-b}$  could be identified, which could not be present in absence of zinc coordination. In the following calculations, a  $\text{Zn}^{2+}$  ion was added and zinc coordination distance restraints were specified to ARIA (2.3 Å for  $\text{Zn}^{2+}\text{-S}_{\gamma}$  and 2.0 Å for  $\text{Zn}^{2+}\text{-N}_{\epsilon 2}$ ). The 20 lowest-energy conformers out of 300 for the final iteration of the calculation were refined in water and submitted to PROCHECK\_NMR (Laskowski et al. 1996) to validate the structural quality. The final structure ensembles of ZF 5 (residues 4-26), ZF 6 (residues 32-54) and ZF 7 (residues 60-83) were deposited in the protein databank (PDB) under identification codes 2m0d, 2m0e and 2m0f, respectively; and all NMR resonance assignments for Miz 5-8 was deposited in the BMRB under accession number 18806.

## Results

### Folding of Miz 5-8

Circular dichroism (CD) was used to monitor and optimize the folding conditions of Miz 5-8, as previously described (Bédard et al. 2012).  $\text{ZnCl}_2$  was added in slight excess (4.4 molar equivalents, or 1.1 molar equivalent per finger) to the sample. The pH was then raised using NaOH until the secondary structure content was optimal as judged by the decrease in molar ellipticity between 210 and 230 nm, as well as the shifting of the minimum from 204 nm to 207 nm (data not shown). The optimal pH was found to be 7; a value larger than the optimal pH found for the Miz 8-10 construct (pH 6.5). We present the spectrum of Miz 5-8 shown along with that of Miz 8-10 (see Fig. 1b). As one can notice, the optimal secondary structure content of Miz 5-8 is comparable to that of Miz 8-10.

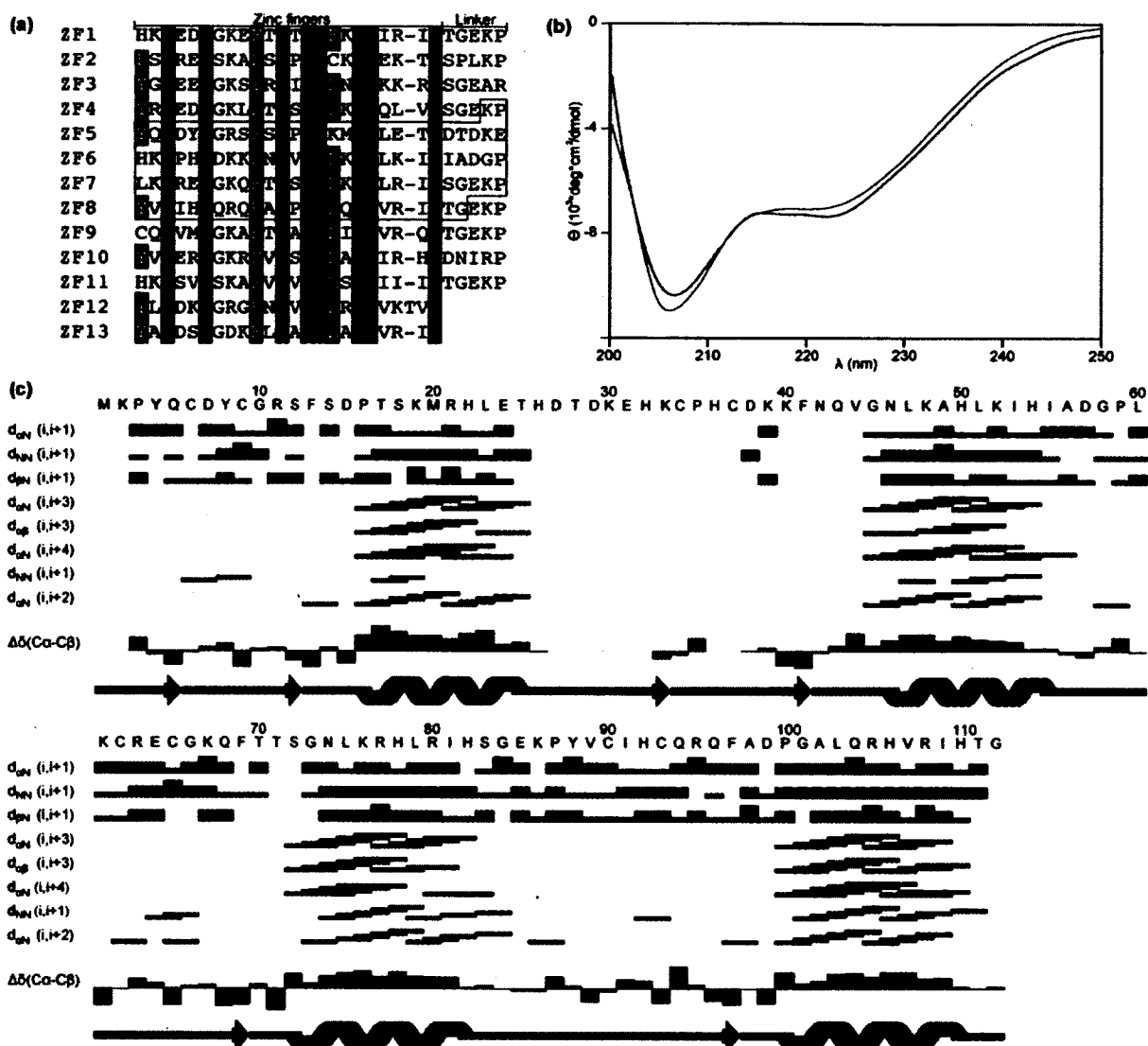


Figure 1: Characterization of Miz 5-8

(a) Primary structure of the 13 ZF domains of Miz-1. In the complete protein, the first 12 ZFs are found within residues 306 to 637, and ZF 13 is within residues 717 to 739. Residues belonging to the Miz 5-8 fragment (416 to 526) are boxed, and are labeled residues 2 to 112 in this paper (residue 1 is an additional methionine). Conserved zinc-coordinating residues are highlighted in green, conserved hydrophobic core residues are highlighted in blue, and putative DNA-contacting residues are highlighted in magenta. (b) Far-UV CD spectrum of Miz 5-8 (red) and Miz 8-10 (blue). Results are shown in mean residue ellipticity and calculated using equation 1. (c) Sequential NOEs, chemical shift variation  $\Delta\delta(C_{\alpha}-C_{\beta})$  and secondary structure profile calculated by DANGLE for Miz 5-8.

## Resonance assignments and secondary structure of Miz 5-8

The dataset recorded (see materials and methods) allowed for the assignment of 86.4% of the backbone  $^{15}\text{N}_\text{H}$ , 91.3% of the  $^1\text{H}_\text{N}$ , 85.5% of the  $^{13}\text{C}'$ , 93.6% of the  $^{13}\text{C}_\alpha$ , 85.0% of the  $^1\text{H}_\alpha$  and 93.1% of the  $^{13}\text{C}_\beta$  of residues 3 to 112. 81.6% of side-chain protons

were assigned as well. Fig. 1c shows the difference in secondary chemical shifts  $\Delta\delta(C_\alpha - C_\beta)$ , sequential NOEs and DANGLE secondary structure predictions as a function of the primary structure for Miz 5-8. Positive and negative  $\Delta\delta(C_\alpha - C_\beta)$  values are indicative of  $\alpha$ -helices and  $\beta$ -strands, respectively. The number of secondary structure elements determined from the secondary chemical shifts indicates that the 4 ZFs adopt the consensus  $\beta\beta\alpha$  fold, except for the first  $\beta$  strand of ZFs 7 and 8.  $^{13}C_\alpha$  and  $^{13}C_\beta$  chemical

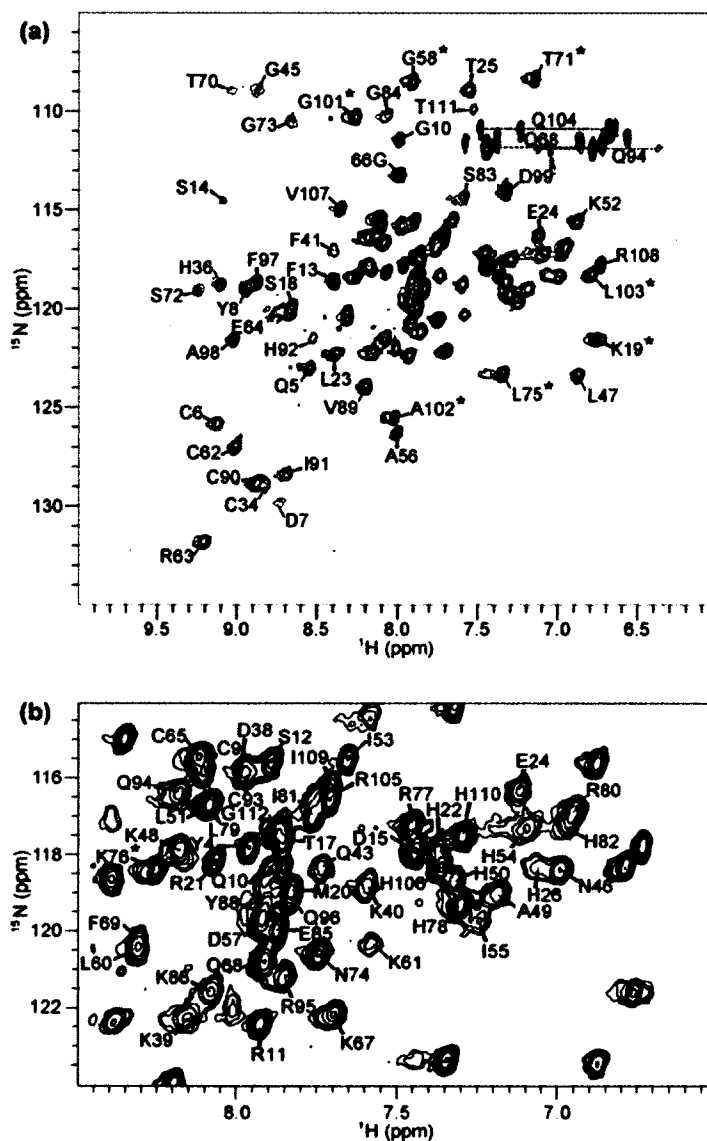


Figure 2:  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Miz 5-8

(a) Full  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Miz 5-8 recorded at pH 7.0, 682  $\mu\text{M}$  and 600 MHz. Resonances are labeled using the one-letter code. (b) Magnification of the crowded region of (a) for clarity. Resonances exhibiting processing-derived cardinal sinus wiggles have been labeled with an asterisk.

shifts of all cysteines (except for those of Cys37 which could not be assigned) were consistent with the zinc-coordinating thiol state (Kornhaber et al. 2006). Likewise, the  $^{13}\text{C}_{\delta 2}$  and  $^{13}\text{C}_{\epsilon 1}$  chemical shifts of all zinc-coordinating histidines (except that of His26, His 106 and His110 which could not be assigned) correlate with a zinc coordination mediated through the  $\text{N}_{\epsilon 2}$  (Barraud et al. 2012).

The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Miz 5-8 is displayed in Fig. 2a and b. Although the spectra show well-dispersed resonances, one can notice that most of the resonances from residues 27 to 44, which include the linker between ZFs 5 and 6 as well as the N-terminal region of ZF 6, are either weak or missing in the spectrum. This indicates that the backbone amide resonances are too weak (or broad) to be measured. This can be the result of conformational exchange on the  $\mu\text{s}$ - $\text{ms}$  time-scale ( $R_{\text{ex}}$ ), which broadens the resonances, and/or exchange between the  $^1\text{H}_{\text{N}}$  and either the solvent deuterium or the suppressed protons from  $\text{H}_2\text{O}$ . The fact that our optimal folding conditions were found to be at a pH of 7 may contribute to the  $\text{H}_{\text{N}}$  exchange. Indeed, the intrinsic rate of amide exchange is faster at pH 7 than the minimal rates obtained around pH 4-5 (Bai et al. 1993; Eriksson et al. 1995).

#### **Evidence of conformational exchange on the $\mu\text{s}$ - $\text{ms}$ timescale located in ZF 5 and 6**

In order to further explore the causes for the vanishing amide resonances, we attempted amide exchange rate measurements. However, we found out it was not possible to resolubilize the lyophilized and folded Miz 5-8 construct to measure amide exchange rates. We have therefore resorted to the analysis of the pH dependence of the backbone amide resonance intensities on the  $^1\text{H}$ - $^{15}\text{N}$  HSQC. It is expected that the exchanging amide resonances will increase in intensity by decreasing the pH from 7.5 to 6.0, since the intrinsic rate of exchange slows down as the pH approaches a value of 4 to 5 (Bai et al. 1993; Eriksson et al. 1995).

We present in Fig. 3a and b the  $^1\text{H}$ - $^{15}\text{N}$  HSQCs of Miz 5-8 recorded at pH 7.5 and 6.0, respectively, and have labeled the resonances with pH-sensitive intensities. We also

map these residues onto the backbone of an homology model of Miz 5-8 in a DNA-bound conformation generated using I-TASSER (Roy et al. 2010) (Fig. 3c and d). In addition, we have labeled in magenta the position of residues for which no resonances could be recorded on the  $^1\text{H}$ - $^{15}\text{N}$  HSQCs at any pH between 7.5 and 6.5. No significant changes in intensity were noted at pH lower than 6.5 and until pH 5.0 where a second set of resonances, most likely originating from the unfolded state (not shown), appeared. Only a few resonances shifted in the spectra between pH 7.5 and 6.5. Those belonged to histidine side-chains not involved in chelating the  $\text{Zn}^{2+}$  ions.

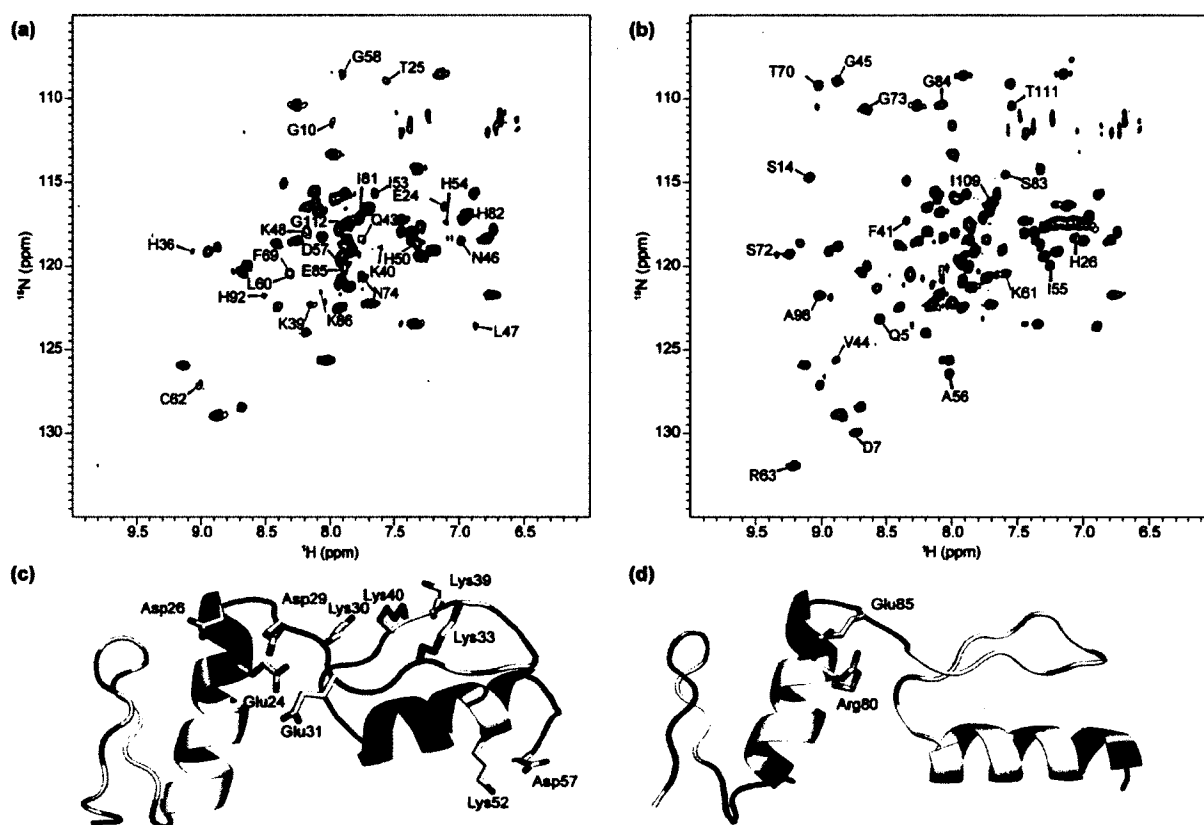


Figure 3: Effect of pH on the intensities of Miz 5-8 backbone amide resonances.

(a)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Miz 5-8 recorded at pH 7.5. Labeled resonances are those whose intensity increases when the pH is decreased. (b)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Miz 5-8 recorded at pH 6.0. Labeled resonances are those that have appeared between pH 7.5 and pH 6.0. (c) and (d) depict homology models of ZFs 5-6 and ZFs 7-8, respectively. Resonances from residues broadened or weakened beyond detection are shown in magenta and the pH-sensitive residues are shown in green. Also shown are the side-chains of the acidic and basic clusters and those of the salt bridges involving the conserved basic and acidic side chains from the helices and the linkers. See text for details. The rendition was done with Pymol ([www.pymol.org](http://www.pymol.org)).

One striking observation that can be made based on the results depicted in Fig. 3 is the fact that of the four helices, only the one in ZF 6 has pH-sensitive amides across its length as well as at the interface with the  $\beta$ -hairpin. Otherwise, and as expected due to end fraying effects, all four helices have pH-sensitive amides mainly located at their N- and C-termini. This indicates that the  $^1\text{H}_\text{N}$  in the middle of the helix of ZF 6 are less protected from exchange with the solvent, and this result suggests a conformational exchange and/or opening of the tertiary structure involving the  $\beta$ -hairpin in this ZF. Additionally, the amide resonances of the unusual linker (DTDKE) between ZF 5 and 6, as well as those from the amides of His32, Lys33 and Cys37 located in the two  $\beta$ -strands are not observable in any HSQC recorded. This suggests that in addition to amide exchange, the resonances of these amides may be broadened by conformational exchange ( $R_{\text{ex}}$ ) occurring on the  $\mu\text{s}$ -ms timescale. In order to assess the presence of such a conformational exchange and characterize the presence of motions in the ps-ns and  $\mu\text{s}$ -ms timescales, we have employed  $^{15}\text{N}$ -spin relaxation experiments.

### $^{15}\text{N}$ spin relaxation

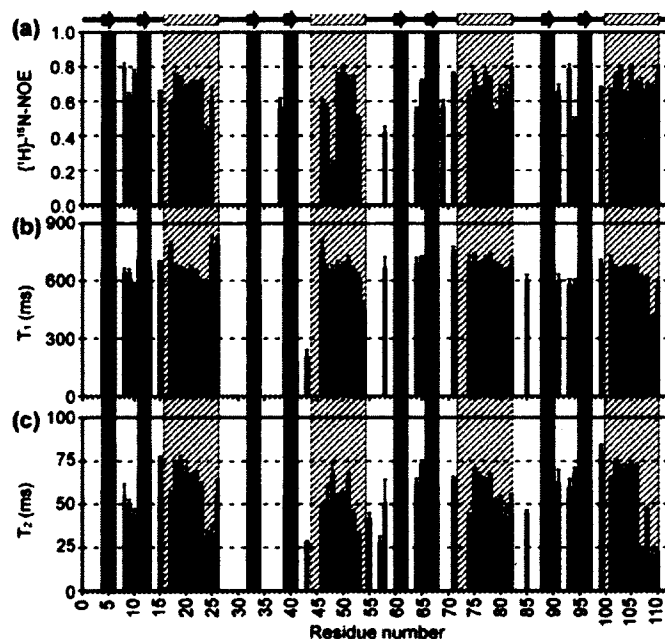


Figure 4:  $^{15}\text{N}$  relaxation measurements of the Miz 5-8 backbone amides on the ps-ns time-scale.

(a)  $\{^1\text{H}\}$ - $^{15}\text{N}$ -NOE, (b)  $T_1$  and (c)  $T_2$  values are plotted against the residue numbers. The  $\beta$ -strands and  $\alpha$ -helices of the 4 ZFs are highlighted in grey and hatched, respectively.

Heteronuclear NOEs ( $\{^1\text{H}\}$ - $^{15}\text{N}$  NOE),  $^{15}\text{N}$ - $T_1$  and  $^{15}\text{N}$ - $T_2$  values recorded are presented on Fig. 4. In accordance with previous studies (Potter 2005; Zandarashvili et al. 2012), the  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOE measured in the strands and helices range between 0.6 and 0.8 and suggest limited librational movements of the backbone amides on the ps-ns timescale. In contrast, and as observed by the same groups, the backbone amides in the linkers between ZF 6 and 7 that could be measured display slightly lower values (approx. 0.4) and indicate movements with slightly larger amplitudes on the ns-ps timescale. However, these (positive) values clearly show that the inter-domain linkers are not completely disordered. Indeed, while the linkers between consecutive ZFs are more flexible than ZFs themselves, they remain somewhat rigid, resulting in the “quasi-ordered” conformation described by Potter (Potter 2005).

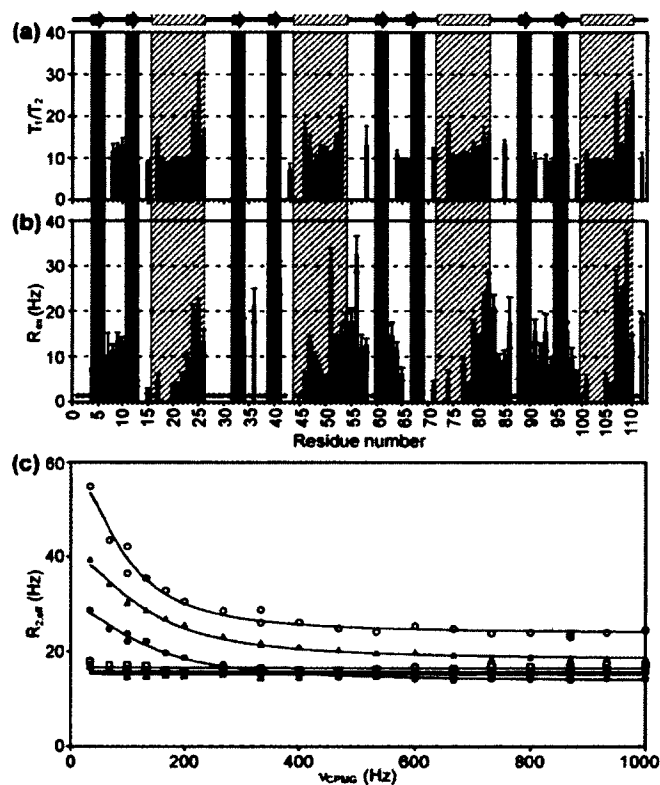


Figure 5:  $T_1/T_2$  and  $^{15}\text{N}$  relaxation dispersion of Miz 5-8 backbone amides.

(a)  $T_1/T_2$  ratios and (b)  $R_{ex}$  plotted against the residue numbers. The  $\beta$ -strands and  $\alpha$ -helices of the 4 ZFs are highlighted in grey and hatched, respectively. Asterisks denote residues for which no dispersion could be measured due to extreme broadening of the resonances. Shown in (c) are representative  $^{15}\text{N}$  relaxation dispersion profiles ( $R_{2,eff}(\nu_{CPMG})$ ) for residues 9 (filled circles), 19 (filled squares), 24 (filled triangles), 40 (open circles), 75 (open squares) and 104 (open triangles). Note the absence of dispersion for residues 19, 75 and 104.



The average  $T_1$  value recorded  $620 \pm 40$  ms is very similar to  $670 \pm 67$  ms, the average measured on a similar construct of 4 ZFs (Potter 2005). The average  $T_2$   $83 \pm 21$  ms recorded is also comparable but longer than those recorded for Miz 5-8 (i.e.  $58 \pm 6$ ). As previously described by others (Brüschweiler et al. 1995; Potter 2005), the rotational diffusion of multi-ZF constructs is highly anisotropic. More precisely, ZFs located towards the center of multi-ZF constructs experience anisotropic diffusion of higher intensity, with the N-H vectors of their  $\alpha$ -helices aligned with the unique axis of the rotational diffusion tensor. This alignment contributes to decrease the effective correlation time experienced by rigid N-H vectors, which increases their  $T_1$  and decreases their  $T_2$  (Sauvé et al. 2004; Tjandra et al. 1997). However, elevated  $T_1/T_2$  ratios are not always a sign of anisotropic rotational diffusion. Indeed, if a decrease in  $T_2$  is not accompanied by an increase in  $T_1$ , there is a probability that the resulting elevated  $T_1/T_2$  ratio is caused by conformational exchange ( $R_{ex}$ ) occurring on the  $\mu$ s-ms time-scale. Inspection of Fig. 5a and b reveals that many of the residues with high  $T_1/T_2$  ratios follow this trend and indicates the presence of such conformational exchange. In order to validate this postulation, we have carried out  $^{15}\text{N}$ -relaxation dispersion or CPMG experiments at 600 MHz. Such experiments reduce the  $R_{ex}$  resonance broadening, and also allow estimation of the  $R_{ex}$  values (Ishima and Torchia 1999; Mittermaier and Kay 2009; Palmer 2004).

$R_{ex}$  values are approximated by  $R_{ex} = R_{2,eff}(\nu_{cpmg} \rightarrow 0) - R_{2,eff}(\nu_{cpmg} \rightarrow \infty)$ , where  $\nu_{cpmg}$  is the frequency of the refocusing  $180^\circ$  pulse. This frequency is given by  $1/4\tau_{cpmg}$ , with  $2\tau_{cpmg}$  representing the time between successive refocusing pulses (see materials and methods for further details). One convenient and rapid way to detect the presence of such conformational exchange is to compare HSQCs recorded with low and high CPMG refocusing rates. We present, in Fig. 6a and b, HSQCs recorded at  $\nu_{cpmg} = 33$  Hz and 1000 Hz, respectively. Resonances undergoing conformational exchange gain in intensity upon the increase of the frequency of the refocusing pulses. Indeed, many low-intensity peaks at 33 Hz (broadened by  $R_{ex}$ ) become more intense at 1000 Hz. Moreover, some peaks broadened beyond detection at 33 Hz acquire intensities similar to the remaining resonances at high field strength. This clearly establishes the presence of  $\mu$ s-ms motions

in the Miz 5-8 construct and a clear contribution of such motions to the shortening of many  $T_2$  values.

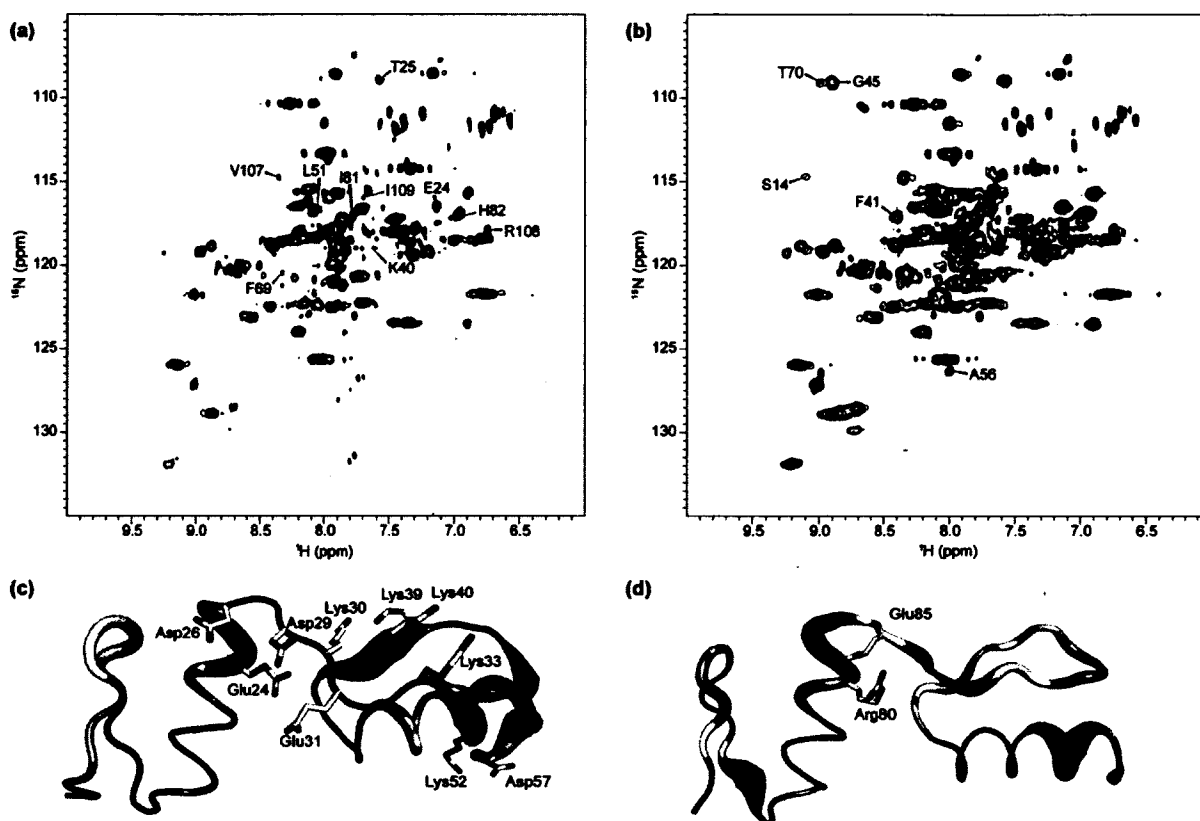


Figure 6:  $^{15}\text{N}$  relaxation dispersion of Miz 5-8 backbone amides.

HSQC spectra recorded with a CPMG frequency of 33Hz (a) and 1000Hz (b), respectively. Labeled resonances in (a) exhibit a large increase of intensity (a) or appear (b) as the CPMG frequency is increased. (c) and (d) depict homology models of ZFs 5-6 and ZFs 7-8, respectively. The width of the backbone worm is proportional to the  $R_{\text{ex}}$  displayed in Fig. 4 (c). Note the color-coding: residues in blue have low  $R_{\text{ex}}$  while residues in red have large  $R_{\text{ex}}$  values. Also shown are the side-chains of the acidic and basic clusters and those of the salt-bridges involving the conserved basic and acidic side-chains from the helices and the linkers. See text for details. The rendition were done with Pymol ([www.pymol.org](http://www.pymol.org)).

Fig. 5b displays the  $R_{\text{ex}}$  values estimated for residues that were not overlapped and for which a relaxation dispersion could be measured by fitting the  $R_{2,\text{eff}}(\nu_{\text{cpmg}})$  curves as described in the materials and methods section. Shown in Fig. 5c are examples of such  $R_{2,\text{eff}}(\nu_{\text{cpmg}})$  fits, and the associated fitted parameters are shown in supplementary material (Table S1). Interestingly, the  $R_{\text{ex}}$  values show a similar dependence on the sequence specific secondary structure as  $T_1/T_2$  (Fig. 5a). Indeed, the largest  $R_{\text{ex}}$  values coincide with the largest  $T_1/T_2$  ratios and are located near the end of the  $\alpha$ -helices. To get a better

appreciation of the phenomenon, we display the determined  $R_{ex}$  graphically on a ribbon representation of an homology model of Miz 5-8 in a DNA bound state (see Fig. 6c and d). The width of the ribbon is proportional to  $R_{ex}$ . As one can clearly notice, the largest  $R_{ex}$  values are located near the end of the helices and in the subsequent linkers, except for the linker between ZFs 7 and 8. On the other hand, it is interesting to note that the positions involved in DNA binding (towards the N-termini of the helices) do not undergo extensive  $\mu$ s-ms time-scale motions.

### Solution structures of Miz-1 ZFs 5 to 7

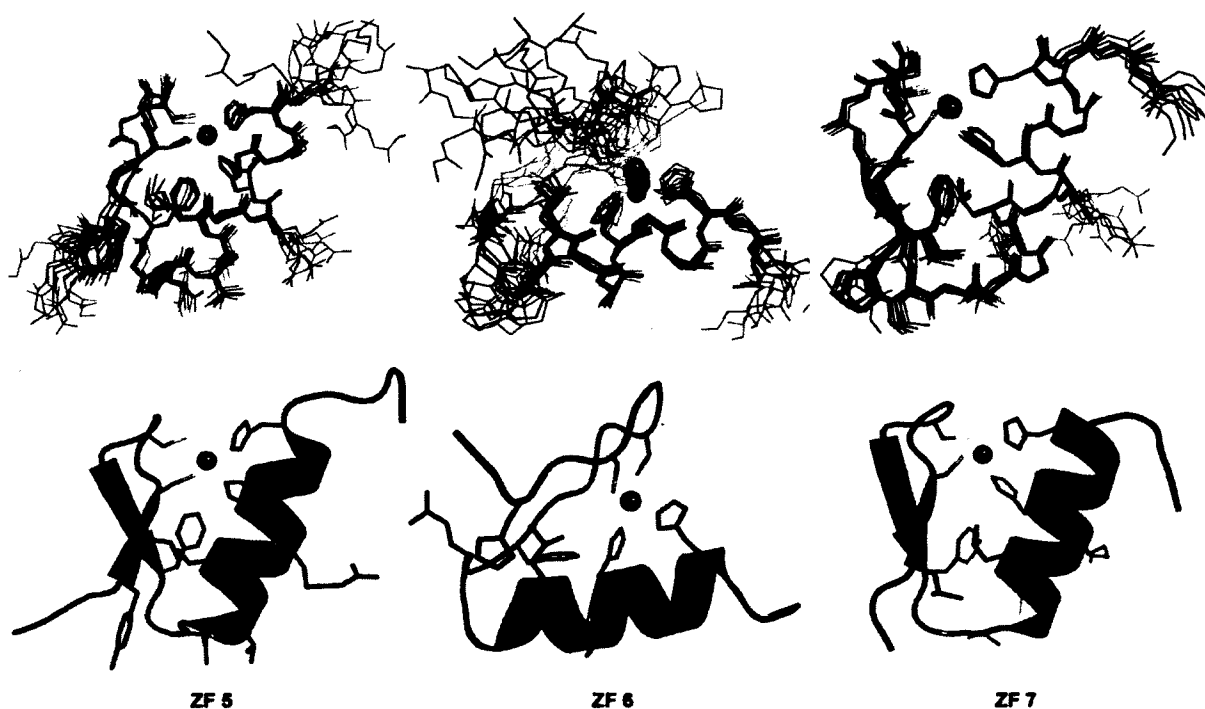


Figure 7: NMR solution structures of (a) ZF 5, (b) ZF 6 and (c) ZF 7.

The backbone atoms, as well as the side chains of zinc-coordinating (carbon atoms in green), hydrophobic core residues (carbon atoms in blue) and residues that would putatively contact DNA (carbon atoms in magenta) are depicted. The  $Zn^{2+}$  ions are shown as grey balls. All three structures are a superposition of the 10 lowest energy structures. The rendition and superposition were done with Pymol ([www.pymol.org](http://www.pymol.org)).

The structure of ZF 8 has already been solved by our group (Bédard et al. 2012) as part of the Miz 8-10 construct (PDB identification code 2LVR). Hence, we will focus on the description of the solution structures of ZFs 5 to 7. The restraints used to calculate the

structure of each ZF (NOEs, dihedral angles and H-bond restraints) are summarized in Table 1. The structures of ZFs 5 to 7 are shown in Fig 7. As can be seen, the selected structures for ZFs 5 and 7 display excellent convergence, with RMSD values of  $0.3\pm 0.1$  Å and  $0.26\pm 0.06$  Å, respectively, for the backbone atoms ( $C_{\alpha}$ ,  $C'$ , O, N) of the consensus ZF sequences. In both ZFs, all conserved side chains (hydrophobic or zinc-coordinating) are

Table 1: Structural statistics for ZFs 5 to 7 of Miz-1

	ZF 5	ZF 6	ZF 7
<b>Restraints for final structure calculations</b>			
<b>NOE distance restraints</b>			
Intraresidue ( $ i - j  = 0$ )	287	219	391
Sequential ( $ i - j  = 1$ )	128	94	163
Medium range ( $1 <  i - j  < 5$ )	101	77	118
Long range ( $ i - j  \geq 5$ )	66	10	113
Total NOE distance restraints	582	400	785
Hydrogen bonds	12 x 2	7 x 2	13 x 2
Zinc ligands	4 x 2	4 x 2	4 x 2
<b>Dihedral angle restraints<sup>a</sup></b>			
$\phi$ and $\psi$ angles	49	51	49
$\chi$ angles	3	1	4
<b>Structure statistics (20 structures)</b>			
Number of NOE violations $> 0.5$ Å	0	0	0
Number of dihedral angle violations $> 5^{\circ}$	0	0	0
<b>RMS deviations from experimental data</b>			
Average distance restraint violation (Å)	$0.035\pm 0.006$	$0.009\pm 0.002$	$0.029\pm 0.004$
Average dihedral restraint violation ( $^{\circ}$ )	$0.26\pm 0.12$	$0.54\pm 0.21$	$0.76\pm 0.10$
<b>RMS deviation to mean coordinates</b>			
Backbone heavy atoms (Å)	$1.70\pm 0.52$	$1.89\pm 0.62$	$0.73\pm 0.24$
All heavy atoms (Å)	$2.32\pm 0.49$	$2.54\pm 0.61$	$1.36\pm 0.30$
<b>For ZF consensus sequences<sup>b</sup></b>			
Backbone heavy atoms (Å)	$0.30\pm 0.10$	$0.95\pm 0.26$	$0.26\pm 0.06$
All heavy atoms (Å)	$0.96\pm 0.12$	$2.00\pm 0.48$	$1.00\pm 0.13$
<b><math>\alpha</math>-helical residues<sup>c</sup></b>			
Backbone heavy atoms (Å)	$0.17\pm 0.07$	$0.23\pm 0.06$	$0.13\pm 0.04$
All heavy atoms (Å)	$1.00\pm 0.19$	$0.88\pm 0.14$	$0.93\pm 0.17$
<b>Ramachandran plot statistics<sup>d</sup> (%)</b>			
Residues in most favored regions	89.4	86.7	91.4
Residues in additionally allowed regions	9.8	12.1	8.6
Residues in generously allowed regions	0.2	0.8	0.0
Residues in disallowed regions	0.6	0.4	0.0

<sup>a</sup>  $\phi$  and  $\psi$  angles were derived from the program DANGLE

<sup>b</sup> ZF consensus sequences comprise residues 4-26 for ZF 5, 32-54 for ZF 6 and 60-82 for ZF 7

<sup>c</sup>  $\alpha$ -helical residues comprise residues 16-27 for ZF5, 45-55 for ZF6 and 72-83 for ZF7

<sup>d</sup> Ramachandran plot statistics were generated using PROCHECK NMR

also well defined by the set of restraints used. As observed for the solution structures reported earlier, the hairpin between the two  $\beta$  strands as well as the region between the second  $\beta$  strand and the  $\alpha$  helix on both ZFs are not quite as well defined, due to a lack of observable NOEs (see supplementary figure 1). The N- and C-terminus of the consensus sequence of ZF 5 and 7 are not as well defined as the remainder of the core residues.

The  $\alpha$  helix of ZF 6 is well folded, with a RMSD value of  $0.23\pm 0.06$  Å for the backbone of helical residues (residues 45-55). The RMSD values for the backbone of helical residues of ZFs 5 and 7 are  $0.17\pm 0.07$  Å and  $0.13\pm 0.04$  Å, respectively (residues 16-27 for ZF 5 and residues 72-83 for ZF 7). In comparison, the backbone RMSD value of ZF 6 consensus sequence (residues 32-54) is  $0.95\pm 0.26$  Å, which is triple that of the other ZFs.

Resonances for amino acids 1-2 and 27-33 in ZF 5 and 6 could not be observed in any spectrum. Partly because of this, too few long-range and tertiary NOEs could be unambiguously assigned in ZF 6 (see table 1) to define the expected  $\beta\beta\alpha$  fold, despite the aforementioned DANGLE predictions. This lack of NOEs and resonance absence is most likely caused by the exchange broadening and/or rapid amide exchange discussed above at pH 7.0.

## Discussion and conclusion

Miz-1 is a BTB/POZ transcription factor that contains 13  $C_2H_2$  ZFs that activates the transcription of cell cycle inhibitors (e.g. p15<sup>INK4b</sup> and p21<sup>CIP1</sup>). Although DNA binding and protein-protein interactions (e.g. with SMAD3) involving the ZFs are important for transcriptional activation, it is not known how these ZFs bind the promoters of activated genes and how they bind to SMAD3/4. In order to provide structural information on these fundamental issues, we report the structural and dynamical characterization by solution-state NMR spectroscopy of a Miz-1 construct composed of residues 416 to 526, which encompasses ZF 5 to 8.

Whereas ZFs 5, 7 and 8 adopt a canonical and stable  $\beta\beta\alpha$  fold at neutral pH, the NMR data collected under these conditions indicate that ZF 6 and the linker between ZFs 5 and 6 are undergoing conformational exchange. Upon inspection of the primary structure of Miz 5-8 (Fig. 1a), many unusual features can be recognized. One of them is the presence of a histidine instead of the conserved aromatic residues Phe and Tyr (Schmidt 2004), two positions before the first zinc-coordinating cysteine of ZF 6. Upon inspection of other ZF solution structures found on the PDB (such as 1VA1 or 2DMD), one can notice that a histidine in this position can cause local conformational fluctuations.

Secondly, and most strikingly, the sequence of the linker between ZF 5 and ZF 6 (DTDKE) is highly uncharacteristic. Indeed, compared to the consensus TGEKP, only the lysine is conserved. Moreover, the sequence contains three negatively charged side chains. In addition, the glutamate (Glu24) found two residues before the last histidine in ZF 5 is also highly unusual. Indeed, arginine (62.2%), lysine (12.4%) or a hydrophobic amino acid (16.9%) are the usual amino acids found at this position in a database consisting of 2435 human ZFs (Schmidt 2004). The importance of a salt bridge between the conserved basic side chain and the conserved glutamate in the TGEKP linker has recently been demonstrated for inter-ZF movements in the three ZFs of Egr-1 (Zandarashvili et al. 2012). In fact, in the WT primary structure of Egr-1, the basic side chain and the glutamate are replaced by a threonine in ZF 1 and a glutamine in the linker, respectively. The engineering of a salt bridge between a lysine and a glutamate in the TGEKP linker was shown to reduce movements experienced by ZF 1 in absence of DNA. As argued by Zandarashvili et al. (Zandarashvili et al. 2012), the presence of such movements in the WT constructs plays a key role in the search process for the specific DNA sequence of Egr-1. But while Egr-1's threonine-glutamine pair merely lacks the stabilizing salt bridge, the equivalent residues between Miz 5-8's ZFs 5 and 6 (Glu24 and Asp29) are both negatively charged residues, which are mutually repulsive. Assuming that the Miz 5-8 construct can adopt a classical DNA-bound state, we have generated a homology model of such a conformation (Fig. 3c). In this conformation, the side chains of Asp27 and Glu31 are also oriented towards Glu24 and Asp29, which should further destabilize this region of the protein. It is also important to note that in a DNA complex,

this negatively charged cluster would be situated close to the phosphodiester backbone. This canonical conformation for the linker between ZF 5 and 6 is therefore expected to be highly unfavorable for DNA binding.

Another destabilizing cluster, formed by Lys30, Lys33, Lys39 and Lys40, is situated in the putative  $\beta$ -sheet of ZF 6. Thus, with these two charged clusters, it is expected that the free energy of the state where ZFs 5 and 6 adopt a classical DNA-bound conformation would be elevated. This results in the low probability of adoption of this conformation by these two ZFs. We propose that this could contribute to the local destabilization of the tertiary structure of ZF 6 and explain the extreme broadening of resonances observed in this region of the construct, as well as the lack of long-range and tertiary NOEs.

ZFs 1 to 4 have been shown to interact with MH1 domain of SMAD3/4. Indeed, examination of the primary sequence of those ZFs reveals a high amount of hydrophobic side chains in lieu of the expected polar side chains in the residues expected recognize DNA (positions -1, 2, 3 and 6 of the ZF  $\alpha$ -helix), supporting their participation in protein-protein interactions. Knowing that ZF 6 undergoes conformational exchange and possesses unusual charged residues in its sequence that could prevent it from binding DNA, we hypothesize that ZF 6 could play a hinge-like role, separating a protein-protein interaction domain from a DNA-interacting domain. This peculiar sequence and resultant dynamical behavior could be important for Miz-1 function by promoting the availability of ZF 1 to 4 to engage protein-protein interaction with SMAD3/4, and potentially other protein-protein interactions involving ZFs 1 to 5.

Interestingly, Zandarashvili et al. have also reported the presence of  $^{15}\text{N-R}_{\text{ex}}$  located at the end of the helix of ZF 1, the first linker and the  $\beta$ -hairpin of the next ZF (Zandarashvili et al. 2012). In addition, residues with significant  $^{15}\text{N-R}_{\text{ex}}$  were located at the end of the helix in ZF 2. As one can notice in Fig. 5,  $T_1/T_2$  ratios are on average higher in helices than in  $\beta$ -sheets, and that they increase strikingly as they approach the C-terminus of the helices. This trend has also been found in all the ZFs of the Miz 8-10

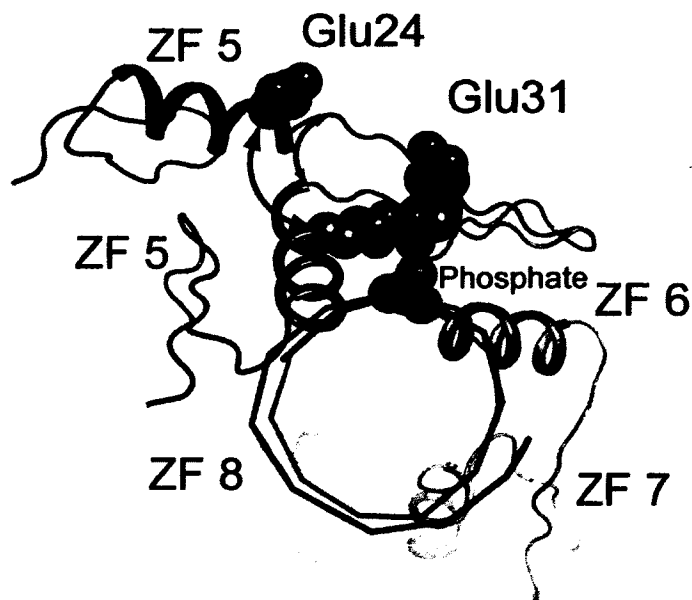


Figure 8: Molecular model for the conformational exchange located in ZFs 5 and 6.

This model displays the putative hinge function of these ZFs in the structural biology of Miz-1. As described in the text, in a canonical DNA complex, electrostatic repulsions between negatively charged side chains (e.g. Glu24 and Glu31) and phosphates (Glu31) can be predicted to occur. Note that many intramolecular repulsions are also expected to be present (see Fig. 6) in a DNA-binding competent configuration of ZFs 5 and 6. We propose that these enthalpically unfavorable states (blue ribbon and carbon atoms) will promote local conformational fluctuations that would bring ZF 5 on and off DNA (magenta strings). The local unfolding of the last helical turn of ZF 5 and transient opening of the tertiary structure between the first strand and the helix of ZF 6 (green ribbon and carbon atoms) are logical possible fluctuations according to our results. As described in the text and on Fig. 7, the presence of a His (instead of the more conserved Tyr and Phe) could contribute to such a movement. The rendition and superposition were done with Pymol ([www.pymol.org](http://www.pymol.org)).

construct (data not shown), as well as in many other poly-ZF proteins (Brüschweiler et al. 1995; Potter 2005; Zandarashvili et al. 2012). To date, this information has been attributed to anisotropy and results obtained by the model-free analysis have suggested that the helices aligned with the unique axis of their poly-ZFs constructs (Brüschweiler et al. 1995; Potter 2005). In light of our results, we propose that this rise in  $T_1/T_2$  ratios at the C-terminus of helices could be attributed to the concerted effects of rotational diffusion anisotropy and  $R_{ex}$ . As can be seen on Fig. 5, high values of  $R_{ex}$  are found in numerous positions in the Miz 5-8 construct, and strikingly coincide with high  $T_1/T_2$  ratios in  $\alpha$ -helices.



## Acknowledgements

This work was supported by the Natural Science and Engineering Research Council (NSERC) of Canada (grant to P.L. and studentships to M.B. and D.B.) and by the Regroupement stratégique sur la fonction, la structure et l'ingénierie des protéines (PROTEO). D.B. also acknowledges PROTEO for the award of graduate studentship. The authors thank Dr. Frank Hänel (Hans-Knöll-Institut für Naturstoff-Forschung e.V, Germany) for kindly providing us with the Miz-1 cDNA, Dr. Yves L. Dory (Université de Sherbrooke, Canada) for giving access and advice to use his HPLC system, Dr. Jean-François Naud for the initial cloning and Dr. Martin Montagne for his helpful comments on the manuscript. Finally, we thank Prof. Jim Omichinski (U. de Montréal) for his help with the refolding protocol of Zinc Fingers.

## References

- Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17:75–86. doi: 10.1002/prot.340170110
- Barraud P, Schubert M, Allain FH-T (2012) A strong <sup>13</sup>C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins. *J Biomol NMR* 53:93–101. doi: 10.1007/s10858-012-9625-6
- Berg JM (1988) Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc Natl Acad Sci USA* 85:99–102.
- Bédard M, Maltais L, Beaulieu M-E, et al. (2012) NMR structure note: solution structure of human Miz-1 zinc fingers 8 to 10. *J Biomol NMR*. doi: 10.1007/s10858-012-9670-1
- Bowen H (2003) Characterization of the Murine Nramp1 Promoter: REQUIREMENTS FOR TRANSACTIVATION BY Miz-1. *Journal of Biological Chemistry* 278:36017–36026. doi: 10.1074/jbc.M304301200
- Brayer KJ, Segal DJ (2008) Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochem Biophys* 50:111–131. doi: 10.1007/s12013-008-9008-5
- Brünger AT, Adams PD, Clore GM, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
- Brüschweiler R, Liao X, Wright PE (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* 268:886–889.
- Delaglio F, Grzesiek S, Vuister GW, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293.

- Emerson RO, Thomas JH (2009) Adaptive evolution in zinc finger transcription factors. *PLoS Genet* 5:e1000325. doi: 10.1371/journal.pgen.1000325
- Eriksson MA, Härd T, Nilsson L (1995) On the pH dependence of amide proton exchange rates in proteins. *Biophysical Journal* 69:329–339. doi: 10.1016/S0006-3495(95)79905-2
- Farrow NA, Muhandiram R, Singer AU, et al. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry* 33:5984–6003.
- Hyre DE, Klevit RE (1998) A disorder-to-order transition coupled to DNA binding in the essential zinc-finger DNA-binding domain of yeast ADR1. *J Mol Biol* 279:929–943. doi: 10.1006/jmbi.1998.1811
- Ishima R, Torchia DA (1999) Estimating the time scale of chemical exchange of proteins from measurements of transverse relaxation rates in solution. *J Biomol NMR* 14:369–372.
- Kornhaber GJ, Snyder D, Moseley HNB, Montelione GT (2006) Identification of Zinc-ligated Cysteine Residues Based on <sup>13</sup>C $\alpha$  and <sup>13</sup>C $\beta$  Chemical Shift Data. *J Biomol NMR* 34:259–269. doi: 10.1007/s10858-006-0027-5
- Laskowski RA, Rullmannn JA, MacArthur MW, et al. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486.
- Lipchock J, Loria JP (2009) Millisecond dynamics in the allosteric enzyme imidazole glycerol phosphate synthase (IGPS) from *Thermotoga maritima*. *J Biomol NMR* 45:73–84. doi: 10.1007/s10858-009-9337-8
- Luz Z, Meiboom S (1963) Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution—Order of the Reaction with Respect to Solvent. *J Chem Phys* 39:366–370. doi: 10.1063/1.1734254
- Mittermaier AK, Kay LE (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* 34:601–611. doi: 10.1016/j.tibs.2009.07.004
- Mulder FA, Mittermaier A, Hon B, et al. (2001a) Studying excited states of proteins by NMR spectroscopy. *Nat Struct Biol* 8:932–935. doi: 10.1038/nsb1101-932
- Mulder FA, Skrynnikov NR, Hon B, et al. (2001b) Measurement of slow (micro-s) time scale dynamics in protein side chains by (<sup>15</sup>N) relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J Am Chem Soc* 123:967–975.
- Palmer AG (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104:3623–3640. doi: 10.1021/cr030413t
- Palmer AG, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Meth Enzymol* 339:204–238.
- Peukert K, Staller P, Schneider A, et al. (1997) An alternative pathway for gene regulation by Myc. *EMBO J* 16:5672–5686. doi: 10.1093/emboj/16.18.5672
- Potter BM (2005) The Six Zinc Fingers of Metal-responsive Element Binding Transcription Factor-1 Form Stable and Quasi-ordered Structures with Relatively Small Differences in Zinc Affinities. *Journal of Biological Chemistry* 280:28529–28540. doi: 10.1074/jbc.M505217200
- Razin SV, Borunova VV, Maksimenko OG, Kantidze OL (2012) Cys2His2 zinc finger protein family: classification, functions, and major members. *Biochemistry Mosc* 77:217–226. doi: 10.1134/S0006297912030017

- Rieping W, Habeck M, Bardiaux B, et al. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23:381–382. doi: 10.1093/bioinformatics/btl589
- Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725–738. doi: 10.1038/nprot.2010.5
- Sauvé S, Tremblay L, Lavigne P (2004) The NMR solution structure of a mutant of the Max b/HLH/LZ free of DNA: insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. *J Mol Biol* 342:813–832. doi: 10.1016/j.jmb.2004.07.058
- Schmidt D (2004) Adaptive Evolution Drives the Diversification of Zinc-Finger Binding Domains. *Molecular Biology and Evolution* 21:2326–2339. doi: 10.1093/molbev/msh246
- Seoane J, Le H-V, Massagué J (2002) Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* 419:729–734. doi: 10.1038/nature01119
- Seoane J, Pouponnot C, Staller P, et al. (2001) TGFbeta influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b. *Nat Cell Biol* 3:400–408. doi: 10.1038/35070086
- Staller P, Peukert K, Kiermaier A, et al. (2001) Repression of p15INK4b expression by Myc through association with Miz-1. *Nat Cell Biol* 3:392–399. doi: 10.1038/35070076
- Stead MA, Trinh CH, Garnett JA, et al. (2007) A beta-sheet interaction interface directs the tetramerisation of the Miz-1 POZ domain. *J Mol Biol* 373:820–826. doi: 10.1016/j.jmb.2007.08.026
- Stogios PJ, Cuesta-Seijo JA, Chen L, et al. (2010) Insights into strand exchange in BTB domain dimers from the crystal structures of FAZF and Miz1. *J Mol Biol* 400:983–997. doi: 10.1016/j.jmb.2010.05.028
- Tjandra N, Garrett DS, Gronenborn AM, et al. (1997) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat Struct Biol* 4:443–449.
- Vranken WF, Boucher W, Stevens TJ, et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696. doi: 10.1002/prot.20449
- Wang C, Grey MJ, Palmer AG (2001) CPMG sequences with enhanced sensitivity to chemical exchange. *J Biomol NMR* 21:361–366.
- Wanzel M, Russ AC, Kleine-Kohlbrecher D, et al. (2008) A ribosomal protein L23-nucleophosmin circuit coordinates Miz1 function with cell growth. *Nat Cell Biol*. doi: 10.1038/ncb1764
- Wishart DS, Bigam CG, Yao J, et al. (1995) <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135–140.
- Wolfe SA, Nekludova L, Pabo CO (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* 29:183–212. doi: 10.1146/annurev.biophys.29.1.183
- Zandarashvili L, Vuzman D, Esadze A, et al. (2012) Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1121500109

## 6. Discussion

Premièrement, il est important de mentionner que la région étudiée de la protéine a été sélectionnée de façon à ce que les coupures dans sa séquence primaire coïncident avec des espaceurs non structurés. Il a déjà été montré dans la littérature que les domaines contenant des ZFs entiers adoptent une conformation similaire à leur véritable conformation dans la protéine entière (POTTER et al., 2005), et donc la structure de Miz 5-8 sera similaire à la structure des ZFs 5 à 8 dans Miz-1 entière. En général, les ZFs séparés par un espaceur canonique sont relativement indépendants les uns des autres (BÉDARD et al., 2012).

L'article rapporte d'abord et avant tout la structure des ZFs 5 à 8 de Miz-1. Les ZFs 5, 7 et 8 adoptent la structure  $\beta\beta\alpha$  traditionnelle des ZFs, mais il est impossible de confirmer cette structure pour le ZF 6, à cause d'un manque important de NOEs dans la région de ce qui devrait être le feuillet  $\beta$ . En fait, tel que montré dans l'article, le ZF 6 subit des mouvements et de l'échange conformationnel extensif, tous deux à l'échelle de la micro-/milliseconde. Les ZFs 5 et 7 subissent eux aussi ce type d'échange conformationnel. De tels échanges avaient été remarqués dans la protéine Egr-1, mais leur intensité était de beaucoup moindre. Ceux présents dans le ZF 1 de Egr-1 avaient été attribués à l'absence du pont salin entre le résidu généralement basique situé deux résidus avant la dernière histidine conservée et le glutamate de l'espaceur TGEKP, qui dans le cas du ZF 1 de Egr-1 étaient remplacé par une thréonine et TGQKP (ZANDARASHVILI et al., 2012). Dans le cas de Miz 5-8, deux des trois espaceurs avaient une séquence inhabituelle, soit DTDKE et IADGP (voir la figure 1 de l'article). Le premier de ces espaceurs serait responsable d'importantes répulsions électrostatiques entre le N-terminal du ZF 6 et le C-terminal du ZF 5, où on retrouve un glutamate au lieu du résidu basique conservé, une occurrence rare selon une base de données de ZFs (SCHIMDT, 2004). Ces répulsions sont la cause la plus probable d'une déstabilisation du microétat correspondant à l'adoption par le ZF 6 d'une structure tertiaire canonique pour un ZF.

La présence des deux groupements déstabilisants sur le ZF 6 et l'espaceur précédent rend hautement improbable la possibilité qu'il puisse se lier à l'ADN, puisque la conformation adoptée par un ZF lié à l'ADN serait beaucoup trop exigeante énergétiquement pour ce ZF en particulier. Ceci signifierait donc que l'une des hypothèses qui avait été formulée au départ, à savoir que le triplet de ZFs 5 à 7 était l'un des triplets responsables de la reconnaissance du promoteur de p15<sup>INK4B</sup>, est vraisemblablement erronée. Elle a été corrigée suite aux découvertes présentées dans l'article, et le modèle présenté à la figure 8 de l'article constitue donc le résultat le plus important rapporté dans celui-ci, exception faite des structures de protéine.

L'article se concentrait majoritairement sur la région allant de l'extrémité C-terminale du ZF 5 à la fin du feuillet  $\beta$  du ZF 6, et en particulier sur les caractéristiques dynamiques de cette région. Toutefois, Miz 5-8 possède quelques autres particularités structurales et dynamiques dignes de mention. Tout d'abord, les histidines coordonnatrices du zinc dans le ZF 5 ont une orientation différente de celle des autres ZFs. Ceci ne devrait avoir aucun effet sur la stabilité de la structure de ce ZF. Dans la protéine Miz 8-10, par exemple, l'histidine finale du ZF 8 a une orientation différente de celle des ZFs 9 et 10 (BÉDARD et al., 2012).

Ensuite, dans le cœur hydrophobe du ZF 5 se retrouve une lysine, en lieu et place de la leucine conservée de l'hélice  $\alpha$ . La figure 29 montre un gros plan de cette Lys19 (encadrée en bleu sur l'image) dans la superposition des vingt structures obtenues pour le ZF 5. Il est possible de remarquer que les carbones  $\alpha$ ,  $\beta$  et  $\gamma$  de cette lysine sont bien définis, ce qui signifie que ces carbones servent à effectuer des interactions de Van der Waals avec les autres résidus du cœur hydrophobe. La figure 29 met également en évidence que les N $\epsilon$ H de cette lysine ont la possibilité de former des ponts hydrogène soit avec le groupement carbonyle de la Gln5 (encadrée en noir), soit avec la chaîne latérale de l'Asp7 (encadré en rouge), soit avec les deux à la fois. Ces ponts hydrogène pourraient servir à stabiliser la structure du ZF 5. Toutefois, en observant la structure cristalline de EGR-1 (PAVLETICH et PABO, 1991), dont la leucine conservée du ZF 3 est remplacée par une arginine, la présence de ponts salins entre cette arginine et les phosphates du

squelette de l'ADN est probable étant donné leur proximité. Il est donc possible que la Lys19 de Miz 5-8 ait une fonction similaire, que ce soit avec un phosphate de l'ADN ou un résidu chargé négativement sur une protéine avec laquelle interagirait Miz-1. Des expériences supplémentaires seront nécessaires pour déterminer si c'est le cas.



Figure 29 : Lys19 dans le cœur hydrophobe du ZF 5

Gros plan de la Lys19 (encadrée en bleu) qui forme des ponts hydrogène avec la Gln5 (encadrée en noir) et l'Asp7 (encadrée en rouge) sur la superposition des vingt structures du ZF 5 de Miz 5-8. Les ponts hydrogène sont représentés par des pointillés jaunes. La structure a été réalisée à l'aide du logiciel PyMOL ([www.pymol.org](http://www.pymol.org)).

Il est fait mention à de très nombreuses reprises dans l'article des échanges conformationnels que subit le ZF 6. L'équation 2 dans l'article est en fait une simplification de l'équation suivante (MULDER et al., 2001) :

$$R_{2,eff}(v_{CPMG}) = \frac{1}{2} \left( R_{2A}^0 + R_{2B}^0 + k_{ex} - v_{CPMG} \cosh^{-1} (D_+ \cosh \eta_+ - D_- \cosh \eta_-) \right)$$

$$D_{\pm} = \frac{1}{2} \left( \pm 1 + \frac{\Psi + 2\Delta\omega^2}{\sqrt{\Psi^2 + \zeta^2}} \right)$$

$$\eta_{\pm} = \frac{1}{v_{CPMG}} \sqrt{\frac{\pm\Psi + \sqrt{\Psi^2 + \zeta^2}}{2}}$$

$$\Psi = \left( R_{2A}^0 - R_{2B}^0 - P_A k_{ex} + P_B k_{ex} \right)^2 - \Delta\omega^2 + 4P_A P_B k_{ex}^2$$

$$\zeta = 2\Delta\omega \left( R_{2A}^0 - R_{2B}^0 - P_A k_{ex} + P_B k_{ex} \right)$$

dans la limite où la population de l'état B ( $P_B$ ) est beaucoup plus faible que celle de l'état A ( $P_A$ ). Dans l'équation 2 de l'article,  $P_A$  et  $P_B$  sont indissociables, ils sont également indissociables de  $\Delta\omega^2$ , et aucune de ces trois variables n'apparaît ailleurs dans l'équation. Ceci signifie que les populations des deux états n'influenceront pas la valeur de  $R_{ex}$  dans cette limite. Dans le cas du ZF 6, les deux états en question sont, selon toute vraisemblance, un état complètement replié et un état où le ZF est plutôt sous forme de « random coil ». Ce dernier serait l'état le moins probable (l'état B), d'où l'absence de l'amalgame de résonances typique de la présence de « random coil » sur le HSQC de Miz 5-8 (figure 2 de l'article). Ceci explique également la faible intensité des résonances appartenant au ZF 6 sur le même HSQC.

L'article démontre l'importance de l'espaceur, et plus particulièrement du pont salin entre l'espaceur et le dernier tour de l'hélice  $\alpha$  le précédant, dans la stabilité de la structure des ZFs. Dans le cas de Miz-1, cette instabilité du ZF 6 pourrait servir à délimiter deux groupes de ZFs : l'un servant à lier l'ADN, l'autre médiant des interactions protéine-protéine. Il a été montré que Miz-1 pouvait se lier à SMAD3 et SMAD4 via les ZFs 1 à 4 (SEOANE et al, 2001). Or, SMAD3 peut aussi se lier directement au promoteur de  $p15^{INK4B}$  ou de  $p21^{CIP1}$  (FENG et al, 2000). Miz-1 doit donc se lier simultanément à SMAD3/4 et sur deux sites très espacés du promoteur de  $p15^{INK4B}$  (SEOANE et al, 2002), et conjointement SMAD3/4 doit se lier au promoteur de  $p15^{INK4B}$ . Ceci n'est possible que si une partie de Miz-1 s'éloigne de l'ADN, ce qui serait rendu possible par l'instabilité du ZF 6. Afin de résumer toutes ces informations, nous proposons un modèle de liaison de Miz-1 au promoteur de  $p15^{INK4B}$  (voir figure 30). Celui-ci inclut l'enroulement du promoteur sur lui-même de façon à permettre à Miz-1 de lier les deux sites qu'il reconnaît, ce qui sera probablement à l'aide des ZFs 7 à 13. Le ZF 6 sert de charnière, permettant aux ZFs 1 à 5 de rester éloignés de l'ADN afin de se lier à SMAD3/4 alors que SMAD3/4 est lui-même lié au promoteur. Des résultats préliminaires obtenus par Mikaël Bédard laissent croire que les ZFs 3 et 4 adoptent une orientation relative assez particulière, probablement due à l'espaceur inhabituel les séparant (SGEAR) et une leucine située entre le deuxième brin  $\beta$  et l'hélice  $\alpha$ . Nous croyons qu'il soit possible que cet arrangement inusité de ZFs soit responsable de la liaison aux protéines SMAD.

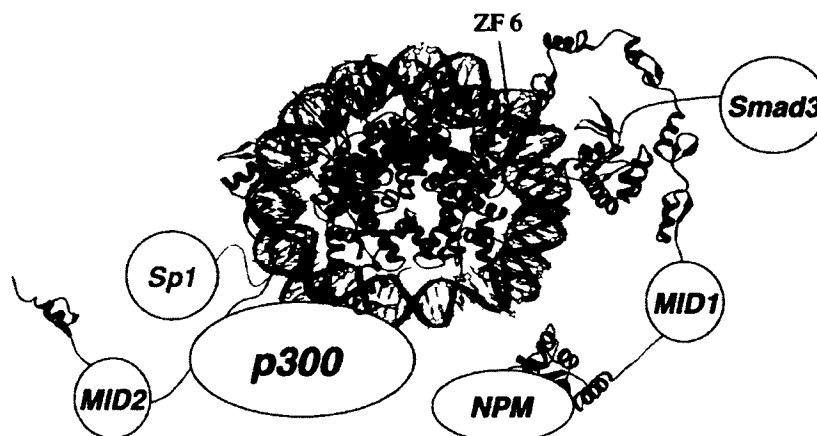


Figure 30 : Modèle proposé de la liaison de Miz-1 à l'ADN avec les partenaires impliqués

Miz-1 (bleu) est activée par la nucléophosmine (NPM) lorsqu'elle se lie à son domaine BTB/POZ. Miz-1 peut alors se lier à l'ADN (rouge) via un ou des triplets de ZFs parmi les ZFs 7 à 13. Le ZF 6 sert de charnière afin de faire diverger le N-terminal de Miz-1 de l'ADN. Ceci permet aux ZFs 1 à 4 de Miz-1 de se lier à SMAD3/4 (vert) qui se lie simultanément à l'ADN. Finalement, la portion C-terminale de Miz-1 ainsi que Sp1 (rose), qui est également lié à l'ADN, recrutent p300.

L'absence du pont salin entre le dernier tour d'hélice  $\alpha$  et l'espaceur suivant est critique dans la flexibilité relative d'un ZF par rapport au ZF suivant. Le groupe de Zandarashvili (ZANDARASHVILI et al., 2012) a montré que la simple absence de ce pont salin permettait à un ZF d'agir en tant que fureteur, de façon à ce qu'un triplet de ZFs puisse changer de site de liaison à l'ADN et ainsi de trouver sa séquence de liaison spécifique plus rapidement. Dans le cas de Miz-1, la présence de répulsion électrostatique en lieu et place du pont salin est à l'origine de l'explication que le ZF 6 sert de charnière à la protéine, de façon à séparer un domaine d'interaction à l'ADN d'un domaine d'interaction protéique. Ceci permet à son tour d'émettre l'hypothèse qu'il en serait de même pour l'ensemble des protéines à poly-ZFs : la présence de résidus se repoussant mutuellement à ces positions spécifiques de la séquence primaire d'un ZF pourrait délimiter deux domaines d'interaction différents sur les nombreuses protéines à poly-ZF. Cette information pourrait ainsi faciliter l'identification de la fonction des différents groupes de ZFs de ces protéines, dont un grand nombre a un rôle encore non identifié. Dans le même ordre d'idées, les ZFs 4 et 11 de Miz-1 présentent tous deux une paire de résidus ne pouvant pas former de pont salin entre ces deux positions clés (voir figure 1a de l'article). Il s'agit toutefois d'une simple absence de pont salin plutôt que de présence de répulsion, donc ces deux ZFs risquent davantage de se comporter comme le ZF 1 de



EGR-1, c'est-à-dire en fureteurs (ZANDARASHVILI et al., 2012). Ceci serait intéressant à vérifier par RMN.

Des rapports  $T_1/T_2$  très élevés ont été remarqués aux extrémités C-terminales des hélices  $\alpha$  dans plusieurs protéines à ZFs auparavant (BRÜSCHWEILER et al., 1995; POTTER, 2005; ZANDARASHVILI et al., 2012), incluant des travaux antérieurs effectués au laboratoire sur Miz 8-10. Ces rapports ont été attribués à de l'anisotropie de diffusion rotationnelle, mais l'article démontre que ces rapports élevés sont en fait dus, au moins en partie, à des échanges conformationnels dans le cas de Miz 5-8 (voir figure 5 de l'article). À notre connaissance, il s'agit de la première fois que cette explication est avancée. Nous proposons en outre que ces échanges aux extrémités C-terminales des hélices  $\alpha$  expliquent la présence de rapports  $T_1/T_2$  élevés pour toutes les autres protéines à ZFs, ce qui a été confirmé au laboratoire pour Miz 8-10. Dans tous les cas que nous avons observés, les rapports  $T_1/T_2$  sont très faibles ou absents dans la partie N-terminale des hélices  $\alpha$  des ZFs, qui est la partie de ces domaines qui sert à reconnaître et lier les bases d'ADN. Nous pensons que l'extrémité N-terminale des hélices soit bien définie de manière à toujours reconnaître la même séquence d'ADN, et que l'extrémité C-terminale subisse des échanges conformationnels de façon à pouvoir s'adapter à la forme de l'ADN pour permettre la liaison de ZFs consécutifs.

Néanmoins, au meilleur de nos connaissances, Miz 5-8 présente un réseau d'échanges conformationnels beaucoup plus important que tout ce qui a été mesuré sur les autres protéines à poly-ZF. Les échanges en C-terminal des hélices  $\alpha$  ne suffisent pas à expliquer cette étendue. Nous suggérons que leur omniprésence dans Miz 5-8 soit en partie due à l'échange conformationnel octroyé par l'instabilité de l'espaceur entre les ZFs 5 et 6 et celle du ZF 6. Les ZFs étant des domaines assez rigides (voir figure 4a de l'article) et partiellement inter-dépendants, les échanges conformationnels autour du ZF 6 pourraient se transmettre aux ZFs adjacents simplement via des mouvements globaux de la protéine.

En perspectives, notre laboratoire veut confirmer ou ajuster le modèle de liaison de Miz-1 à l'ADN présenté à la figure 30. Les études nécessaires à cet objectif comprennent la détermination de la structure et de la dynamique des ZFs 1 à 4 et 11 à 13 de même que celle des MIDs, l'analyse de l'interaction entre Miz-1 et SMAD3/4, et l'identification structurale des ZFs impliqués dans la liaison aux promoteurs de p15<sup>INK4B</sup> ou de p21<sup>CIP1</sup>. Par la suite, l'élaboration du modèle de liaison de c-Myc à Miz-1 nécessitera la détermination structurale du complexe c-Myc/Miz-1. Toutes ces expériences seront effectuées par RMN.

## 7. Conclusion

Ce mémoire présente les résultats de l'analyse structurale et dynamique de Miz 5-8 par RMN en solution. Les ZFs 5, 7 et 8 adoptent une structure classique de doigt de zinc de type Cys<sub>2</sub>His<sub>2</sub>, à savoir un court feuillet  $\beta$  antiparallèle suivi d'une hélice  $\alpha$  d'une dizaine de résidus de longueur, avec un atome de zinc coordonné par deux cystéines dans le feuillet  $\beta$  et deux histidines dans l'hélice  $\alpha$ . Le ZF 6 a une hélice  $\alpha$  bien définie, mais son feuillet  $\beta$  subit d'importants changements conformationnels à l'échelle de la microseconde-milliseconde, empêchant la détection des résonances de cette région de la protéine. Ces échanges sont dus à la présence importante de résidus se repoussant mutuellement, incluant à des positions où sont généralement conservés des résidus formant un pont salin stabilisateur.

Ce mémoire démontre également la présence d'échanges conformationnels distribués sur l'ensemble de la protéine. Des valeurs importantes de  $R_{ex}$  ont été mesurées aux extrémités C-terminales de chaque ZF de Miz 5-8, et nous proposons que ces échanges soient présents dans l'ensemble des ZFs, puisque d'importants rapports  $T_1/T_2$  ont déjà été rapportés dans la littérature à ces positions.

Les informations révélées par la structure et l'analyse dynamique de Miz 5-8 ont permis de déterminer que contrairement à ce qui était formulé dans l'hypothèse initiale, ce n'est vraisemblablement pas le triplet de ZFs 5 à 7 qui servirait à se lier à l'ADN. Au

contraire, le ZF 6 servirait de charnière pour délimiter le domaine d'interaction protéique du domaine d'interaction avec l'ADN dans Miz-1. Ceci nous a permis d'élaborer un nouveau modèle de liaison entre Miz-1, SMAD3/4 et le promoteur de p15<sup>INK4B</sup>, modèle qui sera confirmé ou retravaillé suite aux travaux sur Miz-1 qui seront effectués prochainement au laboratoire.

## **Remerciements**

En premier lieu, je tiens à remercier mon directeur de recherche, Pierre Lavigne, qui m'a d'abord permis d'effectuer mes études dans son laboratoire alors que j'avais un parcours plutôt inhabituel pour quelqu'un désirant étudier en pharmacologie. Il a un don pour motiver ses étudiants qui m'a souvent été utile! Et son habitude de dire toutes les idées qui lui passent par la tête a mené à de nombreuses discussions scientifiques (et parfois non scientifiques) très intéressantes!

J'aimerais remercier ensuite Mikaël Bédard, qui travaillait sur Miz 8-10 à l'époque où j'en étais à l'analyse RMN de ma protéine, et qui m'a énormément aidé en m'enseignant à analyser les différents spectres. Il m'a également aidé pour mettre au point le protocole d'expression des protéines et à me faire comprendre certains principes de RMN que je trouvais nébuleux, surtout en ce qui a trait à la dynamique des protéines.

Je me dois également de remercier Josée Bilodeau, Martin Montagne, Marie-Ève Beaulieu, Loïka Maltais et les membres du laboratoire de Yves Dory pour m'avoir montré l'utilisation des différents appareils et techniques du laboratoire, à savoir la transformation et la culture de bactéries, le SDS-PAGE, le FPLC, le CD, et dans le cas des membres du laboratoire Dory, le HPLC. Sans votre aide, le physicien que je suis n'aurait jamais pu être à l'aise dans un laboratoire de sciences biologiques!

Finalement j'aimerais remercier ma famille de m'avoir encouragé de toutes les façons possibles pendant mes études, même si elles ont été plus longues que prévu!

## LISTE DES RÉFÉRENCES – LIVRES

- ME BEAULIEU (2011) *Biologie structurale de c-Myc et Max: Évidences pour un nouveau mécanisme de transrépression par Myc*, Université de Sherbrooke. Sherbrooke
- J CAVANAGH, WJ FAIRBROTHER, AG PALMER, NJ SKELTON (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, London
- C COHEN-TANNOUDI, B DIU, F LALOË (1998) *Mécanique quantique I*, Hermann, Paris
- C COHEN-TANNOUDI, B DIU, F LALOË (2007) *Mécanique quantique II*, Hermann, Paris
- DJ GRIFFITHS (2005) *Introduction to Quantum Mechanics second edition*, Pearson Prentice Hall, Upper Saddle River
- PJ HORE (2001) *Nuclear Magnetic Resonance*, Oxford University Press, New York
- NE JACOBSEN (2007) *NMR Spectroscopy Explained*, John Wiley & Sons, Hoboken
- JB LAMBERT, EP MAZZOLA (2004) *Nuclear Magnetic Resonance Spectroscopy*, Pearson Prentice Hall, Upper Saddle River
- PK MANDAL, A MAJUMDAR (2004) *A Comprehensive Discussion of HSQC and HMQC Pulse Sequences*, Wiley Periodicals, Hoboken
- M MONTAGNE (2008) *Élucidation des déterminants structuraux impliqués dans la reconnaissance moléculaire entre MXD1 et MAX et la liaison à l'ADN*, Université de Sherbrooke, Sherbrooke
- RM SILVERSTEIN, FX WEBSTER, DJ KIEMLE (2005) *Spectrometric Identification of Organic Compounds seventh edition*, John Wiley & Sons, Hoboken

## LISTE DES RÉFÉRENCES – ARTICLES

- S ADHIKARY, K PEUKERT, H KARSUNKY, V BEUGER, W LUTZ, HP ELSÄSSER, T MÖRÖY, M EILERS (2003) Miz1 is required for early embryonic development during gastrulation, *Mol Cell Biol* (23) 7648-7657
- S BASU, Q LIU, Y QIU, F DONG (2009) Gfi-1 represses CDKN2B encoding p15INK4B through interaction with Miz-1, *Proc Natl Acad Sci USA* (106) 1433-1438
- M BÉDARD, L MALTAIS, M-E BEAULIEU, J BILODEAU, D BERNARD, P LAVIGNE (2012) NMR structure note: solution structure of human Miz-1 zinc fingers 8 to 10, *J Biomol NMR*, doi 10.1007/s10858-012-9670-1
- H BOWEN, A LAPHAM, E PHILLIPS, I YEUNG, M ALTER-KOLTUNOFF, BZ LEVI, VH PERRY, DA MANN, CH BARTON (2003) Characterization of the murine Nramp1 promoter: requirements for transactivation by Miz-1, *J Biol Chem* (278) 36017-36026
- KJ BRAYER, DJ SEGAL, (2008) Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains, *Cell Biochem Biophys* (50) 111-131

- RS BROWN (2005) Zinc finger proteins: getting a grip on RNA, *Curr Opin Struct Biol* (15) 94-98
- R BRÜSCHWEILER, X LIAO, PE WRIGHT (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling, *Science* (268) 886-889
- Y CHOO, A KLUG (1993) A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA, *Nucleic Acids Res* (21) 3341-3346
- VH COWLING, MD COLE (2008) An N-Myc truncation analogous to c-Myc-S induces cell proliferation independently of transactivation but dependent on Myc homology box II, *Oncogene* (27) 1327-1332
- R DALLA-FAVERA, M BREGNI, J ERIKSON, D PATTERSON, RC GALLO, CM CROCE (1982) Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells, *Proc Natl Acad Sci USA* (79) 7824-7827
- L DAUPHINOT, C DE OLIVEIRA, T MELOT, N SEVENET, V THOMAS, BE WEISSMAN, O DELATTRE (2001) Analysis of the expression of cell cycle regulators in Ewing cell line: EWS-FLI-1 modulates p57KIP2 and c-Myc expression, *Oncogene* (20) 3258-3265
- S DEL RIO, SR MENEZES, DR SETZER (1993) The function of individual zinc fingers in sequence-specific DNA recognition by transcription factor IIIA, *J Mol Biol* (233) 567-579
- Y DOI-KATAYAMA, H HIROTA, T TOMIZAWA, S KOSHIBA, T KIGAWA, S YOKOYAMA (2005) Solution Structure of RSGI RUH-035, a Zn-ribbon module in Mouse cDNA, *unpublished*, doi 10.2210/pdb2con/pdb
- M EILERS, RN EISENMAN (2008) Myc's broad reach, *Genes Dev* (22) 2755-2766
- RO EMERSON, JH THOMAS (2009) Adaptative evolution in zinc finger transcription factors, *PLoS Genet* (5) doi 10.1371/journal.pgen.1000325
- XH FENG, X LIN, R DERYNCK (2000) Smad2, Smad3 and Smad4 cooperate with Sp1 to induce p15(ink4B) transcription in response to TGF-beta, *EMBO J* (19) 5178-5193
- A GEBHARDT, M FRYE, S HEROLD, SA BENITAH, K BRAUN, B SAMANS, FM WATT, HP ELSÄSSER, M EILERS (2006) Myc regulates keratinocyte adhesion and differentiation via complex formation with Miz1, *J Cell Biol* (172) 139-149
- JN GLOVER, SC HARRISON (1995) Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA, *Nature* (373) 257-261
- C GRANDORI, N GOMEZ-ROMAN, ZA FELTON-EDKINS, C NGOUENET, DA GOLLOWAY, RN EISENMAN, RJ WHITE (2005) c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I, *Nat Cell Biol* (7) 311-318
- B HERKERT, M EILERS (2010) Transcriptional repression: the dark side of myc, *Genes Cancer* (1) 580-586
- S HEROLD, M WANZEL, V BEUGER, C FROHME, D BEUL, T HILLUKKALA, J SYVAOJA, HP SALUZ, F HAENEL, M EILERS (2002) Negative regulation of the mammalian UV response by Myc through association with Miz-1, *Mol Cell* (10) 509-521
- M ISALAN, Y CHOO, A KLUG (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition, *Proc Natl Acad Sci USA* (94) 5617-5621

- VA JARYMOWYCZ, MJ STONE (2006) Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences, *Chem Rev* (106) 1624-1671
- T KAPLAN, N FRIEDMAN, H MARGALIT (2005) Ab initio prediction of transcription factor targets using structural knowledge, *PLoS Comput Biol* (1) doi 10.1371/journal.pcbi.0010001
- L KIME, SC WRIGHT (2003) Mad4 is regulated by a transcriptional repressor complex that contains Miz-1 and c-Myc, *Biochem J* (370) 291-298
- SS KRISHNA, I MAJUMDAR, NV GRISHIN (2003) Structural classification of zinc fingers: survey and summary, *Nucleic Acids Res* (31) 532-550
- JH LAITY, HJ DYSON, PE WRIGHT (2000) DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers, *J Mol Biol* (295) 719-727
- BC LECHTENBERG, MD ALLEN, TJ RUTHERFORD, SM FREUND, M BYCROFT (2009) Solution structure of the FCS zinc finger domain of the human polycomb group protein L(3)mbt-like 2, *Protein Sci* (18) 657-661
- L LI, C NERLOV, G PRENDERGAST, D MACGREGOR, EB ZIFF (1994) c-Myc represses transcription in vivo by a novel mechanism dependent on the initiator element and Myc box II, *EMBO J* (13) 4070-4079
- KM LEE, WI CHOI, DI KOH, YJ KIM, BN JEON, JH YOON, CE LEE, SH KIM, J OH, MW HUR (2011) The proto-oncoprotein KR-POK represses transcriptional activation of CDKN1A by MIZ-1 through competitive binding, *Oncogene*, doi: 10.1038/onc.2011.331
- JD LICCHESI, L VAN NESTE, VK TIWARI, L COPE, X LIN, SB BAYLIN, JG HERMAN (2010) Transcriptional regulation of Wnt inhibitory factor-1 by Miz-1/c-Myc, *Oncogene*, doi: 10.1038/onc.2010.322
- TD LITTLEWOOD, P KREUZALER, GI EVAN (2012) All things to all people, *Cell* (151) 11-13
- J LOVÉN, DA ORLANDO, AA SIGOVA, CY LIN, PB RAHL, CB BURGE, DL LEVENS, TI LEE, RA YOUNG (2012) Revisiting global gene expression analysis, *Cell* (151) 476-482
- Z LUZ, S MEIBOOM (1963) Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution—Order of the Reaction with Respect to Solvent, *J Chem Phys* (39) 366-370
- PC MA, MA ROULD, H WEINTRAUB, CO PABO (1994) Crystal structure of MyoD bHLH domain – DNA complex: perspectives on DNA recognition and implications for transcriptional activation, *Cell* (77) 451-459
- SB MCMAHON, MA WOOD, MD COLE (2000) The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc, *Mol Cell Biol* (20) 556-562
- F MORRISH, N NERETTI, JM SEDIVY, DM HOCKENBERY (2008) The oncogene c-Myc coordinates regulation of metabolic networks to enable rapid cell cycle entry, *Cell Cycle* (7) 1054-1066
- FJ MOY, AP SEDDON, EB CAMPBELL, P BÖHLEN, R POWERS (1995) <sup>1</sup>H, <sup>15</sup>N, <sup>13</sup>C and <sup>13</sup>CO assignments and secondary structure determination of basic fibroblast growth factor using 3D heteronuclear NMR spectroscopy, *J Biomol NMR* (6) 245-254
- FA MULDER, A MITTERMAIER, B HON, FW DAHLQUIST, LE KAY (2001) Studying excited states of proteins by NMR spectroscopy, *Nat Struct Biol* (8) 932-935

- MS NIEDERHUT, BJ GIBBONS, S PEREZ-MILLER, TD HURLEY (2001) Three-dimensional structures of the three human class I alcohol dehydrogenases, *Protein Sci* (10) 697-701
- NP PAVLETICH, CO PABO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å, *Science* (252) 809-817
- EA PEROZA, R SCHMUCKI, P GUNTERT, E FREISINGER, O ZERBE (2009) The beta(E)-domain of wheat E(c)-1 metallothionein: a metal-binding domain with a distinctive structure, *J Mol Biol* (387) 207-218
- K PEUKERT, P STALLER, A SCHNEIDER, G CARMICHAEL, F HÄNEL, M EILERS (1997) An alternative pathway for gene regulation by Myc, *EMBO J* (16) 5672-5686
- BM POTTER, LS FENG, P PARASURAM, VA MATSKEVICH, JA WILSON, GK ANDREWS, JH LAITY (2005) The six zinc fingers of metal-responsive element binding transcription factor-1 form stable and quasi-ordered structures with relatively small differences in zinc affinities, *J Biol Chem* (280) 28529-28540
- S QUINTAL, A VIEGAS, S ERHARDT, EJ CABRITA, NP FARRELL (2012) Platinated DNA affects zinc finger conformation. Interaction of a platinated single-stranded oligonucleotide and the C-terminal zinc finger of nucleocapsid protein HIVNCp7, *Biochemistry* (51) 1752-1761
- S SAUVÉ, L TREMBLAY, P LAVIGNE (2004) The NMR solution structure of a mutant of the Max b/HLH/LZ free of DNA: insights into the specific and reversible DNA binding mechanism of dimeric transcription factors, *J Mol Biol* (342) 813-832
- J SEOANE, C POUPONNOT, P STALLER, M SCHADER, M EILERS, J MASSAGUÉ (2001) TGFbeta influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b, *Nat Cell Biol* (3) 400-408
- J SEOANE, HV LE, J MASSAGUÉ (2002) Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage, *Nature* (419) 729-734
- L SOUCEK, JR WHITFIELD, NM SODIR, D MASSÓ-VALLÉS, E SERRANO, AN KARNEZIS, LB SWIGART, GI EVAN (2013) Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice, *Genes Dev* (27) 504-513
- P STALLER, K PEUKERT, A KIERMAIER, J SEOANE, J LUKAS, H KARSUNKY, T MÖRÖY, J BARTEK, J MASSAGUÉ, F HÄNEL, M EILERS (2001) Repression of p15INK4b expression by Myc through association with Miz-1, *Nat Cell Biol* (3) 392-399
- PJ STOGIOS, JA CUESTA-SEIJO, L CHEN, NC POMROY, GG PRIVÉ (2010) Insights into strand exchange in BTB domain dimers from the crystal structures of FAZF and Miz-1, *J Mol Biol* (400) 983-997
- R TAUB, I KIRSCH, C MORTON, G LENOIR, D SWAN, S TRONICK, S AARONSON, P LEDER (1982) Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells, *Proc Natl Acad Sci USA* (79) 7837-7841
- RY TSAI, RR REED (1998) Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz, *Mol Cell Biol* (18) 6447-6456
- RB TURNER, DL SMITH, ME ZAWROTNY, MF SUMMERS, MC POSEWITZ, DR WINGE (1998) Solution structure of a zinc domain conserved in yeast copper-regulated transcription factors, *Nat Struct Biol* (5) 551-555

M WANZEL, AC RUSS, D KLEINE-KOHLBRECHER, E COLOMBO, PG PELICCI, M EILERS (2008) A ribosomal protein L23-nucleophosmin circuit coordinates Miz1 function with cell growth, *Nat Cell Biol*, doi: 10.1038/ncb1764

A WIKSTRÖM, H BERGLUND, C HAMBÆUS, S VAN DEN BERG, T HÄRD (1999) Conformational dynamics and molecular recognition: backbone dynamics of the estrogen receptor DNA-binding domain, *J Mol Biol* (289) 963-979

JM WOJCIAK, MA MARTINEZ-YAMOUT, HJ DYSON, PE WRIGHT (2009) Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains, *EMBO J* (28) 948-958

SA WOLFE, HA GREISMAN, EI RAMM, CO PABO (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code, *J Mol Biol* (285) 1917-1934

SA WOLFE, L NEKLUDOVA, CO PABO (2000) DNA recognition by Cys<sub>2</sub>His<sub>2</sub> zinc finger proteins, *Annu Rev Biophys Biomol Struct* (29), 183-212

L ZANDARASHVILI, D VUZMAN, A ESADZE, Y TAKAYAMA, D SAHU, Y LEVY, J IWAHARA (2012) Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1, *Proc Natl Acad Sci USA* (109) E1724-E1732

KI ZELLER, AG JEGGA, BJ ARONOW, KA O'DONNELL, CV DANG (2003) An integrated database of genes responsive to the Myc oncogenic factor: identification of direct genomic targets, *Genome Biol* (4) doi 10.1186/gb-2003-4-10-r69, <http://www.myc-cancer-gene.org/site/mycTargetDB.asp>



## Annexe 1 – Matériel supplémentaire à l'article

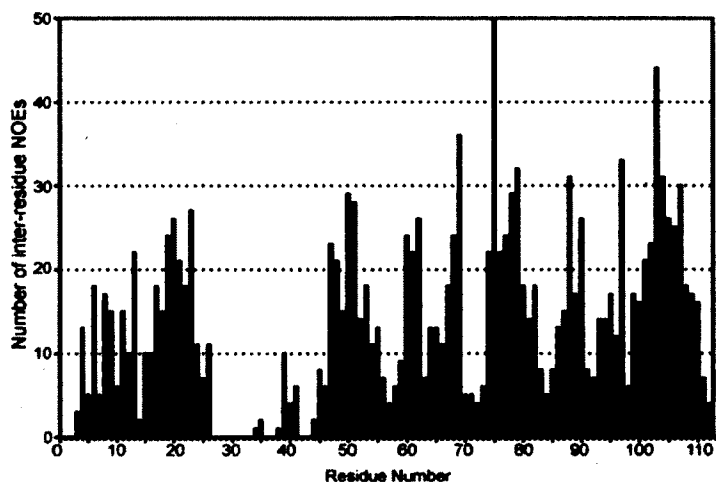


Figure S1 : Distribution de NOEs inter-résidus en fonction de la structure primaire de Miz 5-8

Histogramme donnant le nombre de NOEs inter-résidus en fonction de la position du résidu dans la structure primaire de la protéine. La faible quantité de tels NOEs dans la séquence 27-46 saute aux yeux.

Tableau supplémentaire 1: Paramètres calculés de relaxation de dispersion

Residue	$R_{2,eff}(v \rightarrow \infty)$	$\sigma R_{2,eff}(v \rightarrow \infty)$	$k_{ex}$	$\sigma k_{ex}$	$R_{ex}$	$\sigma R_{ex}$
Met1	---	---	---	---	---	---
Lys2	---	---	---	---	---	---
Pro3	---	---	---	---	---	---
Tyr4	12,8	0,2	760	60	8	1
Gln5	15,5	0,3	500	100	7	2
Cys6	15,9	0,1	300	100	2	1
Asp7*	17,4	0,7	---	---	---	---
Tyr8	13,8	0,2	280	50	18	4
Cys9	14	0,2	350	40	21	3
Gly10	17,3	0,3	420	50	17	3
Arg11	16,92	0,08	160	70	2	1
Ser12	13,37	0,09	350	40	6	1
Phe13	13,46	0,06	260	20	5,2	0,7
Ser14	---	---	---	---	---	---
Asp15	13,72	0,07	310	50	2,8	0,7
Pro16	---	---	---	---	---	---
Thr17	16	0,2	400	80	6	2
Ser18	14,3	0,07	350	90	1,6	0,7
Lys19*	15,7	0,08	---	---	---	---
Met20	17,1	0,08	200	40	4	1
Arg21	15,43	0,06	310	30	4,1	0,7
His22	16	0,2	320	80	6	2
Leu23	16,9	0,2	680	60	11	2
Glu24	18,7	0,5	300	100	40	20
Thr25	17,5	0,5	650	60	29	4
His26	18,4	0,6	500	100	18	6
Asp27	---	---	---	---	---	---
Thr28	---	---	---	---	---	---
Asp29	---	---	---	---	---	---
Lys30	---	---	---	---	---	---
Glu31	---	---	---	---	---	---

His32	---	---	---	---	---	---	---	---
Lys33	---	---	---	---	---	---	---	---
Cys34	20,6	0,8	500	100	17	7		
Pro35	---	---	---	---	---	---	---	---
His36*	24	1	---	---	---	---	---	---
Cys37	---	---	---	---	---	---	---	---
Asp38*	27	1	---	---	---	---	---	---
Lys39	19,4	0,1	300	70	4	2		
Lys40*	24	1	---	---	---	---	---	---
Phe41	19,8	0,6	240	80	100	70		
Asn42	---	---	---	---	---	---	---	---
Gln43*	24,8	0,5	---	---	---	---	---	---
Val44	---	---	---	---	---	---	---	---
Gly45*	22	1	---	---	---	---	---	---
Asn46	18,1	0,1	170	40	10	3		
Leu47*	20,5	0,3	---	---	---	---	---	---
Lys48	18,7	0,2	260	40	16	3		
Ala49	19,66	0,08	280	20	10	1		
His50	19,5	0,1	130	50	10	6		
Leu51*	34	2	---	---	---	---	---	---
Lys52	20,5	0,2	410	30	19	2		
Ile53	21,8	0,4	410	60	21	5		
His54	23,9	0,6	500	100	23	7		
Ile55	19,5	0,8	300	200	30	30		
Ala56	16	1	800	200	39	6		
Asp57	12	0,4	380	80	16	5		
Gly58	8,4	0,4	530	70	15	3		
Pro59	---	---	---	---	---	---	---	---
Leu60*	10,1	0,5	---	---	---	---	---	---
Lys61	16,9	0,3	260	70	14	5		
Cys62	16,9	0,5	380	90	22	7		
Arg63*	14,7	0,6	---	---	---	---	---	---
Glu64	12,6	0,3	270	70	19	7		
Cys65	14,7	0,2	370	60	7	2		
Gly66	16,3	0,2	400	100	3	2		
Lys67	18	0,2	330	80	4	2		
Gln68	15	0,2	240	40	18	4		
Phe69*	16	2	---	---	---	---	---	---
Thr70	---	---	---	---	---	---	---	---
Thr71	13,6	0,4	900	200	4	2		
Ser72*	28	2	---	---	---	---	---	---
Gly73*	24	2	---	---	---	---	---	---
Asn74	16,9	0,3	200	100	8	7		
Leu75*	16,5	0,1	---	---	---	---	---	---
Lys76*	15,8	0,2	---	---	---	---	---	---
Arg77	15,1	0,2	240	40	13	3		
His78	16,3	0,2	200	70	5	2		
Leu79	18,5	0,3	300	100	30	20		
Arg80	16,7	0,2	440	40	18	3		
Ile81	16,2	0,7	300	200	50	50		
His82*	16	1	---	---	---	---	---	---
Ser83*	12	0,8	---	---	---	---	---	---
Gly84	12,4	0,4	450	90	11	5		
Glu85	9,5	0,3	240	70	15	6		
Lys86	10	1	500	100	26	9		
Pro87	---	---	---	---	---	---	---	---
Tyr88	16,4	0,5	300	100	40	20		
Val89	12,3	0,4	380	80	17	5		
Cys90	16,2	0,3	300	60	25	4		
Ile91	16,1	0,5	440	80	22	5		
His92	15,9	0,6	400	100	11	5		
Cys93	19,7	0,9	300	200	30	20		

Gln94	14,1	0,4	210	90	14	9
Arg95	16	0,2	250	50	19	5
Gln96	14,2	0,1	350	40	8	1
Phe97	14,1	0,5	370	80	23	7
Ala98	17,1	0,5	300	100	13	7
Asp99	12	0,2	350	50	8	2
Pro100	---	---	---	---	---	---
Gly101	14,4	0,2	500	100	5	2
Ala102*	14,1	0,2	---	---	---	---
Leu103	14,1	0,1	180	60	5	2
Gln104*	15,2	0,1	---	---	---	---
Arg105	14,5	0,2	280	50	7	2
His106	14,7	0,1	140	60	6	4
Val107	21	0,6	740	60	38	5
Arg108*	15,3	0,7	---	---	---	---
Ile109*	18	2	---	---	---	---
His110	16,4	0,3	520	60	16	3
Thr111	---	---	---	---	---	---
Gly112	9,4	0,6	300	200	40	40

\* Ces résidus ne présentent pas de relaxation de dispersion

RE : Autorisation

13-08-24 13:24

**RE : Autorisation****Montagne, Martin****Envoyé : 23 août 2013 19:25****À : Bernard, David**

ok. une belle phrase de 3 lignes pleins de virgules qui rendrait Cynthia jalouse!! J'espère que c'est ok! A+!

Par la présente, en prévision du dépôt final de son mémoire en Pharmacologie, j'autorise David Bernard, étudiant dans le laboratoire du Pr. Pierre Lavigne Ph.D., a utiliser la figure "activation de la transcription par c-Myc/Max" située à la page 15 de ma thèse de Doctorat en Pharmacologie, déposée en juin 2008 et intitulée: "Élucidation des déterminants structuraux impliqués dans la reconnaissance moléculaire entre MXD1 et MAX et la liaison à l'ADN".

Martin Montagne Ph.D.  
Associé de Recherche  
Laboratoire du Pr. Pierre Lavigne  
et du Pr. Jean-Guy Lehoux  
Institut de Pharmacologie,  
Département de Biochimie  
Université de Sherbrooke  
Faculté de Médecine et  
des Sciences de la Santé (CHUS)  
819-820-6868 poste 15463

---

**De : Bernard, David****Envoyé : 22 août 2013 11:29****À : Montagne, Martin****Objet : Autorisation**

Salut Martin!

J'aurais besoin de ton autorisation écrite afin d'utiliser la figure "activation de la transcription par c-Myc/Max" située à la page 15 de ta thèse "Élucidation des déterminants structuraux impliqués dans la reconnaissance moléculaire entre MXD1 et MAX et la liaison à l'ADN" afin de satisfaire aux exigences de l'université. Ta thèse est citée comme il se doit dans le titre de la figure.

David

Re: Authorization

13-08-22 12:04

**Re: Authorization**

John Cavanagh [jcavana@ncsu.edu]

Envoyé : 22 août 2013 12:03

À : Bernard, David

David:

I am certainly fine with that.

Good luck with your thesis

Best

John

On Aug 22, 2013, at 12:00 PM, Bernard, David wrote:

Greetings Dr. Cavanagh,

I am currently writing my thesis, and in my university there is a rule that I must acquire the written authorization of the author of every illustration I reproduce in the aforementioned thesis. I intended to use the figure depicting the HSQC sequence from p.412 of your book "Protein NMR Spectroscopy: Principles and Practice", published by Academic Press in 1996. The book is properly cited in the figure legend. I hope this will be satisfying enough for you to grant me your authorization.

Yours truly,

David Bernard, B.Sc.  
Pierre Lavigne laboratory  
Department of pharmacology  
Université de Sherbrooke

---

John Cavanagh  
William Neal Reynolds Distinguished Professor  
Department of Molecular & Structural Biochemistry  
and  
Assistant Vice Chancellor for Research Development  
North Carolina State University  
Raleigh, NC 27695

Phone: 919-513-4349  
FAX: 919-515-2047  
<http://cavanagh-lab.bch.ncsu.edu>

---