



**HAL**  
open science

# Etude de complexes télomériques par Résonance Magnétique Nucléaire

Yann Bilbille

► **To cite this version:**

Yann Bilbille. Etude de complexes télomériques par Résonance Magnétique Nucléaire. domain\_other.  
Université d'Orléans, 2007. Français. NNT: . tel-00385677

**HAL Id: tel-00385677**

**<https://theses.hal.science/tel-00385677>**

Submitted on 19 May 2009

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

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



**THESE**  
**PRESENTEE**  
**A L'UNIVERSITE D'ORLEANS**  
**POUR OBTENIR LE GRADE DE**  
**DOCTEUR DE L'UNIVERSITE D'ORLEANS**

Discipline : Biophysique Moléculaire

Par

**Yann BILBILLE**

Etude de complexes télomériques par

Résonance Magnétique Nucléaire

MEMBRES DU JURY :

|                            |                    |
|----------------------------|--------------------|
| Dr. Françoise KEITA-PAQUET | Directeur de Thèse |
| Dr. Brigitte HARTMANN      | Rapporteur         |
| Dr. Olivier MAUFFRET       | Rapporteur         |
| Pr. Françoise VOVELLE      | Présidente         |
| Dr. Yves BOULARD           | Examineur          |
| Dr. Gérard LANCELOT        | Examineur          |



## Remerciements

*Ce travail de thèse a été réalisé dans l'équipe « Structure et dynamique des acides nucléiques et de leurs complexes par RMN » au sein du Centre de Biophysique Moléculaire, unité propre de recherche appartenant au CNRS d'Orléans. Je tiens à remercier le Docteur Jean-Claude BELOEIL de m'avoir accueilli au sein de ce laboratoire.*

*Je remercie le docteur Gérard LANCELOT de m'avoir accueilli au sein de son équipe, ainsi que ma directrice de thèse, Françoise KEITA-PAQUET, pour son soutien et pour sa gentillesse.*

*Mes remerciements les plus sincères aux Docteurs Brigitte HARTMANN et Olivier MAUFFRET pour avoir accepté d'évaluer ce manuscrit. Vos remarques m'ont permis d'améliorer ce manuscrit.*

*Je remercie le professeur Françoise VOVELLE ainsi que le docteur Yves BOULARD d'avoir accepté la tâche ingrate d'examineur.*

*Je remercie le professeur Jean Louis LEROY, de l'ICSN de Gif sur Yvette, de m'avoir permis de réaliser la synthèse d'oligonucléotides au sein de son équipe.*

*Mes pensées les plus sincères vont à toute l'équipe « Oligonucléotides modifiés : chimie et applications » du CBM dirigée par Ulysse ASSELINE pour sa précieuse aide lors de la synthèse et la purification des oligonucléotides.*

*Je remercie l'équipe de « RMN et modélisation moléculaire de protéines et de peptides » et plus particulièrement Hervé, Laurence, Christian et Céline.*

*Enfin n'oublions pas toutes les personnes que j'ai pu rencontrer et qui ont participé au bon déroulement de cette thèse: Fabien, Oumarou, Olivier, Yann-Vaï, Brice-loïc, Hervé, Pierre, Julien, Sonia, Gilles et tous les autres... .*



|  |           |
|--|-----------|
| <b>INTRODUCTION</b>  | <b>11</b> |
| <b>LES TELOMERES</b>   | <b>13</b> |
| CHAPITRE 1: LES TELOMERES  | 15        |
| 1 <i>Structure des télomères</i>   | 15        |
| 1.1 Complexité de l'ADN télomérique  | 15        |
| 1.2 Le modèle de la boucle-T/boucle-D (T-loop/D-loop)                                | 17        |
| 1.3 Les protéines télomériques   | 19        |
| 2 <i>Les domaines de fixation à l'ADN de type Myb</i>                                | 24        |
| 3 <i>Fonction des télomères</i>  | 26        |
| 3.1 La protection du génome  | 27        |
| 3.2 Le rôle d'horloge biologique   | 27        |
| 3.3 Rôle dans le positionnement des chromosomes pendant la méiose                    | 28        |
| 4 <i>Une enzyme particulière: la télomérase</i>                                      | 28        |
| <b>MATERIELS ET METHODES</b>   | <b>31</b> |
| CHAPITRE 2 : ETUDE DES MACROMOLECULES PAR RESONANCE MAGNETIQUE NUCLEAIRE             | 33        |
| 1 <i>Principes généraux de la résolution de structures de macromolécules par RMN</i> | 34        |
| 2 <i>Les acides nucléiques</i>   | 35        |
| 2.1 Présentation   | 35        |
| 2.2 La synthèse chimique des acides nucléiques                                       | 36        |
| 2.3 Etude RMN des oligonucléotides   | 38        |
| 2.4 L'effet Overhauser nucléaire   | 39        |
| 2.5 Les liaisons hydrogène   | 40        |
| 3 <i>Les protéines</i>   | 43        |
| 3.1 Présentation   | 43        |
| 3.2 Etude RMN des protéines  | 45        |
| 4 <i>Etude de complexes ADN-Protéines</i>  | 51        |
| 5 <i>Les spectromètres RMN utilisés</i>  | 53        |
| CHAPITRE 3 : LES COUPLAGES DIPOLAIRES RESIDUELS                                      | 54        |
| 1 <i>L'Interaction Dipôle-Dipôle Directe</i>   | 54        |
| 2 <i>Couplages Dipolaires Résiduels en milieux orientés</i>                          | 55        |
| 3 <i>Les milieux anisotropes utilisés</i>  | 59        |
| CHAPITRE 4 : ETUDE DE LA DYNAMIQUE DES PROTEINES                                     | 61        |
| 1 <i>Introduction</i>  | 61        |

|  |  |           |
|--|--|-----------|
| 2  | <i>La fonction de densité spectrale</i>  | 62        |
| 3  | <i>La relaxation nucléaire</i>   | 63        |
| 3.1  | La relaxation longitudinale  | 64        |
| 3.2  | La relaxation transversale   | 64        |
| 3.3  | La relaxation croisée ; l'effet NOE $\eta$   | 64        |
| 4  | <i>Paramètres expérimentaux mesurés</i>  | 65        |
| 5  | <i>Le phénomène d'échange</i>  | 66        |
| 6  | <i>Mesure de la relaxation longitudinale <math>R_1</math> et transversale <math>R_2</math></i> | 66        |
| 7  | <i>Mesure de l'effet NOE</i>   | 67        |
| 8  | <i>Interprétation des paramètres de relaxation en terme de mouvement</i>                       | 67        |
| 9  | <i>Détermination du temps de corrélation globale de la protéine <math>\tau_c</math></i>        | 70        |
| <b>CHAPITRE 5 : LA MODELISATION MOLECULAIRE SOUS CONTRAINTES RMN</b> |  | <b>71</b> |
| 1  | <i>L'énergie potentielle</i>   | 71        |
| 2  | <i>La fonction cible</i>   | 73        |
| 3  | <i>Le logiciel ARIA (Ambiguous Restraints for Iterative Assignment)</i>                        | 74        |
| 3.1  | Contraintes de distance ambiguë  | 74        |
| 3.2  | Calibration des pics NOEs  | 75        |
| 3.3  | Attribution des pics NOEs ambigus  | 75        |
| 3.4  | Principe de fonctionnement du logiciel ARIA  | 77        |
| 4  | <i>Le logiciel CYANA</i>   | 78        |
| 4.1  | Principe de l'algorithme CANDID de CYANA   | 78        |
| 5  | <i>Analyse des structures obtenues</i>   | 80        |
| <b>RESULTATS</b>   |  | <b>83</b> |
| <b>CHAPITRE 6 : ETUDE RMN DE L'ADN TELOMERIQUE</b>                   |  | <b>85</b> |
| 1  | <i>Etude RMN de l'ADN télomérique contenant deux répétitions télomériques</i>                  | 85        |
| 1.1  | Présentation de la séquence  | 85        |
| 1.2  | Etude des brins seuls par RMN  | 86        |
| 1.3  | Formation du duplexe d'ADN   | 86        |
| 1.4  | Etude RMN  | 87        |
| 2  | <i>Etude de l'ADN télomérique contenant trois répétitions télomériques</i>                     | 89        |
| 3  | <i>Conclusion</i>  | 91        |
| <b>CHAPITRE 7 : ETUDE DU DOMAINE MYB DE LA PROTEINE TRF2</b>         |  | <b>92</b> |
| 1  | <i>Présentation du domaine Myb</i>   | 92        |

|     |   |     |
|-----|---|-----|
| 2   | <i>Etude RMN de la protéine non-marquée et de la protéine marquée uniformément</i>      |     |
|     | $^{15}\text{N}$ et $^{13}\text{C}$  | 93  |
| 2.1 | Attribution du squelette peptidique   | 94  |
| 2.2 | Attribution des chaînes latérales   | 95  |
| 2.3 | Collection des informations structurales  | 95  |
| 2.4 | Analyse de la structure du domaine Myb de TRF2  | 98  |
| 2.5 | Comparaison avec le domaine Myb de la protéine TRF1                                     | 99  |
| 2.6 | Etude de la dynamique du domaine Myb de TRF2  | 100 |
| 3   | <i>Conclusion</i>   | 106 |
|     | CHAPITRE 8 : REALISATION D'UN PREMIER COMPLEXE TELOMERIQUE                              | 107 |
| 1   | <i>Présentation</i>   | 107 |
| 2   | <i>Réalisation du complexe</i>  | 107 |
| 3   | <i>Etude RMN de l'ADN télomérique complexé</i>  | 108 |
| 4   | <i>Limitation de cette étude</i>  | 111 |
| 5   | <i>Conclusion et discussion</i>   | 111 |
|     | CHAPITRE 9 : REALISATION D'UN SECOND COMPLEXE TELOMERIQUE                               | 113 |
| 1   | <i>Présentation du complexe</i>   | 113 |
| 2   | <i>Etude RMN du duplexe d'ADN</i>   | 114 |
| 2.1 | Préparation du duplexe  | 114 |
| 2.2 | Etude RMN du duplexe  | 114 |
| 3   | <i>Formation du complexe</i>  | 115 |
| 4   | <i>Etude RMN de la protéine complexée</i>   | 116 |
| 4.1 | Attribution de la chaîne peptidique de Myb complexée                                    | 116 |
| 4.2 | Analyse des variations de déplacements chimiques du $^1\text{HN}$ et du $^{15}\text{N}$ | 117 |
| 4.3 | Attribution des chaînes latérales   | 119 |
| 4.4 | Modélisation moléculaire du domaine Myb de TRF2 complexé                                | 119 |
| 4.5 | Structure obtenue   | 120 |
| 4.6 | Analyse de la structure du domaine Myb de TRF2 complexé à l'ADN télomérique.            | 122 |
| 5   | <i>Etude de l'ADN complexé</i>  | 123 |
| 5.1 | Etude des connectivités intermoléculaires   | 124 |
| 6   | <i>Modélisation du complexe</i>   | 125 |
| 7   | <i>Analyse de la structure du complexe</i>  | 127 |

|     |   |            |
|-----|---|------------|
| 8   | <i>Raffinement de la structure du domaine Myb de TRF2 en utilisant les couplages dipolaires résiduels</i> | 129        |
| 8.1 | Synthèse du gel et préparation de l'échantillon   | 129        |
| 8.2 | Etude RMN en milieu orienté   | 131        |
| 8.3 | Détermination du tenseur d'alignement et des angles d'Euler   | 133        |
| 8.4 | Protocole de raffinement de la structure du domaine Myb de TRF2 à l'aide de RDCs                          | 134        |
| 8.5 | Structure du domaine Myb de TRF2 obtenue par raffinement  | 135        |
| 9   | <i>Conclusion et discussion</i>   | 136        |
|     | CHAPITRE 10 : ETUDE DE MODELES DE BOUCLE-D PAR RMN  | 137        |
| 1   | <i>Introduction</i>   | 137        |
| 2   | <i>Présentation des études réalisées</i>  | 139        |
| 3   | <i>Boucle-D : modèle A</i>  | 142        |
| 3.1 | Présentation  | 142        |
| 3.2 | Etude RMN   | 142        |
| 3.3 | Conclusion  | 145        |
| 4   | <i>Boucle-D : modèle B</i>  | 146        |
| 4.1 | Présentation  | 146        |
| 4.2 | Etude RMN et Conclusion   | 146        |
| 5   | <i>Boucle-D : modèle C</i>  | 147        |
| 5.1 | Présentation  | 147        |
| 5.2 | Etude RMN   | 148        |
| 5.3 | Conclusion  | 151        |
| 6   | <i>Boucle-D : modèle D</i>  | 152        |
| 6.1 | Présentation  | 152        |
| 6.2 | Etude RMN   | 153        |
| 6.3 | Conclusion  | 153        |
| 7   | <i>Perspectives : Etude des modèles A et B en milieu orienté</i>  | 154        |
|     | <b>CONCLUSIONS ET PERSPECTIVES</b>  | <b>155</b> |
|     | ANNEXE 1 : SYNTHÈSE CHIMIQUE DES DIFFÉRENTS BRINS D'ADN   | 159        |
|     | ANNEXE 2 : TABLEAU DE DÉPLACEMENTS CHIMIQUES DES DIFFÉRENTES MACROMOLÉCULES<br>ÉTUDIÉES                   | 161        |
|     | ANNEXE 3 : L'ARTICLE  | 175        |

# Abréviations

**1D** : unidimensionnel(le)

**2D** : bidimensionnel(le)

**3D** : tridimensionnel(le)

**A** : Adénine

**ADN** : Acide DésoxyriboNucléique

**ADC** : Analogic Digital Converter

**ADR** : Ambiguous Distance Restraint

**ARN** : Acide RiboNucléique

**ARIA** : Ambiguous distance Restraints for Iterative Assignments

**C** : Cytosine

**RDC** : Residual Dipolar Coupling

**CHARMM** : Chemistry at HARvard Macromolecular Mechanics

**CNS** : Crystallography NMR System

**COSY** : COrrrelation SpectroscopY

**COSY DQF** : COrrrelation SpectroscopY Double Quantum Filter

**CSA** : Chemical Shift Anisotropy

**CYANA** : Combined assignment and dynamics Algorithm for NMR Applications

**D-loop** : Displacement loop

**DO** : Densité Optique

**G** : Guanine

**HOHAHA** : HOmonuclear HArtman HAhn spectroscopy

**HPLC** : High Performance Liquid Chromatography

**HSQC** : Heteronuclear Single Quantum Coherence

**hTER** : human TElomerase RNA

**hTERC** : human TElomerase RNA Component

**INEPT** : Insensitive Nuclei Enhanced Polarisation Transfer

**LB** : Lennox Luria Broth

**Ni – NAT** : Nickel-Nitrilotriacétate

**NMR** : Nuclear Magnetic Resonance

**NOE** : Nuclear Overhauser Effect

**NOESY** : Nuclear Overhauser Effect SpectroscopY

**PCR** : Polymerisation Chain Reaction

**PDB** : Protein Data Bank

**POT1** : Protection Of Telomeres 1  
**PPM** : Partie Par Million  
**PSVS** : Protein Structure Validation Suite  
**RDC** : Residual Dipolar Coupling  
**RMN** : Résonance Magnétique Nucléaire  
**RMSD** : Root Mean Square Deviation  
**RX** : Rayons X  
**SDS-PAGE** : Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis  
**T** : Thymine  
**T<sub>1</sub>** : Temps de relaxation longitudinale  
**T<sub>2</sub>** : Temps de relaxation transversale  
**T-loop** : Telomeric loop  
 **$\tau_c$**  : temps de corrélation  
**TIN2** : TRF1 Interacting Nuclear protein 2  
**t<sub>m</sub>** : température de fusion ou de demi-dénaturation  
**TOCSY** : Total Correlation Spectroscopy  
**TRF1** : Telomeric Repeat binding Factor 1  
**TRF2** : Telomeric Repeat binding Factor 2  
**TRFH** : TRF Homology  
**TROSY** : Transverse Relaxation Optimized Spectroscopy  
**VdW** : Van der Waals

# INTRODUCTION

Les télomères, du grec « telos » (terminal) et « meros » (segment), correspondent aux extrémités des chromosomes. Ce sont des structures hautement conservées au cours de l'évolution. En 1938, Hermann J. Muller montra que les extrémités naturelles des chromosomes nommées « télomères » ne peuvent se lier entre elles pour former des chromosomes réarrangés, contrairement à des extrémités produites par coupure de l'ADN aux rayons X (Muller 1938). En 1941, Barbara McClintock suggéra que les télomères avaient un rôle de protection des chromosomes contre les phénomènes de fusion chromosomique (McClintock 1941).

Les télomères sont des complexes nucléoprotéiques formés d'ADN et de protéines. L'ADN télomérique est constitué de la répétition de séquences non codantes comportant de 5 à 8 paires de bases. Il a la particularité de comporter une partie double brin et une extrémité simple brin riche en guanine (Zakian 1995).

Les télomères sont souvent caractérisés d'horloge mitotique (ou biologique) puisque leur taille est liée à l'espérance de vie de la cellule (Blackburn 1991). Lors de la réplication de l'ADN, la réplication des extrémités est incomplète, aboutissant à une dégradation progressive des répétitions télomériques. Lorsque cette taille devient trop petite, les télomères ne jouent plus leur rôle protecteur, la cellule rentre en sénescence et meurt par apoptose (Hemann, Strong et al. 2001; Karlseder, Smogorzewska et al. 2002).

Il a été montré que dans 85% des cas de cancer la taille des télomères reste constante, l'érosion naturelle des chromosomes étant contrebalancée par l'élongation due à une enzyme, la télomérase (Kim, Piatyszek et al. 1994). Dans beaucoup de types cellulaires, le gène TERT, codant la télomérase, est réprimé. Au contraire, dans les tissus à multiplication cellulaire intense, le gène TERT est exprimé et la longueur des télomères reste constante, ce qui permet à la cellule de devenir immortelle.

Pour expliquer le rôle protecteur des télomères, notamment contre les ligases et les nucléases présentes dans la cellule, plusieurs modèles ont été proposés. Dans le modèle de la boucle-T/boucle-D (T-loop/D-loop), l'ADN télomérique simple brin vient s'insérer à l'intérieur du double brin télomérique, formant ainsi une boucle télomérique (loop). La formation de ces structures particulières fait intervenir de nombreuses protéines dont la protéine TRF2 qui a retenu toute notre attention.

Le but de ce travail de thèse a été de contribuer à la compréhension des mécanismes de fonctionnement des télomères en utilisant la résonance magnétique nucléaire.

La RMN est l'une des deux principales techniques permettant d'obtenir la structure tridimensionnelle d'une macromolécule, l'autre étant la radiocristallographie. Depuis quelques années, les récents progrès instrumentaux (cryosondes, aimants supraconducteurs de plus de 21 T, techniques de marquage isotopique) combinés à des progrès méthodologiques (RMN multidimensionnelle, TROSY ...) rendent les tailles de molécules étudiables par RMN toujours plus grandes. Le prix Nobel de Chimie décerné à Kurt Wüthrich en 2002 est un bel exemple de la réussite de cette technique.

L'objectif premier de ce travail de thèse a consisté en l'étude du domaine de fixation à l'ADN de la protéine télomérique TRF2 (domaine Myb) ainsi que de son complexe avec l'ADN télomérique. Nous verrons que ce sujet initial a évolué, pour diverses raisons, vers d'autres objectifs. Nous avons alors réalisé l'étude d'un second complexe impliquant toujours le domaine Myb de la protéine TRF2 et un ADN comportant une partie double brin et une partie simple brin. Enfin, nous avons essayé de réaliser des assemblages particuliers d'ADN pour essayer de mimer les boucles télomériques, et ainsi essayer de mieux comprendre la formation de ces structures particulières qui peuvent être observées à l'extrémité des télomères.

# Les télomères



# CHAPITRE 1: Les télomères

## 1 Structure des télomères

### 1.1 Complexité de l'ADN télomérique

La taille de l'ADN télomérique varie selon l'âge, la nature cellulaire ou l'espèce (tableau 1). Chez l'homme, la taille des télomères varie de 2000 à 10 000 paires de bases (Cech 2000; Blackburn 2001; Neidle and Parkinson 2003). L'ADN télomérique possède une partie double brin majoritaire, et une partie simple brin d'une longueur variable de 50 à 200 nucléotides riches en guanine (Makarov, Hirose et al. 1997; Wright, Tesmer et al. 1997; Cimino-Reale, Pascale et al. 2001; Stewart, Ben-Porath et al. 2003). La longueur des télomères varie selon l'état de différenciation de la cellule (cellule germinale ou somatique) et son histoire répllicative (Moyzis, Buckingham et al. 1988).

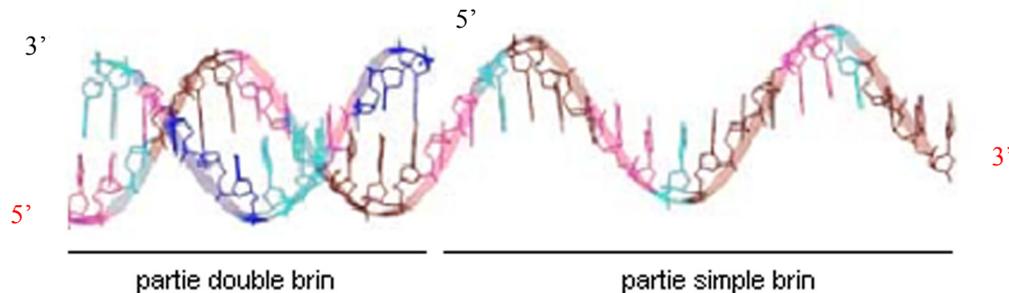


Figure 1: Structure de l'ADN télomérique

| Organisme                     | Motif télomérique              |
|-------------------------------|--------------------------------|
| Homo sapiens                  | T <sub>2</sub> AG <sub>3</sub> |
| Kluyveromyces lactis (levure) | ACGGATTTGATTAGGTATGTGGTGT      |
| Arabidopsis (plante)          | T <sub>3</sub> AG <sub>3</sub> |

Tableau 1 : Séquences télomériques terminales

L'ADN télomérique humain est composé de la répétition de 6 nucléotides formant la séquence (TTAGGG) $_n$ . La composition en nucléotides de l'ADN télomérique le rend capable d'adopter différentes conformations stables (Phan and Mergny 2002). De nombreuses études ont montré que le brin riche en guanine était capable de former seul des structures particulières appelées quadruplexes (Li, Miyoshi et al. 2003). Le brin riche en cytosine peut former des motifs-i à des pH acides (<5) (Gehring, Leroy et al. 1993).

### 1.1.1 Les quadruplexes

Les quadruplexes sont encore très étudiés aujourd'hui et ont fait l'objet de nombreuses publications (Wang and Patel 1993; Patel and Hosur 1999; Luu, Phan et al. 2006); (Gavathiotis and Searle 2003). Dans les conformations de type quadruplexe, quatre guanines peuvent s'apparier entre elles par des liaisons hydrogène de type « Hoogsteen » (Sundquist and Klug 1989) pour former des tétrades de guanines aussi appelées G-quartet. Ces structures sont stabilisées par des ions positifs, notamment les ions potassium ou sodium, qui se placent au cœur de la tétrade (Parkinson, Lee et al. 2002). L'empilement de ces tétrades de guanines forme les quadruplexes (Rhodes and Giraldo 1995). De nombreux types de quadruplexes ont pu être observés : parallèle, antiparallèle...(Simonsson 2001). In vitro, il a été montré que la séquence répétée simple brin T<sub>2</sub>AG<sub>3</sub> était capable de former un quadruplexe de guanines, ce qui pourrait empêcher l'action de la télomérase ou d'autres nucléases (Zahler, Williamson et al. 1991). Même si leur existence in vivo n'a pas pu encore être mise en évidence, certains auteurs ont envisagé une implication de ces structures durant la mitose et la méiose (Muniyappa and Kironmai 1998). Quoiqu'il en soit, les conditions physiologiques du milieu intracellulaire sont parfaitement compatibles avec la formation et la stabilité des structures quadruplexes.

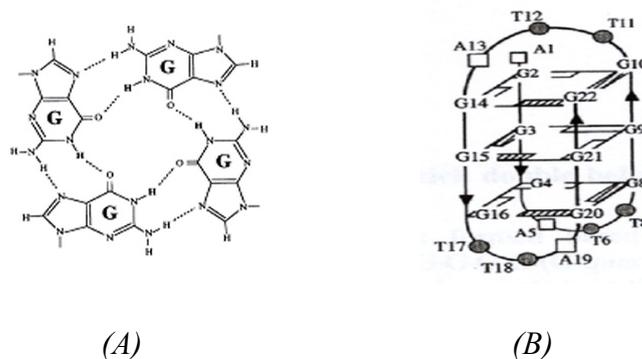


Figure 2 : (A) structure d'un G-quartet; (B) empilement des G-quartet formant le quadruplexe.

Sur les spectres RMN, la présence de conformations quadruplexes est observable. Les protons iminos des guanines impliquées dans les G-quartet ont des fréquences de résonances situées vers 11 ppm.

### 1.1.2 Le motif-i

Le motif-i (Ahmed, Kintanar et al. 1994; Leroy, Gueron et al. 1994; Gueron and Leroy 2000; Phan, Gueron et al. 2000) est observé pour les brins riches en cytosine pour des pH inférieurs à 5. On peut alors observer la protonation de cytosines et l'appariement de deux cytosines entre elles. Ces appariements C.CH<sup>+</sup> sont très résistants (trois liaisons hydrogène) et sont à l'origine de la formation intramoléculaire du motif-i. Sur les spectres RMN, les hydrogènes des cytosines protonées résonnent à plus de 15 ppm et sont donc facilement repérables.

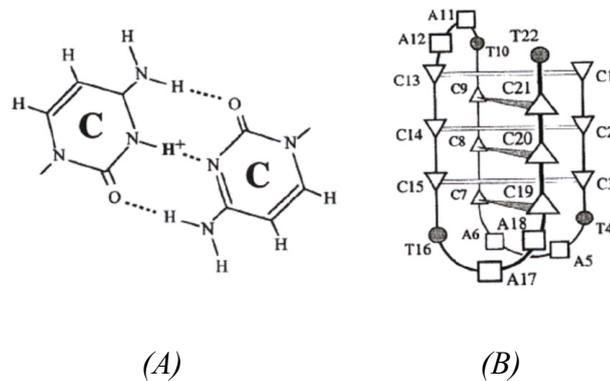


Figure 3 : (A) appariement C-CH<sup>+</sup> ;(B) le motif-i

## 1.2 Le modèle de la boucle-T/boucle-D (T-loop/D-loop)

Un des rôles majeurs des télomères est de préserver l'intégrité des chromosomes contre différentes enzymes présentes dans notre corps comme les nucléases et les ligases. Un des modèles proposés par l'équipe de Titia de Lange pour expliquer cette protection est le modèle de la boucle-T (boucle télomérique) boucle-D (boucle de déplacement) (Greider 1999; Griffith, Comeau et al. 1999; Shay 1999). Ce modèle est fondé sur l'observation des télomères par microscopie électronique (figure 4).



Figure 4: Photographie obtenue par microscopie électronique montrant les télomères terminés par une boucle télomérique.

La boucle-T est formée par le repliement de l'ADN télomérique double brin et par l'insertion de l'extrémité 3' de l'ADN télomérique simple brin au niveau des répétitions télomériques. La boucle-D correspond à la portion du brin G qui, suite à l'insertion du simple brin, n'est plus appariée. La formation de la boucle-T/boucle-D fait appel à de nombreuses protéines (figure ci-dessous).

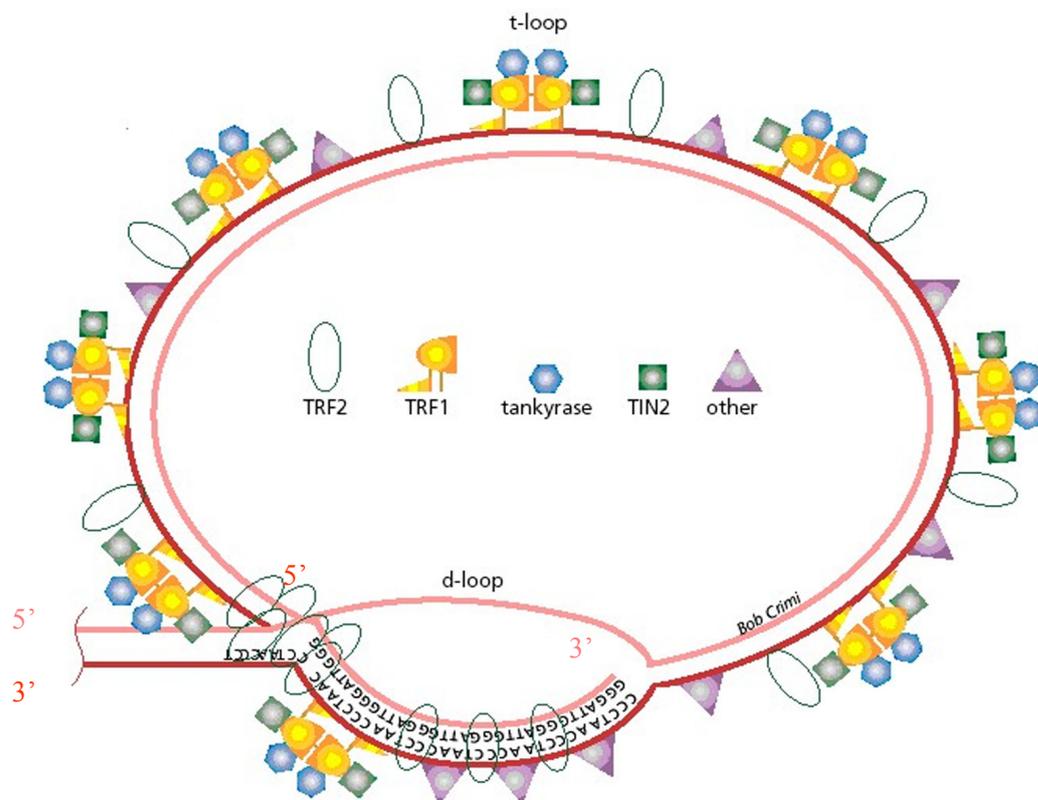


Figure 5: Présentation du modèle de la boucle-T/boucle-D et des protéines favorisant sa formation

La protéine TRF1 permettrait, en se fixant sous forme d'homodimère à l'ADN télomérique double brin, la courbure de celui-ci d'un angle de 60° (Bianchi, Smith et al. 1997). La protéine TRF2 serait à l'origine de l'insertion de l'extrémité simple brin dans l'extrémité double brin de l'ADN télomérique. En effet, la protéine TRF2 se retrouve sur le double brin mais on la retrouve aussi dans la zone d'insertion du simple brin dans le double brin d'ADN télomérique (Chong, van Steensel et al. 1995; van Steensel and de Lange 1997; Smogorzewska, van Steensel et al. 2000).

Les boucles-T pourraient permettre de différencier les extrémités des chromosomes des cassures double brin dans l'ADN, et éviter ainsi la détection des télomères par la machinerie de réparation des cassures.

Pour que les télomères soient fonctionnels, il est impératif d'avoir un nombre minimal de répétitions télomériques T<sub>2</sub>AG<sub>3</sub> notamment au niveau du simple brin, et la présence de protéines télomériques est essentielle (TRF1, TRF2, tankyrase, Ku, TIN2, hRap1, Mre11, Pot1, PINX1 et hnRNPA1). Ces protéines télomériques forment des liaisons non covalentes avec les régions simple brin et double brin des télomères, se liant soit directement soit indirectement à l'ADN télomérique (Blasco 2002).

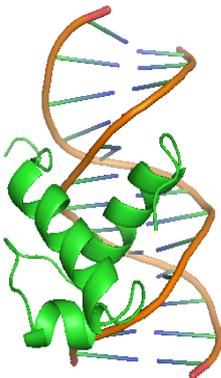
### 1.3 Les protéines télomériques

La chromatine télomérique n'est pas ou peu associée à des nucléosomes (Tommerup, Dousmanis et al. 1994) mais est associée, en contre-partie, à un grand nombre d'autres protéines que l'on ne retrouve pas dans la chromatine normale. Ainsi, de nombreuses protéines interagissant avec l'ADN télomérique ont pu être identifiées, les plus importantes sont détaillées ci-dessous.

#### 1.3.1 TRF1 (Telomeric Repeat binding Factor 1)

La protéine TRF1 est capable de se lier spécifiquement, sous forme d'homodimère, au double brin d'ADN télomérique et permet la courbure de l'ADN télomérique d'un angle de 60° (Bianchi, Smith et al. 1997; Konig, Fairall et al. 1998; Nishikawa, Nagadoi et al. 1998; Bianchi, Stansel et al. 1999). Cette protéine de 439 acides aminés contient une région N-terminale acide riche en aspartate et en glutamate servant de domaine de reconnaissance pour des protéines de régulation, un domaine d'homodimérisation TRFH, une séquence de

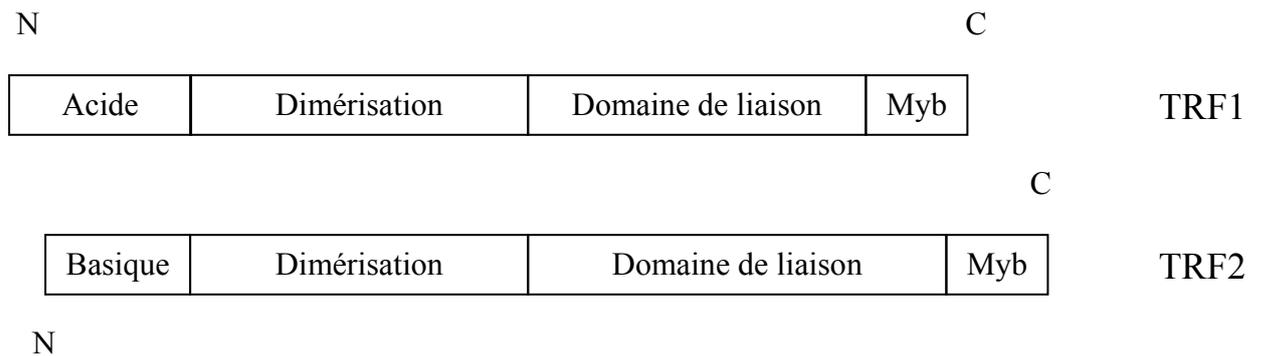
localisation nucléaire liant le domaine d'homodimérisation à un motif de liaison à l'ADN de type Myb situé dans la région C-terminale (figure 7). Des protéines comme la Tankyrase 1 (TRF1-Interacting, ankyrin-related ADP-ribose polymerase) et la Tankyrase 2 se lient à TRF1 et inhibent ses interactions avec l'ADN télomérique par ADP-ribosylation de TRF1 (Smith, Giriat et al. 1998; Smith and de Lange 2000; Cook, Dynek et al. 2002). L'ADN télomérique libéré de TRF1 peut alors être allongé par la télomérase (Ancelin, Brunori et al. 2002; Chang, Dynek et al. 2003). La protéine TIN2 (TRF1-Interacting protein 2) peut au contraire, en modifiant la structure de TRF1, inhiber l'action des tankyrases (Ye and de Lange 2004). TRF1 régule négativement la télomérase (van Steensel and de Lange 1997; Smogorzewska, van Steensel et al. 2000) en permettant la fixation de POT1 (Protection Of Telomeres 1) sur le simple brin télomérique. L'inactivation de TIN2 permet l'activité de la tankyrase 1 sur TRF1, ce qui diminue l'affinité de cette dernière pour l'ADN télomérique double brin et empêche indirectement la fixation de POT1 au simple brin télomérique. Le simple brin devient alors accessible à la télomérase (Ye and de Lange 2004).



*Figure 6: Structure obtenue par RMN correspondant à un complexe entre le domaine myb de TRF1 et un ADN double brin contenant deux répétitions télomériques (Nishikawa, Okamura et al. 2001) (code pdb :liv6) .*

### 1.3.2 TRF2 (Telomeric Repeat binding Factor 2)

TRF2 est une protéine de 460 acides aminés avec une structure proche de TRF1 (Broccoli, Chong et al. 1997). Cette protéine contient un domaine N-terminal, un domaine de dimérisation TRFH, une séquence de localisation nucléaire et un motif de liaison à l'ADN de type Myb (Bilaud, Koering et al. 1996).



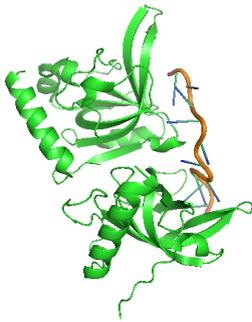
*Figure 7 : Les différents domaines structuraux des protéines TRF1 et TRF2*

Le domaine N-terminal diffère entre les deux protéines. Il est composé de résidus majoritairement acides chez TRF1, alors que chez TRF2, les résidus sont plutôt basiques. Récemment, il a été montré que le domaine N-terminal de TRF2 pourraient reconnaître certains motifs particuliers (Fouché, Cesare et al. 2006). Le domaine de dimérisation de TRF1, qui ne présente que 27% d'identité de séquence avec celui de TRF2, possède néanmoins une structure très proche de ce dernier (Fairall, Chapman et al. 2001). Cependant ces protéines ne possèdent pas la capacité de s'hétérodimériser (Bianchi, Smith et al. 1997; Broccoli, Chong et al. 1997). Le domaine Myb C-terminal est très conservé entre les deux protéines avec 56% d'identité de séquence.

La protéine TRF2 est associée à la protection des extrémités télomériques en maintenant la structure des télomères. Il a été montré que l'absence de TRF2 induit l'entrée des cellules en apoptose (van Steensel, Smogorzewska et al. 1998; Karlseder, Smogorzewska et al. 2002). La protéine TRF2 est localisée sur l'ADN télomérique double brin mais on la retrouve aussi au niveau de la jonction avec l'extrémité simple brin et elle semble jouer un rôle important dans la formation des boucles télomériques (Griffith, Comeau et al. 1999; Wei and Price 2003). Il a été montré que l'expression d'une protéine mutante TRF2 conduit rapidement à la perte de l'extrémité simple brin, à l'apparition de fusions chromosomiques et enfin à la sénescence et/ou l'apoptose de la cellule (Karlseder, Smogorzewska et al. 2002). TRF2 forme un complexe avec la protéine hRAP1 et est associée à de nombreuses protéines impliquées dans la réparation de l'ADN notamment RAD50/MER11/NBS1, Ku86 et ERCC1/XPF (Zhu, Kuster et al. 2000; Zhu, Niedernhofer et al. 2003; O'Connor, Safari et al. 2004). Enfin TRF2 peut interagir avec la protéine TIN2 associée à TRF1 (Kim, Beausejour et al. 2004). De récentes études montrent que TRF2 permet de surenrouler l'ADN télomérique double brin. Ce surenroulement créerait une tension qui serait à même d'ouvrir l'ADN double brin et de permettre l'insertion du simple brin télomérique (Amiard, Doudeau et al. 2007).

### 1.3.3 La protéine POT1 (Protection Of Telomeres 1)

POT1 est une protéine de 634 acides aminés. Son extrémité N-terminale, un domaine "Oligonucleotide Binding fold" (OB fold), reconnaît spécifiquement la séquence simple brin TTAGGGTTAG. Elle joue un rôle dans la protection de l'intégrité des télomères (Baumann and Cech 2001) ainsi que dans la régulation de leurs longueurs (Colgin, Baran et al. 2003). POT1 semble liée indirectement à TRF1 et induit une régulation négative des télomères.

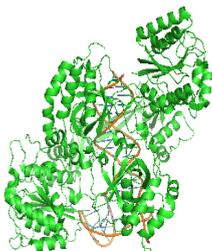


*Figure 8: Structure obtenue par radiocristallographie d'un complexe entre la protéine POT1 sous forme dimérique et un fragment d'ADN télomérique simple brin (Lei, Podell et al. 2004)(code pdb : 1xjv).*

### 1.3.4 Les autres protéines télomériques

=> La protéine Ku

La protéine Ku fait partie du complexe DNA PK (DNA dependent Protein Kinase) en association avec la sous-unité catalytique d'une protéine kinase ADN-dépendante, intervenant dans la réparation des cassures d'ADN double brin par un mécanisme de recombinaison non homologue. Ku est une protéine hétérodimérique composée d'une sous-unité d'environ 70 kDa (Ku70) et d'une sous-unité d'environ 80 kDa (Ku80 ou Ku86). Cette protéine interagit avec TRF1 et TRF2 et a pour fonction de prévenir les phénomènes de fusions chromosomiques (Bianchi and de Lange 1999; Hsu, Gilley et al. 2000; Song, Jung et al. 2000). Des études ont montré que chez les souris déficientes en protéines Ku le nombre de fusions chromosomiques augmente (Hsu, Gilley et al. 2000). Il semble que la protéine Ku puisse jouer un rôle de régulation de la télomérase en interagissant avec sa sous-unité catalytique, permettant ainsi de moduler l'accès de l'enzyme aux télomères (Chai, Ford et al. 2002).



*Figure 9: Structure obtenue par radiocristallographie du complexe entre la protéine Ku et un fragment d'ADN comportant trois brins (Walker, Corpina et al. 2001).*

=> La protéine tankyrase (TRF1-Interacting, ankyrin-related ADP-ribose polymerase)

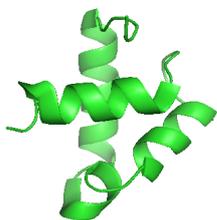
La tankyrase est une protéine proche des ankyrines. Elle présente à son extrémité C-terminale une activité poly-(ADP)-polymérase. Elle interagit avec TRF1 et détecte d'éventuelles interruptions dans la double hélice d'ADN grâce à sa structure en doigt de zinc. Elle agit comme un régulateur positif de la longueur des télomères, en diminuant la liaison de TRF1 sur l'ADN télomérique et en permettant ainsi l'accès de la télomérase au complexe télomérique (Smith and de Lange 2000; Azzalin and Lingner 2004).

=> La protéine TIN2 (TRF1-Interacting Nuclear protein 2)

TIN2 est une protéine de 354 acides aminés qui s'oppose à l'élongation des télomères en interagissant avec TRF1 (Kim, Kaminker et al. 1999). Pour cela, deux mécanismes sont possibles. Soit TIN2 recrute des protéines inhibant la télomérase, soit elle modifie la conformation des télomères empêchant ainsi la télomérase d'accéder à son substrat (Kim, Han et al. 2003). Cette protéine peut aussi interagir avec TRF2.

=> La protéine hRap1

La protéine hRap1 s'associe à l'ADN par l'intermédiaire de la protéine TRF2 (Li, Oestreich et al. 2000). Chez la levure, la taille des télomères est régulée par les molécules scRap1 qui lui sont associées. Comme scRap1, hRap1 joue un rôle dans la régulation de la taille des télomères, une surexpression de hRap1 entraînant une extension de la longueur des télomères (Li, Oestreich et al. 2000).



*Figure 10 : Domaine Myb de la protéine hRap1 (Hanaoka, Nagadoi et al. 2001) (code pdb : 1fex).*

=> Le complexe Mre11

Ce complexe est composé des protéines Mre11 (Meiotic recombinaison 11), Rad50 et NBS1 (Nijmegen Breakage Syndrome). Il est impliqué dans la réparation des cassures d'ADN double brin par un mécanisme de recombinaison homologue. Ce complexe interagit avec TRF1 (Wu, Lee et al. 2000) et TRF2 (Zhu, Kuster et al. 2000).

=> La protéine hnRNPA1 (heterogenous nuclear RibNucleoProtein A1)

La protéine hnRNPA1 appartient à la famille des ribonucléoprotéines nucléaires et est impliquée dans les processus d'épissage alternatif et de maturation des ARN messagers. hnRNPA1 se lie au simple brin de l'ADN télomérique et régule positivement la longueur des télomères en interagissant avec la télomérase.

=> La protéine PINX1 (PIN2/TRF1-Interacting Protein X1)

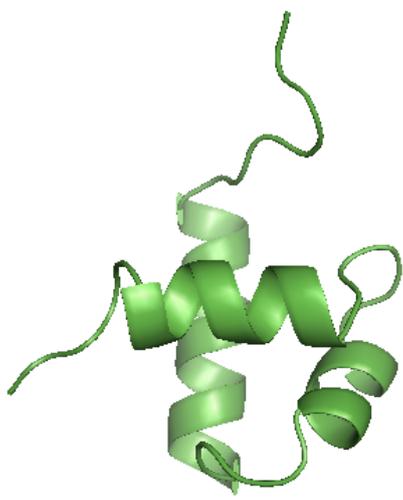
PINX1 est une protéine de 328 acides aminés se liant à TRF1 et à la télomérase. Cette protéine a un rôle d'inhibition de la télomérase (Zhou and Lu 2001).

=> La protéine PIP1 (POT1-Interacting Protein 1)

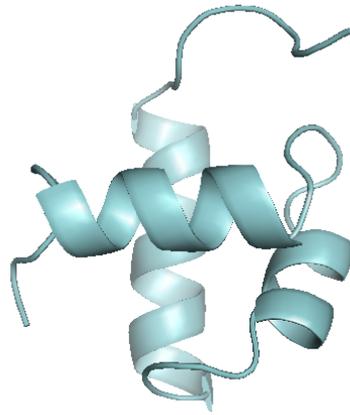
La protéine PIP1 (encore dénommé PTP1 ou TINT1) se lie à POT1 ainsi qu'à TIN2 complexé à TRF1. La réduction de l'activité de PIP1 conduit à l'élongation des télomères (Ye, Hockemeyer et al. 2004).

## 2 Les domaines de fixation à l'ADN de type Myb

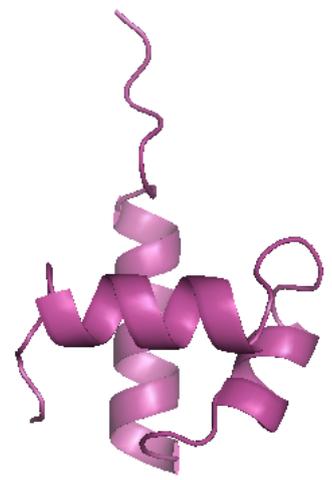
Les protéines Myb sont codées par des gènes « myb » qui peuvent appartenir à différentes classes : a, b ou c. Ces gènes sont homologues du gène ancestral V-myb, oncogène que l'on retrouve dans un retrovirus provoquant une rapide et fatale leucémie chez le poulet, l'AMV (Avian Myeblastosis Virus). Myb est donc un acronyme de MyeBlastosis. Les protéines Myb sont définies par une séquence d'environ 50 acides aminés et présentent généralement un motif caractéristique hélice-tour-hélice. On retrouve le plus souvent dans ces protéines 3 tryptophanes espacés d'une vingtaine de résidus, permettant entre autre la formation d'un cœur hydrophobe. Sur le site de la « Protein Data Bank » ([www.pdb.org](http://www.pdb.org)) peuvent être trouvés un ensemble de structures de ces protéines seules ou liées à l'ADN. Par exemple, le domaine de fixation à l'ADN de la protéine c-Myb est constitué de trois différents domaines (R1, R2 et R3) composé chacun d'un motif hélice-tour-hélice (figure ci-dessous). La fixation de ces protéines à l'ADN se veut relativement spécifique.



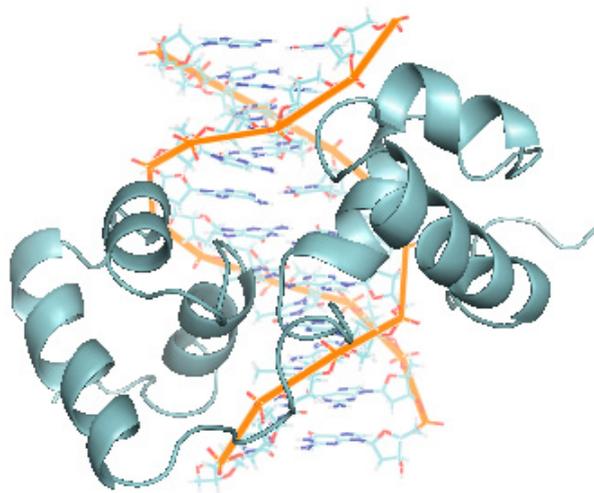
R1



R2



R3



Complexe R2/R3-ADN

*Figure 11: Structures tridimensionnelles des domaines de fixation R1 (code PDB :1mbg), R2 (code PDB :1mbj) et R3 (code PDB :1mbe) de la protéine c-Myb et structure tridimensionnelle du complexe entre les domaines R2 et R3 et un fragment d'ADN (code PDB : 1msf) (Ogata, Morikawa et al. 1995).*

Concernant les télomères, les protéines TRF1 et TRF2 possèdent à leur extrémité C-terminal des domaines de fixation de type Myb très conservés entre les deux protéines avec environ 56% d'identité de séquence. Le domaine Myb de TRF1 a été très étudié et ces études ont montré que ce domaine adopte un motif structural de type hélice-tour-hélice comparable au domaine d'interaction avec l'ADN du facteur c-Myb (d'où le nom de domaine Myb). Grâce à son motif hélice-tour-hélice, le domaine C-terminal de TRF1 reconnaît spécifiquement le site TAGGGTTAG et a, par conséquent, besoin de deux répétitions télomériques pour se fixer à l'ADN. En revanche, au début de ce travail de thèse, le domaine Myb de TRF2 était

relativement peu étudié, c'est pourquoi notre équipe a entrepris des études structurales par RMN de ce domaine et de ces interactions avec l'ADN télomérique.

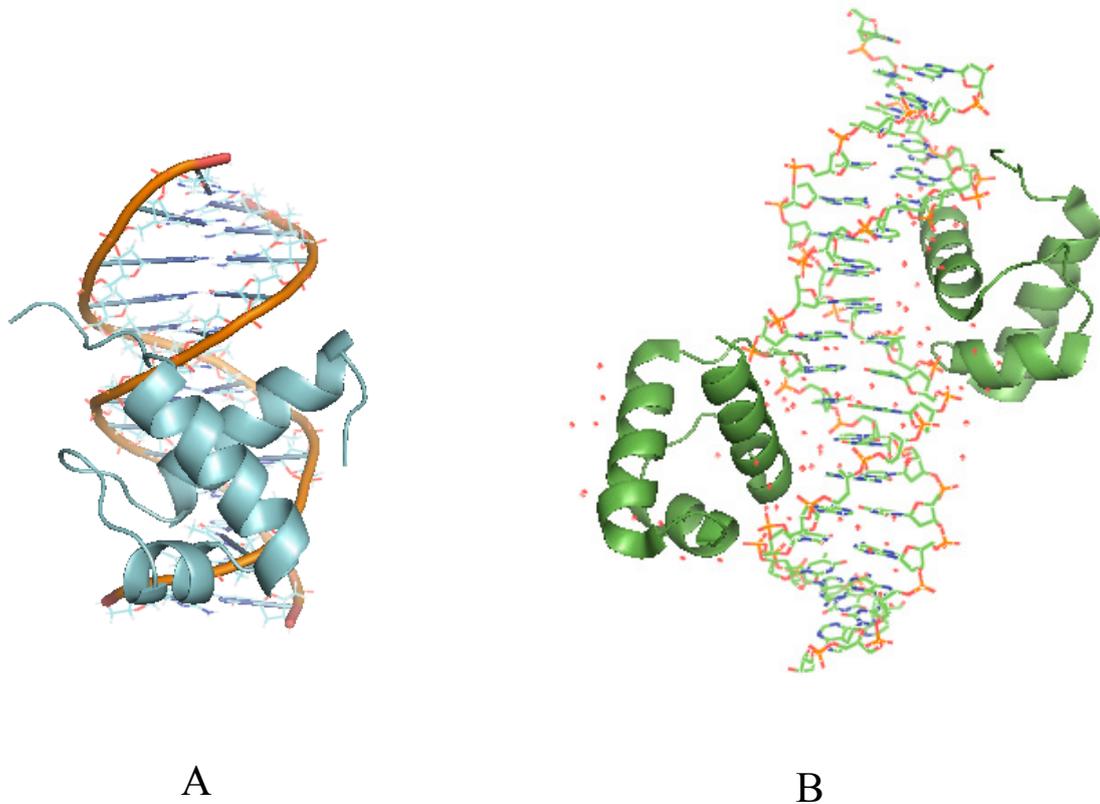


Figure 12 : (A) Structure tridimensionnelle du complexe entre le domaine Myb de la protéine TRF1 et l'ADN télomérique obtenue par RMN (Nishikawa, Okamura et al. 2001) (code pdb : 1iv6) (B) Structure tridimensionnelle du complexe entre le domaine Myb de TRF1 et l'ADN télomérique obtenue par radiocristallographie (Court, Chapman et al. 2005) (code pdb : 1w0t).

### 3 Fonction des télomères

L'ADN télomérique ne correspond pas à un gène. Il s'agit de la répétition d'une même séquence nucléotidique courte, mais répétée un grand nombre de fois. La présence des télomères empêche les gènes de l'organisme de disparaître peu à peu sous l'effet de réplifications incomplètes successives. Les télomères sont impliqués dans diverses fonctions cellulaires qui seront détaillées par la suite :

- Ils préservent l'intégrité des chromosomes contre différentes enzymes présentes dans notre corps.
- Ils jouent le rôle d'horloge mitotique.
- Ils aident au placement des chromosomes pendant la méiose.
- Ils sont impliqués dans la régulation transcriptionnelle de certains gènes.

### 3.1 La protection du génome

Les télomères ainsi que certaines protéines qui leur sont associées empêchent les extrémités des chromosomes d'activer le système d'alarme cellulaire. En l'absence de télomères, les extrémités d'une molécule d'ADN seraient reconnues comme des cassures de brin double, ce qui pourrait déclencher les mécanismes de conversion-amplification du stimulus conduisant à l'arrêt du cycle cellulaire ou à la mort de la cellule (apoptose).

De plus, les télomères protègent le génome des fusions (van Steensel, Smogorzewska et al. 1998) et autres recombinaisons chromosomiques (Slijepcevic 1998) mais aussi de l'action d'exonucléases qui dégraderaient rapidement les extrémités chromosomiques.

### 3.2 Le rôle d'horloge biologique

Du fait de la réplication incomplète de l'ADN, la taille des télomères diminue à chaque nouveau cycle cellulaire (Bryan and Cech 1999; Huffman, Levene et al. 2000). Lorsque les télomères ont atteint une taille trop courte pour stabiliser correctement le génome, la cellule rentre en sénescence répllicative, on parle alors de limite de Hayflick (Hayflick 1997). Une cellule sénescence ne peut plus entrer en phase S et reste bloquée en phase G1. La cellule montre des caractéristiques morphologiques particulières comme un cytoplasme plus grand et granulaire, un niveau d'enzyme élevé dans la matrice cellulaire ainsi que de nombreux changements de métabolisme, notamment une activité accrue de l'activité  $\beta$ -galactosidase. De plus, ce système a aussi un rôle dans la prévention de cellules cancéreuses en éliminant les cellules qui se sont répliquées un trop grand nombre de fois et qui sont donc plus susceptibles d'être mutées.

### 3.3 Rôle dans le positionnement des chromosomes pendant la méiose

Les télomères ont un rôle fondamental dans le bon déroulement de la méiose, notamment dans la prophase de la première division méiotique. Au cours de cette phase, les chromosomes homologues se rapprochent et s'apparient sur toute leur longueur permettant ainsi les échanges d'informations génétiques par recombinaisons. Les télomères vont alors interagir avec des protéines de l'enveloppe nucléaire et se regrouper à un pôle du noyau selon une organisation dite en bouquet qui serait impliquée dans la reconnaissance des chromosomes homologues (Muniyappa and Kironmai 1998).

## 4 Une enzyme particulière: la télomérase

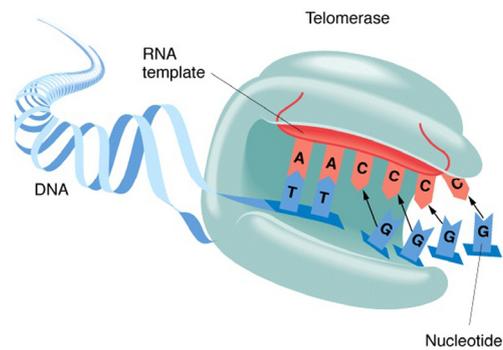
La télomérase, enzyme ADN-polymérase ARN-dépendante, a été initialement caractérisée en 1985 chez un organisme protozoaire cilié *Tétrahymena* (Greider and Blackburn 1985; Lingner, Hughes et al. 1997). La télomérase est une ribonucléoprotéine formée par l'association d'une sous-unité ribonucléique ARN, d'une sous-unité catalytique et de plusieurs protéines associées à ces différentes sous-unités (Morin 1989). Elle présente une activité de transcriptase inverse en synthétisant des répétitions télomériques à l'extrémité 3' des chromosomes grâce à son ARN matrice interne (Collins 1996).

L'ARN de la télomérase est une molécule de 451 nucléotides comportant une séquence de 11 nucléotides complémentaires de la séquence de l'ADN télomérique (5'-CUAACCCUAAC-3'), servant de molécule matrice lors de la synthèse des répétitions télomériques T<sub>2</sub>AG<sub>3</sub> (Feng, Funk et al. 1995). L'ARN de la télomérase est appelé htr (human telomerase RNA), hTER ou hTERC (human telomerase RNA component). L'élongation de l'ADN télomérique est réalisée par un mécanisme de « processivité » représenté par l'addition successive et répétitive des séquences télomériques (Greider 1991).

La sous-unité hTERT (human telomerase reverse transcriptase) ou sous-unité catalytique (Kilian, Bowtell et al. 1997; Meyerson, Counter et al. 1997) est une transcriptase inverse car elle synthétise l'ADN à partir d'une matrice ARN. Cette sous-unité est un polypeptide de 1132 acides aminés (~120 kDa).

La télomérase est présente sous la forme d'un complexe tétramérique composé de 2 sous-unités ARN et de 2 sous-unités catalytiques (Beattie, Zhou et al. 2001).

La télomérase est présente, chez l'homme, pendant le développement embryonnaire mais disparaît après la période néonatale. Chez l'homme adulte, elle est absente de la plupart des cellules somatiques mais présente dans les cellules de lignée germinale, dans certaines cellules souches et dans une large majorité des cellules cancéreuses. Cette enzyme est suffisante pour rendre les cellules immortelles (Bodnar, Ouellette et al. 1998; Halvorsen, Leibowitz et al. 1999). Dans de telles cellules, l'érosion naturelle des télomères est contrebalancée par l'élongation due à la télomérase, la taille de l'ADN télomérique demeure constante et la cellule devient immortelle (Larson, Spangler et al. 1987; Greider 1996).



*Figure 13 : Représentation de la télomérase allongeant l'ADN télomérique*



## Matériels et méthodes



## Chapitre 2 : Etude des macromolécules par Résonance Magnétique Nucléaire

Il existe deux techniques principales permettant d'obtenir la structure des macromolécules biologiques: la Radiocristallographie et la Résonance Magnétique Nucléaire (RMN).

Ces deux techniques sont complémentaires. La radiocristallographie consiste à étudier une molécule sous sa forme cristalline et à obtenir, à partir des diagrammes de diffraction, la position des différents atomes lourds qui la constituent. Elle permet, contrairement à la RMN, de déterminer la structure de grosses molécules. Dans le cas d'études de macromolécules biologiques, elle a permis de résoudre la structure du ribosome. Néanmoins, cette technique présente un certain nombre de limites. Par exemple, certaines régions possèdent une forte flexibilité conformationnelle rendant ainsi leur étude par cristallographie très difficile voir impossible.

La RMN apparaît comme une technique particulièrement adaptée à l'étude de telles régions. La RMN s'applique aux molécules en solution, dans des conditions plus proches des milieux biologiques que la radiocristallographie. Elle permet également d'étudier la dynamique interne des molécules et constitue un excellent complément aux études cristallographiques. Cependant, la résonance magnétique nucléaire est une technique relativement peu sensible nécessitant, pour l'étude structurale, une grande quantité de produit.

# 1 Principes généraux de la résolution de structures de macromolécules par RMN

Les grandes étapes nécessaires pour réaliser l'étude structurale d'une macromolécule par RMN sont détaillées dans la figure suivante :

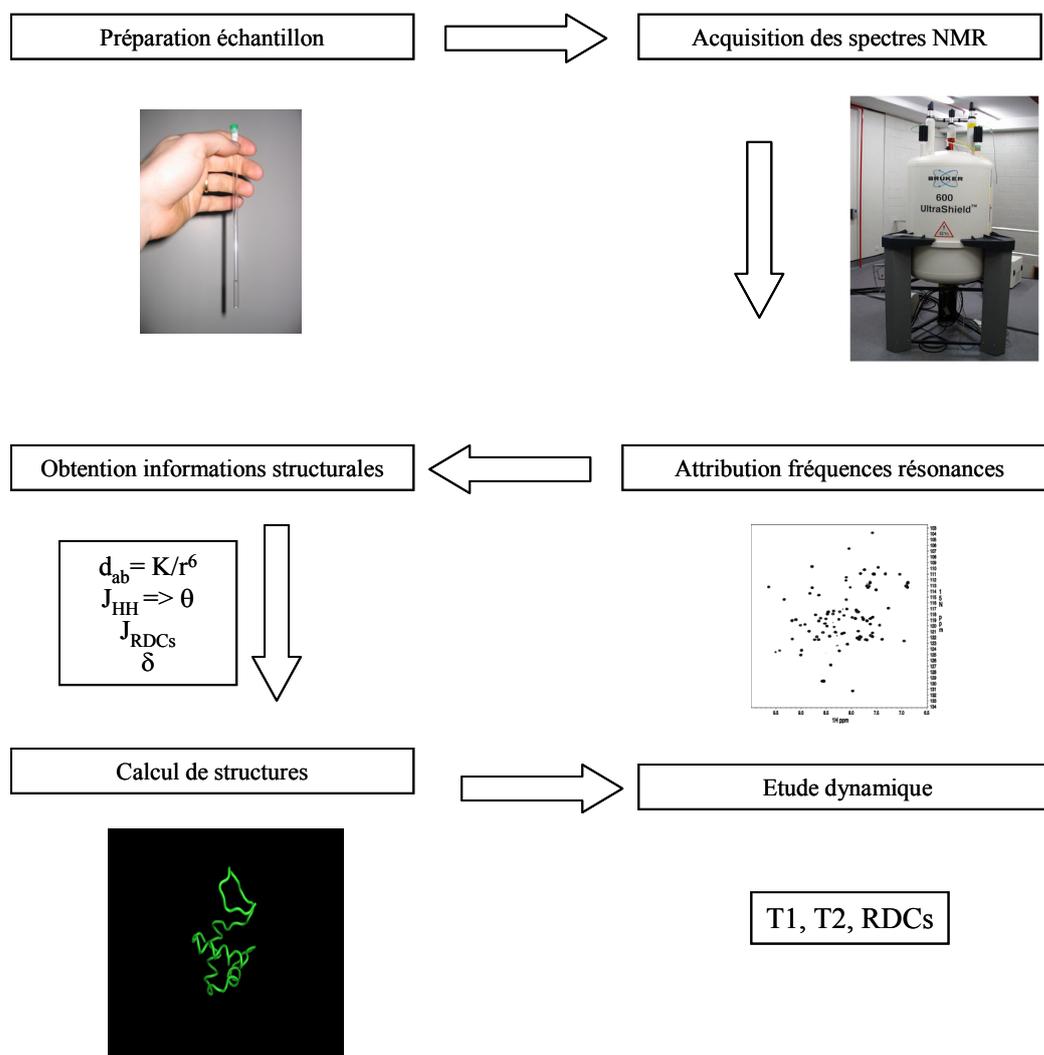


Figure 14 : Principales étapes de l'étude d'une macromolécule par RMN

Dans la suite de ce chapitre sera évoquée l'étude de deux familles de macromolécules que sont les acides nucléiques et les protéines.

## 2 Les acides nucléiques

### 2.1 Présentation

L'ADN (ou Acide DésoxyriboNucléique) est le support de l'information génétique. C'est une longue molécule constituée d'un enchaînement de nucléotides que l'on retrouve dans le noyau des cellules et dans les mitochondries. Un nucléotide est constitué d'un sucre (un désoxyribose) associé à un acide phosphorique et une base azotée. Il existe quatre bases azotées différentes : adénine et guanine qui appartiennent à la famille des Purines; thymine et cytosine qui appartiennent à la famille des Pyrimidines.

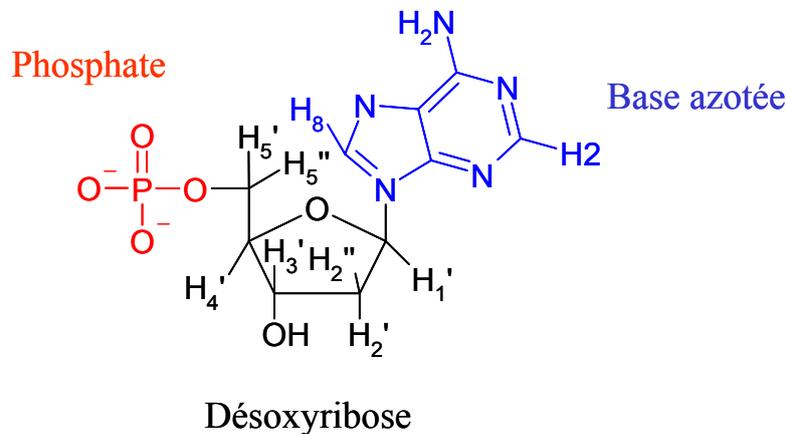


Figure 15 : Structure d'un nucléotide

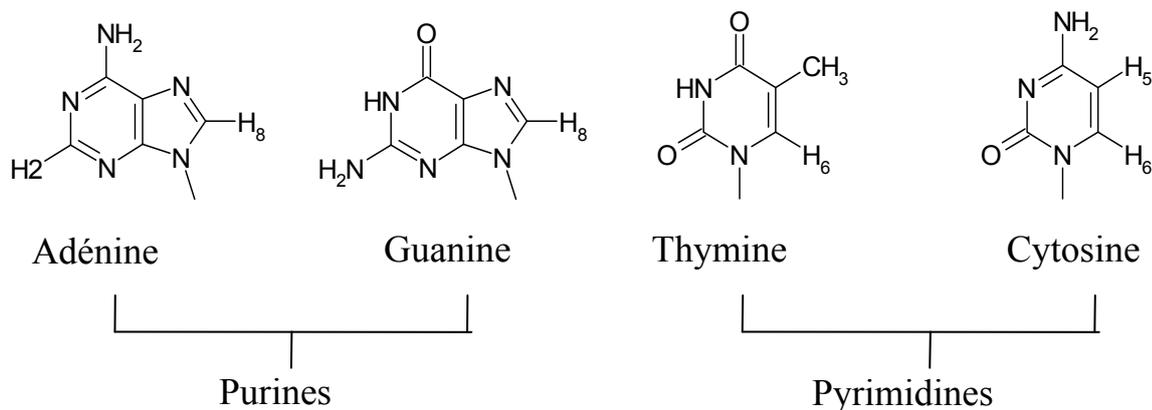


Figure 16 : Les différentes bases azotées constituant l'ADN

Les bases azotées sont complémentaires deux à deux : adénine et thymine, cytosine et guanine qui s'assemblent en formant des liaisons hydrogène. Classiquement, les bases azotées

complémentaires s'apparient en formant des appariements de type Watson et Crick mais d'autres appariements ont pu être mis en évidence : Hoogsteen... .

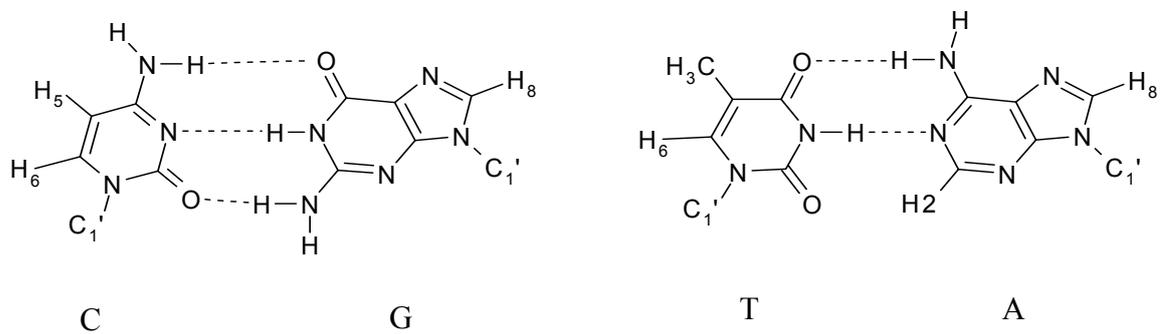


Figure 17 : appariement de types Watson et Crick

L'ADN possède une structure en double hélice dans laquelle deux brins d'ADN s'assemblent de façon antiparallèle. Les groupements phosphates se placent à l'extérieur de la double hélice, alors que les bases azotées, plus hydrophobes, se placent à l'intérieur. Il existe trois grands types de familles de double hélices : A, B et Z (Saenger 1984).

## 2.2 La synthèse chimique des acides nucléiques

Il existe plusieurs méthodes de synthèse chimique d'oligonucléotides. La plupart se déroule en phase solide. Les méthodes fondées sur l'utilisation de phosphonate ou de phosphoramidite sont les plus répandues. La synthèse en phase solide est réalisée automatiquement à l'aide d'un synthétiseur. La synthèse comporte de nombreuses étapes qui commencent par l'accrochage du groupe 3'-hydroxyle du premier nucléoside sur un support constitué soit de verre à porosité contrôlée dit CPG (controlled pore glass) soit d'un copolymère polystyrène. L'oligonucléotide est ensuite synthétisé dans la direction 3' vers 5' suivant un cycle de synthèse (correspondant à l'ajout d'un nucléotide) comprenant les étapes résumées dans la figure suivante :

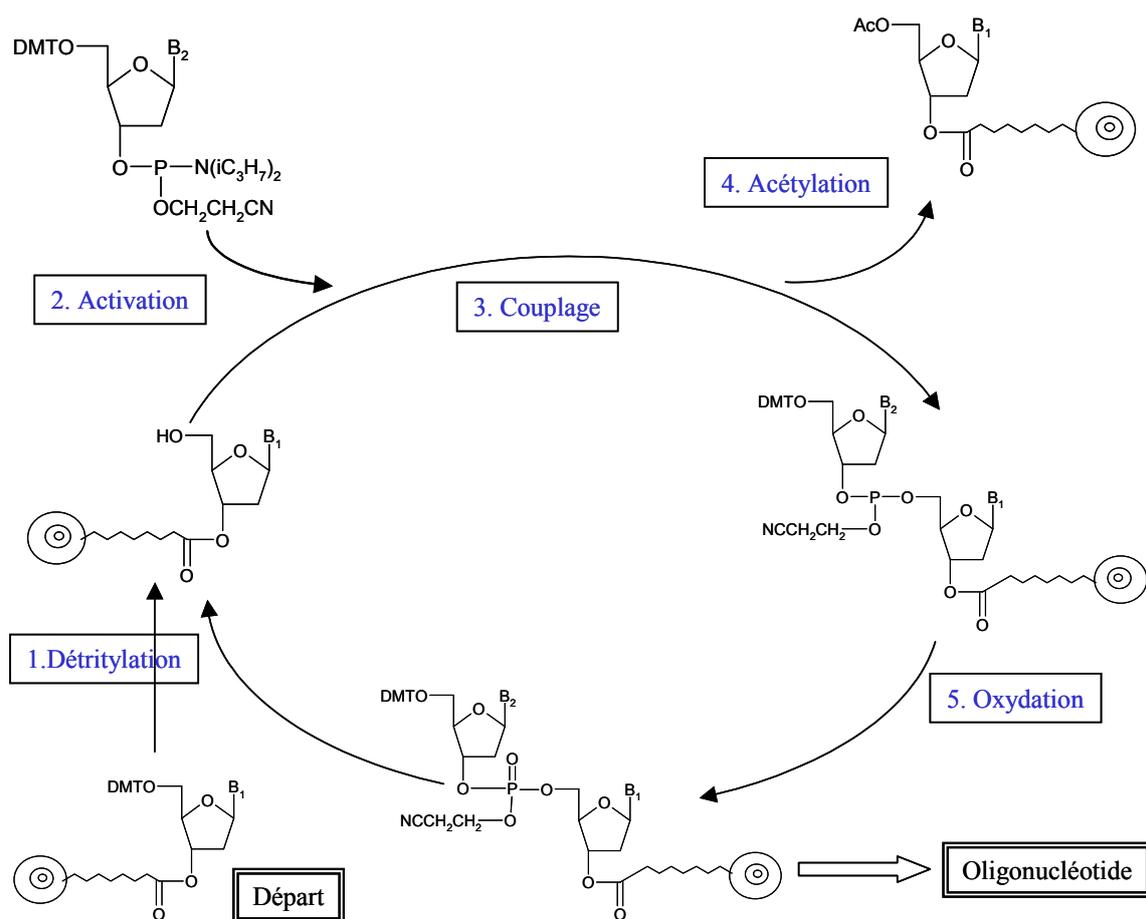


Figure 18 : Cycle de synthèse d'oligonucléotides sur phase solide en utilisant la méthode phosphoramidite.

La détritylation du nucléotide fixé au support se fait en milieu légèrement acide. Le couplage entre le nucléotide déprotégé et le phosphoramidite se fait à l'aide de tétrazole. On réalise aussi des opérations de « capture » pour éliminer les oligonucléotides qui n'auraient pas réagi lors de la réaction de couplage. Enfin l'oxydation est réalisée par du diiode en présence de pyridine. La synthèse des nucléotides peut continuer par l'élimination du groupe trityle (ou diméthoxytrityl) de la position 5' en fin de chaîne et le cycle peut se poursuivre par l'ajout d'un nouveau nucléotide.

Le traitement final suivant la synthèse comprend l'élimination complète du support solide à l'aide d'une solution d'ammoniaque (la déprotection et la coupure du support sont réalisées dans une solution ammoniacale à 28% à une température de 65°C pendant une nuit), ensuite le surnageant est prélevé, concentré dans un volume inférieur à 1,5ml puis est passé sur une colonne d'exclusion afin d'éliminer les groupements protecteurs libres et les restes de support. Pendant la synthèse des oligonucléotides, une quantité non négligeable de séquences indésirables s'accumule. Ces séquences tronquées apparaissent à la suite d'un couplage incomplet, d'une détritylation incomplète ou d'une protection imparfaite. Elles constituent la

principale source d'impuretés. Après la synthèse en phase solide, le mélange brut contient les oligonucléotides de longueur souhaitée, mais aussi des oligonucléotides tronqués et une faible quantité de produits de réactions secondaires ainsi que des sels d'ammonium étrangers. Ces impuretés sont éliminées lors des étapes de purification réalisées par HPLC en phase inverse. Cette technique permet d'obtenir un bon niveau de pureté (> 95 %) et peut être utilisée à l'échelle préparative. Pour analyser la pureté des oligonucléotides, la technique de choix est l'électrophorèse capillaire (CE) couplée à la spectrométrie de masse.

Quelques oligonucléotides utilisés pendant cette thèse ont été préparés au laboratoire mais la plupart proviennent de préparation commerciale (EUROGENTEC).

## 2.3 Etude RMN des oligonucléotides

De nombreuses revues décrivent la démarche générale à employer pour réaliser l'étude d'acides nucléiques par RMN (Wüthrich 1986; Borer, Pappalardo et al. 1997).

### 2.3.1 Le déplacement chimique du proton

Les déplacements chimiques des protons d'un acide nucléique s'étalent sur une gamme de déplacement chimique allant de 0 à 16 ppm.

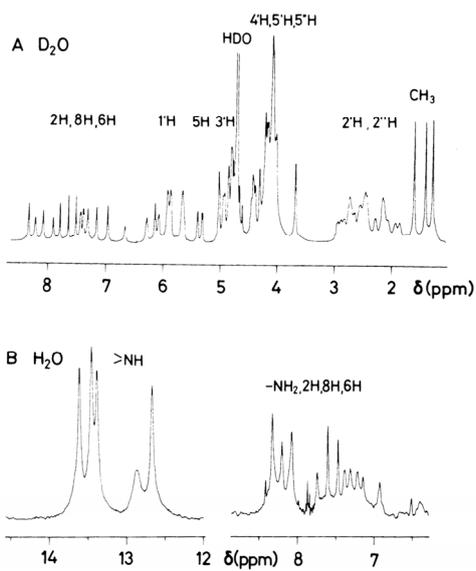


Figure 19 : Spectre RMN 1D d'un fragment d'ADN double brin (A) dans D<sub>2</sub>O on observe les protons non-échangeables; (B) dans l'eau, Les protons échangeables sont observables (Wüthrich 1986)

Les protons H<sub>4'</sub>, H<sub>5'</sub> et H<sub>5''</sub> sont fortement recouverts et sont donc difficilement attribuables. L'attribution des protons H<sub>1'</sub>, H<sub>8</sub>/H<sub>6</sub>, H<sub>2'</sub>, H<sub>2''</sub>, H<sub>3'</sub>, H<sub>2</sub> et CH<sub>3</sub> est réalisable et repose sur la méthode de l'attribution séquentielle.

### 2.3.2 Méthodologie d'attribution séquentielle

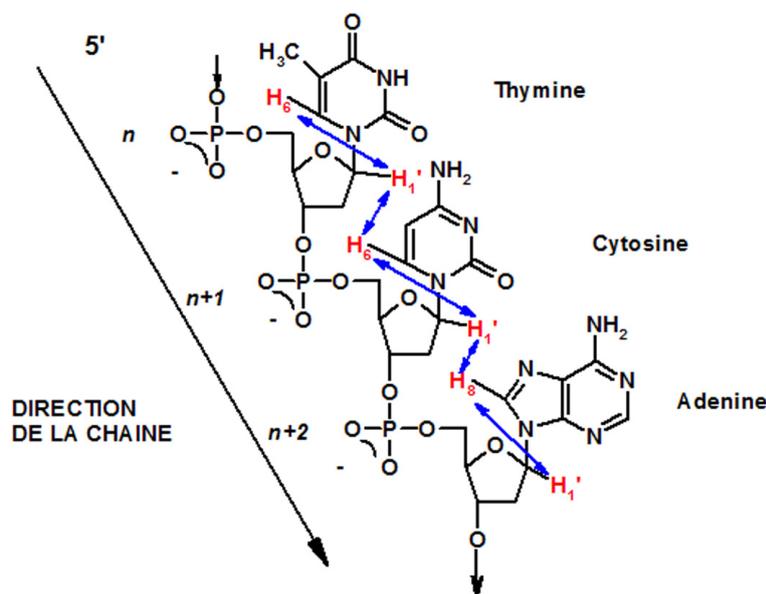


Figure 20 : Principe de l'attribution séquentielle

L'attribution séquentielle consiste à identifier l'ensemble des protons  $H_{6/8}$  et  $H_{1'}$  des différents résidus constituant le fragment d'ADN à l'aide des spectres NOESY. On procède de la façon suivante : le proton  $H_6$  du nucléotide  $n$  est proche de son proton  $H_{1'}$  (distance  $< 5\text{\AA}$ ), une connectivité entre ces deux protons sera donc observable sur le spectre NOESY. De plus le proton  $H_{1'}$  du nucléotide  $n$  est proche du  $H_6$  du nucléotide  $n+1$  qui est lui-même proche de son proton  $H_{1'}$  et ainsi de suite. On peut alors remonter l'ensemble de la séquence nucléotidique et attribuer l'ensemble des protons  $H_6 / H_8$  et  $H_{1'}$ .

Les protons  $H_6/H_8$  des bases azotées sont proches dans l'espace des protons  $H_2'/H_2''$  du sucre de leur nucléotide mais sont aussi proches dans l'espace des protons  $H_2'/H_2''$  du sucre voisin en 5'. Cela permet d'attribuer théoriquement l'ensemble des protons  $H_2'$  et  $H_2''$ . Les résonances  $H_5$  des cytosines et  $CH_3$  des thymines peuvent être identifiées à l'aide d'un spectre TOCSY.

### 2.4 L'effet Overhauser nucléaire

L'effet Overhauser nucléaire (NOE) représente la principale source d'information pour la détermination structurale en RMN. Ce phénomène repose sur le transfert d'aimantation à travers l'espace entre 2 protons. La vitesse de transfert est directement reliée à la distance entre ces protons. L'expression du NOE est donnée par la formule suivante :

$$\eta_i = \frac{K}{d_i^6}$$

K est une constante qui dépend du temps de corrélation de la molécule. Cette relation n'est plus valable pour des temps de mélange longs, où le proton j va transférer l'aimantation issue du proton i sur un troisième proton k qui est proche de j. Il apparaît alors un NOE entre les protons i et k sans que ceux-ci soient nécessairement proches dans l'espace, tandis que l'intensité du NOE(i/j) est diminuée et ne satisfait plus la relation.

Si on connaît une distance de référence  $r_{ref}$ , par exemple la distance entre  $H_{\beta 1}$  et  $H_{\beta 2}$  des chaînes latérales des acides aminés dans une protéine ou la distance  $H_5-H_6$  d'une cytosine dans un acide nucléique, on pourra obtenir une estimation de la distance  $d_i$  en utilisant la relation :

$$d_i = d_{ref} \left( \frac{\eta_{ref}}{\eta_i} \right)^{\frac{1}{6}}$$

où  $\eta_{ref}$  est l'intensité NOE du pic de référence.

Cette étape de « calibration » des NOE est essentielle. Cependant, compte tenu de l'incertitude sur la mesure de l'intensité et de la dynamique interne de certaines régions de la molécule qui peut entraîner une diminution de l'effet NOE, on utilisera des intervalles de valeurs permises pour les distances inter protons :  $d_{min} < d_i < d_{max}$ .

Seules les informations de courtes distances, c'est-à-dire inférieures à 5 Å, sont recueillies. Les effets NOE permettent d'avoir des données locales sur la molécule. La multiplication de ces données permettra d'obtenir la conformation globale de la molécule. Il est donc très important de recueillir un maximum de données pour affiner la structure de la molécule.

## 2.5 Les liaisons hydrogène

Les protons acides s'échangent rapidement avec les protons de l'eau, ils résonnent alors à la fréquence de l'eau. Néanmoins, si ceux-ci sont engagés dans des liaisons hydrogène, leur échange est ralenti et ils peuvent alors être observés par RMN. Les protons amides des protéines sont très importants pour l'étude structurale des protéines du fait de leur dispersion en déplacement chimique et du grand nombre de couplages dipolaires que l'on peut obtenir grâce à eux. De même dans les oligonucléotides, les protons iminos et aminos sont très importants, ils nous permettent d'avoir des informations essentielles sur les appariements entre

les bases azotées complémentaires. Le pH est un paramètre essentiel dans les phénomènes d'échanges.

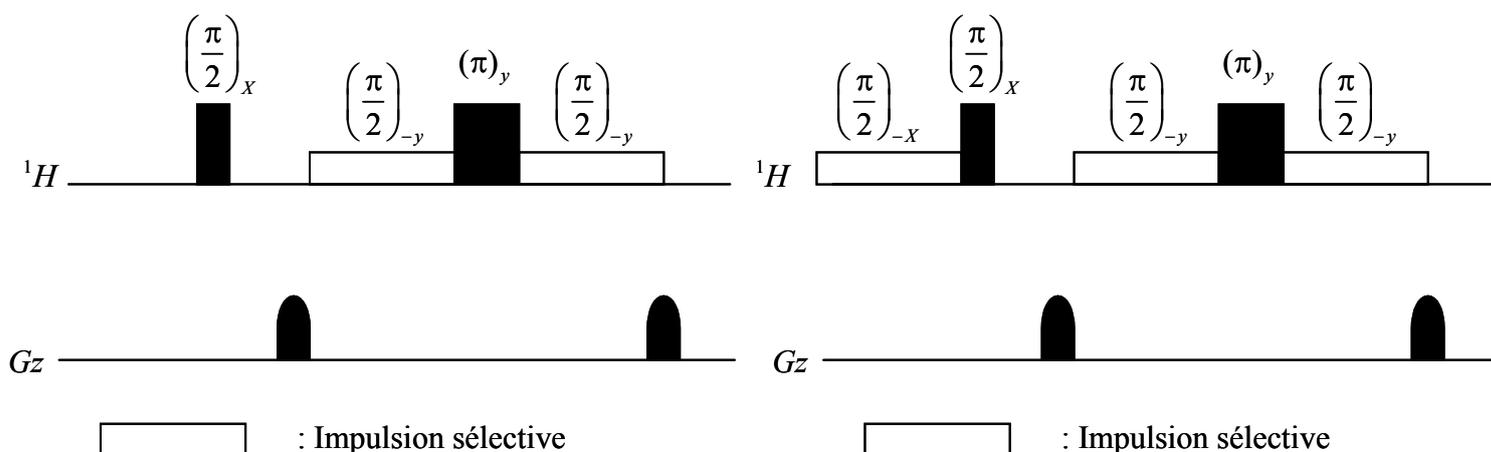
Les spectres de molécules biologiques dans l'eau posent de nombreux problèmes. Dans le tube RMN, la concentration en eau (H<sub>2</sub>O) est d'environ 110 M et celle du composé biologique entre 0.5 et 2mM. En RMN impulsionnelle, le signal analogique reçu dans la bobine est converti en signal digital par l'ADC (convertisseur analogique digital). Les intervalles de temps entre deux points successifs enregistrés ne doivent pas dépasser le pas de Nyquist défini comme étant égal à 1/2SW. Le signal très intense de l'eau va saturer le convertisseur analogique digital, au point que les raies de résonances de notre molécule apparaîtront comme du bruit. Il est donc indispensable de réaliser l'élimination du signal du solvant. Pour cela différentes techniques d'élimination de l'eau ont été mises au point dont quelques-unes sont détaillées ci-dessous.

### 2.5.1 Utilisation de gradients

On appelle gradients des impulsions de gradient de champ magnétique. Ces impulsions ont pour but de rendre le champ B<sub>0</sub> inhomogène et d'entraîner un déphasage complet des cohérences. On réalise en fait une défocalisation de l'aimantation (M=0). L'intensité du champ B<sub>0</sub> varie alors linéairement selon l'axe du gradient. La fréquence de résonance est alors (si le gradient est selon l'axe z) :  $\omega(z) = -\gamma[B_0 + Bg(z)] = -\gamma[B_0 + zGz]$ . Néanmoins l'application d'un second gradient (inverse au premier) pourra refocaliser l'aimantation. L'utilisation des gradients permet la sélection des chemins de cohérence, l'élimination du signal, notamment celui de l'eau, ou encore de réaliser la détection en quadrature dans la dimension 1.

### 2.5.2 La séquence Watergate (Piotto, Saudek et al. 1992)

La séquence Watergate est présentée à la figure 19. La première impulsion  $\frac{\pi}{2}$  n'est pas sélective, elle affecte donc tous les spins. Suit ensuite le premier gradient qui défocalise l'aimantation dans le plan xoy. Ensuite les protons de l'eau vont subir une rotation de  $2\pi$  (2 impulsions sélectives  $\frac{\pi}{2}$  plus une impulsion  $\pi$ ) alors que les autres protons vont subir simplement une inversion due à l'impulsion  $\pi$ . L'aimantation de ces derniers sera alors refocalisée par le second gradient alors que le signal de l'eau sera de nouveau défocalisé.



*Figure 21: Sequence Watergate avec ou sans flipback*

Notons qu'à la place du train classique d'impulsions Watergate, on préfère généralement utiliser la séquence 3-9-19, représentée par le train d'impulsions  $3\alpha-9\alpha-19\alpha-19\alpha-9\alpha-3\alpha$  avec  $26\alpha \sim 180^\circ$  (Sklenar, Piotto et al. 1993).

### 2.5.3 Waterflipback (Grzesiek and Bax 1993)

On ajoute une impulsion sélective avant la première impulsion  $\frac{\pi}{2}$  non sélective. L'intérêt de cette méthode est de garder l'aimantation due aux protons de l'eau selon l'axe z. Si le signal de l'eau n'est pas totalement selon z, il y aura défocalisation de l'aimantation grâce aux gradients. L'avantage de cette méthode est que le signal de l'eau n'est plus dans le plan xoy, il n'y a donc pas de risque d'avoir des phénomènes de transfert de saturation entre l'eau et la molécule étudiée.

Enfin d'autres techniques comme la présaturation, le jump and return (Plateau and Gueron 1982) ou certains traitements post-acquisitions permettent l'élimination du signal de l'eau.

### 3 Les protéines

#### 3.1 Présentation

Une protéine est une macromolécule composée d'une séquence d'acides aminés reliés entre-eux par des liaisons peptidiques (figure 20). Il existe 20 acides aminés dits essentiels. L'enchaînement des acides aminés forme la structure primaire de la protéine.

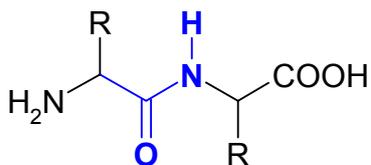


Figure 22: Liaison peptidique planaire CONH

On a coutume de définir les angles de torsion  $\psi$  (Psi),  $\phi$  (Phi),  $\omega$  (Omega) :

$\psi$  : N(i-1) - C $\alpha$ (i-1) - CO(i-1) - N(i)

$\phi$  : CO(i-1) - N(i) - C $\alpha$ (i) - CO(i)

$\omega$  : C $\alpha$ (i-1) - CO(i-1) - N(i) - C $\alpha$ (i)

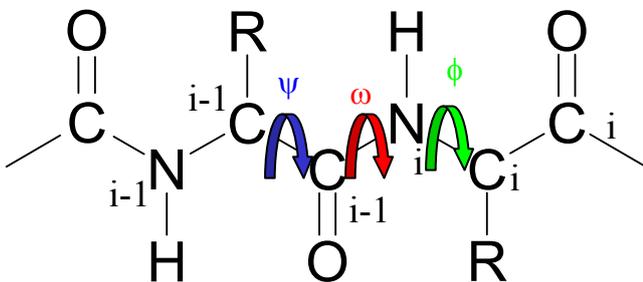
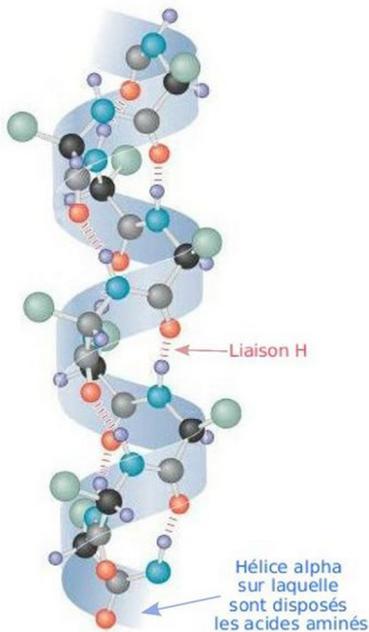


Figure 23: Les angles peptidiques dans une protéine

La séquence en acides aminés est toujours donnée du N-terminal vers le C-terminal. Cette chaîne primaire se replie et s'organise pour former des structures secondaires comme les hélices, les feuilletts ou les coudes. Ces structures sont stabilisées par différentes interactions entre les atomes des différents acides aminés.

### 3.1.1 Les hélices



Il existe deux grands types d'hélice: l'hélice  $\alpha$  et l'hélice 3.10. L'hélice  $\alpha$  a 3.6 acides aminés par tour d'hélice et il existe des liaisons hydrogène entre les atomes d'oxygène des groupements carbonyles CO d'un acide aminé  $n$  et les protons amides d'un acide aminé  $n+4$ . L'hélice 3.10 se retrouve dans certaines protéines globulaires. Elle est moins stable que l'hélices  $\alpha$  et ne possède que trois résidus par tour d'hélice.

### 3.1.2 Les feuillets

Les feuillets  $\beta$  sont des structures regroupant plusieurs chaînes d'acides aminés reliés par des liaisons hydrogène. Des liaisons hydrogène impliquant les groupements carbonyles et les protons amides se forment entre les deux brins. Lorsque l'orientation des brins est dans la même direction, il s'agit d'un feuillet  $\beta$  parallèle. Lorsque les acides aminés dans les brins successifs ont des directions opposées, c'est un feuillet  $\beta$  antiparallèle.

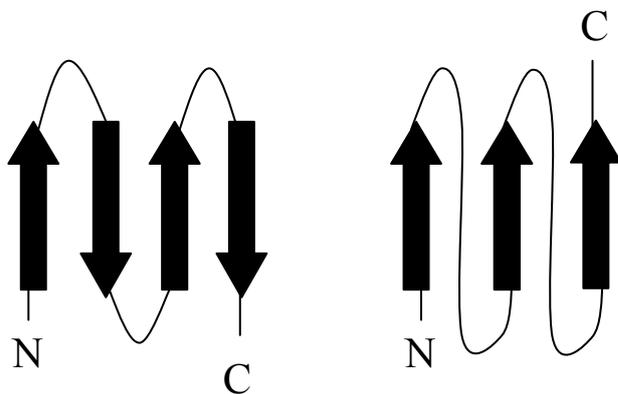


Figure 24 :a) feuillet  $\beta$  antiparallèle b) feuillet  $\beta$  parallèle

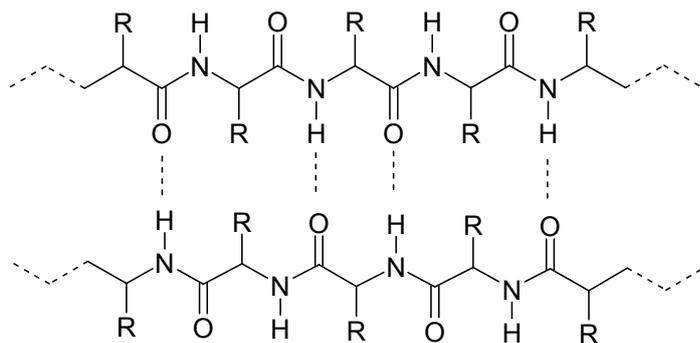


Figure 25 : Liaisons hydrogène liant deux brins  $\beta$

### 3.1.3 Les coudes

Les coudes permettent à la protéine de se replier afin de favoriser un changement de direction. Un coude  $\beta$  est défini par rapport à quatre atomes successifs ( $i, i+1, i+2, i+3, i+4$ ). Si la distance entre les  $C\alpha$  du résidu  $i$  et le  $C\alpha$  du résidu  $i+3$  est inférieure à 7 Å et si les résidus  $i+2$  et  $i+3$  ne sont pas dans des hélices, nous aurons un coude  $\beta$ . Le coude est classé dans une des neuf catégories suivant les angles  $\phi$  et  $\psi$  des résidus  $i+1$  et  $i+2$ .

Les coudes  $\gamma$  sont définis, pour trois résidus  $i, i+1, i+2$ , s'il existe une liaison hydrogène entre les résidus  $i$  et  $i+2$  et suivant les angles  $\psi$  et  $\phi$  du résidu  $i+1$ . Il y a alors deux classes de coudes  $\beta$ , classique ou inversé.

Enfin la structure tertiaire des protéines correspond aux repliements des structures secondaires.

## 3.2 Etude RMN des protéines

L'étude d'une protéine par Résonance Magnétique Nucléaire est différente de l'étude d'un acide nucléique. La première structure 3D de protéine par RMN a été publiée en 1985 (Williamson, Havel et al. 1985). Depuis d'énormes progrès technologiques et méthodologiques ont été réalisés pour rendre la spectroscopie RMN performante dans la détermination de structures tridimensionnelles de macromolécules (protéines, acides nucléiques, sucres, complexes protéine/protéine ou protéine/acide nucléique). Ces progrès ont permis d'augmenter la résolution et la sensibilité de la RMN, points faibles de la technique. Ils ont consisté dans le développement de spectromètres à haut champ (22,3T actuellement) et de sondes cryogéniques. Ces progrès technologiques ont été accompagnés de développements méthodologiques : avec l'apparition du marquage isotopique ( $^{15}\text{N}$ ,  $^{13}\text{C}$  et  $^2\text{H}$ ), le

développement de séquences RMN 3D, et 4D (Cavanagh, Fairbrother et al. 1996);(Sattler, Schleucher et al. 1999), l'utilisation de séquences TROSY (Transverse Relaxation Optimized Spectroscopy) (Pervushin, Riek et al. 1997) et l'utilisation des constantes de couplages dipolaires résiduels (Tjandra, Omichinski et al. 1997). Ces développements ont permis de réaliser l'étude structurale de biomolécules de taille de plus en plus importante. L'enzyme malate synthase de E. coli de 723 résidus (Tugarinov, Hwang et al. 2004) est actuellement la plus grosse protéine dont les déplacements chimiques sont connus ( $^1\text{H}$ ,  $^{13}\text{C}$  et  $^{15}\text{N}$ ).

### 3.2.1 Méthodes d'attribution RMN des protéines

La structure 3D de l'inhibiteur de trypsine de taureau de 53 résidus, publiée en 1985 (Williamson, Havel et al. 1985) par l'équipe de Kurt Wüthrich, prix Nobel de chimie 2002, a été la première structure 3D de protéine en solution. L'attribution des déplacements chimiques de ses protons, seuls noyaux alors observables en abondance naturelle de façon exploitable à l'époque, a été obtenue par la méthode d'attribution séquentielle (Wüthrich 1986) basée sur l'effet NOE.

### 3.2.2 Méthode d'attribution séquentielle

Cette méthode exploite les informations de corrélations scalaires révélées par les expériences de type COSY et TOCSY. Ces expériences qui mettent en évidence des corrélations entre protons au travers des liaisons permettent d'obtenir les signatures caractéristiques des systèmes de spins de chaque acide aminé.

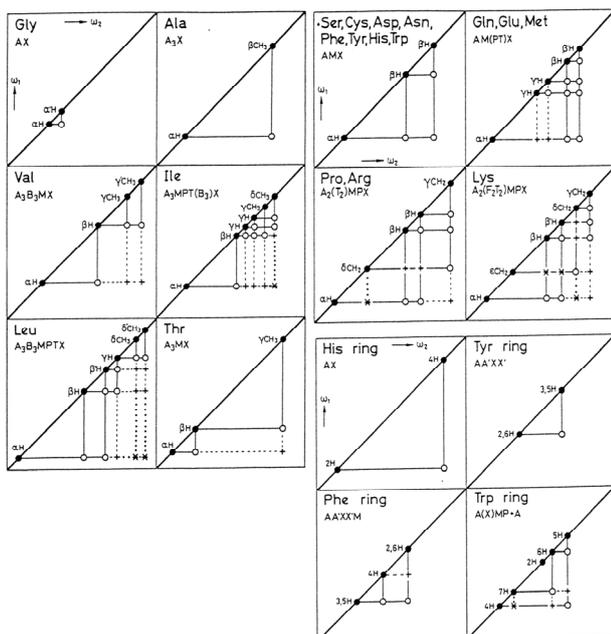


Figure 26 : Diagramme de connectivités observables dans les spectres COSY et TOCSY pour les différents acides aminés.

L'étape suivante consiste à déterminer la position des acides aminés ainsi identifiés dans la séquence de la protéine. C'est ce que l'on appelle l'attribution séquentielle. Celle-ci est fondée sur des expériences NOESY qui permettent d'observer des protons distants de moins de 5Å. Pour l'attribution séquentielle, on cherche les corrélations dipolaires entre le résidu i et le résidu i+1, et plus particulièrement les corrélations H $\alpha$ (i)-HN(i+1) qui permettent de reconstituer l'enchaînement des systèmes de spins. La méthode d'attribution séquentielle est applicable aux protéines de petite taille, généralement inférieure à une centaine de résidus.

### 3.2.3 Utilisation de la RMN hétéronucléaire

L'étude structurale de macromolécules de taille de plus en plus importante par la RMN du proton seul connaît certaines limites. La gamme relativement réduite des déplacements chimiques des résonances du proton entraîne le recouvrement de résonances conduisant à des indéterminations. De plus, le temps de relaxation transversale T2 est diminué entraînant un

élargissement des raies de résonances ( $\nu_{\frac{1}{2}} \propto \frac{1}{\pi T2}$ ). La RMN hétéronucléaire permet de palier

ces difficultés. Tout d'abord elle apporte une (ou plusieurs) échelle de déplacements chimiques supplémentaires portant de nouvelles informations concernant l'environnement de l'hétéronoyau considéré. Par ailleurs, les constantes de couplage mettant en jeu un hétéronoyau ( $J_{CH}=170\text{Hz}$ ,  $J_{NH}=90\text{Hz}$ ) sont pour la plupart plus élevées que les constantes de couplage d'une paire  $^1\text{H}-^1\text{H}$ . Le transfert de polarisation le long de la liaison X-H proportionnel à  $1/2J$  se fera donc plus rapidement, évitant ainsi une perte de sensibilité due aux processus de relaxation transversale pour les grosses molécules. De plus, le marquage isotopique uniforme ou sélectif des macromolécules biologiques a permis de contourner le problème de la faible abondance naturelle des isotopes ( $^{15}\text{N}=0.37\%$ ,  $^{13}\text{C}=1.1\%$ ). Enfin les mécanismes de relaxation des hétéronoyaux de spin  $\frac{1}{2}$  engagent essentiellement l'hétéronoyau et les protons auxquels il est directement lié. Dans un système de spins XH, la distance entre ces deux noyaux étant fixe, seule l'orientation du vecteur XH va varier au cours du temps. La relaxation hétéronucléaire reliée aux mouvements moléculaires de ce vecteur apporte des informations dynamiques beaucoup plus directes que la relaxation du proton.

La méthode d'attribution séquentielle présentée précédemment est facilitée par un marquage  $^{15}\text{N}$  (ou  $^{13}\text{C}$  ou les deux). L'attribution se réalise alors à partir de spectres multidimensionnels hétéronucléaires, comme par exemple les expériences 3D TOCSY-ou NOESY-HSQC éditées  $^{15}\text{N}$  ou  $^{13}\text{C}$ . Ces expériences 3D sont la combinaison de deux expériences 2D : la première

TOCSY ou NOESY  $^1\text{H}$ , et la deuxième HSQC (pour Heteronuclear Single Quantum Coherence) qui permettent d'étaler l'information selon la fréquence de l'hétéronoyau auquel sont liés les protons.

### 3.2.4 Attribution du backbone: les expériences HNCA et HN(CO)CA

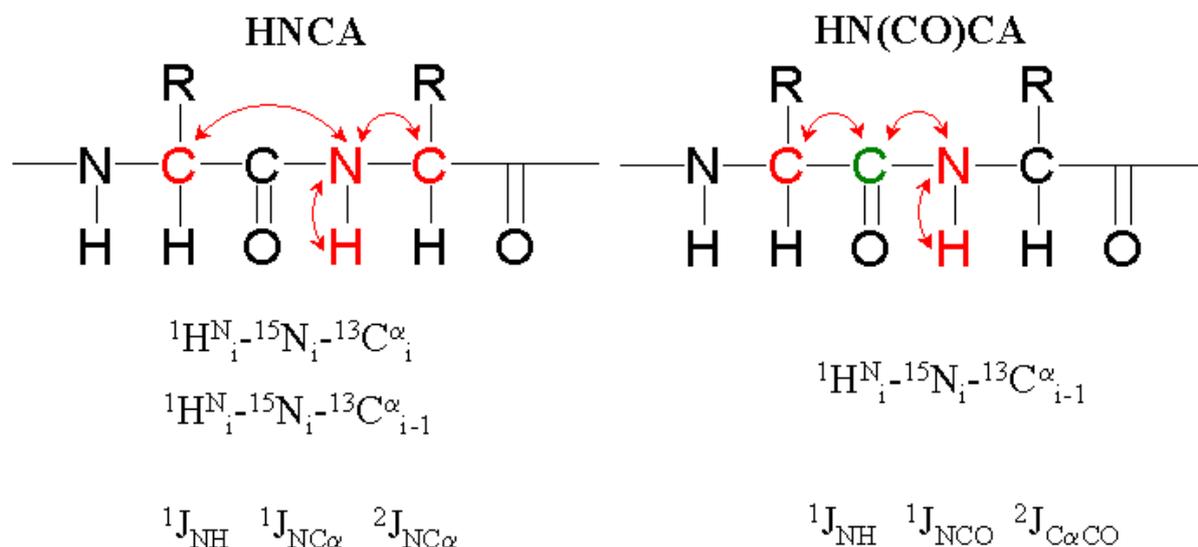
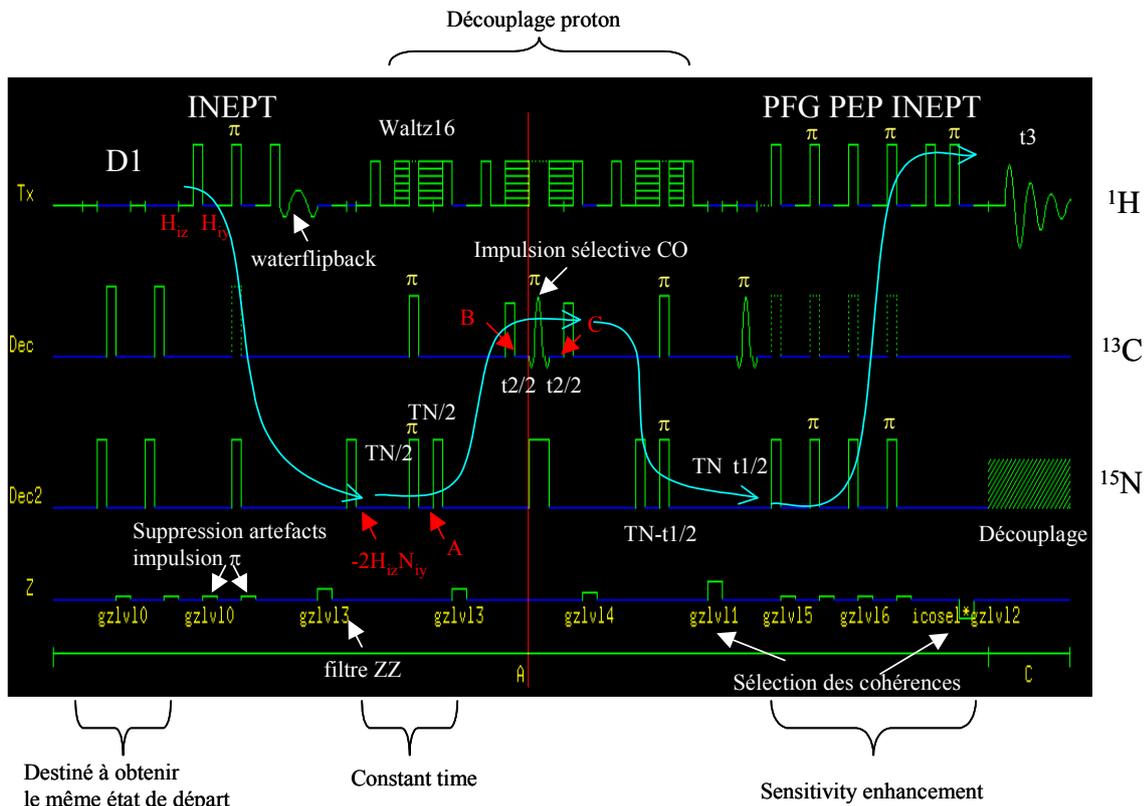


Figure 27 : (a) Schéma des noyaux impliqués dans les expériences HNCA et HN(CO)CA (b) Corrélations pouvant être observées dans ces deux expériences (c) Constantes de couplages mises en jeu dans chacune des expériences

Ces deux expériences permettent, en théorie, l'attribution des résonances des  $^{13}\text{C}\alpha$ ,  $^1\text{HN}$  et  $^{15}\text{N}$  de la chaîne peptidique de la protéine. L'expérience HNCA (Kay, Ikura et al. 1990) corrèle les résonances  $^1\text{HN}$  et  $^{15}\text{N}$  à la résonance  $^{13}\text{C}\alpha$  du résidu  $i$  et à la résonance  $^{13}\text{C}\alpha$  du résidu  $i-1$ . L'expérience HN(CO)CA (Bax and Ikura 1991) permet de relier les résonances  $^1\text{HN}$  et  $^{15}\text{N}$  d'un même résidu avec seulement la résonance  $^{13}\text{C}\alpha$  du résidu  $i-1$  via le carbonyle CO, ce dernier n'étant pas édité au cours de l'expérience. La figure ci-dessous présente la séquence ghn\_ca, provenant de biopack, que nous avons utilisée sur le spectromètre VARIAN.



$$A: \left\{ 4H_{iz}N_{ix}C_{iz}^{\alpha}\Gamma_1(t) + 4H_{iz}N_{ix}C_{(i-1)z}^{\alpha}\Gamma_2(t) \right\}$$

$$\Gamma_1(t) = \sin(\pi J_{C^{\alpha}N}t) \cos(\pi^2 J_{C^{\alpha}N}t)$$

$$\Gamma_2(t) = \cos(\pi J_{C^{\alpha}N}t) \sin(\pi^2 J_{C^{\alpha}N}t)$$

$$B: \left\{ 4H_{iz}N_{iz}C_{iy}^{\alpha}\Gamma_1(t) + 4H_{iz}N_{iz}C_{(i-1)y}^{\alpha}\Gamma_2(t) \right\}$$

$$C: \left\{ 4H_{iz}N_{iz}C_{iy}^{\alpha}\Gamma_1(t) + 4H_{iz}N_{iz}C_{(i-1)y}^{\alpha}\Gamma_2(t) \right\} \cos(\Omega_C t^2) \cos(\pi J_{CaC\beta} t^2)$$

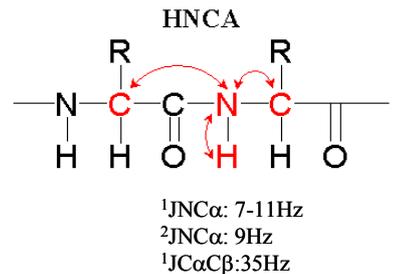


Figure 28 : Séquence ghn\_ca utilisée

Néanmoins il peut y avoir des superpositions de corrélations ne permettant pas d'identifier tous les résidus. Une solution possible pour lever ces ambiguïtés est l'utilisation d'une autre combinaison d'expériences fournissant des corrélations alternatives: le couple HN(CA)CO (Clubb, Thanabal et al. 1992) et HNCO (Kay, Ikura et al. 1990). L'expérience HNCO fournit uniquement les corrélations interrésiduelles tandis que l'expérience HN(CA)CO, en plus de ces corrélations, donne celles qui sont intrarésiduelles. Pour lever d'éventuelles ambiguïtés persistantes, on a recours au dernier couple d'expériences: CBCANH (Grzesiek and Bax 1992) et CBCA(CO)NH (Grzesiek and Bax 1992).

De toutes ces expériences, on extrait des données complémentaires qui permettent de réaliser l'attribution séquentielle plus facilement, en limitant les risques d'erreurs.

### 3.2.5 Attribution des chaînes latérales

Outre les expériences 3D TOCSY-HSQC éditées  $^{15}\text{N}$  et  $^{13}\text{C}$ , les expériences HCCH-COSY ( $^1\text{H}$ - $^{13}\text{C}$ - $^{13}\text{C}$ - $^1\text{H}$  COrrrelation SpectroscopY) (Bax, Clore et al. 1990) et HCCH-TOCSY ( $^1\text{H}$ - $^{13}\text{C}$ - $^{13}\text{C}$ - $^1\text{H}$  TOtal Correlation SpectroscopY) (Bax, Clore et al. 1990) permettent d'identifier tous les protons et les carbones aliphatiques des chaînes latérales. Le résultat de l'expérience 3D HCCH-TOCSY contient la même information qu'une expérience 3D TOCSY-HSQC éditée  $^{13}\text{C}$ . Néanmoins ces expériences sont beaucoup plus sensibles car le transfert HCCH-TOCSY est bien plus efficace qu'un transfert par les protons. Les constantes de couplage  $^2\text{J}$  ou  $^3\text{J}$  entre les protons sont de l'ordre de quelques hertz (4-10 Hz) alors que les constantes de couplages  $^1\text{J}_{\text{CH}} \sim 170\text{Hz}$  et  $^1\text{J}_{\text{CC}} \sim 40\text{Hz}$  sont beaucoup plus élevées.

Finalement un certain nombre d'autres expériences permet de lever les ambiguïtés restantes. Elles se dénomment HBHA(CBCACO)NH (Grzesiek and Bax 1993), HBHA(CBCA)NH (Wang, Lodi et al. 1994), HCC(CO)NH-TOCSY (Montelione, Lyons et al. 1992; Grzesiek, Anglister et al. 1993; Logan, Olejniczak et al. 1993), HN(CO)CACB (Yamazaki, Lee et al. 1994), HNCACB (Wittekind and Mueller 1993) et HCCNH-TOCSY (Lyons and Montelione 1993).

### 3.2.6 Obtention des distances

La procédure est la même que pour les acides nucléiques. Néanmoins, il faut noter que contrairement aux acides nucléiques pour lesquels les NOEs sont pour la plupart séquentiels, c'est-à-dire impliquant les résidus  $i$  et  $i \pm 1$ , les NOEs dans les protéines peuvent être classés en trois catégories :

- Les NOEs à courtes distances, c'est-à-dire impliquant des résidus  $i$  et  $j$  tel que  $|i - j| \leq 1$ . On distingue ensuite les NOEs intrarésiduels des NOEs interrésiduels.
- Les NOEs à moyennes distances, c'est-à-dire impliquant des résidus  $i$  et  $j$  tel que  $1 \leq |i - j| \leq 5$ .
- Les NOEs à longues distances impliquant les résidus  $i$  et  $j$  tel que  $|i - j| \geq 5$ .

### 3.2.7 Obtention des angles de torsion

Les angles de torsion du squelette peptidique peuvent être déterminés soit à partir des valeurs de couplage  $^3\text{J}_{\text{HNH}\alpha}$  déduites de l'expérience HNHA, soit en utilisant des méthodes combinant différents types d'informations comme le logiciel TALOS (Cornilescu, Delaglio et al. 1999) disponible à l'adresse internet suivante <http://spin.niddk.nih.gov/bax/software/talos>). Ce

logiciel permet d'obtenir un intervalle de valeurs possibles pour les angles ( $\phi, \psi$ ) à partir des valeurs des déplacements chimiques de cinq atomes d'un résidu:  $H\alpha, C\alpha, C\beta, CO, N$ . TALOS considère chaque résidu dans l'environnement de ses 2 voisins et compare, pour chaque tripeptide, les déplacements chimiques mesurés à ceux de tripeptides dans une banque de données. Cette banque regroupe les données de 78 protéines de structures connues, représentant environ 9000 triplets d'acides aminés. Les dix tripeptides comportant à la fois une bonne similarité de séquence et les déplacements chimiques les plus proches de ceux de la protéine cible sont conservés. Les angles  $\phi$  et  $\psi$  correspondants sont représentés sur une carte de Ramachandran. L'utilisateur valide le résultat : si dans au moins 9 cas sur 10, ceux-ci se situent au niveau de la même région favorable du diagramme de Ramachandran, la prévision est considérée comme satisfaisante. La méthode de TALOS, bien que particulièrement efficace, demeure empirique et il a été montré que les prédictions étaient fausses pour seulement 3% des angles. Les valeurs des angles de torsion ne sont pas déterminées de façon précise et on définit un intervalle de valeurs possibles pour l'angle dièdre :  $\phi_{\min} < \phi < \phi_{\max}$

## 4 Etude de complexes ADN-Protéines

L'étude de complexe ADN-protéine par RMN peut être réalisée grâce à des séquences RMN utilisant des filtres si l'un des deux partenaires est marqué  $^{13}C$  et/ou  $^{15}N$  (Otting and Wüthrich 1989; McIntosh 1999).

Les expériences demi-filtre, bien que très peu sensibles (à cause des délais pendant lesquels la relaxation agit), permettent la séparation des résonances des protons liés à un hétéroatome X de tous les autres protons.

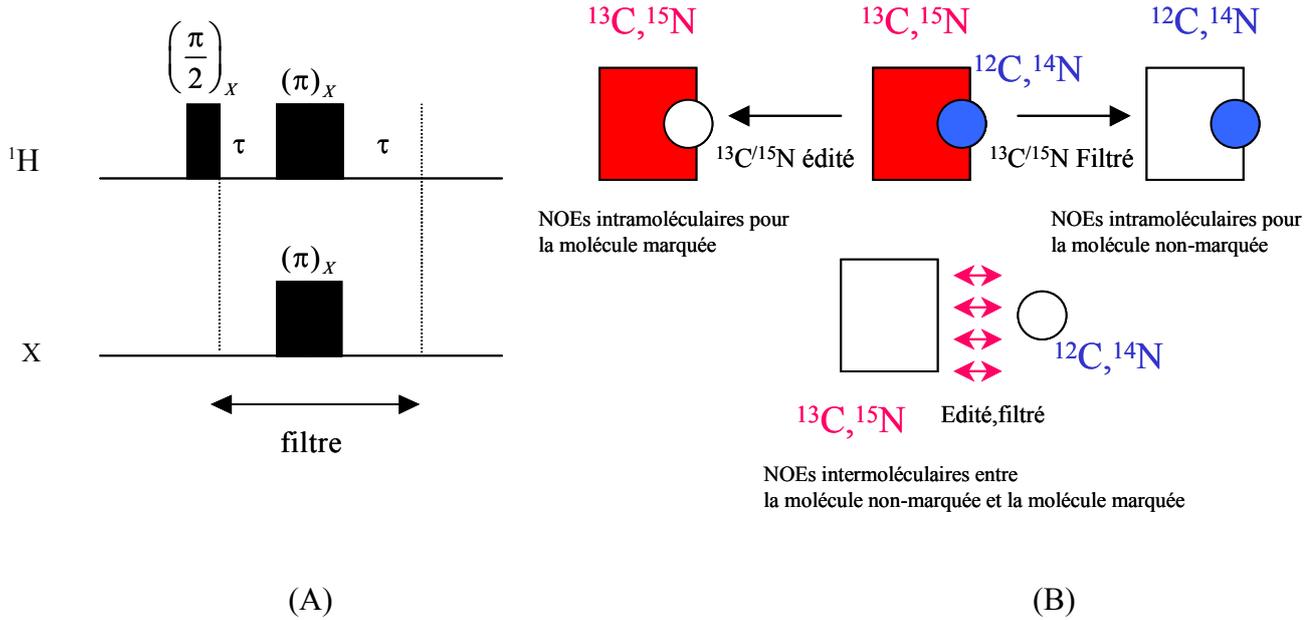


Figure 29 : (A) Présentation d'une séquence RMN demi-filtre; (B) Utilisation des expériences filtrées/éditées

Les filtres correspondent à des impulsions  $\pi$  non sélectives encadrées de deux délais  $\tau = \frac{1}{2J_{I^A I^X}}$  où  $J_{I^A I^X}$  est la constante de couplage entre le proton et son hétéroatome (figure 27 A).

La première impulsion  $\frac{\pi}{2}$  sur le canal du proton fait basculer l'aimantation dans le plan xoy, nous avons ensuite évolution sous l'effet du déplacement chimique et de la constante de couplage  $J_{AX}$ :

$$I_z^A \xrightarrow{\left(\frac{\pi}{2}\right)_X} I_y^A$$

$$I_y^A \xrightarrow{\frac{\tau}{J_{AX}}} I_y^A \cos \omega_A t \cos \pi J_{AX} t - 2I_x^A I_z^X \cos \omega_A t \sin \pi J_{AX} t - I_x^A \sin \omega_A t \cos \pi J_{AB} t - 2I_y^A I_z^X \sin \omega_A t \sin \pi J_{AX} t$$

La présence ou non de l'impulsion  $\pi$  à la fréquence de l'hétéroatome est déterminante. Si l'impulsion  $\pi$  est présente, l'opérateur  $-2I_y^A I_z^X$  deviendra  $+2I_y^A I_z^X$ . Cet opérateur évoluera, pendant le second délai, pour donner l'opérateur  $I_y^A$ . Si l'impulsion  $\pi$  n'est pas appliquée,  $-2I_y^A I_z^X$  reste inchangé et évolue ensuite en  $-I_y^A$ . Pour un proton lié à un  $^{12}\text{C}$ , l'impulsion  $\pi$  sur le carbone n'a pas d'influence sur l'évolution de l'opérateur  $I_y^A$ . En jouant ainsi sur 2 impulsions  $\pi$  exercées sur le canal de l'hétéroatome, quatre différentes expériences peuvent

être enregistrées. La combinaison linéaire de ces quatre expériences permet d'obtenir les informations souhaitées :

- connectivités intramoléculaires dans la molécule marquée.
- connectivités intramoléculaires dans la molécule non-marquée.
- connectivités intermoléculaires entre la molécule marquée et non-marquée (figure 27 B).

## 5 Les spectromètres RMN utilisés

Les spectres RMN ont été enregistrés sur les deux spectromètres suivants:

Varian Inova 14.13 Tesla (600MHz pour  $^1\text{H}$ ) équipé d'une sonde triple résonance 5 mm ou d'une sonde triple résonance 3 mm. Le logiciel était d'abord VNMR 6.1b qui a ensuite été remplacé par VNMRJ version 2.1b. Le spectromètre est équipé de gradients de champ impulsionnel selon l'axe Z. Les séquences, modifiées et optimisées au sein du laboratoire utilisées proviennent de Biopack qui a récemment été mis à jour.

Bruker Avance 11.75 Tesla (500MHz pour  $^1\text{H}$ ) équipé d'une sonde double résonance 5 mm. Le logiciel utilisé est Xwinnmr version 1.1 couplé avec le logiciel xwinplot pour le tracé des spectres.

Quelques spectres RMN ont été enregistrés sur un spectromètre Varian Inova à 18,8 Tesla à l'IBS de Grenoble.

# Chapitre 3 : Les couplages dipolaires résiduels

## 1 L'Interaction Dipôle-Dipôle Directe

Chaque spin nucléaire est associé à un moment magnétique de spin et génère un champ magnétique autour de lui. Deux spins nucléaires proches dans l'espace interagissent mutuellement l'un avec l'autre car contrairement aux couplages scalaires, l'interaction dipolaire ne nécessite pas de liaison électronique entre les deux atomes. Elle est aussi appelée « interaction dipôle-dipôle directe » car l'interaction dipolaire s'effectue sans intermédiaire. Les couplages dipolaires sont donc aussi bien intramoléculaires qu'intermoléculaires.

L'expression complète de l'interaction dipôle-dipôle entre les spins  $I_j$  et  $I_k$  est donnée par l'Hamiltonien de spin suivant (Levitt 2001) :

$$H_{jk}^{DD} = b_{jk} \left( 3(\hat{\mathbf{I}}_j \cdot \mathbf{e}_{jk})(\hat{\mathbf{I}}_k \cdot \mathbf{e}_{jk}) - \hat{\mathbf{I}}_j \cdot \hat{\mathbf{I}}_k \right)$$

où  $\mathbf{e}_{jk}$  est le vecteur unitaire colinéaire à la droite joignant le noyau  $j$  au noyau  $k$ . L'amplitude dans l'interaction dipolaire est donnée par la constante de couplage dipolaire :

$$b_{jk} = -\frac{\mu_0}{4\pi^2} \frac{\gamma_j \gamma_k \hbar}{r_{jk}^3}$$

où  $\gamma_j$  et  $\gamma_k$  sont les rapports gyromagnétiques des deux spins (en  $\text{rad}\cdot\text{s}^{-1}$ ),  $\mu_0$  la permittivité du vide,  $\hbar = \frac{h}{2\pi}$  avec  $h$  la constante de Planck et  $r_{jk}$  la distance entre les deux spins (en m). Pour la paire de spins  $^{15}\text{N}-^1\text{H}$ , considérant la distance  $r_{1\text{H}-15\text{N}} = 1.02 \text{ \AA}$ , l'amplitude de l'interaction est égale à  $\frac{b_{jk}}{2\pi} = 23 \text{ kHz}$ .

Lorsque les spins sont soumis à une intensité de champ magnétique  $B_0$  intense, seule la composante des spins selon la direction du champ est significative dans l'Hamiltonien de l'interaction dipôle-dipôle directe. Ainsi, dans l'approximation des hauts champs, seule cette composante est prise en compte, les autres sont négligées. Si l'on suppose que le champ  $B_0$  est colinéaire à l'axe  $z$  alors l'Hamiltonien  $H_{DD}$  se réécrit de la façon suivante :

$$H_{jk}^{DD}(\theta_{jk}) = 2d_{jk} \hat{I}_{jz} \hat{I}_{kz}$$

où  $d_{jk}$  vaut :

$$d_{jk} = \frac{1}{2} b_{jk} (3 \cos^2 \theta_{jk} - 1)$$

et  $\theta_{jk}$  est l'angle que fait le vecteur internucléaire avec la direction du champ magnétique c'est-à-dire :

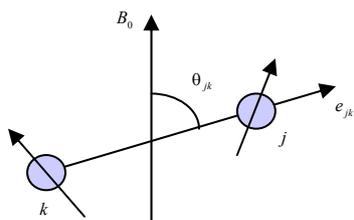


Figure 30 : Représentation des deux spins, de l'orientation du champ magnétique  $B_0$  et de l'angle  $\theta$ .

$$\cos \theta_{jk} = \mathbf{e}_{jk} \cdot \mathbf{e}_z$$

Dans un liquide isotrope, les molécules se réorientent uniformément et avec une vitesse suffisamment grande pour que seule la moyenne des couplages dipolaires intramoléculaires soit mesurable. L'isotropie de la réorientation rend cette moyenne nulle. Aucun couplage dipolaire intramoléculaire n'est donc mesurable dans ces milieux.

$$\langle H_{jk}^{DD}(\theta_{jk}) \rangle = 0$$

De la même façon, les interactions dipôle-dipôle intermoléculaires à courte distance sont moyennées à zéro du fait du mouvement de translation des molécules. Les interactions dipôle-dipôle à plus grande distance ne sont pas moyennées, mais sont trop faibles pour être mesurées. Ainsi, dans un liquide isotrope les couplages dipolaires ne sont pas mesurables. Les raies de résonance d'un spectre RMN en solution sont donc plus fines et donc plus lisibles, contrairement aux spectres RMN du solide où les couplages dipolaires sont si intenses que les spectres sont rapidement illisibles. En contre-partie, nous avons une perte d'information importante en solution. Entre ces deux extrêmes, les liquides anisotropes permettent l'exploitation des couplages dipolaires directs en RMN en solution.

## 2 Couplages Dipolaires Résiduels en milieux orientés

Dans des milieux orientants, les protéines sont partiellement orientées. L'anisotropie induite doit être suffisamment faible pour ne pas perturber le comportement des protéines mais doit

aussi permettre la mesure de couplages dipolaires résiduels (RDCs) (Prestegard, al-Hashimi et al. 2000; Bax, Kontaxis et al. 2001; de Alba and Tjandra 2002).

La molécule sera considérée comme rigide. Les distances ainsi que les orientations en son sein sont supposées constantes et seule son orientation par rapport au champ magnétique  $B_0$  varie.

La valeur du couplage dipolaire entre deux spins  $I_j$  et  $I_k$  de la molécule s'écrit:

$$D_{jk} = b_{jk} \left\langle \frac{3\cos^2 \theta_{jk} - 1}{2} \right\rangle$$

où les crochets représentent la moyenne temporelle ou d'ensemble de la réorientation de la molécule. L'orientation de chaque vecteur internucléaire  $e_{jk}$ , dans n'importe quel repère lié à la molécule, peut être définie par les angles  $\alpha_x$ ,  $\alpha_y$  et  $\alpha_z$  que fait le vecteur avec les axes du système de coordonnées. De la même manière, les angles  $\beta_x$ ,  $\beta_y$  et  $\beta_z$  définissent l'orientation instantanée du champ  $B_0$  dans le repère lié à la molécule.

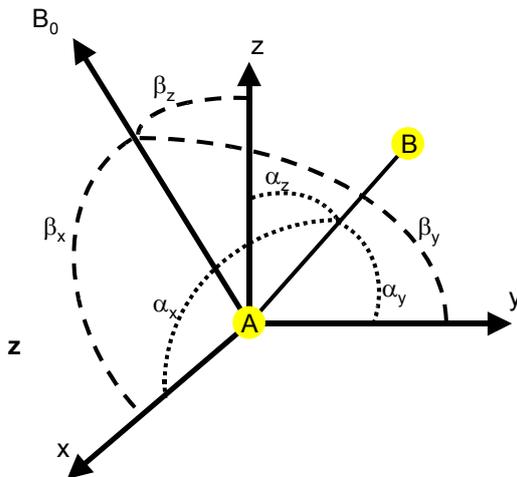


Figure 31 : Orientation du champ magnétique  $B_0$  et d'un vecteur internucléaire  $AB$  dans le repère lié à la molécule.

$\cos\theta$  peut être réécrit sous la forme suivante :

$$\cos\theta = \begin{pmatrix} \cos\alpha_x \\ \cos\alpha_y \\ \cos\alpha_z \end{pmatrix} \cdot \begin{pmatrix} \cos\beta_x \\ \cos\beta_y \\ \cos\beta_z \end{pmatrix} = \cos\alpha_x \cos\beta_x + \cos\alpha_y \cos\beta_y + \cos\alpha_z \cos\beta_z$$

Et alors :

$$\left\langle \frac{3\cos^2\theta_{jk} - 1}{2} \right\rangle = \frac{3}{2} \left\langle \left( \cos\alpha_x \cos\beta_x + \cos\alpha_y \cos\beta_y + \cos\alpha_z \cos\beta_z \right)^2 \right\rangle - \frac{1}{2}$$

Avec  $c_i = \cos\alpha_i$  et  $C_i = \cos\beta_i$ , on obtient :

$$\left\langle \frac{3\cos^2\theta_{jk} - 1}{2} \right\rangle = \frac{3}{2} \left[ \langle C_x^2 \rangle c_x^2 + \langle C_y^2 \rangle c_y^2 + \langle C_z^2 \rangle c_z^2 + 2\langle C_x C_y \rangle c_x c_y + 2\langle C_x C_z \rangle c_x c_z + 2\langle C_y C_z \rangle c_y c_z \right] - \frac{1}{2}$$

Soit S la matrice définie par :  $S_{ij} = \frac{3}{2} \langle C_i C_j \rangle - \frac{1}{2} \delta_{ij}$  où  $\delta_{ij}$  est la fonction de Kronecker. Alors

l'équation ci-dessus peut aussi s'écrire :

$$\left\langle \frac{3\cos^2\theta_{jk} - 1}{2} \right\rangle = \sum_{i,j=\{x,y,z\}} S_{ij} \cos\alpha_i \cos\alpha_j$$

La matrice 3×3 S est appelée matrice de Saupe ou matrice d'ordre. La matrice S est de trace nulle car  $\langle C_x^2 \rangle + \langle C_y^2 \rangle + \langle C_z^2 \rangle = 1$ . De plus, S est aussi symétrique comme  $\langle C_x C_y \rangle = \langle C_y C_x \rangle$ . Elle ne contient donc que cinq termes indépendants. Si la structure de la molécule est connue, c'est-à-dire que les  $\cos\alpha_i$  sont connus, la valeur des cinq termes indépendants de la matrice peut généralement être déterminée à partir de la valeur du couplage dipolaire d'au moins cinq vecteurs internucléaires différents. Dans l'étude d'une macromolécule biologique, beaucoup plus de couplages sont généralement mesurés et S est surdéterminée.

La matrice S est réelle et symétrique, il est donc toujours possible de trouver un repère lié à la molécule dans lequel S est diagonale. Dans la suite, on se placera toujours dans le repère qui diagonalise cette matrice, repère dont on caractérisera la position dans le repère de la molécule par les angles d'Euler ( $\xi_1, \xi_2, \xi_3$ ). L'équation devient alors :

$$D_{jk}(\alpha_x, \alpha_y, \alpha_z) = b_{ij} \left\{ \frac{3}{2} \left[ \langle C_x^2 \rangle c_x^2 + \langle C_y^2 \rangle c_y^2 + \langle C_z^2 \rangle c_z^2 \right] - \frac{1}{2} \right\}$$

où  $\langle C_i^2 \rangle$  correspond à la probabilité de trouver l'axe  $i$  colinéaire au champ magnétique. Seule la différence relative entre les valeurs  $\langle C_i^2 \rangle$  contribue au couplage dipolaire résiduel.

En écrivant  $\langle C_i^2 \rangle = \frac{1}{3} + A_{ii}$  et en coordonnées polaires  $\theta = \alpha_z$  ;  $c_z = \cos\theta$  ;  $c_x = \sin\theta \cos\phi$  ;

$c_y = \sin\theta \sin\phi$  ; l'équation peut être réécrite :

$$D_{jk}(\theta, \phi) = \frac{3}{2} b_{ij} \left[ A_{zz} \cos^2\theta + A_{xx} \sin^2\theta \cos^2\phi + A_{yy} \sin^2\theta \sin^2\phi \right]$$

On suppose que  $|A_{xx}| < |A_{yy}| < |A_{zz}|$ . Il est toujours possible de se ramener à ce cas,

l'hypothèse n'enlève donc rien à la généralité du raisonnement.

En utilisant le fait que :  $A_{xx} + A_{yy} = -A_{zz}$  ;  $2\cos^2\phi = 1 + \cos 2\phi$  ;  $2\sin^2\phi = 1 - \cos 2\phi$  ; la formule se réécrit :

$$D_{jk}(\theta, \phi) = \frac{3}{2} b_{ij} \left[ A_{zz} \left( \frac{3\cos^2\theta - 1}{2} \right) + (A_{xx} - A_{yy}) \frac{1}{2} \sin^2\theta \cos 2\phi \right]$$

Si on définit la composante axiale  $A_a$  et rhombique  $A_r$  du tenseur A comme :

$$A_a = \frac{3}{2} A_{zz} \text{ et } A_r = A_{xx} - A_{yy}$$

l'équation devient :

$$D_{jk}(\theta, \phi) = b_{ij} \left[ A_a \left( \frac{3\cos^2\theta - 1}{2} \right) + \frac{3}{4} A_r \sin^2\theta \cos 2\phi \right]$$

Les couplages dipolaires résiduels sont dégénérés, c'est-à-dire qu'une valeur donnée d'un couplage dipolaire ne correspond pas à une orientation unique du vecteur internucléaire étudié. On lève ces dégénérescences en introduisant une cohérence entre les différentes données en considérant des motifs structuraux particuliers ou en utilisant des mesures relativement à deux tenseurs dont les orientations et les valeurs propres diffèrent.

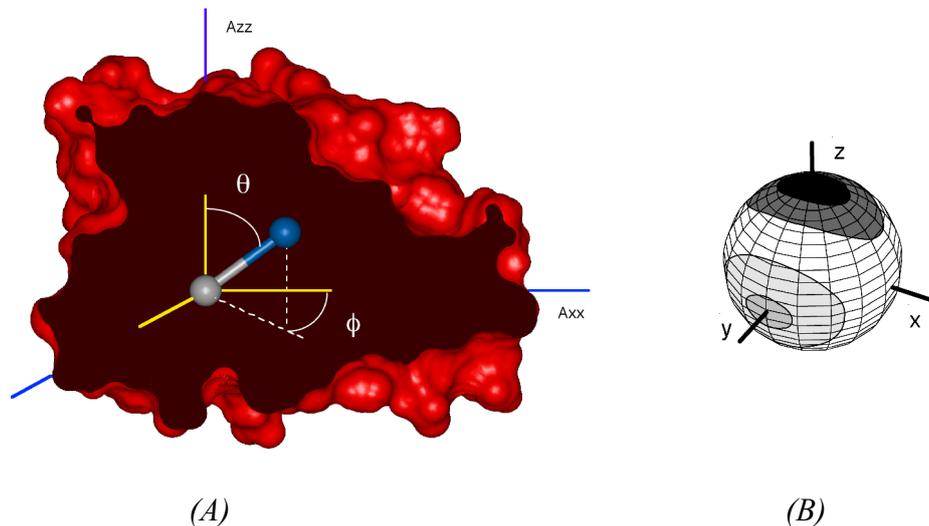


Figure 32 : (A) Coordonnées d'un vecteur internucléaire dans le repère principal du tenseur d'alignement ; (B) La valeur des RDCs va croissant du gris au noir en passant par le blanc.

Le calcul du tenseur d'alignement a pu être réalisé à partir de la structure de la protéine et des couplages dipolaires disponibles à l'aide du logiciel MODULE (Dosset, Hus et al. 2001). La structure apporte les coordonnées des vecteurs internucléaires associées aux valeurs des RDCs. Cinq couplages et leur vecteur internucléaire associé doivent être connus pour

déterminer complètement le tenseur. Dans la pratique, on préfère utiliser le plus grand nombre de couplage afin de minimiser les erreurs. L'algorithme du programme MODULE effectue une régression par la méthode des moindres carrés. Il minimise la fonction  $\chi^2$  suivante (en jouant sur les paramètres de la matrice) :

$$\chi^2(\xi_1, \xi_2, \xi_3, A_a, A_r) = \sum_i \left( \frac{D_i^{\text{exp}} - D_i^{\text{calc}}}{\sigma_i} \right)^2$$

où  $D_i^{\text{exp}}$  est la valeur expérimentale du couplage dipolaire  $i$ ,  $D_i^{\text{calc}}$  est la valeur du couplage dipolaire  $i$  calculée et  $\sigma_i$  représente l'erreur sur la mesure du couplage dipolaire  $i$ . Ainsi, plus l'erreur sur un couplage sera grande moins celui-ci aura de poids dans le  $\chi^2$ . La méthode de minimisation s'appuie sur deux algorithmes : un algorithme de recuit simulé et l'algorithme de Levenberg-Marquardt. Les deux algorithmes travaillent de manière complémentaire : l'algorithme de Levenberg-Marquardt est particulièrement bien adapté à la recherche d'un minimum de l'espace paramétrique et le recuit simulé permet d'identifier les différents minima de ce même espace. L'association des deux est donc particulièrement efficace pour déterminer le minimum de l'espace paramétrique défini par le  $\chi^2$ .

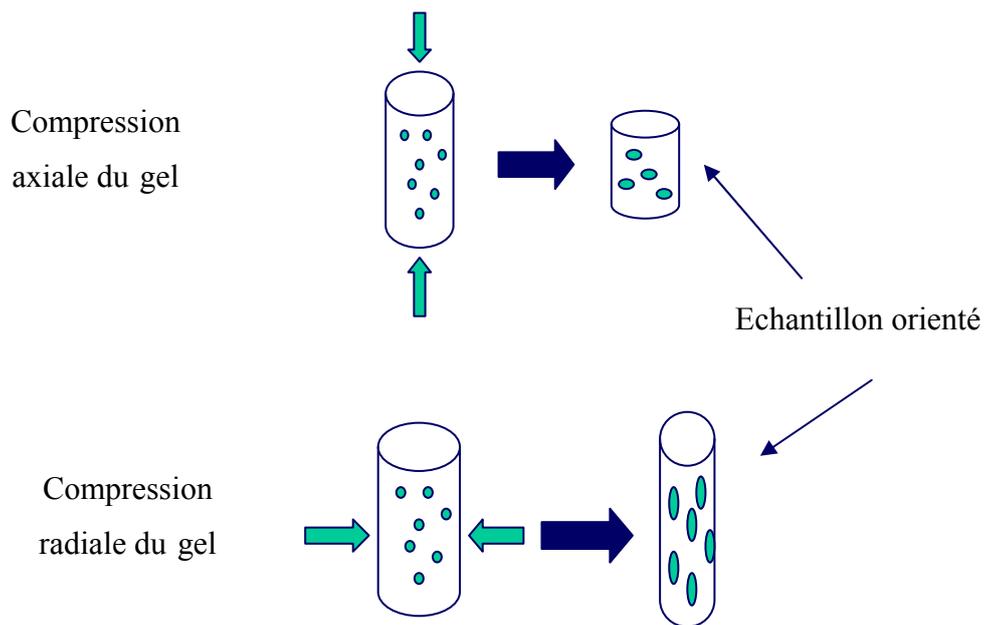
### 3 Les milieux anisotropes utilisés

Il existe beaucoup de milieux orientants (Prestegard and Kishore 2001) qui peuvent être classés selon leur forme, leur mode d'interaction et leur comportement physique. Pour qu'il soit possible d'étudier une macromolécule dans ces milieux, l'ordre imposé à la molécule doit être très petit. Les interactions des macromolécules avec le cristal liquide doivent être faibles et les milieux utilisés doivent être aqueux pour que les protéines soient dans un environnement proche de leur milieu naturel. Nous n'envisagerons que le cas des gels de polyacrylamide, car il s'agit du milieu utilisé pour nos mesures de couplages dipolaires. Néanmoins, il est important de savoir que d'autres milieux sont généralement utilisés comme les cristaux liquides minéraux (ex: pentoxide de Vanadium) qui présentent l'avantage de ne pas posséder de noyaux observables en RMN ( $^1\text{H}$ ,  $^{13}\text{C}$  ou  $^{15}\text{N}$ ) ou organiques comme les bicelles, les phages ou les polyéthylène glycols (PEG).

Les gels de polyacrylamide sont couramment utilisés dans l'électrophorèse pour séparer des protéines ou des oligonucléotides. Les gels de polyacrylamide possède une anisotropie

spatiale résultant de l'application d'une contrainte mécanique ((Sass, Musco et al. 2000; Ackerman and Shortle 2002). Les gels de polyacrylamide permettent l'étude de biomolécules sur une large gamme de température (5-45°) et de pH (2-8,5) et permettent de travailler à de forte force ionique ( $\leq 200\text{mM}$ ). Un point très important est la possibilité de récupérer l'échantillon après les mesures de RDCs en faisant diffuser le soluté dans le solvant.

L'orientation du gel de polyacrylamide se fait par l'intermédiaire d'une contrainte mécanique, soit une compression, soit une élongation verticale ou horizontale du gel.



*Figure 33 : Contraintes mécaniques appliquées au gel permettant l'orientation de l'échantillon*

Le degré d'alignement dépend de trois paramètres : La concentration en acrylamide, la densité de polymérisation et le taux de compression.

# Chapitre 4 : Etude de la dynamique des protéines

## 1 Introduction

L'étude de la dynamique des molécules apporte énormément d'informations sur les molécules, en particulier pour celles d'intérêt biologique. En effet, la connaissance de leur dynamique permet une meilleure compréhension de leurs propriétés, que ce soit lors de la formation d'un complexe, du repliement d'une protéine ou bien pour un phénomène de catalyse enzymatique. Par exemple pour une protéine, la mobilité d'une région peut traduire une adaptabilité conformationnelle qui conduirait à une plus grande facilité à se lier pour former des complexes. La RMN se révèle être un outil de choix pour accéder à ces échelles de temps, notamment par la mesure de la vitesse de relaxation transverse de l'azote  $^{15}\text{N}$  ou par l'utilisation des couplages dipolaires résiduels.

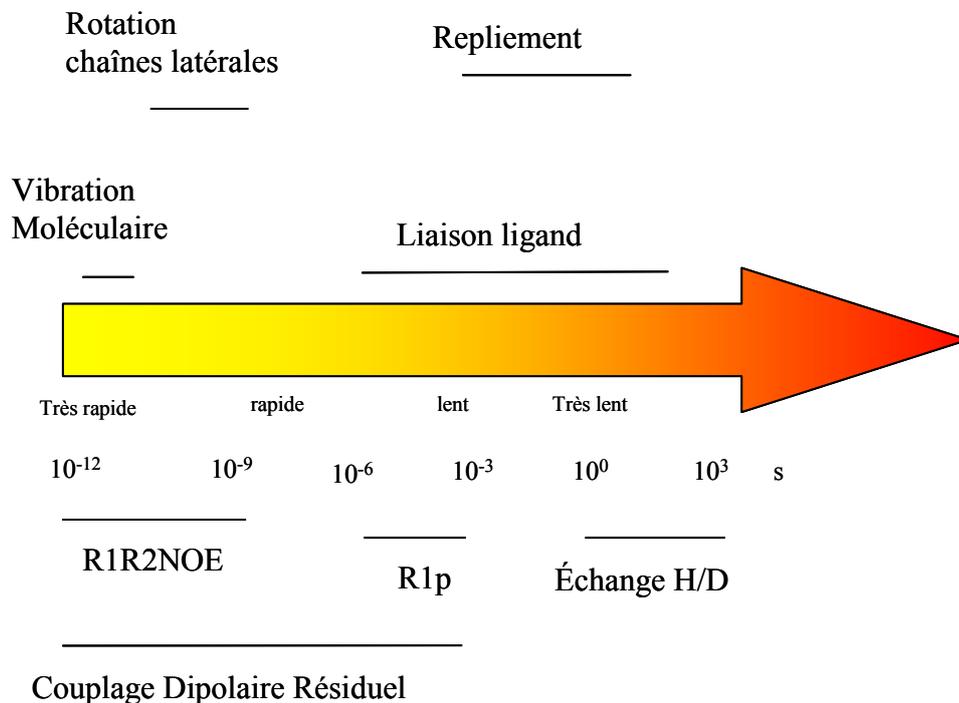


Tableau 2 : Echelle de temps des différents mouvements moléculaires

Les mouvements peuvent être classés en deux familles: lents et rapides. Les mouvements lents sont compris entre des temps de l'ordre de la microseconde à plusieurs milliers de

secondes pour les phénomènes les plus lents. Les phénomènes rapides comme les vibrations moléculaires ou les mouvements des chaînes latérales sont compris entre la picoseconde et la microseconde. L'étude de la relaxation du  $^{15}\text{N}$  permet d'étudier les phénomènes compris entre  $10^{-12}\text{s}$  et  $10^{-9}\text{s}$  alors que l'étude de la dynamique d'une protéine à l'aide des couplages dipolaires résiduels permet d'obtenir des informations sur des temps compris entre  $10^{-12}\text{s}$  et  $10^{-3}\text{s}$ .

La théorie de la relaxation des spins nucléaires a été introduite dans les années 1950-1960 par de nombreux auteurs (Abragam 1961). Sur la base de cette théorie, des techniques expérimentales de RMN ont été mises en œuvre afin de mesurer les paramètres de relaxation. L'agitation thermique animant les protéines fait que celles-ci ne sont pas rigides mais qu'elles présentent des mouvements moléculaires. Les mouvements de faible amplitude, nécessitant peu d'énergie et rapides (pico à nanoseconde), vont pouvoir être étudiés à l'aide de la relaxation nucléaire (Fischer, Majumdar et al. 1998).

## 2 La fonction de densité spectrale

Le vecteur X-H ( $X = ^{15}\text{N}$  ou  $^{13}\text{C}$ ) se réoriente suivant les mouvements de la molécule. Les variations angulaires de ce vecteur sont décrites par la fonction d'autocorrélation  $C(t)$ . Cette fonction représente la probabilité pour le vecteur X-H de retrouver son orientation originelle après un intervalle de temps donné. Si on considère une protéine comme une sphère rigide adoptant un mouvement brownien isotrope, la fonction d'autocorrélation  $C(t)$  peut être associée à une fonction monoexponentielle décroissante (Abragam 1961):

$$C(t) = \frac{1}{5} \left[ e^{-\left(\frac{t}{\tau_c}\right)} \right]$$

où  $\tau_c$  représente le temps de corrélation d'un vecteur X-H. Ce temps correspond à la durée nécessaire au vecteur pour tourner d'un angle de 1 radian.

La transformée de Fourier de la fonction d'autocorrélation  $C(t)$  est la fonction de densité spectrale  $J(\omega)$  qui fournit une analyse en fréquence des mouvements des vecteurs X-H. Elle caractérise ainsi la « quantité » de mouvements à une fréquence définie pour un vecteur donné. Dans le cas d'une sphère rigide en mouvement brownien, la densité spectrale est décrite par une fonction Lorentzienne :

$$J(\omega) = \frac{2}{5} \left[ \frac{\tau_c}{1 + (\omega\tau_c)^2} \right]$$

où  $\omega$  correspond à la fréquence de Larmor du noyau considéré  $^{15}\text{N}$  ou  $^{13}\text{C}$ .

### 3 La relaxation nucléaire

Lorsque les spins nucléaires sont soumis à un champ magnétique  $B_0$ , ils se répartissent sur différents niveaux d'énergie, le système est alors à l'équilibre thermodynamique. Après une excitation provoquée par un champ magnétique  $B_1$ , le système est déplacé hors de sa position d'équilibre. Le retour à la position d'équilibre se réalise grâce à la relaxation nucléaire. De nombreux mécanismes sont à l'origine de la relaxation nucléaire, les deux plus importants étant l'interaction dipôle-dipôle (DD) et l'anisotropie de déplacement chimique (CSA). Les deux phénomènes se traduisent par l'existence de champs magnétiques locaux autour des noyaux de la protéine. Ces champs magnétiques fluctuent à cause des mouvements globaux et internes de la protéine. Parmi tous ces mouvements de la protéine, certains présentent des fréquences qui correspondent à celles des transitions des spins nucléaires. Plus ils sont nombreux, plus la relaxation est rapide. Les mouvements rapides avec une fréquence moyenne de 100MHz sont à l'origine de l'efficacité de la relaxation du système.

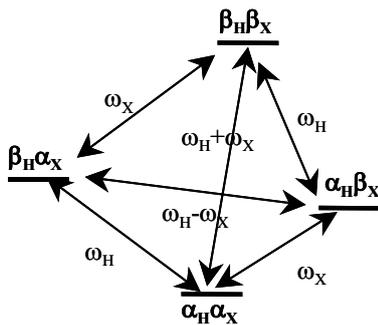


Figure 34: Niveaux d'énergie et transitions correspondantes pour un couple de spins XH.

Les vitesses de relaxation d'un système de spins couplés dépendent donc des valeurs adoptées par la fonction de densité spectrale  $J(\omega)$  aux fréquences de transition décrites dans la figure ci-dessus, c'est à dire :  $J(\omega_X)$ ,  $J(\omega_H + \omega_X)$ ,  $J(\omega_H)$ ,  $J(\omega_H - \omega_X)$ .

### 3.1 La relaxation longitudinale

Souvent appelée relaxation de type Spin-réseau, celle-ci est liée aux échanges d'énergie entre les spins et leur environnement. Elle décrit le retour à l'équilibre de l'aimantation  $M_z$  le long de l'axe  $z$ , associé au passage des spins de l'état  $\beta$  à l'état  $\alpha$  (et réciproquement). La vitesse de relaxation sera notée  $R_1(X_z)$  où  $X_z$  désigne l'aimantation longitudinale du noyau  $X$ . Cette relaxation est favorisée par des fluctuations de champ magnétique à haute fréquence ( $\omega \approx \omega_0$ ). Seules celles qui fluctuent à des fréquences de l'ordre de  $\omega_0$  sont notables dans le repère tournant. Le temps de relaxation longitudinale  $T_1$  est donc minimum pour  $\tau_c = 1/\omega_0$ .

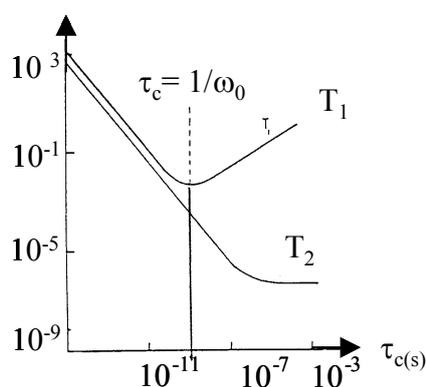


Figure 35: Variations des temps de relaxation  $T_1$  et  $T_2$  en fonction du temps de corrélation  $\tau_c$

### 3.2 La relaxation transversale

Elle décrit le retour à zéro de l'aimantation transversale  $M_t$  située dans le plan  $xoy$  perpendiculairement au champ magnétique  $B_0$ . La vitesse de relaxation est notée  $R_2(X_{x,y})$ . Elle correspond au déphasage des aimantations nucléaires sous l'effet des fluctuations du champ local à toutes les fréquences : les mouvements « lents » (devant les fréquences de transition RMN) y contribuent cette fois aussi bien que les mouvements haute fréquence.  $T_2$  est inférieur à  $T_1$  puisqu'un plus grand nombre de phénomènes contribuent au  $T_2$ . La courbe décrivant les variations de  $T_2$  en fonction de  $\tau_c$  est décroissante, indiquant que la relaxation est d'autant plus efficace que les mouvements sont lents.

### 3.3 La relaxation croisée ; l'effet NOE $\eta$

La relaxation croisée est une interaction dipôle-dipôle à travers l'espace. La relaxation croisée résulte en un gain ou une perte de polarisation du noyau  $^{15}\text{N}$  ou  $^{13}\text{C}$  suite à la modification de polarisation du proton avec lequel il interagit. L'expérience donne accès au gain ou à la perte

d'aimantation par mesure de l'effet NOE qui dépend de  $R(X_Z)$  et de la vitesse de relaxation croisée  $R(H_Z \rightarrow X_Z)$ . Le facteur NOE  $\eta$  s'écrit de la manière suivante :

$$\eta = \frac{\gamma_H}{\gamma_X} \times \frac{R(H_Z \rightarrow X_Z)}{R(X_Z)}$$

où  $\gamma_H$  et  $\gamma_X$  sont les rapports gyromagnétiques respectifs du proton et de l'hétéroatome  $^{15}\text{N}$  ou  $^{13}\text{C}$ .

## 4 Paramètres expérimentaux mesurés

La relaxation longitudinale ( $R_{1(X_Z)} = 1/T_1$ ), la relaxation transversale ( $R_{2(X_{x,y})} = 1/T_2$ ) et le NOE hétéronucléaire ( $NOE = 1 + \eta$ ) à l'état stationnaire ont été mesurés. Ces trois paramètres dépendent de la fonction de densité spectrale  $J(\omega)$  de la façon suivante :

$$R_1 = R(X_Z) = \frac{d^2}{4} \{J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\} + c^2 J(\omega_N)$$

$$R_2 = R(X_{x,y}) = \frac{d^2}{8} \{4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\} + \frac{c^2}{6} \{3J(\omega_N) + 4J(0)\} + R_{ex}$$

$$NOE = 1 + \left( \frac{\gamma_H}{\gamma_N} \right) \times d^2 \{6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\} \times \frac{d^2}{R_1}$$

$$c^2 = \left( \frac{\omega_N (\sigma_{//} - \sigma_{\perp})}{\sqrt{3}} \right)^2$$

$$d^2 = \left( \frac{\mu_0 h \gamma_H \gamma_N}{8\pi^2 r_{NH}^3} \right)^2$$

Le terme  $d^2$  représente la constante d'interaction dipolaire et  $c^2$  la constante d'anisotropie de déplacement chimique. Les fréquences  $\omega_H$ ,  $\omega_N$  représentent respectivement les fréquences de résonances des noyaux  $^1\text{H}$  et  $^{15}\text{N}$ ,  $\gamma_H$  et  $\gamma_N$  sont leurs rapports gyromagnétiques avec  $\gamma_H = 26,756 \times 10^7 \text{ rad.s}^{-1}.\text{T}^{-1}$  et  $\gamma_N = -2,711 \times 10^7 \text{ rad.s}^{-1}.\text{T}^{-1}$ ,  $h$  est la constante de Planck dont la valeur est  $6,626 \times 10^{-34} \text{ J.s}$ ,  $r_{NH}$  est la distance interatomique de  $1,02 \text{ \AA}$  entre  $^1\text{H}$  et  $^{15}\text{N}$  et enfin  $\sigma_{//}$  et  $\sigma_{\perp}$  sont les composantes parallèles et perpendiculaires du tenseur de déplacement chimique de  $^{15}\text{N}$ . La quantité  $\sigma_{//} - \sigma_{\perp}$  qui constitue l'anisotropie de déplacement chimique de  $^{15}\text{N}$  est proche de  $-170 \text{ ppm}$ . Le terme  $R_{ex}$  permet de tenir compte d'éventuels phénomènes d'échange au sein de la protéine.

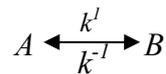
## 5 Le phénomène d'échange

La présence d'échange chimique pour un résidu particulier se traduit expérimentalement par une valeur de  $R_2$  élevée par rapport à la valeur moyenne mesurée. Le paramètre de relaxation transversale mesuré expérimentalement  $R_2^{\text{exp}}$  est en fait la somme de la vitesse de relaxation transversale réelle  $R_2^{\text{réel}}$  et du terme d'échange  $R_{\text{ex}}$  :

$$R_2^{\text{exp}} = R_2^{\text{réel}} + R_{\text{ex}}$$

$$\text{avec } R_{\text{ex}} = \frac{4\pi^2 \Delta\nu^2 P_A P_B \tau_{\text{ex}}}{1 + \omega^2 \tau_{\text{ex}}^2}$$

dans le cas d'un échange chimique entre deux états  $A$  et  $B$  de populations  $P_A$  et  $P_B$  dont les fréquences sont  $\nu_A$  et  $\nu_B$ .  $\Delta\nu$  correspond à la différence de déplacement chimique ( $\Delta\nu = \nu_A - \nu_B$ ).  $\tau_{\text{ex}}$  est égale à l'inverse de la somme des constantes de vitesse d'échange entre ces deux états  $A$  et  $B$ ,  $k_1$  et  $k_{-1}$ ,  $\tau_{\text{ex}} = (k_1 + k_{-1})^{-1}$  :



## 6 Mesure de la relaxation longitudinale $R_1$ et transversale $R_2$

Les séquences d'impulsions pour mesurer les vitesses de relaxation longitudinale  $R_1$  et transversale  $R_2$  sont basées sur la séquence HSQC. Au cœur de la séquence se trouve un délai  $tr$  de relaxation, avec une série d'impulsions qui permettent de conserver le simple processus de relaxation en supprimant les effets de relaxation croisée dipôle-dipôle et d'anisotropie de déplacement chimique. Une série d'expériences est enregistrée en faisant varier le délai  $tr$  afin d'échantillonner au mieux la courbe de relaxation. On obtient donc une série de spectres de type  $^1\text{H}-^{15}\text{N}$  HSQC pour les différentes valeurs de  $tr$ . L'intensité des pics croisés décroît de manière exponentielle et tend vers zéro quand les temps de relaxation tendent vers l'infini. Les valeurs de  $R_1$  et  $R_2$  sont déterminées en réalisant un ajustement de l'intensité des pics croisés à l'aide d'une fonction monoexponentielle décroissante :

$$I(tr) = I_0 e^{(-R \times tr)}$$

où  $I(tr)$  est l'intensité du pic croisé au délai de relaxation  $tr$  et  $I_0$  l'intensité au temps  $tr = 0$  et  $R$  correspond à la constante de vitesse de relaxation  $R_1$  ou  $R_2$ . L'optimisation des valeurs de  $I_0$  et  $R$  est obtenue en minimisant la fonction cible  $\chi^2$  dont l'expression est la suivante :

$$\chi^2 = \sum_{i=1}^N \frac{\left( I(t_R)_{calc} - I(tr)_{exp} \right)^2}{\sigma_i^2}$$

où  $I(t_R)_{calc}$  et  $I(tr)_{exp}$  sont respectivement les intensités des pics calculées et mesurées expérimentalement, et  $\sigma_i$  correspond à l'incertitude de mesure sur l'intensité du pic. Plus la valeur du  $\chi^2$  est faible plus la correspondance entre le modèle et les valeurs expérimentales est correcte. Les incertitudes sur les valeurs des paramètres de relaxation  $R_1$  et  $R_2$  sont évaluées par la méthode de Monte Carlo. Les intensités calculées sont perturbées aléatoirement selon une distribution normale avec un écart type correspondant à celui de l'erreur sur la mesure.

## 7 Mesure de l'effet NOE

La mesure de l'effet NOE nécessite la réalisation de deux expériences : une avec irradiation des  $^1\text{H}$  et l'autre sans. Cette irradiation est assurée par un train d'impulsions qui permet de saturer les protons. Lorsque les protons sont saturés, le système se trouve à l'état stationnaire. Il relaxe alors vers l'équilibre thermique aux transitions zéro quantum et double quanta avec un transfert de polarisation du proton vers le noyau  $^{15}\text{N}$ . On obtient deux spectres de type  $^1\text{H}$ - $^{15}\text{N}$  HSQC correspondant aux deux expériences avec et sans irradiation. Il faut noter que la sensibilité de mesure de l'effet NOE est moins importante que celle des paramètres de relaxation  $R_1$  et  $R_2$  car on ne peut bénéficier du premier transfert de polarisation du  $^1\text{H}$  vers  $^{15}\text{N}$  réalisé par la première série INEPT.

Les valeurs du NOE hétéronucléaire  $^1\text{H} - ^{15}\text{N}$  sont déterminées par le rapport suivant :

$$NOE = \frac{I_{irr}}{I_{eq}}$$

où  $I_{irr}$  et  $I_{eq}$  correspondent respectivement aux intensités des pics croisés dans l'expérience avec et sans irradiation du proton. L'incertitude sur la mesure du NOE est estimée à  $\pm 10\%$  de l'intensité du pic mesuré.

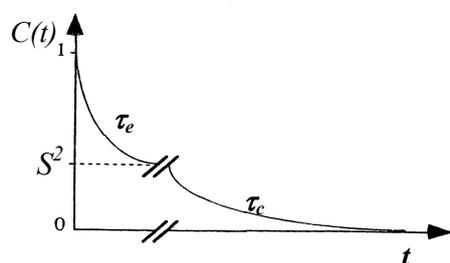
## 8 Interprétation des paramètres de relaxation en terme de mouvement

La découverte de la RMN hétéronucléaire a rendu possible la mesure des temps de relaxation des noyaux hétéronucléaires. De nombreuses approches théoriques ont été réalisées pour

modéliser la fonction de densité spectrale à partir des constantes de relaxation  $R_1$  et  $R_2$  et du  $NOE$ . Le modèle le plus couramment utilisé dans l'interprétation des paramètres de relaxation en terme de mouvements est l'approche « model-free » introduite par Lipari et Szabo en 1982 (Lipari and Szabo 1982a; Lipari and Szabo 1982b).

### *Le formalisme de Lipari et Szabo*

On considère deux types de mouvements présents dans la molécule ; un mouvement de rotation globale de la molécule caractérisé par un temps de corrélation unique  $\tau_c$  et des mouvements internes plus rapides décrits par deux paramètres. La première hypothèse posée par Lipari et Szabo est que ces deux mouvements, globaux et internes, sont indépendants. Ceci est vrai si les deux types de mouvement s'effectuent à des échelles de temps très différentes et il est donc possible de les différencier et de définir un temps de corrélation pour chacun d'eux. L'allure de la fonction de corrélation  $C(t)$  est donnée à la figure 34. Après une décroissance rapide due aux mouvements internes rapides, la courbe atteint un plateau. A l'échelle de la nanoseconde, le mouvement global de la molécule provoque la décroissance jusqu'à 0 de la fonction de corrélation.



*Figure 36 : Evolution schématique de la fonction de corrélation au cours du temps.*

La fonction de corrélation se met alors sous la forme du produit de deux fonctions d'autocorrélation, l'une correspondante au mouvement globale de la molécule, l'autre représentant les mouvements internes :

$$C(t) = C_c(t) \times C_i(t)$$

La seconde hypothèse de Lipari Szabo est que le mouvement global est isotrope. Un seul temps de corrélation intervient et la fonction d'autocorrélation de ce mouvement est une fonction monoexponentielle représentée par :

$$C_c(t) = \frac{1}{5} \times e^{\left(\frac{-t}{\tau_c}\right)}$$

Lorsqu'un vecteur H-X a des mouvements totalement libres dans une sphère de rayon rHX, son mouvement est dit isotrope (c'est-à-dire que toutes les directions sont équiprobables). Mais dans une molécule il existe des contraintes stériques entre les atomes qui vont faire que les mouvements du vecteur internucléaire vont être restreints. Le paramètre spatial, encore appelé paramètre d'ordre ou  $S^2$ , va décrire cette restriction. Le second paramètre est le temps de corrélation effectif ou  $\tau_e$  des mouvements internes.  $\tau_e$  est véritablement un temps de corrélation lorsque la fonction de corrélation des mouvements internes est monoexponentielle, sinon il représente l'aire située sous la fonction d'autocorrélation. La fonction d'autocorrélation des mouvements internes est :

$$C_i(t) = S^2 + (1 - S^2) \times e^{\left(\frac{-t}{\tau_e}\right)}$$

Cette fonction prend une valeur comprise entre 0 et 1 qui dépend du degré de restriction spatiale. Lorsque les mouvements sont très contraints  $S^2$  est proche de 1, alors que pour des mouvements totalement isotropes,  $S^2$  tend vers 0.

La fonction de corrélation totale est donc :

$$C(t) = \frac{1}{5} \left\{ S^2 e^{\left(\frac{-t}{\tau_c}\right)} + (1 - S^2) \times e^{\left(\frac{-t}{\tau_i}\right)} \right\}$$

avec

$$\frac{1}{\tau_i} = \frac{1}{\tau_c} + \frac{1}{\tau_e}$$

La transformée de Fourier ( $J(\omega) = 2 \int_0^{+\infty} C(t) \cos(\omega t) dt$ ) de cette fonction d'autocorrélation est

la fonction de densité spectrale  $J(\omega)$  suivante.

$$J(\omega) = \frac{2}{5} \left\{ \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{(1 - S^2) \tau_i}{1 + (\omega \tau_i)^2} \right\}$$

Cette fonction est la somme de deux lorentziennes. L'une dépend du temps de corrélation  $\tau_c$  et l'autre du temps de corrélation  $\tau_i$ . Si les mouvements internes sont très rapides devant le mouvement global, alors  $\tau_e \ll \tau_c$  et donc  $\tau_i \approx \tau_e$ . Le deuxième terme dépend donc des mouvements internes.

D'autre part, le modèle de Lipari Szabo se simplifie si on considère que les mouvements internes sont très rapides avec  $\tau_e$  inférieur à quelques picosecondes. L'expression de la fonction de densité spectrale  $J(\omega)$  devient alors égale à :

$$J(\omega) = \frac{2}{5} \left\{ \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} \right\}$$

Les mouvements des vecteurs N-H ne sont plus dépendants que du mouvement global de la protéine caractérisé par le temps de corrélation  $\tau_c$ .

Malheureusement, ce formalisme souffre de plusieurs limites. Tout d'abord l'hypothèse d'un mouvement isotrope global de la molécule implique que celle-ci soit sphérique ou globulaire. D'autre part, ce formalisme n'est pas applicable lorsque les mouvements internes ont des temps caractéristiques de l'ordre de ceux du mouvement global. Or ces mouvements peuvent avoir une grande importance d'un point de vue biologique. En 1990, Clore (Clore, Szabo et al. 1990) et ses collaborateurs ont proposé d'introduire une troisième lorentzienne et un deuxième paramètre d'ordre rendant compte de ces mouvements intermédiaires. Ce nouveau modèle porte le nom de modèle de Lipari Szabo étendu. La détermination des paramètres de dynamique peut être réalisée à l'aide du programme Model-free (Palmer III, Rance et al. 1991; Mandel, Akke et al. 1995) en interface graphique avec le programme Fast-Modelfree (Cole and Loria 2003).

## 9 Détermination du temps de corrélation globale de la protéine $\tau_c$

Une estimation du temps de corrélation globale de la protéine  $\tau_c$  peut être obtenue à l'aide des rapports  $R_2/R_1$  des résidus dont les vecteur N-H présentent une mobilité interne réduite et pas de phénomène d'échange (ce qui pourrait conduire à surestimer la valeur de  $R_2$ ). Ces résidus sont généralement impliqués dans des structures secondaires. Pour ces résidus, les mouvements internes sont négligeables et le terme  $(1-S^2)$  tend vers 0. La fonction de densité spectrale est alors :

$$J(\omega) = \frac{2}{5} \left[ \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} \right]$$

Le rapport  $R_2/R_1$  devient indépendant de  $S^2$  et permet de déterminer le temps de corrélation  $\tau_c$  (Kay, Torchia et al. 1989; Barbato, Ikura et al. 1992).

# Chapitre 5 : La Modélisation moléculaire sous contraintes RMN

La RMN est une technique de détermination structurale indirecte, c'est-à-dire qu'elle ne permet pas d'obtenir directement la structure tridimensionnelle d'une macromolécule mais qu'elle permet d'obtenir un ensemble de contraintes expérimentales qui permettront de calculer la structure 3D de la macromolécule.

Le but de la modélisation va donc être d'obtenir un ensemble de structures, ou de jeux de coordonnées, répondant aux données RMN et correspondant à un minimum énergétique, tout en sachant que le nombre de contraintes expérimentales est bien inférieur aux  $3N-6$  degrés de liberté du système (3 coordonnées de translation globale et 3 coordonnées de rotation de la molécule).

La méthode de calcul par recuit simulé (simulated annealing) est à l'heure actuelle la meilleure pour explorer l'espace conformationnel (Guntert, Mumenthaler et al. 1997). Il s'agit d'une méthode de dynamique moléculaire à haute température permettant le franchissement de barrières énergétiques importantes. Cette dynamique peut être effectuée dans l'espace des coordonnées cartésiennes ou dans l'espace des angles dièdres et consiste à trouver la solution de l'équation du mouvement de Newton:

$$m_i a = m_i \frac{d^2 r_i}{dt^2} = -\nabla_i E$$

où  $m_i$  représente la masse du système et  $E$  l'énergie potentielle du système.

## 1 L'énergie potentielle

L'énergie potentielle dépend des contraintes expérimentales ainsi que des informations de topologie définies dans le champ de force. Celui-ci regroupe tous les paramètres nécessaires à la fonction d'énergie potentielle. Il contient la liste des constantes de forces, des valeurs des longueurs de liaisons, des angles et des charges atomiques partielles nécessaires à la description de la molécule. Ainsi l'énergie potentielle du système correspond à un ensemble de termes

pouvant être regroupé en trois classes, les termes liés, non-liés et les termes expérimentaux. L'énergie potentielle de notre système sera alors la somme de ces différents termes :

$$E_{Totale} = E_{Liés} + E_{Non-liés} + E_{Experimentales}$$

### Termes liés

#### contribution des atomes liés de manière covalente

##### Energie de liaison covalente

Variation de la longueur des liaisons (extension ou compression)

$$E_{liaison} = \frac{1}{2} \sum_{liaisons} K_B (r_i - r_0)^2$$

$K_B$  : constante de force  
 $r_i$  : longueur de liaison  
 $r_0$  : valeur moyenne

##### Energie de valence

Variation des angles de valence entre trois atomes

$$E_{angles} = \frac{1}{2} \sum_{angles} K_\theta (\theta - \theta_0)^2$$

$K_\theta$  : constante de force pour la liaison considérée  
 $\theta$  : angle de valence entre trois atomes  
 $\theta_0$  : valeur à l'équilibre

##### Energie de torsion

Rotation d'une liaison selon l'angle dièdre formé par quatre atomes

$$E_{dihe} = \frac{1}{2} \sum_{angles\ de\ torsion} K_\varphi [(1 + \cos(n\varphi - \delta))]$$

$K_\varphi$  : constante de force liée à la courbure  
 $n$  : périodicité de la rotation ( $n=2$  pour  $sp^2$ )  
 $\varphi$  : angle de torsion  $\delta$  : phase à l'origine

##### Energie des angles dièdres impropres

Déformation en considérant trois atomes liés à un quatrième atome

$$E_{impr} = \frac{1}{2} \sum_{impropres} K_\xi (\xi - \xi_0)^2$$

$K_\xi$  : constante de force associée à la déformation d'un groupe d'atomes par rapport à une conformation donnée  
 $\xi$  : angle dièdre impropre

### Termes non liés

#### Contribution des atomes non-liés de façon covalente

##### Energie de Van der Waals

Somme des forces attractives et répulsives entre deux atomes :  
Prend la forme d'un potentiel de Lennard-Jones

$$E_{VDW} = \sum_{i,j} \left[ D_{ij} \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - C_{ij} \left( \frac{r_{ij}^0}{r_{ij}} \right)^6 \right]$$

$r_{ij}^0$  : somme des rayons de van der Waals des deux atomes  
 $r_{ij}$  : distance séparant les deux atomes  
 $D_{ij}$  et  $C_{ij}$  : constantes de Van der Waals

##### Energie électrostatiques

Interaction entre les distributions de charges ponctuelles  $q_i$  centrées sur l'atome  $i$  de la molécule

$$E_{ELEC} = \frac{1}{2} \sum_{i,j} \left[ \frac{1}{4\pi\epsilon} \cdot \frac{q_i q_j}{r_{ij}} \right]$$

$\epsilon = \epsilon_0 \times \epsilon_r$   
 $\epsilon_r$  : permittivité relative  
 $\epsilon_0$  : permittivité du vide

##### Energie de liaisons hydrogène

Interaction entre un hydrogène déficient en électron et un atome de forte densité électronique portant un doublet libre

$$E_{HBOND} = \sum_{i,j} \left[ \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{10}} \right]$$

$A$  et  $B$  : constantes  
 $r_{ij}$  : distance entre les deux atomes

### Termes de contraintes expérimentales

##### Energie NOE

Représente les distances interprotons. Elle est représentée sous la forme d'une fonction de potentiel harmonique dont la plus couramment utilisée est la fonction harmonique square-well

$$E_{noe} = \min(\text{ceil}, K) \cdot \Delta$$

$$\text{avec } \Delta = \begin{cases} (R - d_{\text{sup}})^2 & \text{si } R > d_{\text{sup}} \\ 0 & \text{si } d_{\text{inf}} < R < d_{\text{sup}} \\ (R - d_{\text{inf}})^2 & \text{si } R < d_{\text{inf}} \end{cases}$$

$\text{ceil}$  : valeur limite imposée à la fonction  
 $K$  : constante dépendant du type de potentiel utilisé

$$E_{CDIH} = S \sum C \cdot well(\text{modulo}_{2\pi}(\phi - \phi_0), \Delta\phi)^{ed}$$

Energie dièdre

$$\text{avec } well(\phi - \phi_0) = \begin{cases} \phi - (\phi + \Delta\phi_0) & \text{si } \phi > (\phi + \Delta\phi_0) \\ 0 & \text{si } (\phi - \Delta\phi_0) < \phi < (\phi + \Delta\phi_0) \\ \phi - (\phi + \Delta\phi_0) & \text{si } \phi < (\phi - \Delta\phi_0) \end{cases}$$

« well » : potentiel square-well  
S et C: constants d'énergies associées au type de potentiel employé

Energie de plan

Dans le cas d'acides nucléique, ce terme permet d'associer les bases d'une même paire

$$E_{PLAN} = \sum_{g \in \text{groupe}} K_{plan} \sum_{i \in g} g_i^2$$

gi : distance orthogonale de l'atome i par rapport au plan défini par tous les atomes du groupe

Trouver les solutions de l'équation de Newton va donc consister en la minimisation de l'énergie potentielle du système. La méthode du recuit simulé (chauffage intense et brutal puis refroidissement du système moléculaire) permet au système d'acquérir suffisamment d'énergie cinétique pour passer les barrières de potentiels énergétiques, évitant ainsi au système de se retrouver dans un minimum énergétique local. Durant l'étape à haute température, une structure est calculée puis est minimisée lorsque la température diminue. La température détermine la barrière de potentielle maximum pouvant être surmontée, si bien que les protocoles de recuit simulé sont très complexes car ils doivent permettre d'éviter les minimums locaux.

## 2 La fonction cible

L'ensemble des logiciels développé par l'équipe de Wüthrich puis P. Günter (Guntert and Wuthrich 1991; Guntert, Mumenthaler et al. 1997; Guntert 2004) utilise la notion de fonction cible V plutôt que celle d'énergie potentielle :

$$V = \sum_{c=u,l,v} w_c \sum_{(\alpha,\beta) \in l_c} (d_{\alpha\beta} - b_{\alpha\beta})^2 + w_a \sum_{i \in l_a} \left[ 1 - \frac{1}{2} \left( \frac{\Delta_i}{\Gamma_i} \right)^2 \right] \Delta_i^2$$

Si la contrainte de distance est vérifiée entre les atomes  $\alpha$  et  $\beta$  séparés d'une distance  $d_{\alpha\beta}$ , alors la contribution est nulle, sinon l'expression est calculée.  $b_{\alpha\beta}$  correspond alors soit aux contraintes supérieures (u), inférieures (l), ou de van der Waals (v) s'il n'y a pas de contrainte expérimentale.  $\Delta_i$  représente la violation de contrainte angulaire et  $\Gamma_i = \pi - (\phi_{\max} - \phi_{\min})/2$ , demi-intervalle de valeurs interdites pour un angle de torsion  $\phi_i$ . Les facteurs w sont des facteurs de pondération et V s'exprimera en  $\text{\AA}^2$ .

Afin d'éviter les mauvais contacts stériques, pour les couples d'atomes pour lesquels il n'y a pas de données NOE, on choisit  $d_{\min} = D_i$ , somme des rayons de van der Waals des 2 atomes et  $d_{\max}$  est fixé arbitrairement à 999 Å.

### 3 Le logiciel ARIA (Ambiguous Restraints for Iterative Assignment)

ARIA (Linge, Habeck et al. 2003) est un logiciel d'attribution semi-automatique des pics NOEs et de calcul de structures fonctionnant en interface avec le logiciel CNS développé par Brünger (Brünger, Adams et al. 1998). Le concept le plus important introduit par le logiciel ARIA concerne l'utilisation de contraintes de distances ambiguës (ambiguous distance restraint ou ADR). Ce concept permet de prendre en compte toutes les attributions possibles pour un pic de corrélation. Le logiciel ARIA a été développé dans le but de réduire de façon itérative le nombre de paires contribuant à chaque NOE observé, en éliminant les paires ne contribuant pas physiquement à l'intensité. De plus l'utilisation des contraintes de distance ambiguë permet d'éliminer le problème de pseudo-atomes car les protons équivalents (méthyles, protons aromatiques) sont traités comme des NOEs ambigus.

#### 3.1 Contraintes de distance ambiguë

Avec la notion d'ADR, chaque pic NOE est traité comme la superposition des signaux de corrélation de plusieurs paires de protons dont les fréquences de résonance sont compatibles avec la position du pic. Un pic NOE avec une attribution multiple ( $n > 1$ ,  $n$  : nombre d'attributions) est considéré comme la somme de  $n$  pics NOEs superposés. Ce pic est alors converti en contrainte de distance ambiguë  $\bar{d}$  d'après l'équation suivante :

$$\bar{d} = \left( \sum_{k=1}^n d_k^{-6} \right)^{\frac{1}{6}}$$

où chaque distance  $d_k$  représente la distance entre une paire  $k$  de protons et  $n$  représente le nombre total d'attributions possibles. La valeur de  $\bar{d}$  est toujours inférieure à  $d_k$ , quelle que

soit  $k$ , une mauvaise attribution est sans grande influence, aussi longtemps que la bonne attribution est présente ; c'est la notion de contribution  $C_k$ , avec  $1 \leq k \leq n$ .

### 3.2 Calibration des pics NOEs

Aria prend comme constante de calibration  $C=1/K$ , où  $K$  est la constante qui intervient dans la relation  $\eta_i=K/d_i^6$  reliant l'intensité NOE  $\eta_i$  à la distance  $d_i$  entre les deux atomes.  $C$  est déterminée de façon simplifiée par l'expression :

$$C = \frac{1}{K} = \frac{1}{n} \sum_{NOEs} \frac{d_{obs}^{-6}}{\eta_i}$$

avec  $n$  le nombre total d'attributions.

Pour chaque paire de protons correspondant à un pic NOE de volume  $\eta_i$  on calcule une distance moyenne  $\langle d_i \rangle$  sur l'ensemble des structures générées. On en déduit les contraintes de distance observées ( $d_i^{obs}$ ) :

$$d_i^{obs} = \frac{1}{(C \times \eta_i)^{1/6}}$$

L'erreur sur l'évaluation des contraintes de distance ( $\Delta$ ) est estimée à 12,5 % du carré de la distance observée  $d_i^{obs}$ . On définit donc un intervalle de distance faisant intervenir les limites  $d_{min}$  et  $d_{max}$ :

$$d_{min} = d_i^{obs} - \varepsilon(d_i^{obs})^2$$

$$d_{max} = d_i^{obs} + \varepsilon(d_i^{obs})^2$$

Avec  $\varepsilon=0,125 \text{ \AA}^{-1}$

### 3.3 Attribution des pics NOEs ambigus

Pour chaque pic NOE ambigu, c'est-à-dire dont une coordonnée de déplacement chimique correspond à la fréquence de résonance de plusieurs protons, une contrainte de distance ambiguë est créée. Le pic NOE est donc la somme de  $n$  contributions  $C_k$  correspondant aux différentes attributions possibles.

Chaque contribution  $C_k$  est calculée de la manière suivante :

$$C_k = \frac{d_k^{-6}}{\sum_{i=1}^n d_i^{-6}}$$

avec  $0 < C_k < 1$  et  $\sum C_k = 1$

où  $d_k$  est la distance moyenne de la paire de protons  $k$  calculée sur les 7 structures de plus basse énergie de l'itération précédente. On classe ensuite les contributions  $C_k$  par ordre décroissant et on exclut les contributions  $C_k$  de faible poids dès lors que :

$$\sum_{i=1}^{n_p} C_k > p$$

où  $p$  constitue la valeur du seuil d'attribution et  $n_p$  est le nombre de contributions  $C_k$  participant majoritairement à la définition du pic NOE.

La valeur du seuil d'attribution est proche de 1 à la première itération et décroît progressivement au fur et à mesure des itérations pour atteindre la valeur de 0,8 à l'itération 8.

De plus, au fur et à mesure des itérations, le traitement ambigu des pics NOEs permet d'exclure de la modélisation des pics dont les contraintes correspondantes sont systématiquement violées sur plus de la moitié des structures analysées. Ces pics exclus sont vérifiés manuellement après chaque cycle itératif et peuvent résulter de situations diverses : problèmes de correction de ligne de base, bruit t1, mauvaise suppression du solvant, diffusion de spin.

L'utilisateur d'ARIA peut régler à volonté tous les paramètres du calcul. Cependant, seuls deux paramètres semblent être réellement dépendants de la protéine : le seuil  $p$  (voir ci-dessus) et la tolérance  $\Delta\delta_{\text{ppm}}$ .

### 3.4 Principe de fonctionnement du logiciel ARIA

L'utilisation simultanée des logiciels ARIA et CNS permet de réaliser l'attribution des pics NOEs et le calcul des structures 3D de manière itérative:

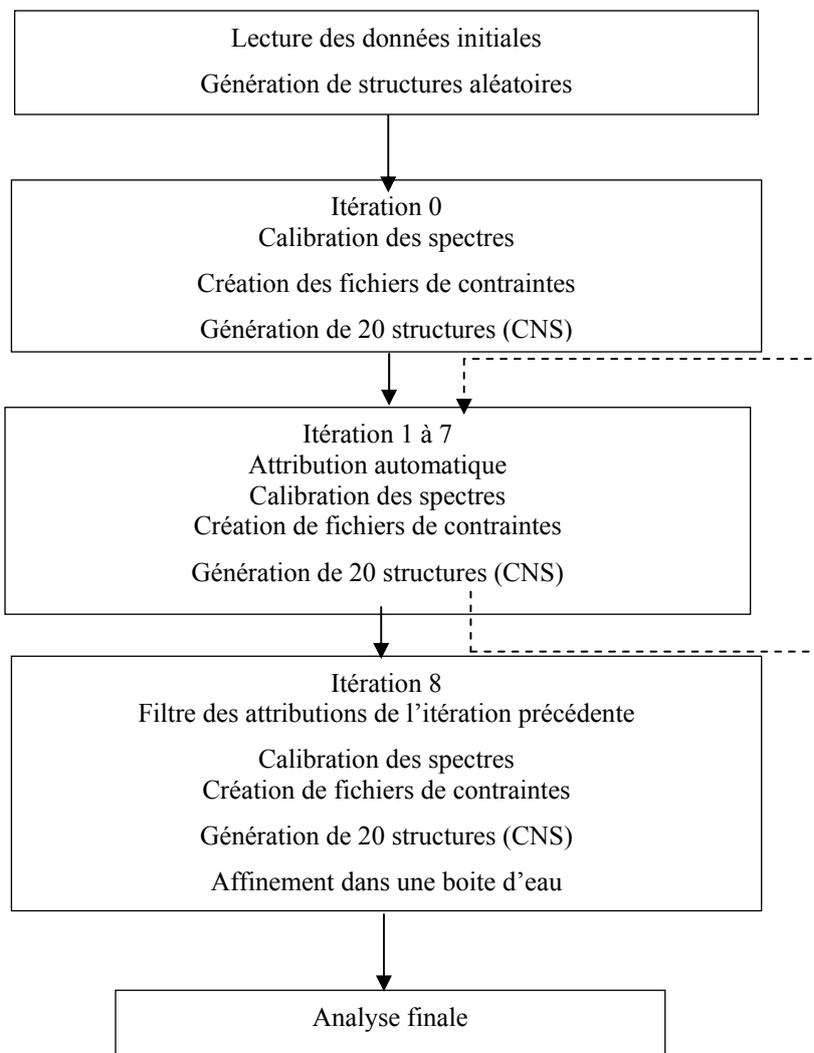


Figure 37: Principe de fonctionnement du logiciel ARIA

Les données initiales comportent les tables de déplacements chimiques et les listes de pics NOEs, éventuellement partiellement attribués. D'autres contraintes comme les angles dièdres, les constantes de couplages J, les RDCs ou encore des liaisons hydrogène peuvent être aussi utilisées. A chaque itération, ARIA calibre, attribue les spectres NOESY et génère des listes de contraintes permettant le calcul d'un ensemble de structures en utilisant le logiciel CNS. Les X structures (nombre défini par l'utilisateur) présentant les énergies les plus basses sont sélectionnées et les distances moyennes de chaque paire de protons sont calculées à partir de ces structures. Les autres itérations sont fondées sur les distances moyennes calculées lors de

l'itération n-1. Enfin l'itération 8 inclut une dynamique dans une boîte d'eau utilisant un champ de force réaliste.

## 4 Le logiciel CYANA

Le logiciel CYANA (Guntert 2004), comme ARIA, permet de réaliser simultanément l'attribution des pics NOEs et le calcul de structures 3D. Ce logiciel est fondé sur la suite de programmes DIANA puis DYANA qui introduit la notion de dynamique dans l'espace des angles de torsion. Les calculs font intervenir les notions de fonction cible et de contraintes ambiguës. CYANA utilise un processus itératif voisin de celui d'ARIA mais n'inclut pas l'étape finale d'affinement dans une boîte d'eau. Le logiciel CYANA est fondé sur l'algorithme CANDID qui combine les particularités des méthodes NOAH et ARIA en apportant un certain nombre de nouveaux concepts comme la notion de « network-anchoring » ou de « constraint combination ».

### 4.1 Principe de l'algorithme CANDID de CYANA

Les données initiales comprennent les tables de déplacements chimiques et les listes des NOEs. A chaque itération, l'attribution est effectuée sur la base des déplacements chimiques en utilisant un intervalle de tolérance et utilise le jeu de structures générées à l'itération précédente (sauf à la 1<sup>ère</sup> itération). Ces attributions sont également pondérées par l'utilisation de la notion de « Network Anchoring » qui vérifie la cohérence des attributions et leur compatibilité avec la géométrie covalente. Un score  $N_{\alpha\beta}$  est calculé pour chaque attribution d'un NOE à une paire d'atomes ( $\alpha$ ,  $\beta$ ), en considérant toutes les connectivités mettant en jeu les atomes  $\alpha$  et  $\beta$  et éventuellement un 3<sup>ème</sup> atome  $\gamma$ . Tous ces NOEs doivent être compatibles sur le plan géométrique et doivent bien sûr être compatibles avec la géométrie covalente.  $N_{\alpha\beta}$  est calculé en utilisant des expressions analytiques assez complexes .

## Network-anchoring

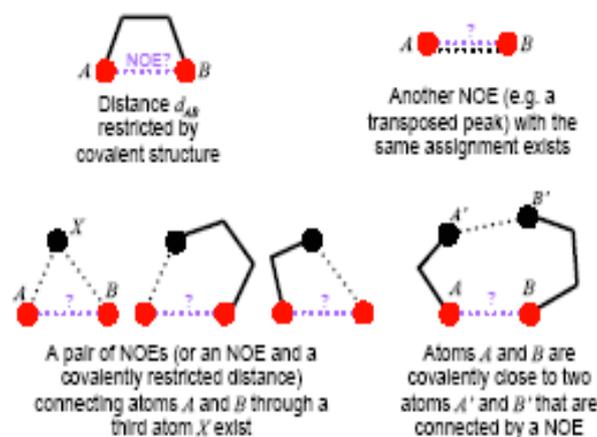


Figure 38 : Chaque attribution doit être confirmée par l'existence d'autres données expérimentales compatibles avec elle et pouvant correspondre à l'une des 4 situations illustrées sur le schéma ci-contre (d'après Guntert, cours EMBO 2003). En violet apparaît la connectivité évaluée.

Trois critères sont pris en compte pour qu'une attribution soit validée. Les attributions ayant un  $N_{\alpha\beta}$  trop faible sont éliminées.

### Conditions for valid NOESY assignments

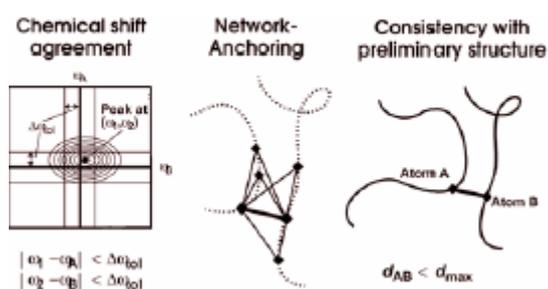
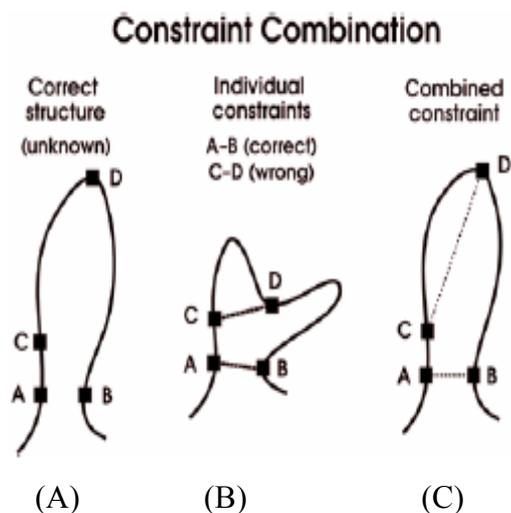


Figure 39: conditions pour qu'une attribution soit prise en compte (d'après Guntert).

Les contraintes ainsi validées sont calibrées et utilisées dans les calculs de structure. Toutefois, des contraintes erronées peuvent provenir de divers artefacts spectraux (bruit, pics en troncature...) confondus avec des pics de corrélation. Ceci peut se produire en particulier pour les NOEs longue distance lorsque aucune structure 3D n'est disponible pour filtrer les attributions. Pour minimiser leur impact, lors des deux premières itérations, un processus de combinaison de contraintes est utilisé. Deux modes sont possibles : dans le premier, deux contraintes longues distances sont remplacées par une seule contrainte ambiguë ; dans le second, 4 contraintes longues distances sont remplacées par 4 contraintes ambiguës

combinant ces contraintes deux à deux. Dans ce cas, la contrainte erronée n'aura pas d'influence sur le calcul de structure.



Combinaison de 2 contraintes (Guntert) : Effet de la contrainte erronée entre C et D ; si elle est prise individuellement (B), la structure est déformée car les contraintes sont satisfaites simultanément ; si la contrainte erronée entre dans une combinaison de contraintes sous forme de contrainte ambiguë (C), la structure est correcte.

Lors de la dernière itération, contrairement à ARIA, un filtre supplémentaire ne conserve que les NOEs ayant une attribution unique à une paire de protons.

Pour la calibration, plusieurs options sont possibles. Elle peut être effectuée manuellement par l'utilisateur ou automatiquement en utilisant ou non un jeu de structures .

Le logiciel ARIA est efficace pour compléter l'attribution des cartes NOESY lorsqu'un repliement initial de la protéine est déjà connu ou lorsque qu'une partie de la carte a été attribué manuellement. L'algorithme CANDID, intégré à CYANA, s'avère extrêmement efficace dans l'attribution *ab initio* des cartes NOESY grâce à l'utilisation du « Network anchoring » et des combinaisons de contraintes. Cette méthode a été évaluée dans des calculs test pour l'attribution *de novo* et la détermination de structures de protéines, et a fourni d'excellents résultats .

Une stratégie intéressante consiste à réaliser un calcul de structures en utilisant le logiciel CYANA et à réaliser ensuite une minimisation dans l'eau en utilisant le logiciel ARIA et les informations obtenues lors du calcul à l'aide du logiciel CYANA.

## 5 Analyse des structures obtenues

Les structures obtenues sont analysées suivant un certain nombre de critères : valeur de la fonction cible, nombre de violation des contraintes, écart quadratique moyen (RMSD), carte

de Ramachandran ... . Un RMSD faible indique certes un ensemble de structures précis mais pas forcément exact. La différence entre exactitude et précision d'un ensemble de structures est décrite dans le schéma ci-dessous:

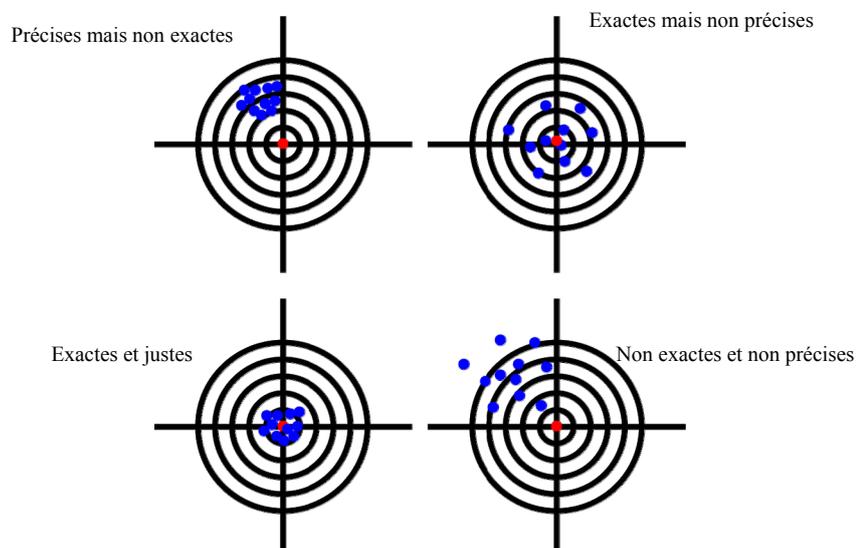


Figure 40 : Exactitude et précision pour un ensemble de structures

Certains critères, énoncés par Günter, permettent d'estimer la validité d'une structure RMN. 90% des protons non-échangeables doivent être attribués. Le RMSD doit être inférieur à 3 lors du premier cycle (pour le squelette peptidique), et moins de 20% des pics longues distances doivent être rejetés lors du calcul de structures.

La carte de Ramachandran représente un diagramme constitué des couples d'angles ( $\psi, \phi$ ). Pour un résidu donné, toutes les combinaisons d'angles ( $\psi, \phi$ ) ne sont pas permises en raison des encombrements stériques défavorables entre les atomes du squelette peptidique et des chaînes latérales, ce qui se traduit sur les cartes de Ramachandran par des zones limitées d'angles ( $\psi, \phi$ ) possibles.

Enfin d'autres logiciels comme PROCHEK ou PROCHECK-AQUA (Laskowski, Rullmann et al. 1996), Promotif (Hutchinson and Thornton 1996), Verify3D (Luthy, Bowie et al. 1992), PROSAIL (Sippl 1993), ou PSVS (Protein Structure Validation Software)(Bhattacharya, Tejero et al. 2007), qui regroupe un ensemble de programmes, permettent de valider les structures.



# RESULTATS



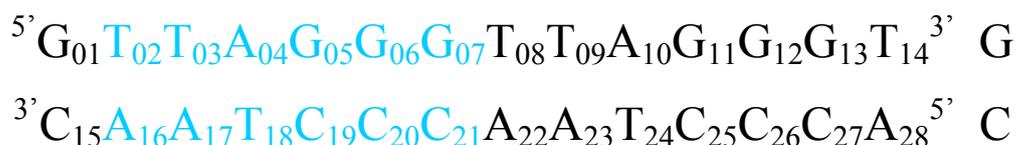
# Chapitre 6 : Etude RMN de l'ADN télomérique

Comme mentionné précédemment, l'ADN télomérique est constitué de la répétition de 6 nucléotides formant la séquence T<sub>2</sub>AG<sub>3</sub>. L'ADN télomérique étant répétitif, nous pouvions nous demander si l'étude par RMN de fragment d'ADN télomérique serait possible, si les recouvrements entre raies de résonance appartenant à des répétitions télomériques différentes n'allaient pas être trop nombreux. L'étude d'un premier fragment d'ADN télomérique comportant deux répétitions télomériques a été réalisée. L'observation de déplacements chimiques particuliers a montré que les deux répétitions étaient différentes. Nous avons alors réalisé l'étude d'un second fragment double brin comportant trois répétitions télomériques afin d'observer si ces différences seraient aussi observables.

## 1 Etude RMN de l'ADN télomérique contenant deux répétitions télomériques

### 1.1 Présentation de la séquence

La séquence utilisée est la suivante :



*Figure 41: Séquence d'ADN télomérique étudiée*

Ce fragment d'ADN, comportant deux répétitions télomériques T<sub>2</sub>AG<sub>3</sub>, est constitué de 28 résidus soit une masse moléculaire d'environ 9 kDa. Celui-ci étant non-marqué, son étude a été réalisée par RMN homonucléaire du proton <sup>1</sup>H.

Dans la suite de cette thèse, le brin riche en guanine sera dénommé « Brin G » alors que le brin riche en cytosine correspondra au « brin C ».

## 1.2 Etude des brins seuls par RMN

Les deux brins seuls proviennent de la société Eurogentec. Ils ont été purifiés par HPLC en phase inverse et par chromatographie échangeuse d'ions et leur masse moléculaire a été vérifiée par spectrométrie de masse. Les différents brins seuls, livrés sous forme lyophilisée, ont été dissous dans un volume approprié d'une solution contenant 90% d'eau légère (H<sub>2</sub>O) et 10% d'eau lourde (D<sub>2</sub>O). Le volume de solvant utilisé pour dissoudre l'échantillon dépend du tube RMN utilisé :

| Tube RMN       | Volume (μl) |
|----------------|-------------|
| 5 mm classique | 600         |
| 5 mm shigemi   | 280         |
| 3 mm classique | 250         |
| 3 mm shigemi   | 120         |

*Tableau 3: Différents tubes RMN utilisés lors de nos études*

Le pH a été ajusté à environ 6. Des spectres RMN 1D des différents brins seuls ont été enregistrés à différentes températures et étudiés avec une grande attention.

Nous avons alors observé, pour le brin riche en guanines, la présence de nombreuses raies de résonance vers 11 ppm, caractéristiques de la présence de structures quadruplexes dans l'échantillon. Les raies sont plus intenses lorsque la température diminue et ces structures sont stabilisées par des ions positifs tels Na<sup>+</sup> ou K<sup>+</sup>. Dans tous les fragments d'ADN comportant des successions de trois guanines que nous avons eu à étudier, la présence de structures quadruplexes a été observée. Dans le but d'obtenir plus d'informations sur ces structures, des spectres 2D Noesy ont été enregistrés mais aucune information réellement intéressante n'a pu être extraite des spectres RMN, du fait que les pics observés correspondent à un ensemble de structures quadruplexes différentes. Dans nos conditions d'études, la stabilité de ces structures est faible, possédant des *t<sub>m</sub>* (point de fusion) d'environ 15°C.

## 1.3 Formation du duplexe d'ADN

Le protocole de formation du duplexe d'ADN défini ci-dessous est celui que nous avons mis en oeuvre pour toutes les formations de duplexes d'ADN étudiés dans cette thèse.

Pour la formation du duplexe, nous avons suivi l'apparition des raies de résonance des protons iminos appartenant aux thymines et aux guanines du fragment d'ADN. La formation du duplexe d'ADN, via des liaisons hydrogène entre les bases complémentaires, fait que les protons iminos des guanines et des thymines se retrouvent protégés du solvant. Leur échange avec les protons du solvant (ici l'eau) est ralenti, les rendant ainsi observables en RMN. Le suivi d'une résonance en particulier (comme par exemple le H<sub>5</sub> d'une cytosine) permet de former un duplexe équimolaire 1 : 1 entre les deux brins. Une fois le duplexe formé, celui-ci est lyophilisé et dissous dans le solvant, eau légère et eau lourde (90%/10%) pour l'étude des protons échangeables ou eau lourde seule pour l'étude des protons non-échangeables. Le pH est alors ajusté à environ 6 et la force ionique est ajustée entre 10 et 100mM de NaCl. Le duplexe est chauffé jusqu'à 90°C au bain marie avant d'être très lentement refroidi.

#### 1.4 Etude RMN

Une série de spectres NOESY et TOCSY à différentes températures (10, 15 et 25°C) et avec différents temps de mélange ( $\tau_m = 50, 100, 150$  et  $300$ ms pour les spectres NOESY et  $\tau_m = 50$  et  $80$  ms pour les spectres TOCSY) a été enregistrée.

L'attribution des raies de résonances a été réalisée de façon conventionnelle. Les spectres TOCSY ont permis de localiser les corrélations entre les protons H<sub>5</sub> et H<sub>6</sub> des cytosines ainsi que certaines connectivités entre les protons H<sub>6</sub> et CH<sub>3</sub> des thymines. L'attribution séquentielle a ensuite été réalisée. Celle-ci, présentée dans la figure ci-dessous, a permis l'attribution des raies de résonances aromatiques H<sub>6</sub>, H<sub>8</sub> et H<sub>1'</sub>, mais aussi H<sub>2'</sub>, H<sub>2''</sub> et CH<sub>3</sub> de tous les résidus de l'oligonucléotide. Enfin l'attribution des H<sub>3'</sub>, H<sub>4'</sub> a été réalisée, celle des H<sub>5'</sub> et H<sub>5''</sub> étant beaucoup plus délicate du fait des recouvrements. L'attribution séquentielle correspond à une double hélice droite.

L'attribution de la région des protons iminos a ensuite été réalisée. Celle-ci a montré que les appariements sont de type Watson et Crick classique. Les spectres du phosphore n'ont quant à eux montré aucune inhomogénéité, l'ensemble des phosphores de l'oligonucléotide résonnant dans une gamme de déplacements chimiques très limitée. Tous ces éléments permettent d'affirmer que le duplexe formé correspond à un ADN de type B classique.

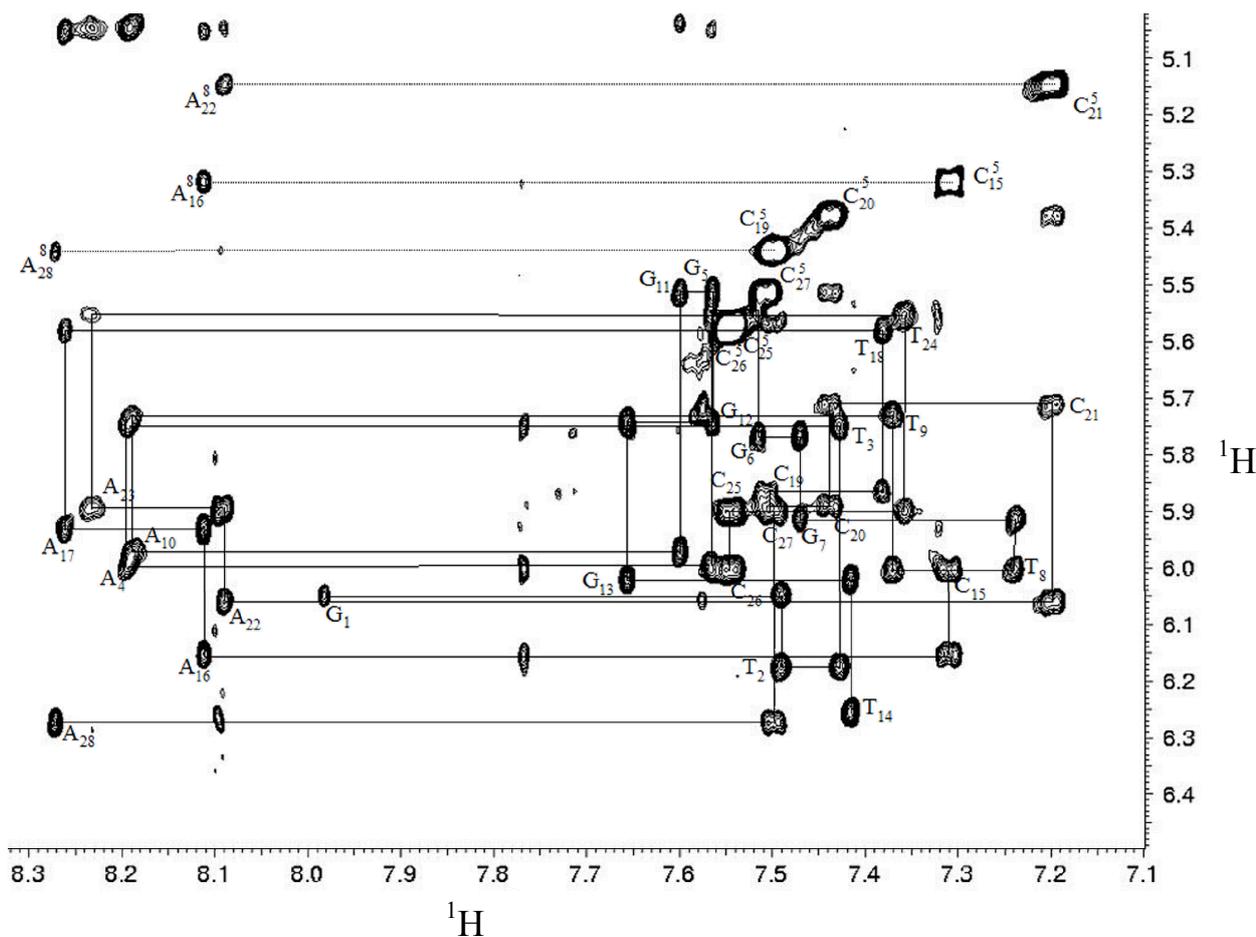


Figure 42 : Attribution séquentielle du fragment d'ADN télomérique comportant deux répétitions télomériques. Agrandissement de la région  $H_6/H_8-H_1/H_5$  d'un spectre Noesy enregistré dans  $D_2O$  avec un  $\tau_m=300ms$  à une température de  $25^\circ C$ .

Néanmoins la simple observation des déplacements chimiques, qui sont le reflet de l'environnement chimique des atomes, montre des différences entre les deux répétitions télomériques. Par exemple, les connectivités  $H_5/H_6$  des cytosines formant les deux répétitions télomériques ( $C_{25}, C_{26}, C_{27}$  et  $C_{19}, C_{20}, C_{21}$ ) ont des déplacements chimiques différents alors que leur environnement chimique est le même. Est-ce là la preuve d'une structure particulière de l'ADN télomérique ou un simple effet lié à la séquence du fragment d'ADN étudié ? Nous avons alors décidé de réaliser l'étude par RMN d'un second duplexe contenant cette fois-ci trois répétitions télomériques.

## 2 Etude de l'ADN télomérique contenant trois répétitions télomériques

La séquence étudiée est la suivante :



Figure 43: Séquence d'ADN télomérique étudié contenant trois répétitions télomériques

Ce fragment d'ADN comporte trois répétitions télomériques et est constitué de 40 résidus. Sa masse moléculaire est proche de 13 kDa et il est non marqué. La formation du duplexe a été réalisée suivant la méthode décrite ci-dessus. Un certain nombre de spectres RMN NOESY et TOCSY a été enregistré à différentes températures (15, 23 et 30°C) et pour différents temps de mélange (80, 150 et 300ms pour les NOESY, 80ms pour les TOCSY). L'attribution séquentielle est présentée sur la figure ci-dessous. Elle correspond à une double hélice droite.

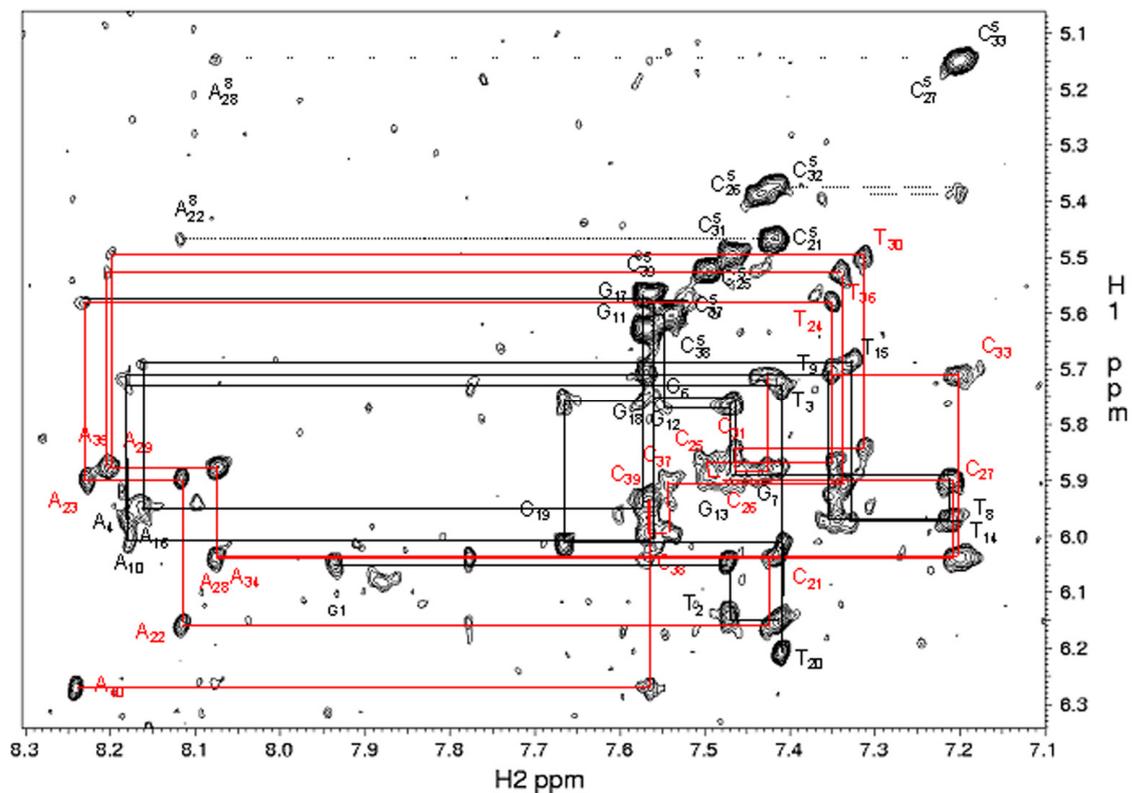


Figure 44: Agrandissement de la région H<sub>6</sub>/H<sub>8</sub>-H<sub>1</sub>' d'un spectre NOESY avec  $\tau_m=300\text{ms}$  à 23°C.

L'analyse des déplacements chimiques confirme les observations faites pour le fragment d'ADN télomérique contenant deux répétitions télomériques. On observe, par exemple, que les déplacements chimiques des protons H<sub>5</sub>/H<sub>6</sub> des cytosines C25,C26,C27 et C31,C32,C33 constituant les deux premières répétitions télomériques sont différents des déplacements chimiques des protons H<sub>5</sub>/H<sub>6</sub> des cytosines C37,C38,C39 constituant la troisième répétition télomérique. Ceci se vérifie aussi pour les guanines et les thymines des différentes répétitions. L'étude des protons iminos nous a permis de mettre en évidence que toutes les bases sont appariées et que les appariements sont de type Watson et Crick.

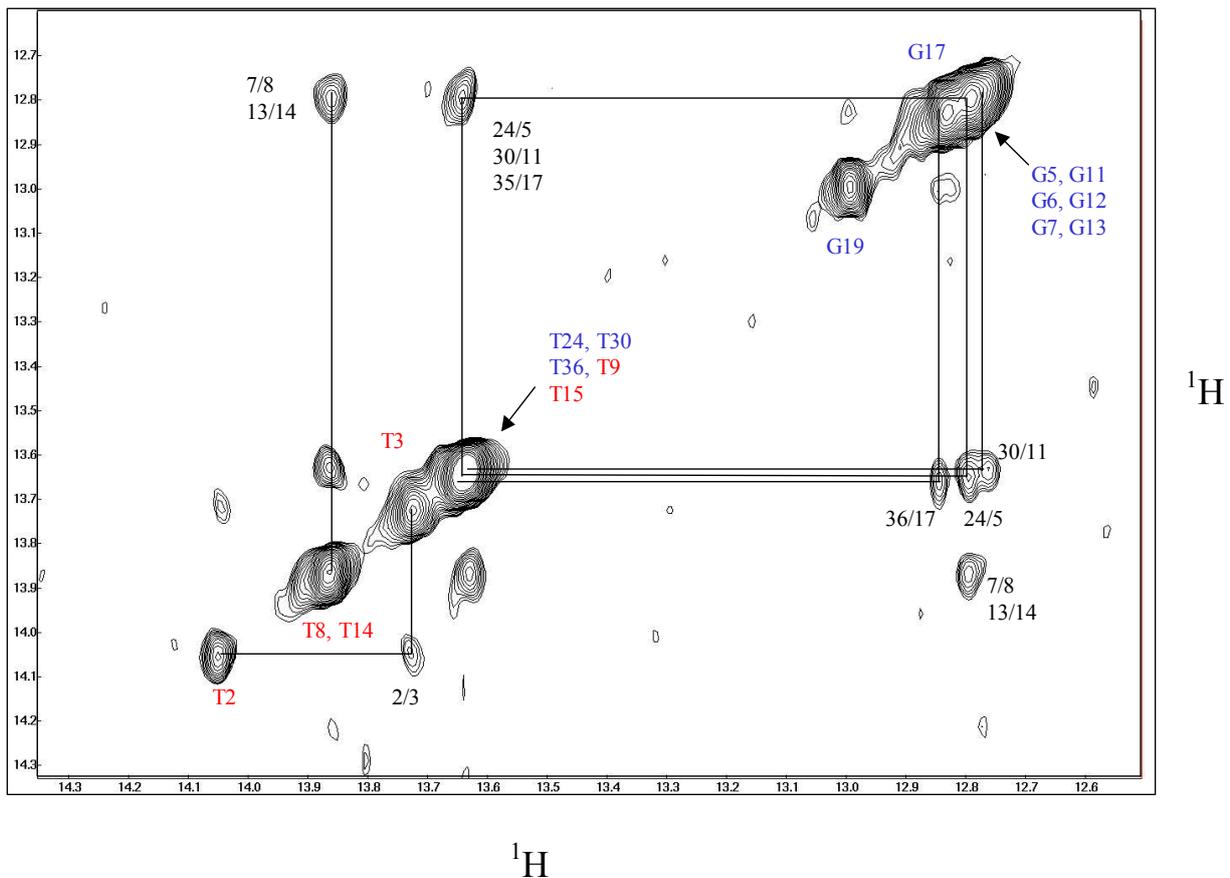


Figure 45: Couplages dipolaires observables entre les protons iminos dans le fragment d'ADN contenant trois répétitions télomériques.

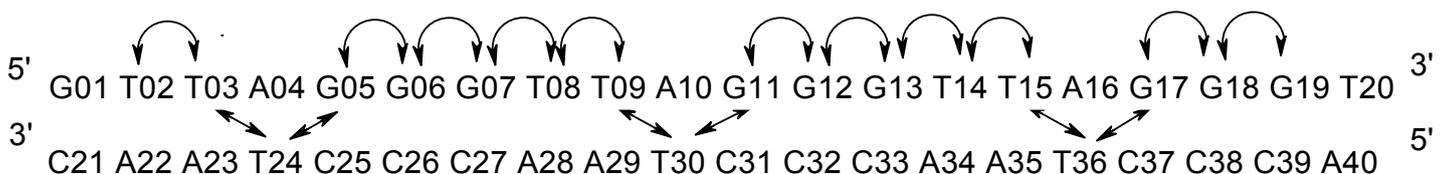


Figure 46 : Schéma présentant les différents couplages dipolaires observés entre les protons iminos

L'observation des spectres TOCSY a permis de déterminer que l'ensemble des 40 sucres possède une conformation proche de C2'-endo. La mesure des distances H<sub>8</sub>/H<sub>6</sub> - H<sub>1</sub>' nous a

permis d'affirmer que l'ensemble des bases azotées de l'oligonucléotide est en conformation anti par rapport à leur sucre.

Les spectres du phosphore montrent que la totalité des raies de résonance du  $^{31}\text{P}$  de l'oligonucléotide est concentrée sur une très faible gamme de déplacements chimiques montrant qu'il n'y a pas d'hétérogénéité structurale du squelette phosphodiester. L'ADN télomérique contenant trois répétitions télomériques adopte donc une conformation en double hélice B classique.

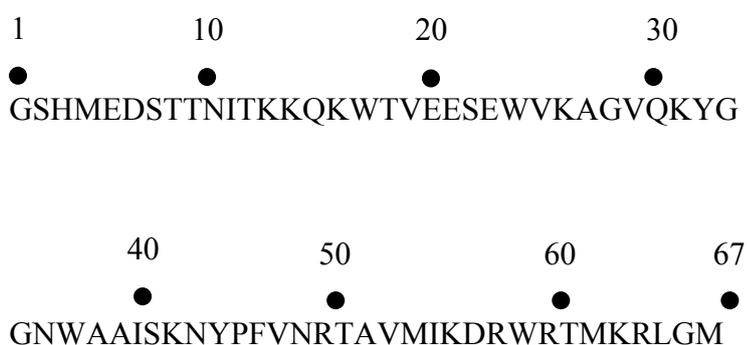
### 3 Conclusion

Plusieurs informations peuvent être déduites de l'étude des ADN double brin comportant deux ou trois répétitions télomériques. Tout d'abord, l'éclatement des raies de résonances des protons constituant les différentes répétitions télomériques est suffisant et rend l'étude de l'ADN télomérique par RMN possible. Nous avons, de plus, observé que l'ADN télomérique adopte une conformation en double hélice B classique. Néanmoins, l'analyse des déplacements chimiques montre des différences au sein des télomères que ce soit pour un ADN comportant deux ou trois répétitions télomériques. La mesure des déplacements chimiques montre que ces différences sont certainement dues à la taille limitée des séquences des fragments d'ADN télomériques étudiés. Enfin notons que lors de la formation des duplexes, nous observons la disparition des raies de résonances appartenant aux structures quadruplexes qui avaient été observées dans les spectres 1D des brin G seuls. La formation du duplexe d'ADN est donc très largement favorisée par rapport aux conformations quadruplexes.

# Chapitre 7 : Etude du domaine Myb de la protéine TRF2

## 1 Présentation du domaine Myb

La protéine étudiée correspond au domaine Myb de la protéine TRF2, c'est-à-dire son domaine de fixation à l'ADN. Sept résidus (EDSTTNI) de la séquence appartiennent au domaine de dimérisation de la protéine, les quatre premiers résidus (GSHM) n'appartiennent pas à la protéine TRF2 mais ont été ajoutés pour faciliter la purification de la protéine. Au niveau de la structure primaire, la protéine comporte 67 résidus pour une masse moléculaire d'environ 7 kDa. La séquence étudiée est la suivante :



*Figure 47: Séquence du domaine Myb de la protéine TRF2 allongé de 11 résidus*

Cette protéine a été préparée au sein du laboratoire dans l'équipe de Marie-Joseph GIRAUD-PANIS. La séquence codant pour le domaine Myb de la protéine TRF2 a d'abord été amplifiée par PCR puis la séquence codante a été clonée dans le plasmide pET14b. Ce plasmide a la particularité de posséder une séquence codante pour une étiquette de 6 histidines qui permettra la purification de la protéine sur colonne Ni-NAT. Cette séquence est suivie par un site de reconnaissance par la thrombine. Après séquençage le plasmide sélectionné est utilisé pour transformer des bactéries BL21 STAR (DE3) pLysS. Ces bactéries sont alors cultivées sur du LB (Lennox Luria Broth) pour la protéine non-marquée. Pour la protéine

marquée  $^{15}\text{N}$  et  $^{13}\text{C}$ , un milieu différent, le Martek 9 (SPECTRA) auquel peut être ajouté un additif Celtone pour accélérer la croissance bactérienne, a été utilisé. Ce milieu contient un hydrolysate d'algues photosynthétiques cultivées sur un milieu synthétique en atmosphère contrôlée ( $^{13}\text{CO}_2$ ) et sur des substrats marqués au  $^{15}\text{N}$  ( $^{15}\text{NH}_4\text{Cl}$ ). Le domaine Myb de TRF2 a été purifié par chromatographie d'affinité sur colonne Ni-NTA grâce à l'étiquette de 6 histidines en N-terminal. L'étiquette a ensuite été clivée par la thrombine. Nous avons d'abord obtenu le domaine Myb de la protéine non-marquée avant d'obtenir le domaine Myb uniformément marqué  $^{15}\text{N}$  et  $^{13}\text{C}$ , c'est pourquoi les premières études effectuées sur le domaine Myb de TRF2 ont été réalisées sur la protéine non-marquée.

## 2 Etude RMN de la protéine non-marquée et de la protéine marquée uniformément $^{15}\text{N}$ et $^{13}\text{C}$

L'étude de la protéine non-marquée a été réalisée de façon classique en utilisant des spectres COSY-DQF, TOCSY et NOESY et le principe de l'attribution séquentielle énoncé plus haut. L'étude de la protéine non-marquée a été réalisée par Hervé Meudal au sein de notre équipe.

Le domaine Myb de TRF2, doublement et uniformément marqué  $^{15}\text{N}$  et  $^{13}\text{C}$ , a été l'une des premières protéines marquées  $^{15}\text{N}$  et  $^{13}\text{C}$  étudiées dans notre équipe.

Un premier spectre  $^{15}\text{N}$ -HSQC a été réalisé et a permis de se rendre compte de l'état parfaitement structuré de la protéine. En effet l'observation de résonances HN (proton amidos de la chaîne peptidique) bien éclatées est plutôt le signe d'une protéine structurée, alors que l'observation de raies de résonances regroupées au centre du spectre est plutôt le signe d'une protéine déstructurée, qui s'agrège. Dans notre cas le spectre  $^{15}\text{N}$ -HSQC est satisfaisant. La qualité du marquage  $^{15}\text{N}$  et  $^{13}\text{C}$  de la protéine a été vérifiée en réalisant un spectre  $^{15}\text{N}$ -HSQC non découplé en  $^{15}\text{N}$ . La présence unique d'un doublet correspondant au couplage  $^{15}\text{N}$ - $^1\text{H}$  est le signe d'une protéine uniformément marquée alors que la présence d'un triplet indique la présence d'un mélange de protéine marquée et non marquée.

## 2.1 Attribution du squelette peptidique

L'attribution du squelette peptidique a été réalisée à l'aide de spectres HNCA, HN(CO)CA, HNCO et HN(CA)CO. Les spectres CBCANH et CBCA(CO)NH ont toutefois permis de lever certaines ambiguïtés et ont permis l'attribution des C $\beta$ . L'utilisation de ces expériences a permis l'attribution des C $\alpha$ , HN, CO, N et C $\beta$  de 62 résidus sur 67. Les cinq résidus non attribués correspondent aux premiers résidus n'appartenant pas au domaine Myb.

Le spectre  $^{15}\text{N}$ -HSQC du domaine myb seul comportant l'attribution des corrélations  $^1\text{H}$ - $^{15}\text{N}$  est représenté sur la figure suivante :

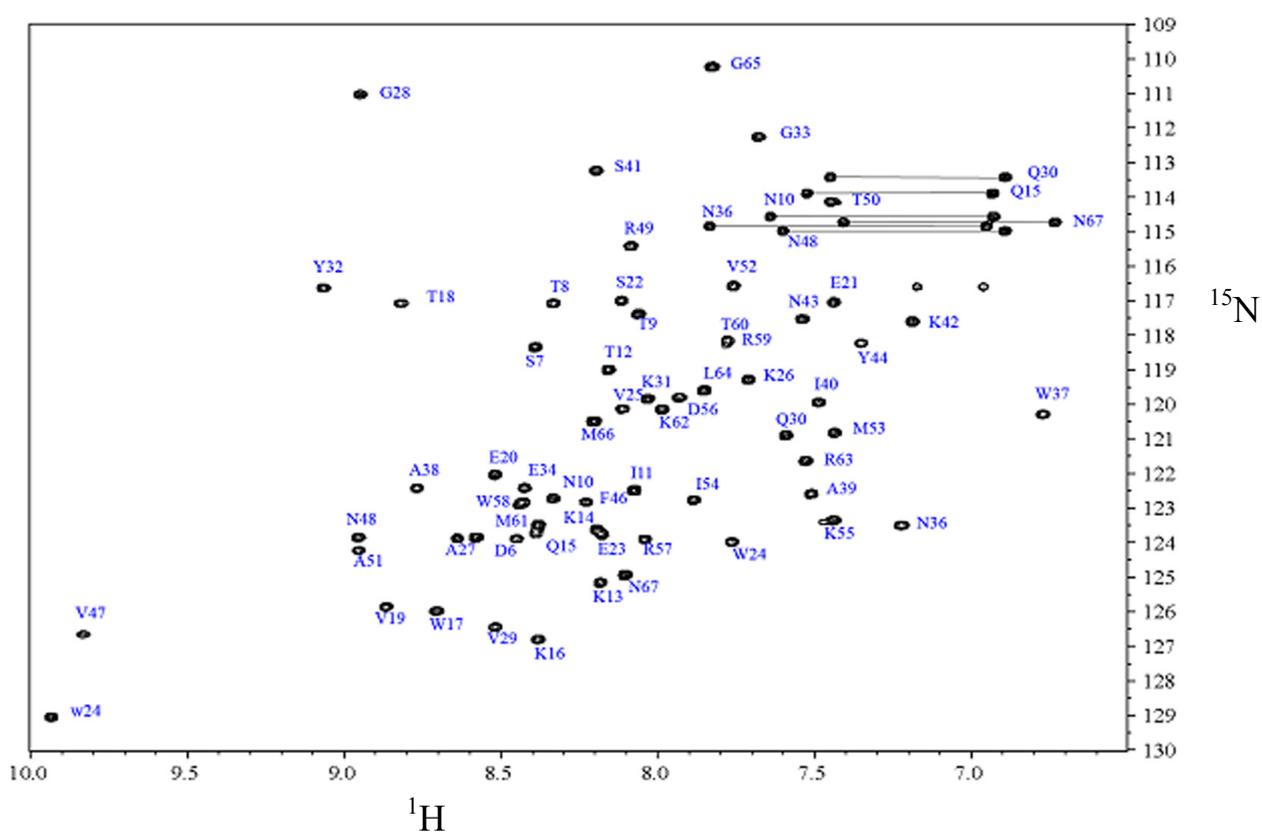


Figure 48: spectre  $^{15}\text{N}$ -HSQC du domaine Myb de HTRF2 seule. En dimension directe, gamme de déplacement chimique du  $^1\text{HN}$ . En direction indirecte, gamme de déplacement chimique de l'azote 15.

## 2.2 Attribution des chaînes latérales

Pour l'attribution des chaînes latérales les expériences 2D HCCH-TOCSY, 2D HCCH-COSY,  $^{15}\text{N}$  TOCSY-HSQC et  $^{15}\text{N}/^{13}\text{C}$  NOESY-HSQC ont été utilisées. La grande majorité des déplacements chimiques des protons des chaînes latérales a pu être attribuée. Les protons non attribués appartiennent pour la plupart aux premiers résidus formant la partie non-structurée de la protéine. De plus, l'étude du spectre  $^{15}\text{N}$ -NOESY HSQC a permis l'attribution des protons aminos des chaînes latérales des Asparagines et Glutamines. Il est à noter le déplacement chimique très particulier du proton amido  $^1\text{HN}$  de la glycine 35 qui résonne près de l'eau ( $\delta^{1\text{H}}=5,45\text{ppm}$ ) et qui n'est donc pas toujours facilement observable. La majeure partie des protons aromatiques a été attribuée. Au total environ 98% des protons non-échangeables de la protéine ont pu être attribués.

## 2.3 Collection des informations structurales

Les spectres  $^{15}\text{N}/^{13}\text{C}$ -NOESY HSQC, NOESY ont été convertis et analysés avec les logiciels NMRPipe et NMRDraw (Delaglio, Grzesiek et al. 1995) puis ont été traités à l'aide du logiciel NMR-VIEW (Johnson and Blevins 1994; Johnson 2004) et les volumes des différentes connectivités NOEs ont été mesurés.

Les angles dièdres  $\phi$  et  $\psi$  ont été estimés à l'aide du logiciel Talos. Ainsi, 46 couples d'angles ( $\psi, \phi$ ) correspondent parfaitement aux prédictions et ont pu être utilisés dans la modélisation moléculaire.

La liste des déplacements chimiques utilisés comprend 98% des protons non-échangeables attribués. La modélisation moléculaire a été réalisée en utilisant, à la fois le logiciel ARIA et le logiciel CYANA version 2.1.

Pour les calculs avec le logiciel Cyana, la liste des déplacements chimiques, les angles dièdres prédits avec le logiciel Talos ainsi que les listes des pics NOEs obtenus par « peak-picking » avec le logiciel NMRView ont été introduits. Les tolérances de déplacements chimiques ont été fixées à 0,05ppm pour la dimension  $^1\text{H}$  et à 0,7ppm pour les dimensions  $^{15}\text{N}$  ou  $^{13}\text{C}$ . Le calcul a été réalisé avec 10000 pas de dynamique moléculaire dans l'espace des angles de torsion et 200 structures ont été calculées. Les 20 structures présentant les fonctions cibles les

plus faibles ont été sélectionnées. La fonction cible finale sur l'ensemble des 20 structures est de  $1,59 \pm 0,11 \text{ \AA}^2$ .

Les fichiers de distances (.upl) et d'angles dièdres (.aco) ont ensuite été convertis au format ARIA à l'aide de programmes écrits dans le langage de programmation Python avant d'être utilisés pour réaliser un raffinement des structures dans une boîte d'eau (dynamique moléculaire dans l'espace cartésien). Les structures obtenues ont été analysées visuellement ainsi qu'avec le serveur web PSVS. Ce serveur réalise un certain nombre d'analyses des structures et des contraintes proposées à l'aide de programmes comme Procheck, Verify 3D, ProsaII, MolProbity et PDB Validation. Il est accessible à l'adresse internet suivante:

<http://www-nmr.cabm.rutgers.edu/PSVS>

Les principaux résultats obtenus sont détaillés dans le tableau ci-dessous :

#### Contraintes

##### *NOE*

|   |                               |
|---|-------------------------------|
| Total   | 1603                          |
| Intra-résidus [ $i = j$ ]                     | 338                           |
| Séquentielles [ $ i - j  = 1$ ]               | 397                           |
| Moyenne distance [ $1 <  i - j  < 5$ ]        | 434                           |
| Longue distance [ $ i - j  > 5$ ]             | 434                           |
| <i>Dièdres</i>                                | 92                            |
| <i>Nombre moyen de contraintes par résidu</i> | 25,7                          |
| <i>Nombre de structures calculées</i>         | 200                           |
| <i>Nombre de structures sélectionnées</i>     | 20                            |
| <i>Fonction cible Cyana</i>                   | $1,59 \pm 0,11 \text{ \AA}^2$ |
| <i>Nombre de violation noe &gt; 0,5Å</i>      | 0                             |

#### Cartes de Ramachandran

|                    |       |
|--------------------|-------|
| Régions favorables | 89,4% |
| Régions autorisées | 10,6% |
| Régions interdites | 0%    |

#### RMSD cyana

|  |                             |
|--|-----------------------------|
| Sur tous les résidus (atomes lourds)                       | $6,11 \pm 0,12 \text{ \AA}$ |
| Sur les résidus de la région ordonnée<br>(Résidus 19 à 64) | $0,50 \pm 0,08 \text{ \AA}$ |

#### Affinement dans une boîte d'eau ARIA

|  |                              |
|--|------------------------------|
| Energie totale moyenne                     | -2330 kcal.mol <sup>-1</sup> |
| RMSD sur les résidus de la région ordonnée | 0,68 ± 0,09Å                 |

Tableau 4: Récapitulatif des différents résultats obtenus lors de la modélisation de la protéine

A

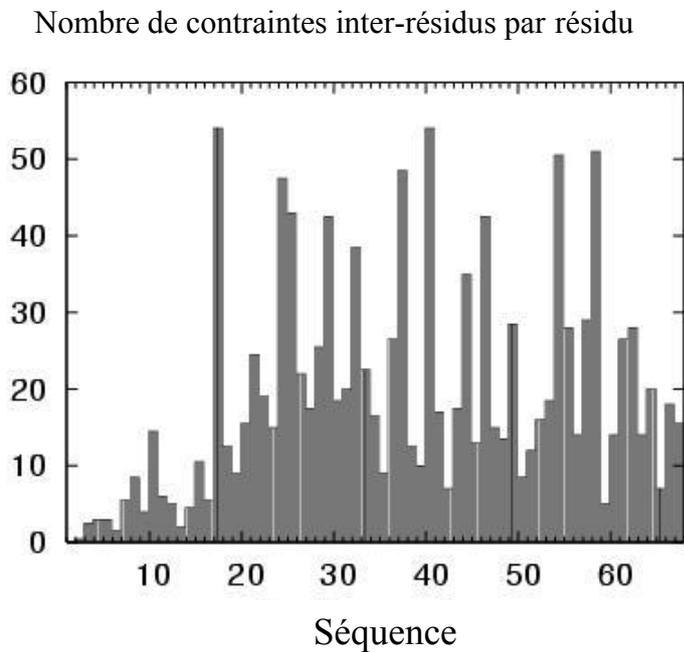


Figure 49 : Nombre de contraintes NOE par résidu

La structure du domaine Myb TRF2 peut être observée dans la figure ci-dessous :

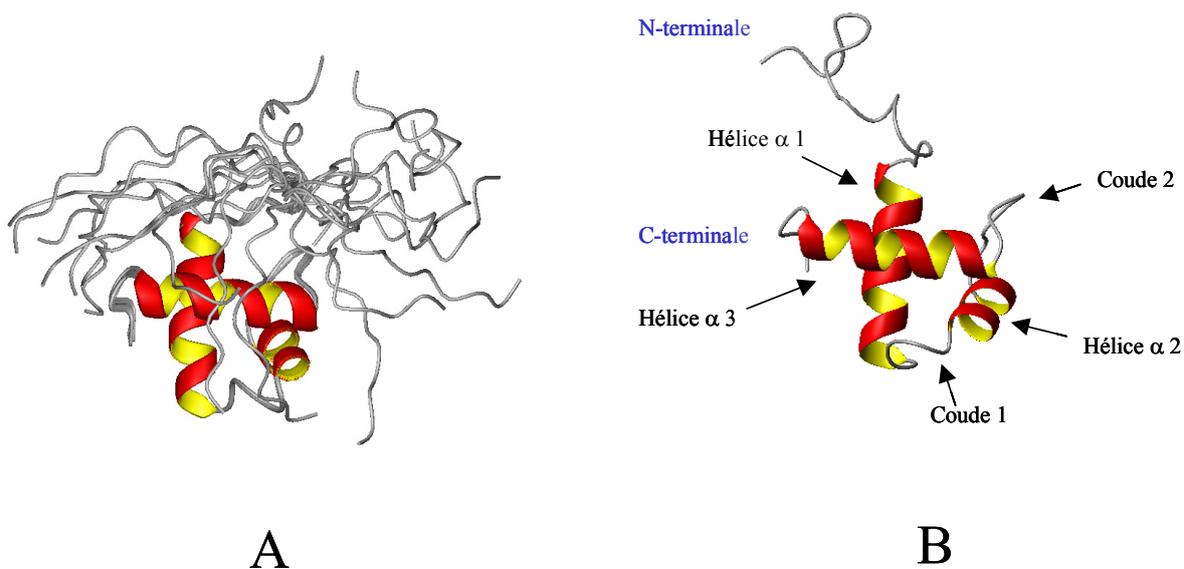


Figure 50 : (A) Superposition des 20 structures du domaine Myb de TRF2 obtenues par modélisation moléculaire avec le logiciel Cyana (B) Une des structures du domaine Myb de TRF2 choisie arbitrairement.

## 2.4 Analyse de la structure du domaine Myb de TRF2

La structure du domaine Myb obtenue par modélisation moléculaire permet d'expliquer quelques déplacements chimiques particuliers qui ont pu être observés sur les différents spectres RMN. Le déplacement particulier du proton amide de la valine 47 vers les bas champs ( $\delta_{1H} = 9,85$  ppm) peut être expliqué par le fait que ce proton est dans le plan du cycle du tryptophane 24, subissant ainsi d'importants courants de cycles (distances  $H\epsilon_{3\text{trp}24} - NH_{\text{val}47} = 3.0\text{\AA}$ ). Le proton amide de la glycine 35 ( $\delta_{1H} = 5,45$  ppm) et le  $C\alpha$  de la lysine 55 sont localisés de chaque côté et relativement proches du cycle du tryptophane 37, expliquant ainsi leur déplacement vers les champs forts. Un des protons  $H\alpha$  de la glycine 28 à cause de sa proximité avec les cycles aromatiques des tyrosines 32 et 44, est déplacé vers les champs faibles à 4,62 ppm. Les protons hydroxyles des thréonines 18 et 60 sont liés par des liaisons hydrogène aux groupements carboxyliques de l'acide glutamique 21 et du groupement carbonyle de l'arginine 57, expliquant que leurs résonances soient situées à 5,66 et 5,12 ppm. La structure présente trois hélices  $\alpha$ . La première hélice correspond aux résidus 19-31, la seconde aux résidus 37 à 43 et la troisième aux résidus 51-64. Ces trois hélices sont reliées par deux coudes constitués des résidus 32-36 et 44-50. On trouve au sein de cette protéine un cœur hydrophobe formé par l'ensemble des résidus suivants : W17, V25, V29, W37, I40, F46, A51, I54, W58 et M61.

Le premier coude, classifié comme un coude  $\beta$  de type 4 est stabilisé par des liaisons hydrogène entre le proton amide de l'asparagine 36 et le carbonyle de la glycine 35 ainsi que par une liaison hydrogène entre le proton amide du résidu tryptophane 37 et le carbonyle de la glycine 35. Le deuxième coude contient la proline 45 qui induit la formation d'un coude  $\gamma$ . Ce coude est stabilisé par l'interaction du noyau aromatique de la phénylalanine 46 avec de nombreux résidus formant le cœur hydrophobe et appartenant aux trois différentes hélices : valine 52, isoleucine 40 et isoleucine 54. Deux liaisons hydrogène entre le proton amide de la phénylalanine 46 et le carbonyle de la tyrosine 44 ainsi qu'entre le proton  $N\epsilon H$  de l'arginine 49 et le carbonyle de la valine 47 participent aussi à la stabilisation de ce coude.

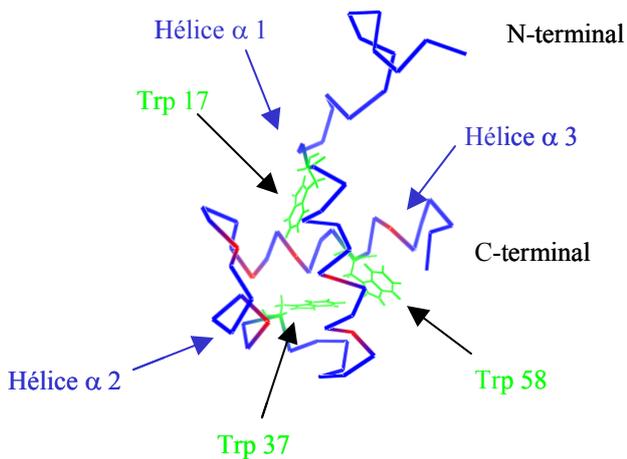


Figure 51 : Description du cœur hydrophobe d'une structure du domaine Myb de TRF2 choisie arbitrairement. En rouge apparaissent les différents résidus impliqués dans la formation de ce cœur hydrophobe, en vert sont représentés les 3 tryptophanes essentiels pour la stabilisation du cœur hydrophobe.

## 2.5 Comparaison avec le domaine Myb de la protéine TRF1

La structure globale de TRF2 est très similaire à celle de TRF1 (figure 52). La présence de trois hélices  $\alpha$  liées par deux coudes se retrouve dans les deux domaines Myb. La majorité des résidus formant le cœur hydrophobe du domaine myb de TRF2 est conservée dans le domaine Myb de TRF1, les seules différences concernent la valine 25 et l'isoleucine 54 qui sont remplacées dans le domaine Myb de TRF1 par des leucines (figure 53). La première hélice diffère un peu entre les deux domaines. Dans le domaine Myb de TRF1, deux faces sont clairement définies, l'une hydrophobe et l'autre chargée. Dans le domaine Myb de TRF2, la première hélice est constituée de résidus de charges différentes. Les deuxième et troisième hélices, formant le domaine hélice-tour-hélice, sont responsables de la fixation du domaine Myb de TRF1 sur l'ADN. L'hélice 2 dans TRF2 est différente, la présence de la proline 45 induisant la formation d'un coude  $\gamma$  (résidus 44 à 46). Après ce coude  $\gamma$ , les domaines Myb de TRF1 et de TRF2 sont de nouveau superposables. L'hélice 3 est la même dans les deux domaines Myb. Ceci est dû à l'homologie importante de séquence entre les deux hélices qui ont en commun 11 des 14 résidus formant l'hélice. Les petites différences qui ont pu être observées ne permettent pas d'expliquer les différences de fonction entre les deux protéines.

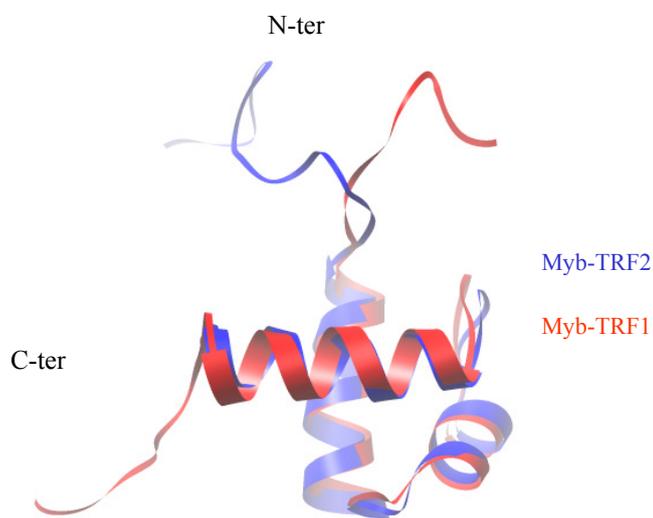


Figure 52: Superposition des structures des domaines Myb des protéines TRF1 et TRF2.

RMSD = 0,822 Å sur 45 paires d'atomes du squelette peptidique

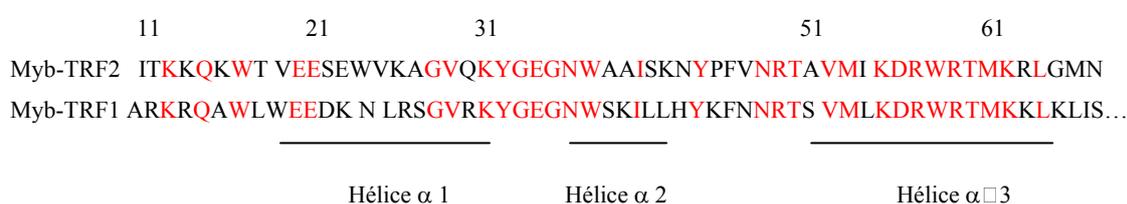


Figure 53: Alignement des séquences des domaines Myb des protéines TRF1 et TRF2. En rouge sont représentés les résidus conservés entre les deux domaines.

## 2.6 Etude de la dynamique du domaine Myb de TRF2

L'étude dynamique du domaine Myb de TRF2 a été réalisée en utilisant la relaxation de l'azote 15. L'échantillon utilisé était uniformément marqué à l'azote 15 et au carbone 13. Sa concentration était de 1,6 mM à un pH de 6,5. Le traitement et l'analyse des données ont été réalisés à l'aide des logiciels NMRPipe (Delaglio, Grzesiek et al. 1995) et NMRView (Johnson and Blevins 1994).

### 2.6.1 Détermination de R1 et de R2

Les valeurs des temps de relaxation longitudinale T1 et transversale T2 ont été déterminées à l'aide d'une série d'expériences de type  $^{15}\text{N}$  HSQC. Ces expériences ont été enregistrées en faisant varier le temps de relaxation  $\tau_r$ . Afin d'ajuster au mieux les courbes de relaxation longitudinale et transversale les délais suivants ont été choisis :

|                    |    |     |     |     |     |     |      |      |     |
|--------------------|----|-----|-----|-----|-----|-----|------|------|-----|
| T1 : $\tau_r$ (ms) | 10 | 100 | 220 | 380 | 700 | 900 | 1100 | 1300 |     |
| T2 : $\tau_r$ (ms) | 10 | 30  | 50  | 70  | 90  | 130 | 190  | 230  | 250 |

Figure 54 : Les différents délais de relaxation  $\tau_r$  utilisés

La figure présente deux spectres  $^{15}\text{N}$ -HSQC utilisés pour déterminer les valeurs des temps de relaxation T1.

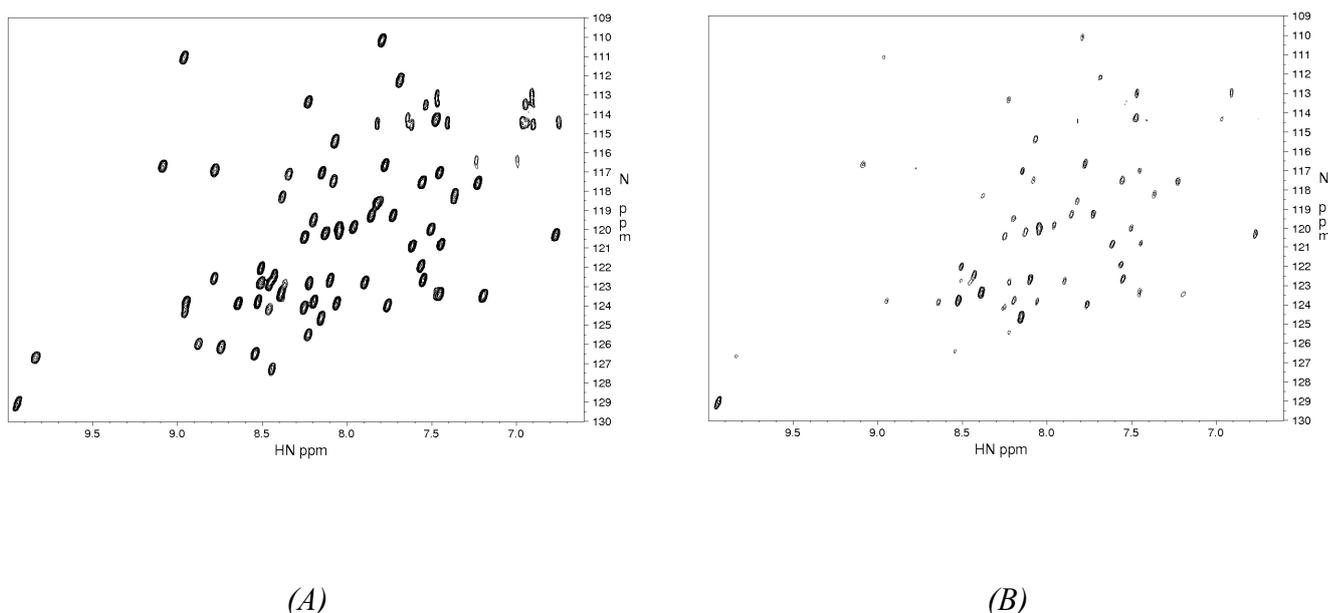


Figure 55: Spectres  $1\text{H}-^{15}\text{N}$  HSQC du domaine Myb de TRF2 avec  $tr=10\text{ms}$  (A) et  $tr=1300\text{ms}$  (B)

Les temps de relaxation T1 et T2 pour chaque résidu ont été extraits à partir des valeurs de l'intensité des pics croisés en réalisant un ajustement de ces intensités à l'aide d'une fonction mono-exponentielle décroissante.

La connaissance des temps de relaxation T1 et T2 pour chaque résidu permet de déterminer les valeurs des vitesses de relaxation longitudinale R1 et transversale R2. Des superpositions de signaux ont empêché d'obtenir certaines valeurs ou du moins des valeurs fiables, ainsi les valeurs de relaxation de seulement 60 résidus sur 67 ont pu être déterminées. Ces résultats sont résumés dans les figures ci-dessous (A et B).

La valeur moyenne des vitesses de relaxation R1 des noyaux  $^{15}\text{N}$  de Myb est de  $2,34\text{ s}^{-1}$ . Elle est de  $6,88\text{ s}^{-1}$  pour R2. Les valeurs de R1 et de R2 sont relativement homogènes sur l'ensemble de la séquence. Néanmoins, il faut noter les valeurs élevées de R1 et R2 pour les résidus A38 et A51 ainsi que les valeurs faibles de R2 pour les résidus Q5 à Q15. Les valeurs

faibles de  $R_2$  indiquent une dynamique interne importante (dans la gamme de la picoseconde), ce qui peut s'expliquer par le caractère non structuré de cette partie de la protéine.

### 2.6.2 Mesure de l'effet NOE hétéronucléaire $^1\text{H}$ - $^{15}\text{N}$

Le délai d'irradiation a été fixé à 5 s. Entre deux expériences, le délai de retour à l'équilibre du système a été placé à 10 s. Les valeurs des intensités NOE sont représentées dans la figure ci-dessous (D). Notons la présence de NOE négatifs, signe d'une grande flexibilité. Pour l'ensemble des résidus, le NOE est situé entre  $-0,62$  et  $+0,82$ . La valeur moyenne reste néanmoins globalement élevée montrant que le domaine Myb de TRF2 est globalement peu flexible, du moins entre les résidus 19 à 64.

### 2.6.3 Temps de corrélation $\tau_c$

Le temps de corrélation d'une protéine globulaire de 7 kDa est d'environ 3,5 ns à 20°C. Cette valeur a été confirmée par le calcul de la moyenne du rapport  $R_2/R_1$  sur les résidus des structures bien définies (dans notre cas les hélices  $\alpha$ ). On trouve un  $\tau_c$  de 3,38 ns à 20°C. Cette valeur est en accord avec le fait que le domaine Myb de la protéine TRF2 est globulaire et qu'en solution (du moins dans les conditions de concentration, pH, force ionique dans lesquelles se trouve l'échantillon), le domaine Myb de TRF2 est sous forme monomère.

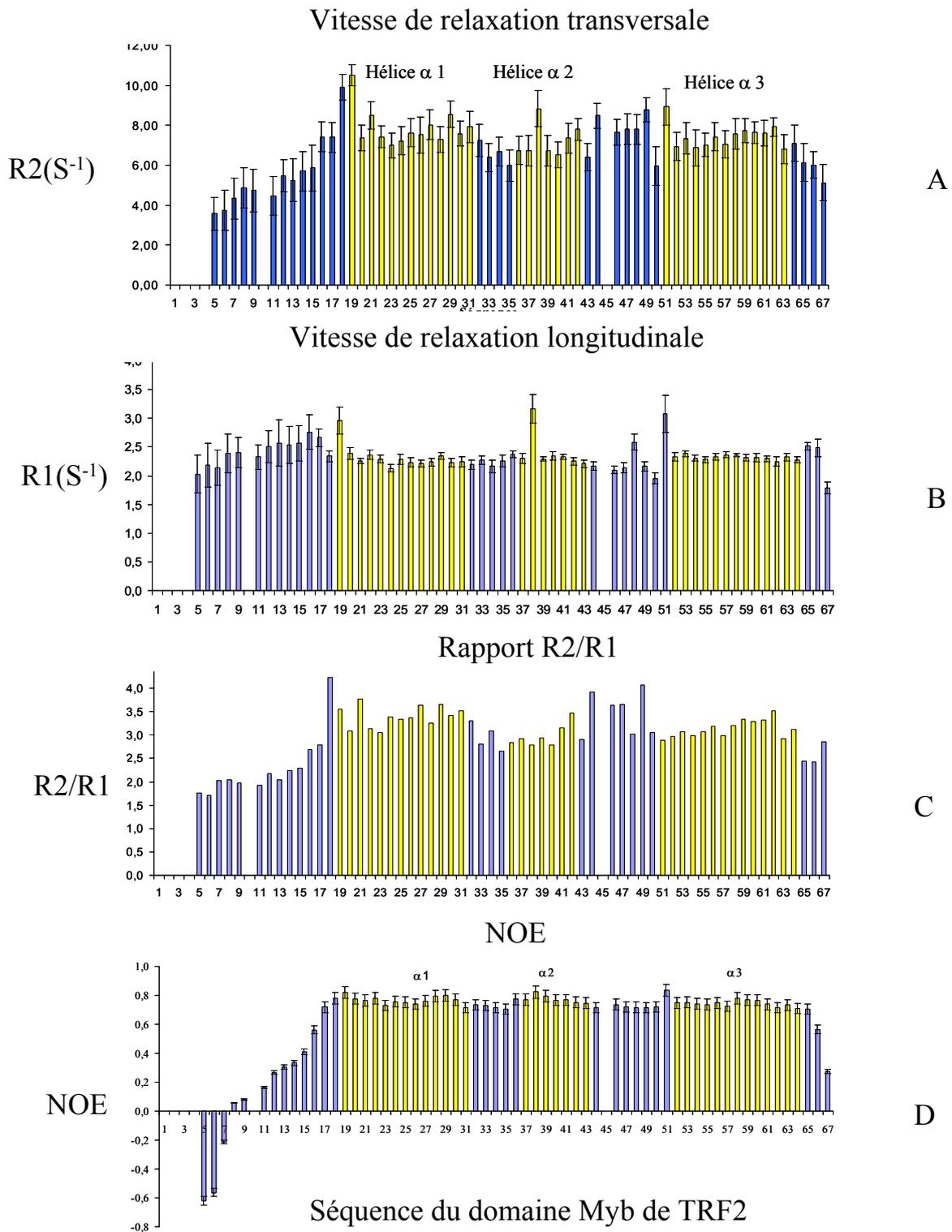


Figure 56 : Valeur de l'effet NOE (D) et des vitesses de relation longitudinale (B) et transversale (A) pour chaque résidu du domaine Myb de la protéine TRF2. En jaune sont représentés les résidus appartenant aux trois hélices  $\alpha$ . (C) Rapport R2/R1 permettant l'estimation du temps de corrélation de la protéine.

#### 2.6.4 Les paramètres d'ordre et les modèles retenus

Les données de relaxation obtenues expérimentalement ont été interprétées par le formalisme de Lipari Szabo à l'aide du logiciel Fast-Modelfree. Les résultats obtenus sont rassemblés dans la figure ci-dessous:

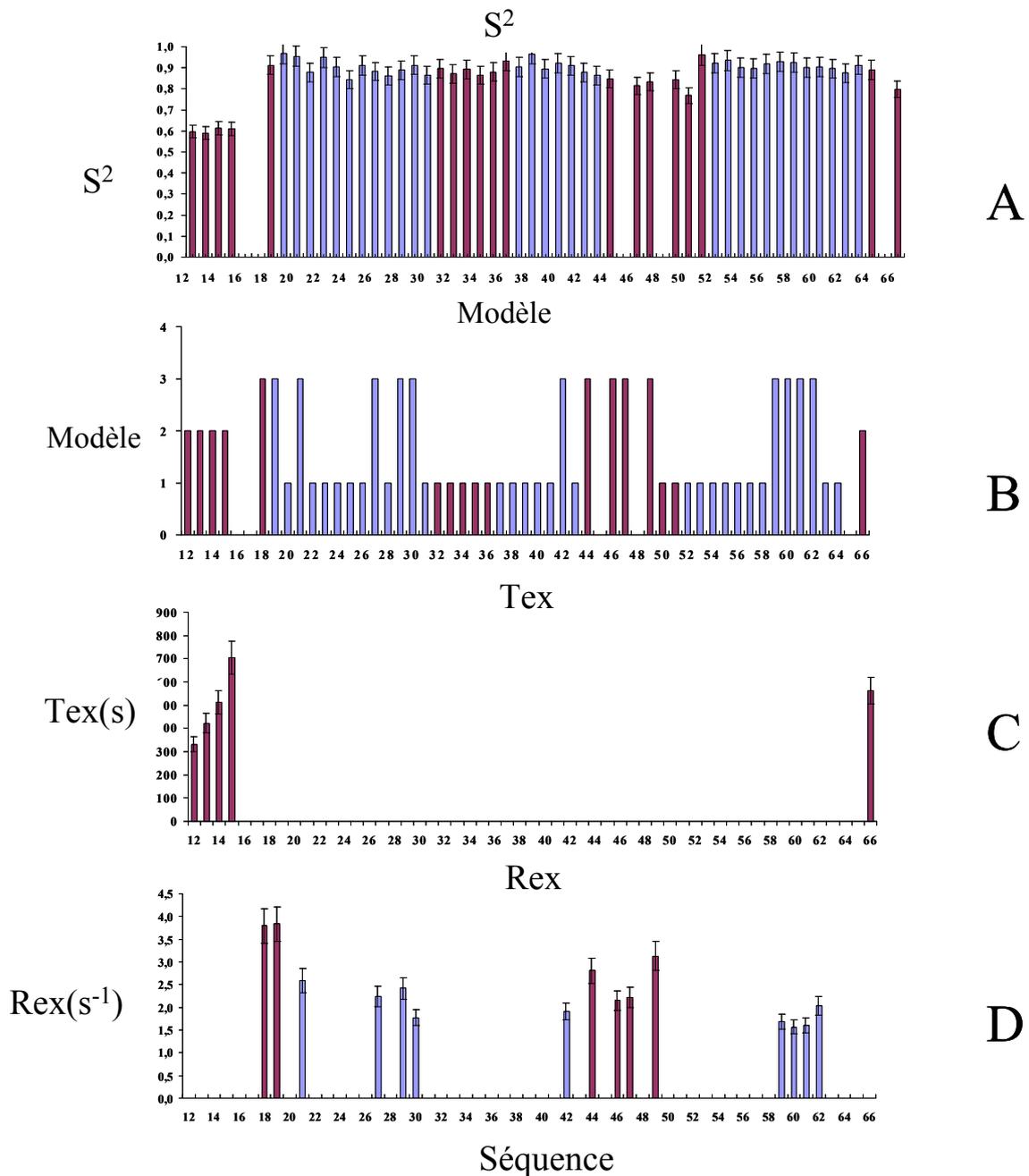


Figure 57 : (A) Paramètre  $S^2$  en fonction de la séquence; (B) Modèle retenu pour chaque résidu; (C)  $Tex$ ; (D)  $Rex$ . En bleu sont représentés les résidus appartenant aux trois hélices  $\alpha$ .

L'analyse effectuée à l'aide du logiciel Fast-Modelfree a permis d'ajuster les paramètres de relaxation de 54 résidus sur 67, les paramètres de certains résidus n'ayant pu être définis par un modèle. 34 résidus sont décrits par le modèle 1. Ces résidus sont uniformément répartis sur l'ensemble de la séquence. Cela confirme la nature isotrope des mouvements du domaine Myb de TRF2, le modèle 1 ne considérant qu'un seul type de mouvement, avec un temps de corrélation  $\tau_c$  unique correspondant à celui de la protéine. Les mouvements des vecteurs N-H de ces résidus sont donc uniquement dépendants du mouvement global de la protéine.

5 résidus ont été interprétés par le modèle 2 correspondant au formalisme de Lipari-Szabo originel « à deux paliers ». Ce modèle ne considère que deux types de mouvements se déroulant à des échelles de temps très différentes, le premier correspondant au mouvement global de la protéine ( $\tau_c=3,38$  ns pour Myb), et le second caractérisant les mouvements internes des vecteurs N-H de ces résidus qui présentent un temps de corrélation interne  $\tau_e$  de l'ordre de la picoseconde. La valeur moyenne de  $\tau_e$  pour les résidus considérés est égale à quelques centaines de picosecondes. Ces résidus présentent une dynamique interne simple, sans intervention d'un terme d'échange  $R_{ex}$ . Ils correspondent aux premiers résidus de la protéine formant la partie non structurée ainsi qu'au résidu 66, avant dernier résidu de la protéine.

Les modèles 3 et 4 correspondent respectivement aux modèles 1 et 2 avec un terme d'échange chimique supplémentaire ( $R_{ex}$ ). Le modèle 3 a permis de modéliser les mouvements internes des vecteurs N-H de 15 résidus. Ces résidus sont situés dans la première hélice (V19, E21, A27, V29 et Q30), dans le premier coude (N42, Y44, F46, V47 et R49) et dans la troisième hélice (R59, T60, M61 et K62).

Les paramètres d'ordre  $S^2$  calculés pour les vecteurs N-H des résidus de Myb montrent que cette dernière est structurée en solution et que ses mouvements internes sont très restreints. En effet, la valeur moyenne des  $S^2$  tout au long de la séquence est élevée et égale à 0,869. Cette valeur est plus importante dans les hélices. La région N-terminale (jusqu'au résidu 16) présente néanmoins des valeurs de  $S^2$  relativement faibles montrant bien la nature non-structurée de cette partie de la protéine.

Les valeurs moyennes du  $S^2$  dans chacune des parties de la protéine sont résumées dans le tableau ci-dessous :

| <i>Région de la protéine Myb</i> | $S^2$ moyen     |
|----------------------------------|-----------------|
| Résidus (1-18)                   | $0,60 \pm 0,05$ |
| Hélice $\alpha$ 1 (19-31)        | $0,90 \pm 0,05$ |
| Résidus 32-36                    | $0,88 \pm 0,05$ |
| Hélice $\alpha$ 2 (37-43)        | $0,91 \pm 0,05$ |
| Résidus 44-50                    | $0,82 \pm 0,05$ |
| Hélice $\alpha$ 3 (51-64)        | $0,90 \pm 0,05$ |

*Figure 58: Moyenne du paramètre d'ordre  $S^2$  sur les différentes régions du domaine Myb de TRF2*

### 3 Conclusion

Cette étude a permis de mettre en avant différents points intéressants. Le domaine Myb de la protéine TRF2 est une protéine globulaire sous forme monomérique en solution. Son temps de corrélation est d'environ 3,38 ns. C'est une protéine peu flexible, bien structurée, comportant trois hélices  $\alpha$  reliées par deux coudes, l'un  $\beta$  et l'autre  $\gamma$ . Enfin, la protéine possède une partie non structurée présentant donc une dynamique importante.

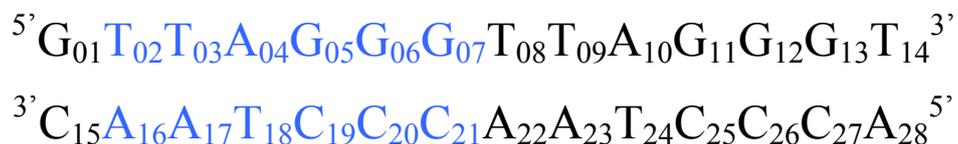
Dans le chapitre suivant sera présenté le premier complexe télomérique étudié entre le domaine Myb de la protéine TRF2 non-marquée et un fragment d'ADN télomérique double brin.

# Chapitre 8 : Réalisation d'un premier complexe télomérique

## 1 Présentation

La protéine TRF2 est connue pour se fixer sur l'ADN télomérique double brin et il a été montré que deux répétitions sont nécessaires pour que la fixation ait lieu. Comment le domaine Myb de la protéine TRF2 se fixe-t-il sur l'ADN télomérique? Quelles modifications peut-on observer au niveau de l'ADN télomérique? C'est ce que nous avons voulu déterminer en réalisant un complexe entre le domaine Myb de la protéine TRF2 (non marqué) et un fragment d'ADN télomérique double brin contenant deux répétitions télomériques également non-marqué.

La séquence du duplexe d'ADN télomérique utilisée est la suivante:



*Figure 59: Séquence du fragment d'ADN utilisée pour la réalisation du complexe*

Cet ADN comprend deux répétitions télomériques et correspond au fragment d'ADN étudié au chapitre 6.

## 2 Réalisation du complexe

La formation du complexe a été suivie par RMN 1D en observant le déplacement chimique des protons iminos des guanines et thymine constituant le fragment d'ADN. En effet ceux-ci sont en échange lent et sont donc caractéristiques d'une forme particulière. Le domaine Myb de TRF2 a été progressivement ajouté à l'ADN. Lors de la formation du complexe, nous observons la disparition de certaines raies de résonance et l'apparition de nouvelles raies de

résonance. La titration de l'ADN télomérique double brin par le domaine MybTRF2 est représentée sur la figure ci-dessous :

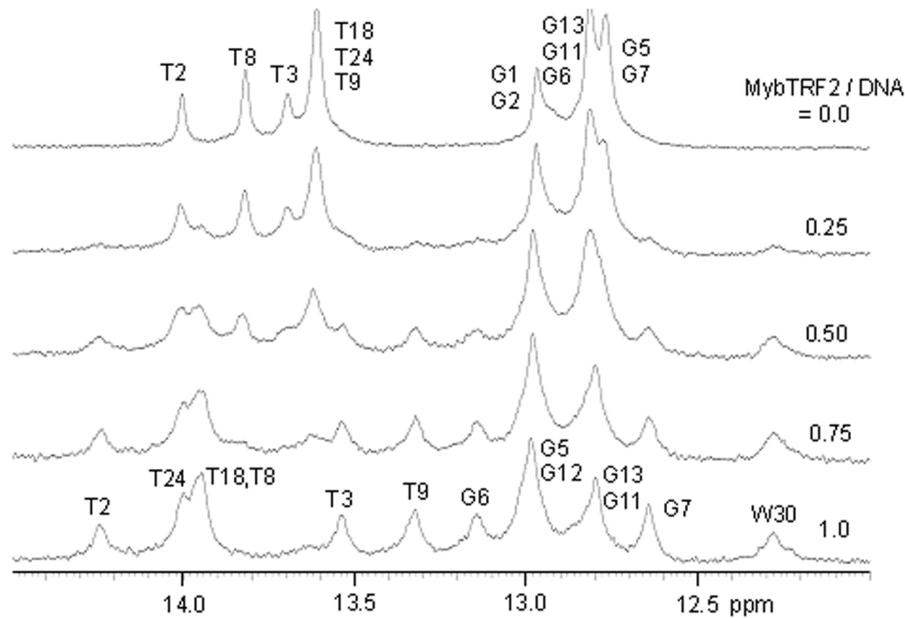


Figure 60: Suivi de l'évolution des variations de déplacements chimiques des protons iminos de l'ADN télomérique lors de l'ajout du domaine Myb de TRF2.

Quatre additions successives de portions de protéine Myb sur l'ADN ont été nécessaires pour obtenir un complexe 1 : 1 comme le montre la figure ci-dessus. La concentration finale du complexe est de 1,5 mM. La concentration en NaCl du complexe est de 100 mM. Un ensemble de spectres NOESY homonucléaires ont alors été enregistrés à 25°C afin de réaliser l'étude RMN du complexe.

### 3 Etude RMN de l'ADN télomérique complexé

L'étude d'un complexe entre une protéine non-marquée et un ADN non-marqué est relativement difficile. Nous sommes obligés de travailler en RMN homonucléaire et les recouvrements entre les raies de résonance de la protéine et de l'oligonucléotide sont nombreux. Fort heureusement il existe une « zone » de déplacement chimique située entre 5 et 8 ppm sur les spectres WgNOESY où il n'existe pas ou très peu de connectivités appartenant à la protéine, celle-ci ne possédant pas de feuillets  $\beta$ . Les connectivités visibles dans cette zone appartiennent à l'ADN sous sa forme complexée. Cette zone correspond de plus à la zone nécessaire à la réalisation de l'attribution séquentielle des protons aromatiques  $H_8/H_6$  et  $H_1'$ .

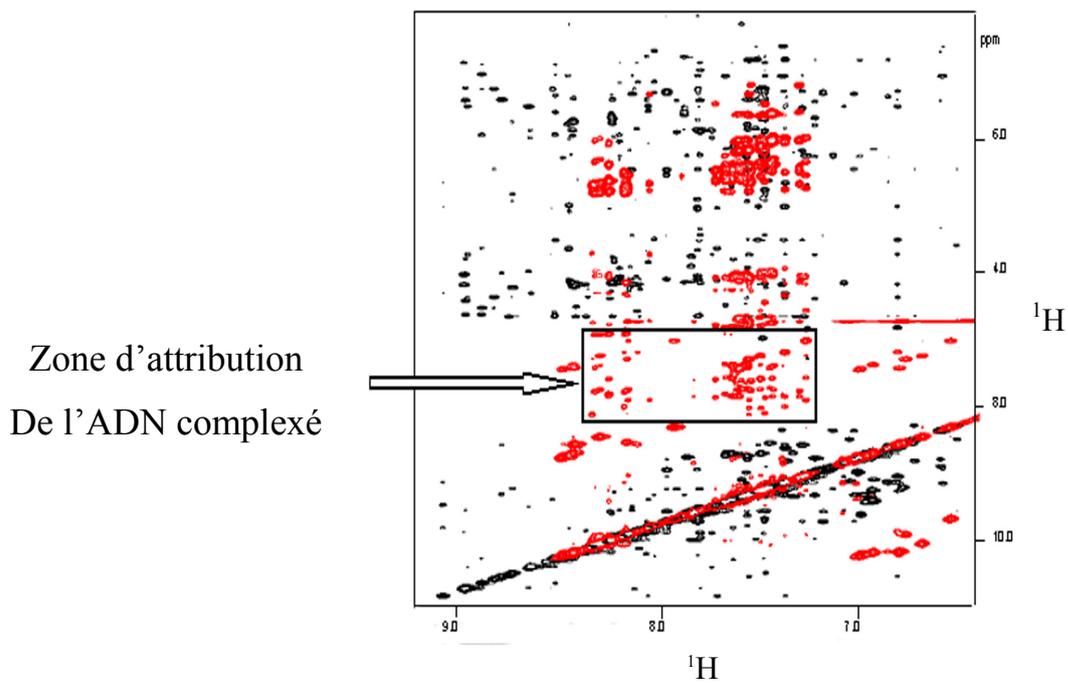


Figure 61: Superposition du spectre NOESY de l'ADN libre (rouge) et du domaine Myb de TRF2 libre (noir) à 25°C avec  $t_m=300ms$ .

L'attribution séquentielle de l'ADN complexé est présentée dans la figure ci-dessous :

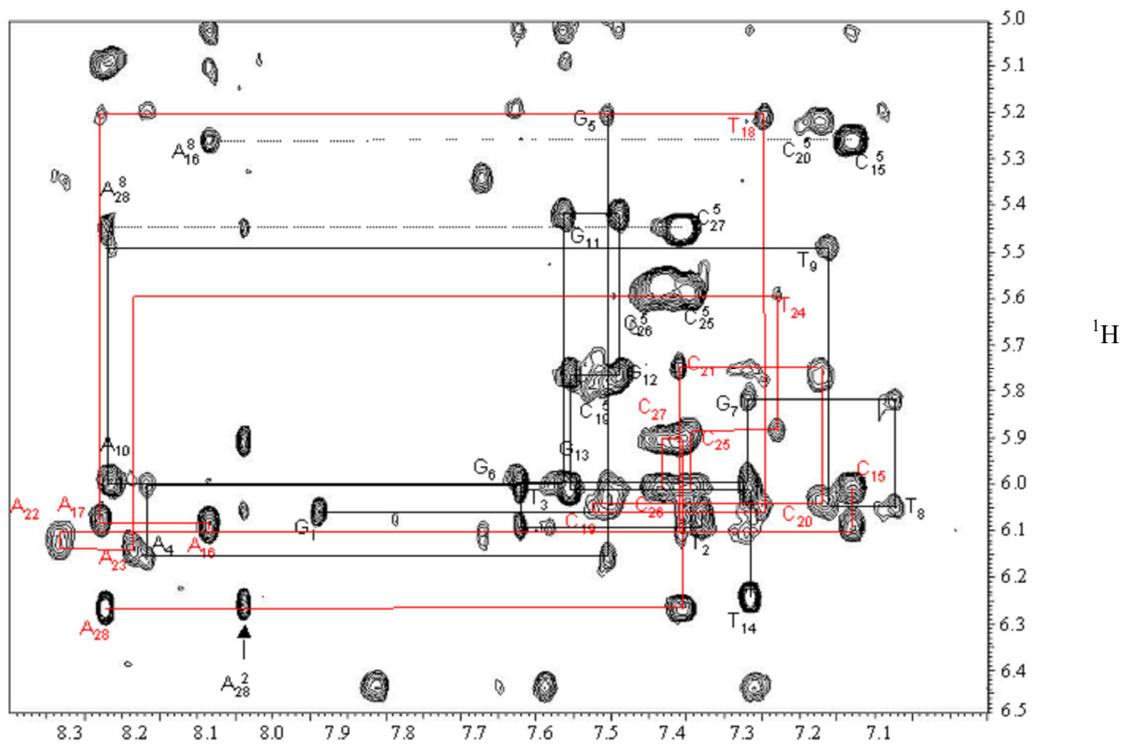


Figure 62 : attribution séquentielle de l'ADN complexé. Agrandissement d'un spectre NOESY avec  $t_m=150ms$  à 25°C.

L'analyse de l'ADN complexé montre que la complexation fait intervenir les deux répétitions de l'ADN télomérique. En effet les déplacements chimiques des résidus T3, A4, G5, G6, A17, T18, C19, C20 et C21 de la première répétition ainsi que les résidus T9, A22 et A23 appartenant à la seconde répétition sont les plus affectés par la complexation. Les différences de déplacements chimiques entre les formes libres et complexées de l'ADN sont reprises dans la figure suivante :

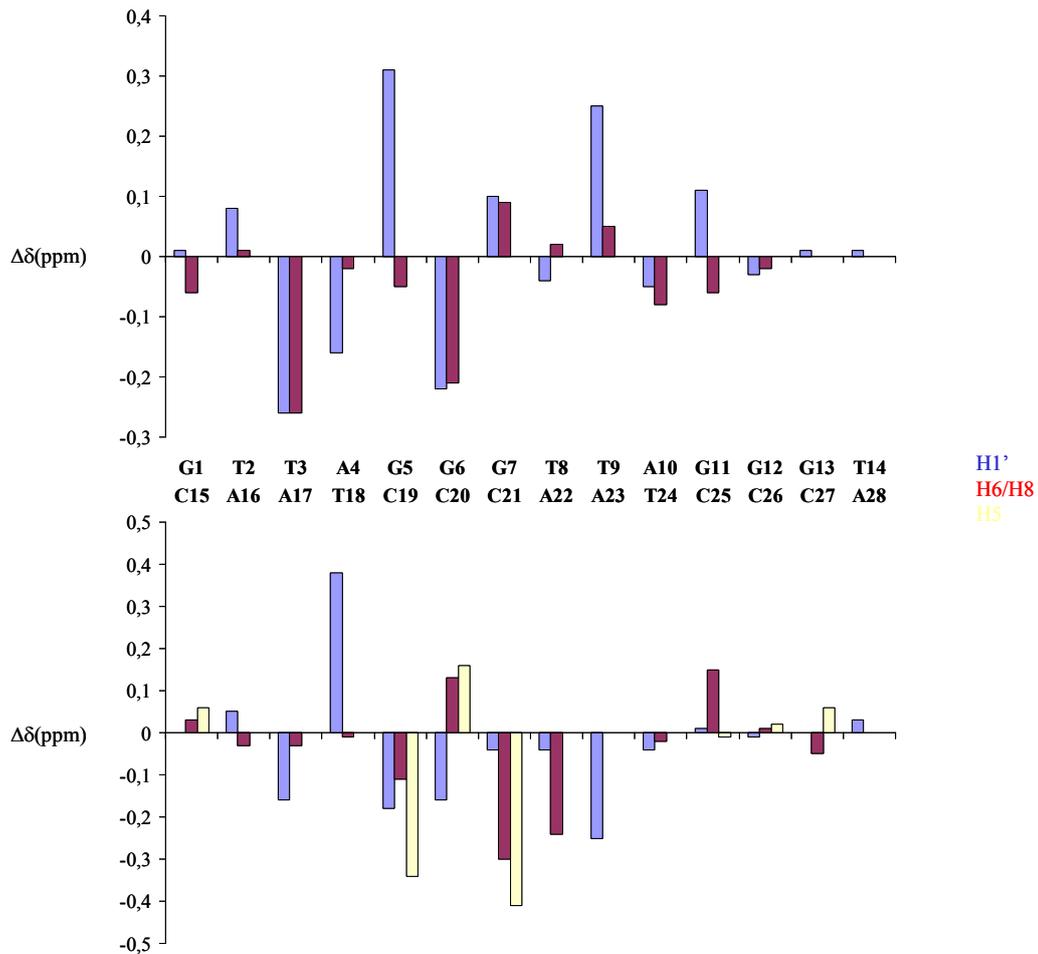


Figure 63 : Différences de déplacement chimique entre les formes libre et complexée de l'ADN télomérique.

Le motif de base pour la fixation du domaine Myb de TRF2 sur l'ADN télomérique est donc constitué de TAG<sub>3</sub>T<sub>2</sub>. La fixation du domaine Myb de TRF2 fait donc intervenir les deux répétitions télomériques.

Pour la protéine, les résidus qui ont subi les plus importantes variations de leurs déplacements chimiques sont les résidus suivants : Trp17, Trp37, Arg49, Val52, Met53, Lys55, Asp56, Arg57, Arg59, Thr60, Arg63. Ces résidus sont exposés au solvant. On peut donc déduire que la fixation fait surtout intervenir la troisième hélice.

#### 4 Limitation de cette étude

L'utilisation de deux molécules non-marquées pose un certain nombre de problèmes pour l'étude RMN. En effet, la superposition des signaux de l'ADN et de la protéine rend toute étude structurale très difficile. De plus sur les spectres NOESY, la différenciation entre les « noe » intra et inter moléculaires ADN-protéine est vraiment très délicate, voire impossible.

#### 5 Conclusion et discussion

La réalisation de ce premier complexe a permis d'étudier la fixation du domaine Myb de la protéine TRF2 sur l'ADN télomérique. La forte fixation de la protéine sur l'ADN fait intervenir les deux répétitions télomériques et semble surtout impliquer la troisième hélice de la protéine.

Nous avons, à cette même époque, pu obtenir le domaine Myb de la protéine TRF2 doublement marquée à l'azote 15 et au carbone 13. Notre première idée était alors l'étude du même complexe faisant intervenir un fragment d'ADN télomérique double brin comportant deux répétitions télomériques ainsi que le domaine Myb-TRF2 mais cette fois-ci doublement marqué  $^{15}\text{N}$  et  $^{13}\text{C}$  afin d'étudier plus en détails la structure de la protéine complexée à l'ADN. La parution de deux articles publiés dans les revues Protein Science (Hanaoka, Nagadoi et al. 2005) et Embo Reports (Court, Chapman et al. 2005) ont largement modifié notre projet. Des études par RMN et Radiocristallographie de complexes impliquant le domaine Myb de TRF2 et des fragments d'ADN télomérique comportant 2 (RMN) ou 2,5 (RX) répétitions télomériques y sont présentées. Une comparaison de la fixation sur l'ADN télomérique du domaine Myb de TRF1 et la fixation du domaine Myb de TRF2 y est faite. Il a été montré que la fixation de Myb-TRF2 sur l'ADN ne pouvait seule permettre de comprendre les différences d'activités entre TRF1 et TRF2 et que l'étude du complexe ne permettait pas de comprendre l'implication de TRF2 dans la formation des loops. Ces résultats sont en accord avec nos observations.

Suite à cela, nous avons tout de même décidé de continuer l'étude des télomères en utilisant le domaine Myb de la protéine TRF2 doublement marqué  $^{15}\text{N}$  et  $^{13}\text{C}$  (c'était notre première protéine doublement marquée) et en étudiant sa fixation sur un ADN particulier, possédant une partie double brin et une extrémité simple brin.

# Chapitre 9 : Réalisation d'un second complexe télomérique

## 1 Présentation du complexe

Nous savons maintenant que le domaine Myb de TRF2 se fixe sur l'ADN télomérique double brin, et que deux domaines Myb peuvent se fixer si l'ADN télomérique comporte 2,5 répétitions télomériques. Ces études ont été réalisées sur des ADN totalement double brin. Néanmoins il a été suggéré par plusieurs équipes de biologistes que TRF2 pourrait se fixer aussi au niveau de la jonction double brin/simple brin comme le propose Rachel M. Stansel: « TRF2 can behave as a structure-dependant telomere binding protein with a strong tendency to interact near or at the telomeric ss/ds junction at the telomere terminus » (Stansel, de Lange et al. 2001). C'est ce que nous avons voulu vérifier avec l'étude de ce complexe. Notre idée était de représenter la jonction existant au passage de l'ADN télomérique double brin à l'ADN télomérique simple brin. L'objectif était de savoir où le domaine Myb de la protéine TRF2 allait se fixer, s'il allait venir se fixer sur la première répétition télomérique ou s'il allait préférer la jonction double brin/simple brin d'ADN, ou allions nous observer la fixation de deux domaines Myb de TRF2 sur l'ADN ?

L'ADN utilisé pour la réalisation du complexe est le suivant :

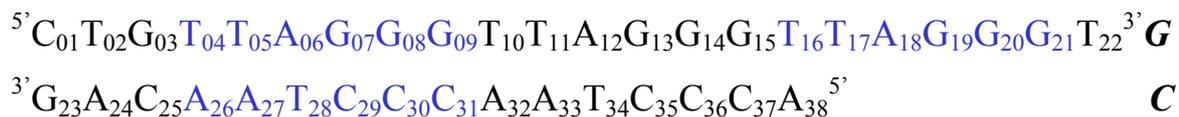


Figure 64: Séquence d'ADN comportant une partie double brin et une partie simple brin

Cet ADN comporte deux répétitions télomériques double brin mais aussi une partie simple brin en 3' comprenant une répétition télomérique. Ce fragment d'ADN est composé de 38 résidus et est totalement non-marqué. Sa masse moléculaire est d'environ 12 kDa. Les deux

premières paires de bases (C1-G23 et T2-G24) empêchent l'obtention de formes non souhaitées comme des polymères d'ADN ou le glissement du brin C sur le brin G. La protéine utilisée est le domaine Myb de la protéine TRF2 comprenant toujours 11 résidus en plus du domaine de fixation à l'ADN, mais doublement et uniformément marquée  $^{15}\text{N}$  et  $^{13}\text{C}$ .

## 2 Etude RMN du duplexe d'ADN

### 2.1 Préparation du duplexe

Les deux brins seuls d'ADN ont été achetés chez Eurogentec. Ceux-ci ont été purifiés par HPLC en phase inverse et par chromatographie échangeuse d'ions. Les brins d'ADN seuls ont ensuite été passés sur Chelex 100 afin d'éliminer les ions divalents. Le duplexe d'ADN a été formé comme précédemment en suivant, par RMN 1D, l'apparition des raies de résonance des protons iminos des thymines et guanines impliquées dans des liaisons hydrogène. Un mélange équimolaire 1 : 1 de chaque brin a ainsi pu être obtenu. Le pH a été ajusté à 5,5 et 100mM de NaCl ont été ajoutés afin de stabiliser le duplexe. L'échantillon a alors été lyophilisé, dissous dans un mélange  $\text{H}_2\text{O}/\text{D}_2\text{O}$  avant d'être chauffé au bain marie à  $90^\circ\text{C}$  et d'être très lentement refroidi jusqu'à température ambiante. La concentration finale du tube RMN était d'environ 0,85mM.

### 2.2 Etude RMN du duplexe

L'étude RMN du duplexe d'ADN a été réalisée de façon classique à l'aide de spectres NOESY, TOCSY et DQF-COSY.

L'attribution séquentielle a pu être réalisée. L'agrandissement de la région des protons aromatiques H6/H8-H1' enregistré à  $25^\circ\text{C}$  avec un temps de mélange de 300ms est présenté dans la figure ci-dessous :

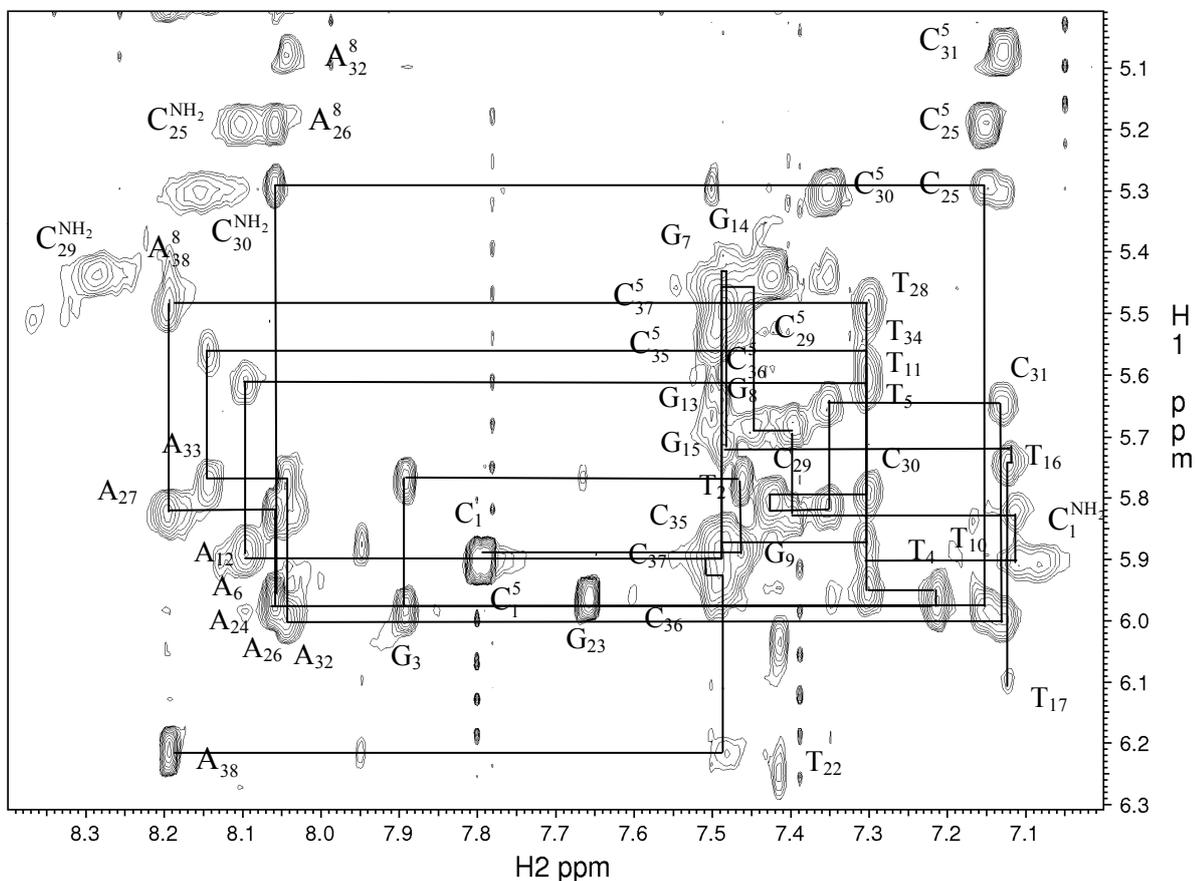


Figure 65: Agrandissement de la région aromatique H8/H6 – H1'/H5 d'un spectre NOESY à 25°C avec un temps de mélange de 300ms.

L'attribution séquentielle réalisée sur la partie aromatique H6/H8 - H1' et sur la partie H6/H8 –H2'/H2'' montre que l'ensemble de la partie double brin du duplexe adopte une conformation en double hélice droite correspondant à un ADN de type B.

L'attribution des protons iminos des guanines et thymines montre que l'ensemble des bases constituant le duplexe est apparié et que les appariements sont de type Watson et Crick. Notons tout de même des intensités NOEs faibles concernant les corrélations entre le H5 et les protons aminos de la cytosine C32 . De même le proton imino de la thymine T16 n'a pu être identifié montrant une certaine dynamique au niveau de cette thymine induite par l'extrémité simple brin de l'ADN. Concernant la partie simple brin , seules quelques connectivités H6/H8 – H1' ont pu être identifiées.

### 3 Formation du complexe

La formation du complexe a été suivie par RMN. Nous avons pour cela réalisé des spectres <sup>15</sup>N-HSQC nous permettant d'observer les modifications de déplacements chimiques des protons amidos de la protéine après chaque ajout d'ADN. Les corrélations <sup>1</sup>H-<sup>15</sup>N, d'abord fines, se sont ensuite élargies, à cause des phénomènes d'échanges, avant de nouveau s'affiner. Quatre ajouts d'ADN ont été réalisés afin d'obtenir un complexe 1-1.

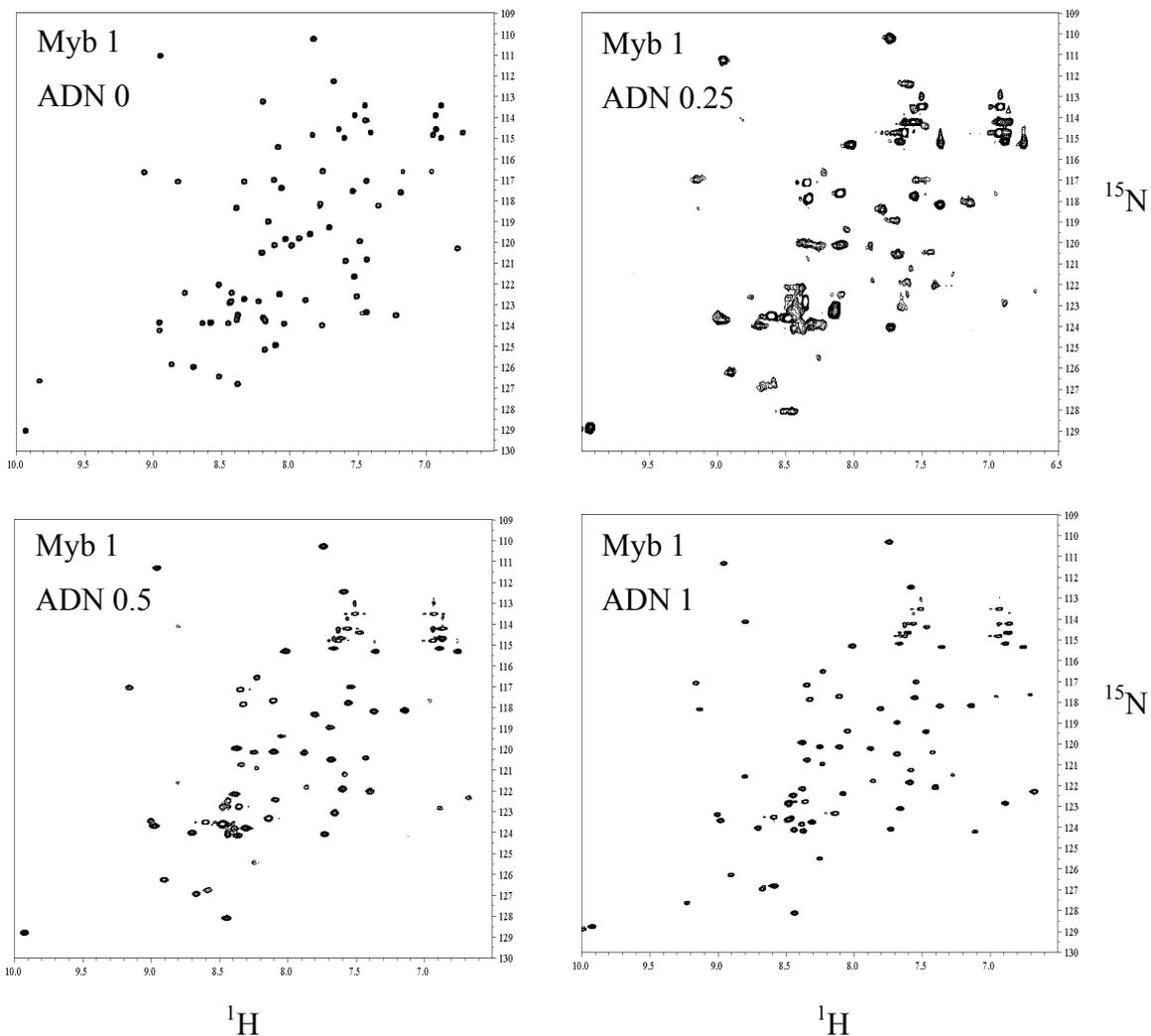


Figure 66: Spectres  $^1\text{H}$ - $^{15}\text{N}$ -HSQC permettant la titration de la protéine par l'ADN

## 4 Etude RMN de la protéine complexée

### 4.1 Attribution de la chaîne peptidique de Myb complexée

La protéine étant marquée  $^{15}\text{N}$  et  $^{13}\text{C}$ , son étude RMN a été réalisée à l'aide de spectres tridimensionnels et de la stratégie évoquée précédemment. Ainsi les expériences  $^{15}\text{N}$ -HSQC, HNCA, HN(CO)CA, CBCA(CO)NH et HNCBCA ont été utilisées pour attribuer les résonances de la chaîne peptidique. L'expérience HBHA(CO)NH s'est révélée particulièrement utile pour l'attribution des protons  $\text{H}\beta$ .

Le traitement et l'analyse de ces expériences ont permis d'attribuer la plus grande partie des déplacements chimiques des atomes du squelette peptidique constituant le domaine Myb complexé. Les résidus dont certaines résonances du squelette peptidique n'ont pu être attribuées se situent au début de la séquence, dans la partie non-structurée de la protéine. Pour les autres résidus, la majeure partie des raies de résonance HN, N, C $\alpha$ , H $\alpha$  a pu être attribuée. Le spectre  $^{15}\text{N}$ -HSQC du domaine Myb de TRF2 est présenté dans la figure ci-dessous :

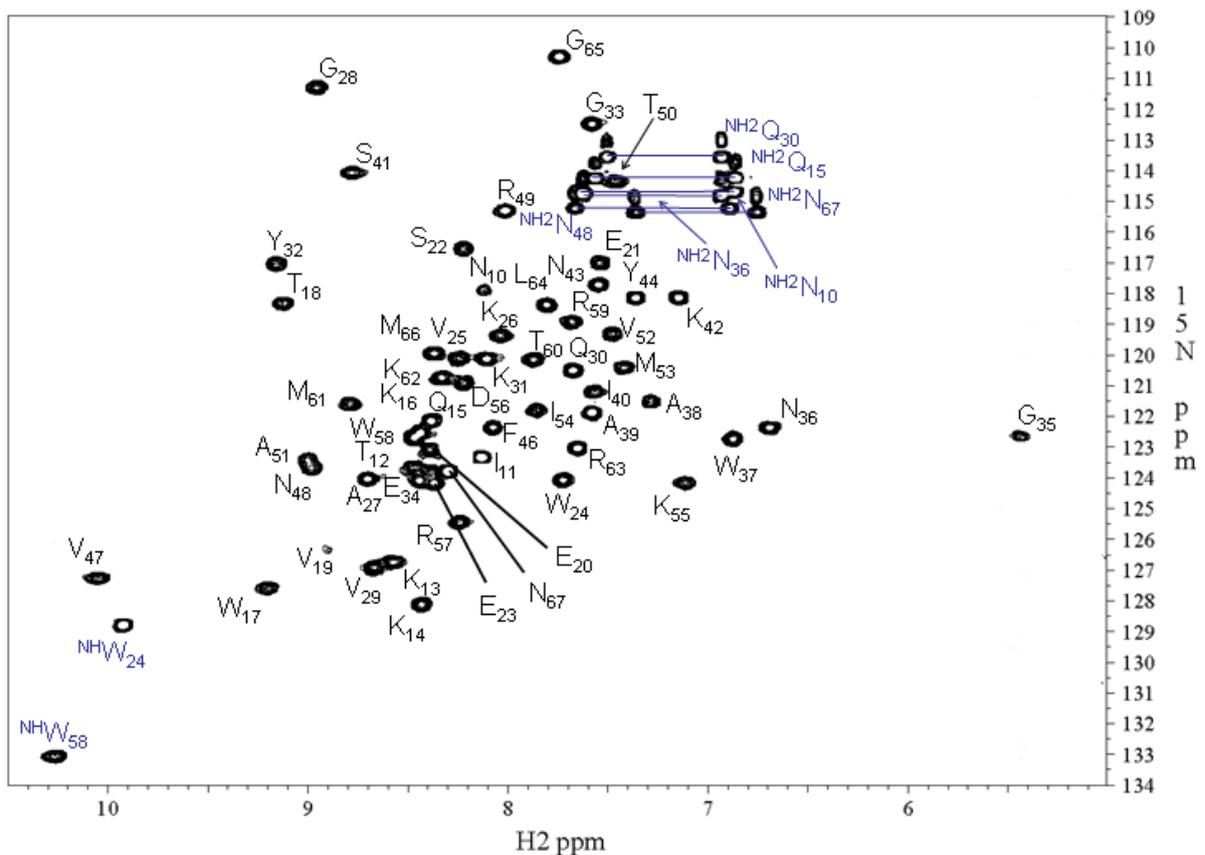


Figure 67: Spectre  $^{15}\text{N}$ -HSQC du domaine Myb de TRF2 lié à l'ADN télomérique.

#### 4.2 Analyse des variations de déplacements chimiques du $^1\text{H}$ N et du $^{15}\text{N}$

La comparaison des spectres  $^{15}\text{N}$ -HSQC avant et après complexation est une technique classique pour détecter la fixation d'un ligand sur une protéine. L'observation des spectres  $^{15}\text{N}$ -HSQC permet de déterminer les acides aminés impliqués subissant une modification de leur environnement en présence de l'ADN télomérique. La variation observée des déplacements chimiques  $^1\text{H}$  et  $^{15}\text{N}$  peut être évaluée et normalisée en utilisant la notion de déplacement quadratique moyen :

$$\sqrt{\frac{\Delta^1HN_{ppm}^2 + \Delta^{15}N_{ppm}^2}{2}}$$

Où  $\Delta^1H_{ppm}$  et  $\Delta^{15}N_{ppm}$  représentent les variations de déplacement chimique pour le proton et l'azote amido.

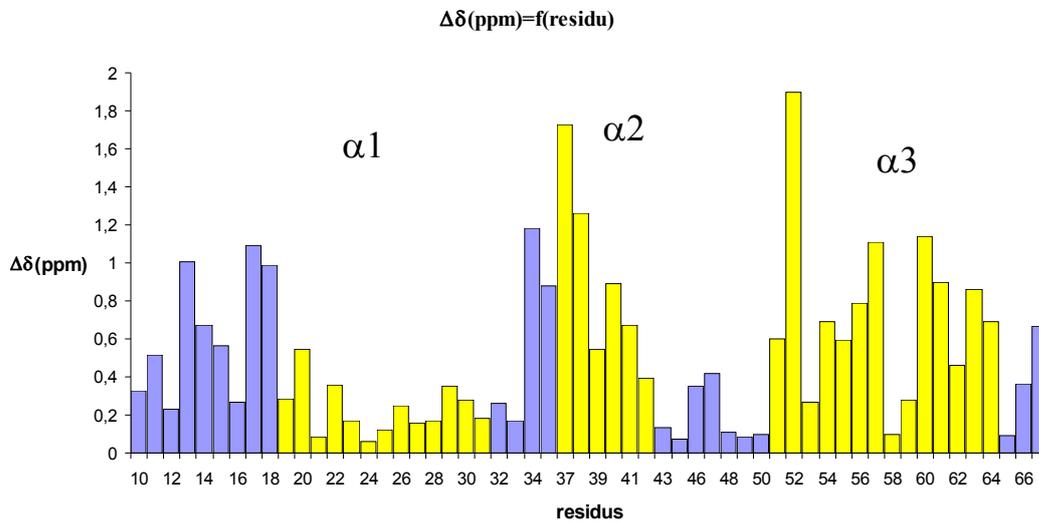


Figure 68 : Variations de déplacements chimiques quadratiques moyens entre les formes libre et complexée à l'ADN du domaine Myb de TRF2. En jaune sont représentés les résidus appartenant aux hélices  $\alpha$ .

Remarquons que le déplacement chimique quadratique moyen n'est relié ni qualitativement ni quantitativement à l'interaction directe avec l'ADN. En effet, ces variations correspondent à des variations d'environnement du squelette peptidique. Les interactions ADN-protéines sont plus susceptibles de se réaliser avec les chaînes latérales des acides aminés. Les modifications observées sont donc des conséquences secondaires de l'interaction.

Néanmoins des changements relativement importants ont pu être observés. Ceux-ci concernent, hormis pour les résidus situés dans la partie non-structurée, des acides aminés appartenant au premier coude (34-36) ainsi qu'aux hélices  $\alpha$  2 (37-42) et  $\alpha$  3 (52 à 64) du domaine Myb de TRF2.

### 4.3 Attribution des chaînes latérales

L'attribution des chaînes latérales a été réalisée à l'aide des expériences HCCH-COSY, HCCH-TOCSY, HCC(CO)NH,  $^{15}\text{N}$ -TOCSY-HSQC,  $^{15}\text{N}/^{13}\text{C}$ -NOESY-HSQC et  $^1\text{H}$ -WgNOESY.

La qualité médiocre des spectres a rendu l'attribution des cartes RMN relativement difficile. Néanmoins, environ 85% des protons non-échangeables ont pu ainsi être attribués. En comptant seulement les résidus 14-67, le pourcentage de protons non-échangeables attribués s'élève à 91%.

### 4.4 Modélisation moléculaire du domaine Myb de TRF2 complexé

Un ensemble d'angles  $\psi$  et  $\phi$  pour le squelette a pu être prédit à l'aide du logiciel Talos. Ainsi si 9 des 10 valeurs observées dans la librairie de protéines coïncident avec la valeur prédite du couple d'angles  $\psi$  et  $\phi$ , alors celui-ci est considéré comme fortement probable. Les valeurs de 42 couples d'angles ( $\psi, \phi$ ) ont été prédites et utilisées dans la modélisation moléculaire de la protéine complexée. L'incertitude sur ces angles a été ajustée à  $\pm 10^\circ$ , le logiciel Talos permettant parfois des incertitudes plus faibles non réalistes.

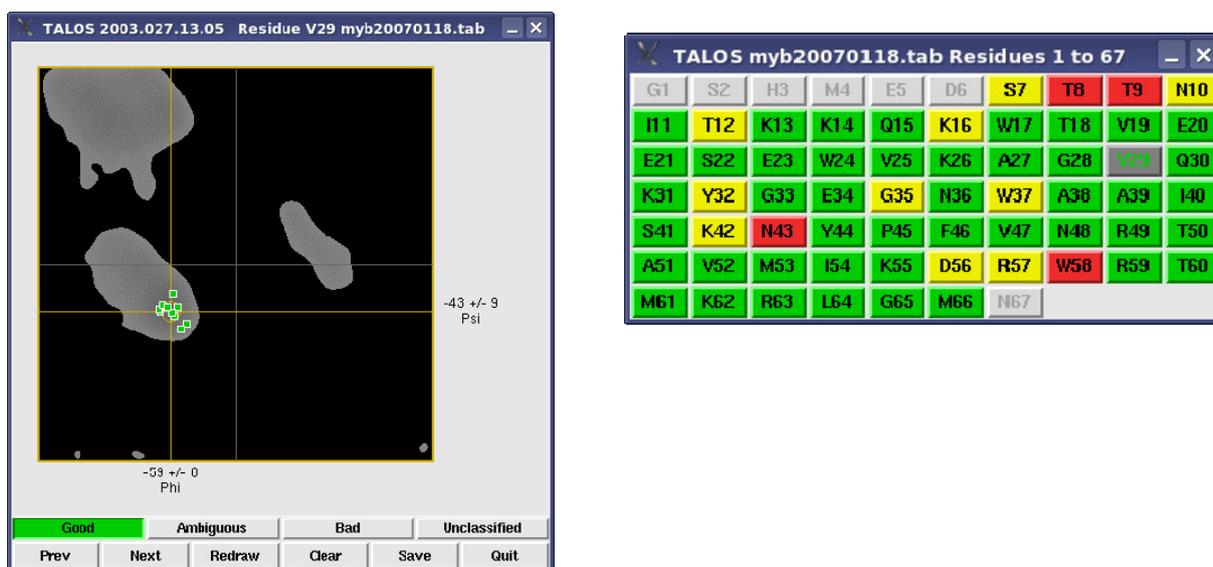


Figure 69: tableaux obtenus avec le logiciel Talos. En haut la carte de Ramachandran illustrant les angles dièdres du résidu central des dix protéines de la banque de données du logiciel Talos ayant des déplacements chimiques similaires à ceux du tripeptide de notre protéine. En bas, tableau montrant la séquence de la protéine, en vert les résidus où TALOS

*suggère un couple d'angles dièdres avec une confiance élevée, en jaune des prédictions ambiguës et en rouge des prédictions contradictoires.*

Le calcul de la structure du domaine Myb de TRF2 a été réalisé avec le logiciel CYANA 2.1. La liste de déplacements chimiques, les angles dièdres prédits avec le logiciels Talos ainsi que les listes de pics NOEs obtenus par « peak-picking » avec le logiciel NMRView ont constitué les données initiales du calcul. Les tolérances de déplacements chimiques ont été fixées à 0,05ppm pour la dimension  $^1\text{H}$  et à 0,7 ppm pour les dimensions  $^{15}\text{N}$  ou  $^{13}\text{C}$ . Le calcul a été réalisé avec 10000 pas de dynamique. Un ensemble de 200 structures a été généré et les 20 structures présentant les fonctions cibles les plus faibles ont été sélectionnées. Nous avons ensuite réalisé, comme pour le domaine Myb seul, un raffinement de la structure dans une boîte d'eau en utilisant le logiciel ARIA.

#### 4.5 Structure obtenue

Les résultats de la modélisation moléculaire peuvent être observés dans le tableau suivant :

##### Contraintes

###### *NOE*

|   |                             |
|---|-----------------------------|
| Total   | 718                         |
| Intra-résidus [ $i = j$ ]                     | 165                         |
| Séquentielles [ $ i - j  = 1$ ]               | 243                         |
| Moyenne distance [ $1 <  i - j  < 5$ ]        | 160                         |
| Longue distance [ $ i - j  > 5$ ]             | 150                         |
| <i>Dièdres</i>                                | 82                          |
| <i>Nombre moyen de contraintes par résidu</i> | 13,7                        |
| <i>Nombre de structures calculées</i>         | 200                         |
| <i>Nombre de structures sélectionnées</i>     | 20                          |
| <i>Fonction cible Cyana</i>                   | $1,3 \pm 0,2 \text{ \AA}^2$ |
| <i>Nombre de violation noe &gt; 0,5\AA</i>    | 0                           |

##### Cartes de Ramachandran

|                    |       |
|--------------------|-------|
| Régions favorables | 91,6% |
| Régions autorisées | 8,4%  |
| Régions interdites | 0%    |

##### RMSD CYANA

|  |                               |
|--|-------------------------------|
| Sur tous les résidus (atomes lourds)                       | $5,11 \pm 0,12 \text{ \AA}$   |
| Sur les résidus de la région ordonnée<br>(Résidus 19 à 64) | $0,80 \pm 0,08 \text{ \AA}$   |
| <u>Affinement dans une boîte d'eau ARIA</u>                |                               |
| Energie totale moyenne                                     | $-2161 \text{ kcal.mol}^{-1}$ |
| RMSD sur les résidus de la région ordonnée                 | $0,86 \pm 0,09 \text{ \AA}$   |

Figure 70: Résultats de la modélisation moléculaire de la protéine dans le complexe

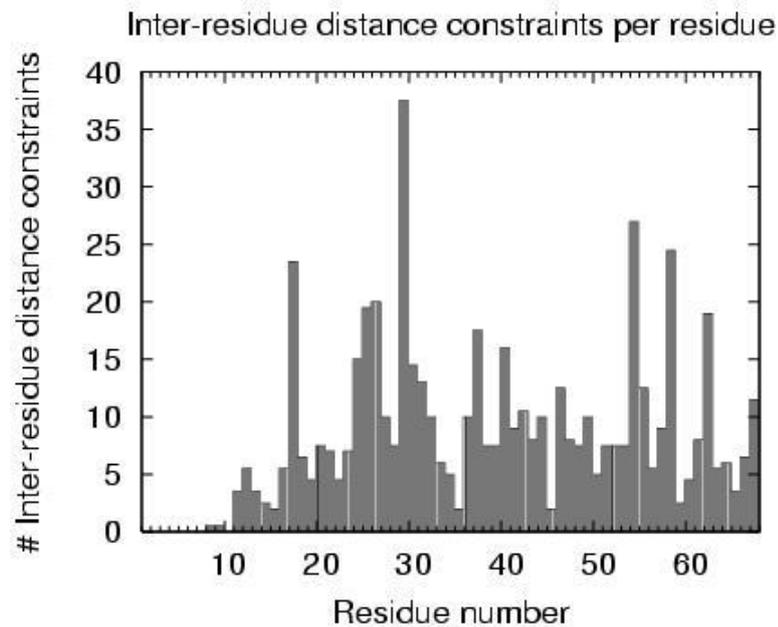


Figure 71: Nombre de contraintes par Résidu

La structure du domaine Myb complexé à l'ADN peut être observée dans les figures ci-dessous :

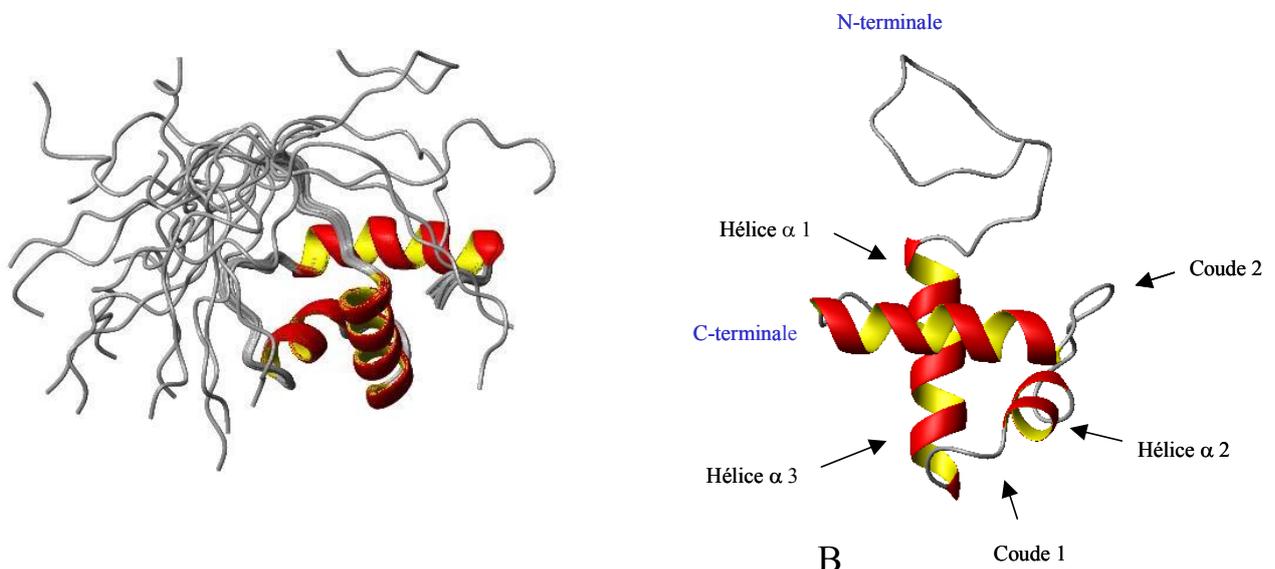


Figure 72: (A) Superposition des 20 meilleures structures obtenues pour le domaine Myb complexé à l'ADN; (B) : Présentation d'une structure du domaine Myb complexé à l'ADN.

#### 4.6 Analyse de la structure du domaine Myb de TRF2 complexé à l'ADN télomérique.

La structure hélice-tour-hélice de la protéine est conservée. On retrouve au centre de la protéine le cœur hydrophobe formé des résidus TRP17, VAL25, ALA29, TRP37, ILE40, PHE46, ALA51, ILE54, TRP58 et MET61. La première hélice  $\alpha 1$  correspond aux résidus 19 à 32. La seconde hélice  $\alpha 2$  correspond aux résidus 37 à 41 et la troisième hélice  $\alpha 3$  aux résidus 51 à 64.

Concernant les variations de déplacements chimiques qui ont pu être observées entre les formes libre et complexée de Myb, certaines peuvent être expliquées. Au niveau des résidus situés dans le coude  $\gamma$ , le déplacement chimique particulier du HN de la glycine 35 est expliqué par les courants de cycle qu'il subit en particulier ceux du cycle aromatique du tryptophane 37. Pour la lysine 55, la variation de déplacement chimique est sûrement due à son changement de position par rapport au cycle aromatique du tryptophane 37. Enfin les résidus dont les variations de déplacements chimiques sont les plus importantes sont situés dans la troisième hélice, ils correspondent aux résidus VAL52, ILE54, LYS55, ASN56, ARG57, THR60 et MET61. Ces résidus sont exposés au solvant et doivent donc participer activement à la complexation.

## 5 Etude de l'ADN complexé

L'étude de l'ADN complexé a pu être réalisée à l'aide de spectres NOESY filtré  $^{15}\text{N}$  et  $^{13}\text{C}$  et de spectres Wg-NOESY  $^1\text{H}$ - $^1\text{H}$  du complexe.

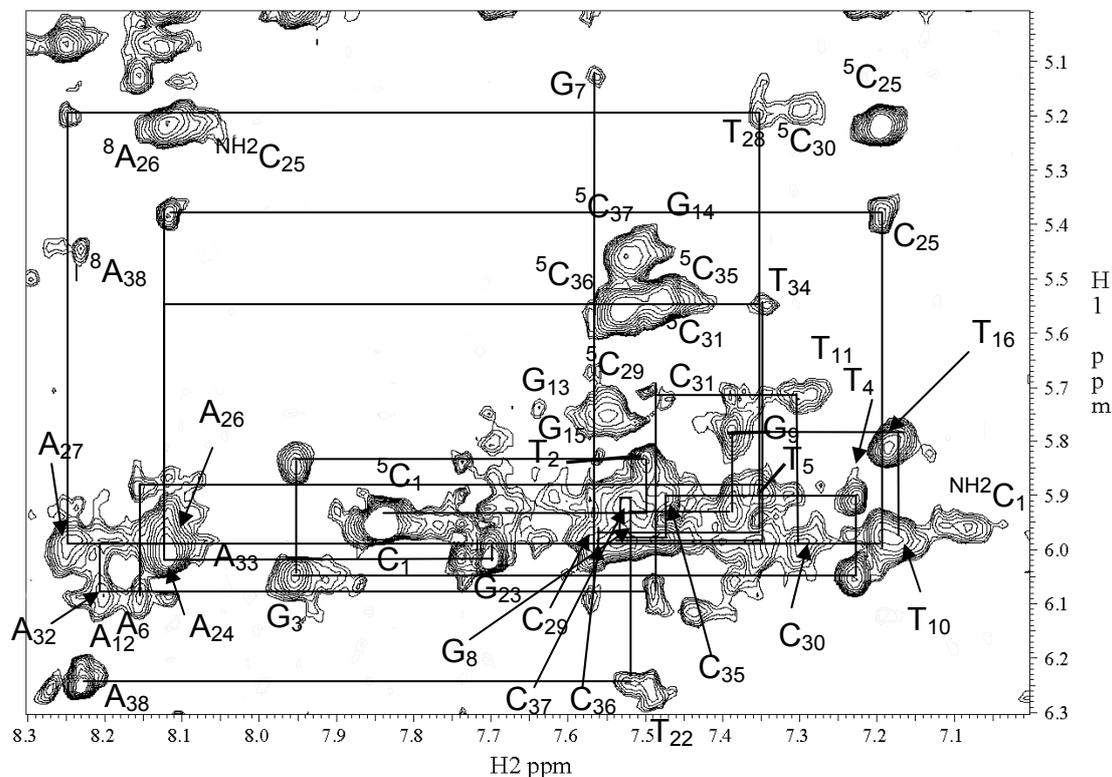


Figure 73: Attribution séquentielle. Agrandissement de la partie Aromatique-H1' d'un spectre NOESY avec  $tm=150ms$

L'attribution séquentielle a été difficile mais de nombreuses informations ont pu être extraites. La mauvaise qualité des spectres s'explique dans notre cas par une concentration trop faible de l'échantillon ( $\sim 0,85\text{mM}$ ), mais aussi par la faible sensibilité des expériences demi-filtre. L'utilisation de spectre NOESY n'utilisant pas de filtre  $^{13}\text{C}$  ou  $^{15}\text{N}$  s'est révélée très utile. Concernant l'attribution, il faut noter l'absence de corrélation aromatique - H1' entre les résidus G7 et G8. L'observation des déplacements chimiques des différents résidus montrent que les résidus les plus perturbés sont : T5, A6, G7, T28, C29, C30, C31 pour la première répétition télomérique ainsi que les résidus A32, A33 et T34 appartenant à la seconde répétition télomérique. La protéine se fixe donc sur la partie double brin exclusivement au niveau de la première répétition, les déplacements chimiques des résidus C35, C36, C37 et

A38 n'étant pratiquement pas modifiés. Les différences de déplacements chimiques entre les formes libre et complexé de l'ADN télomérique sont présentées dans la figure suivante :

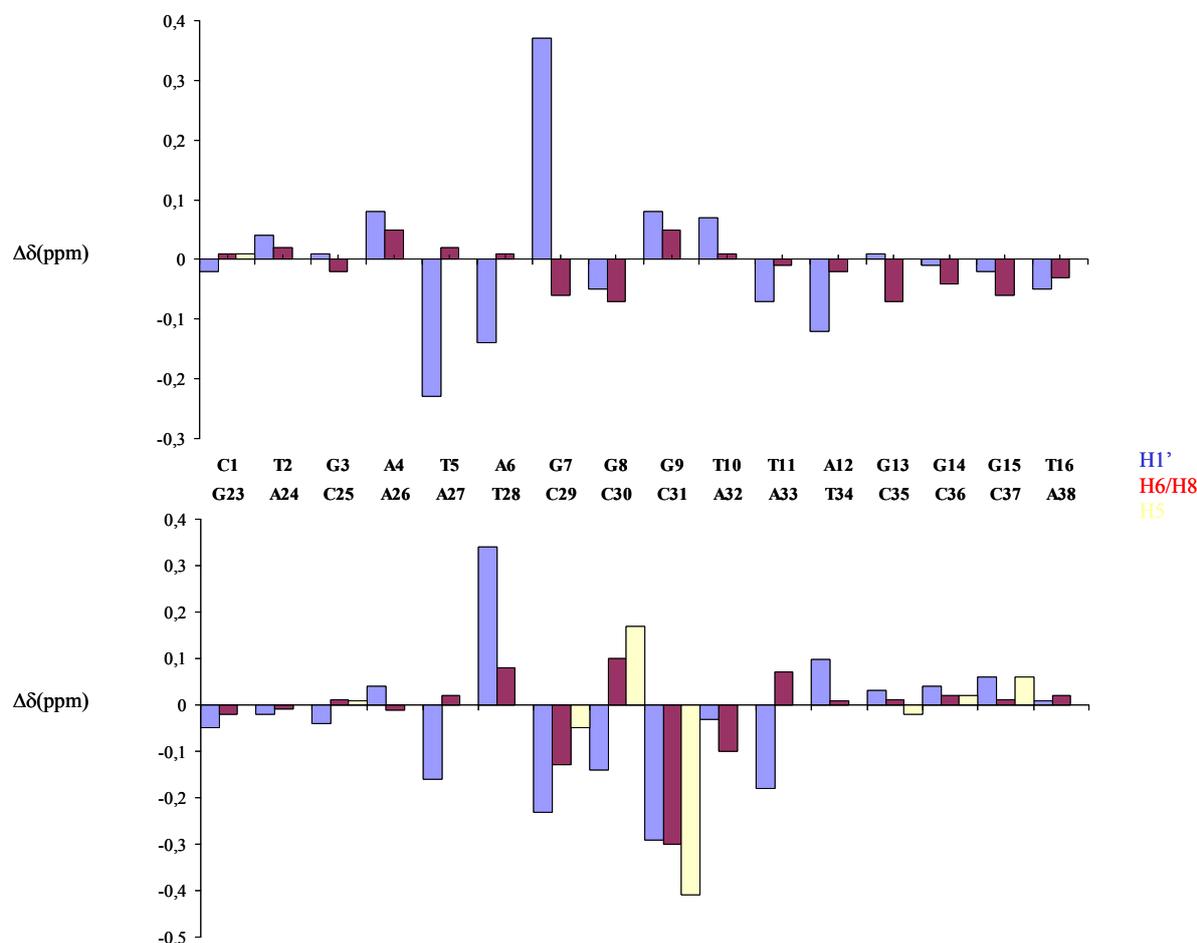


Figure 74: Différences de déplacements chimiques entre les formes libre et complexée de l'ADN

## 5.1 Etude des connectivités intermoléculaires

L'attribution des spectres demi-filtres  $^{13}\text{C}$  et/ou  $^{15}\text{N}$  NOESY a permis de mettre en évidence des connectivités intermoléculaires entre le domaine Myb de TRF2 et l'ADN télomérique et cela malgré la très faible sensibilité des expériences demi-filtres. Ainsi des connectivités entre les résidus LYS16( $\text{H}_\gamma$ ) et A32( $\text{H}_3'$ ), THY5( $\text{H}_3'$ ) et ALA38( $\text{H}_{\beta 1}$ ), entre les résidus CYT30( $\text{H}_5$ ) et VAL52( $\text{H}_{\gamma 1}$ ), entre les résidus THY5( $\text{CH}_3$ ) et VAL52( $\text{H}_{\gamma 2}$ ), les résidus THR60( $\text{CH}_3$ ) et ADE33( $\text{H}_2'$ ) ainsi qu'entre CYT31( $\text{H}_3'$ ) et MET53( $\text{H}_{\epsilon 1}$ ) ont pu être mises en évidence. Ces quelques interactions intermoléculaires ont permis de réaliser les premiers calculs de modélisation.

## 6 Modélisation du complexe

La modélisation de celui-ci a été réalisée à l'aide du logiciel CYANA version 2.1. Les données initiales comprenaient la séquence du complexe, dans laquelle sont définies les trois séquences de résidus constituant le complexe (les deux brins d'ADN et la protéine) liées entre elles par des pseudo-résidus (ou linker), les listes de déplacements chimiques de l'ADN complexé ainsi que de la protéine complexée. Un fichier comprenant les contraintes intermoléculaires a été généré et les distances intermoléculaires ont été classées selon leur intensité observée sur les différents spectres RMN.

Pour la protéine, une liste d'angles dièdres générée par Talos a aussi été utilisée.

Concernant l'ADN, les données introduites pour le calcul comprennent un spectre NOESY à 150 ms « peak-pické », une liste de déplacements chimiques et un fichier de contraintes d'angles dièdres. Les angles dièdres ont été contraints par des valeurs classiques d'une conformation en double hélice de type B.

Les spectres étudiés dans l'eau légère ont permis de montrer que l'ensemble du duplexe est apparié et que les appariements sont de type Watson et Crick. Ces appariements sont introduits, lors de la modélisation, sous forme de contraintes de distances pour chaque liaison hydrogène, soit 6 contraintes pour une paire GC et 4 contraintes pour une paire AT.

Le calcul a été réalisé sur 500 structures et les 10 meilleures structures ont été sélectionnées. Les tolérances de déplacements chimiques ont été fixées à 0,05 ppm pour la dimension  $^1\text{H}$  et à 0,5 ppm pour les dimensions  $^{15}\text{N}$  ou  $^{13}\text{C}$ . Le calcul a été réalisé avec 15000 pas de dynamique. Les dix meilleurs structures du complexe peuvent être visualisées dans la figure ci-dessous :

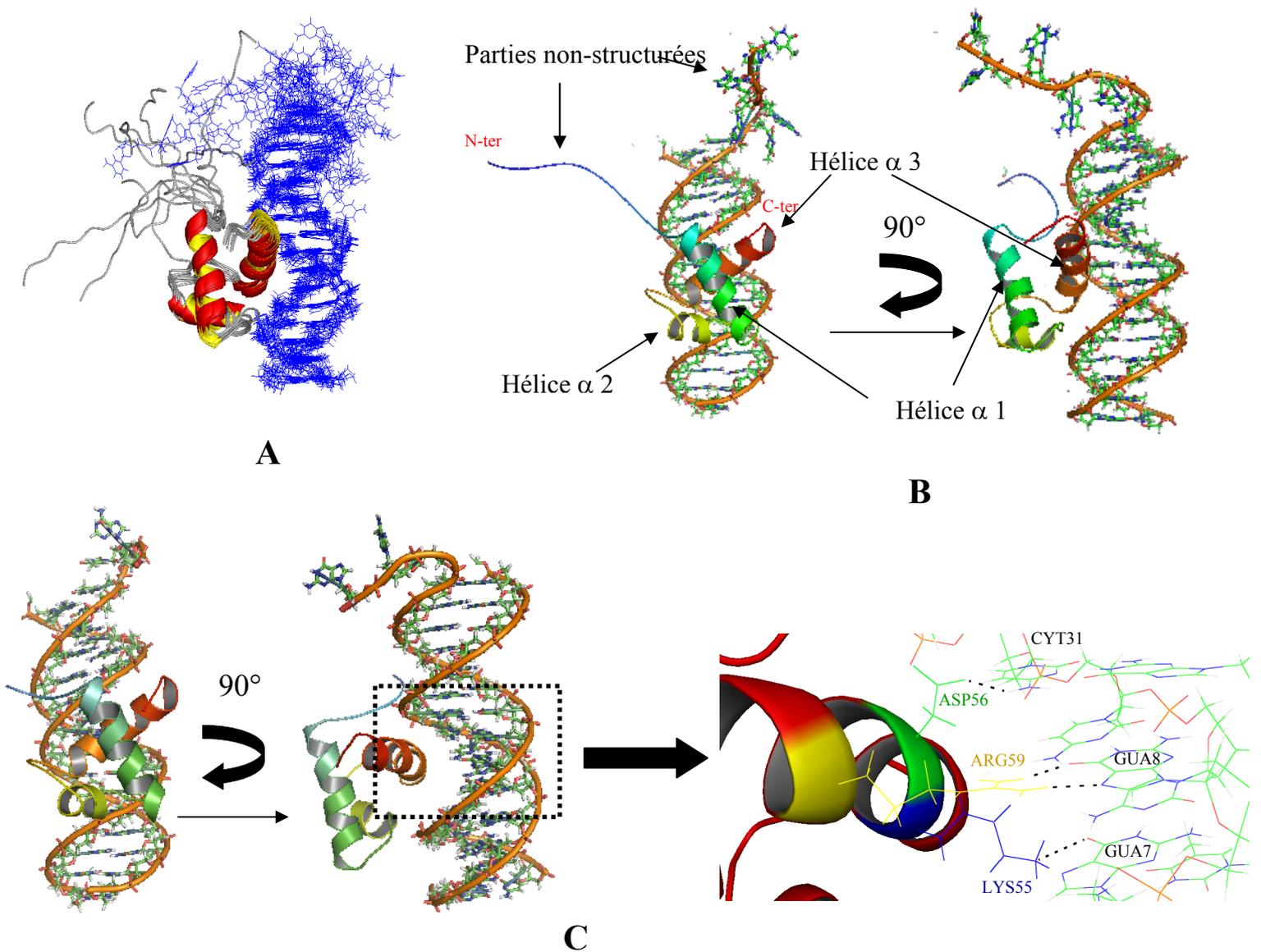


Figure 75 : (A) Superposition des dix meilleures structures obtenues au cours de la modélisation moléculaire (B) Représentation d'une structure du complexe choisie arbitrairement (C) Représentation simplifiée (les parties non-structurées de la protéine et de l'ADN ont été retirées) d'une structure du complexe et agrandissement de la zone d'interaction entre la troisième hélice  $\alpha$  et l'ADN télomérique. En noir sont représentées les liaisons hydrogène directes entre les différents résidus de la protéine et de l'ADN.

La fonction cible sur l'ensemble des 10 structures est de  $6,15 \pm 0,21 \text{Å}^2$  et le RMSD entre les 10 structures sur les résidus 16-67 pour la protéine et 3-15 de l'ADN est de  $1,36 \pm 0,12 \text{Å}$ . Néanmoins la modélisation de ce complexe avec le logiciel HADDOCK2.0 (Dominguez, Boelens et al. 2003) est en cours de réalisation afin de pouvoir intégrer les données RDCs concernant la protéine. Ce logiciel est une extension du logiciel CNS. Il permet de réaliser des opérations de docking entre protéine-protéine, ADN-protéine et protéine-ligand. Différentes

informations peuvent être utilisées comme les variations de déplacements chimiques observées sur les spectres RMN, les distances intermoléculaires ou les contraintes RDCs.

## 7 Analyse de la structure du complexe

La forme globale du complexe est très proche des autres complexes impliquant des domaines Myb (figure 77 A,B et C), la fixation du domaine Myb de TRF2 se réalisant essentiellement grâce à la troisième hélice  $\alpha$ . Celle-ci s'intercale dans le grand sillon de l'ADN télomérique. Les résidus 14 à 16 se lient spécifiquement dans le petit sillon de l'ADN télomérique. On remarque l'existence de liaisons hydrogène directes entre l'arginine 59 (NH1 et NH2) et la guanine 8 (N7 et O6), entre la lysine 14 et l'adénine 12 et l'adénine 107, entre la l'asparagine 56 et la cytosine 31, entre la lysine 55 et la guanine 7. D'autres liaisons hydrogène ont pu être mises en évidence, néanmoins les distances observées entre les atomes semblent indiquer l'implication de molécules d'eau. Ainsi des liaisons hydrogène impliquent la lysine 55 et la guanine 7, entre la lysine 14 et la thymine 11, entre le tryptophane 37 et les groupements phosphates de la thymine 5, entre l'arginine 59 et la guanine 7 (NZ) ainsi qu'entre l'arginine 57 et un groupement phosphate de la cytosine 30.

Au niveau de l'ADN télomérique, comme pour le premier complexe, on observe que la structure de l'ADN télomérique est une double hélice de type B ne présentant que très peu d'irrégularités.

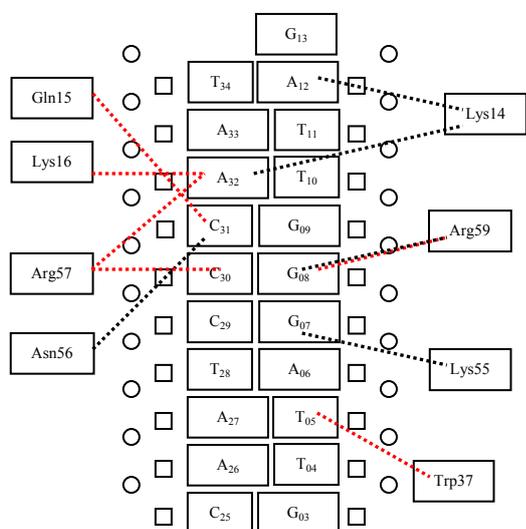


Figure 76 : Représentation des liaisons hydrogène directes (noir, distance :  $2,5\text{\AA}$  et angles autorisés :  $135^\circ$ ) et indirectes (rouge) entre le domaine Myb de TRF2 et le fragment d'ADN télomérique.

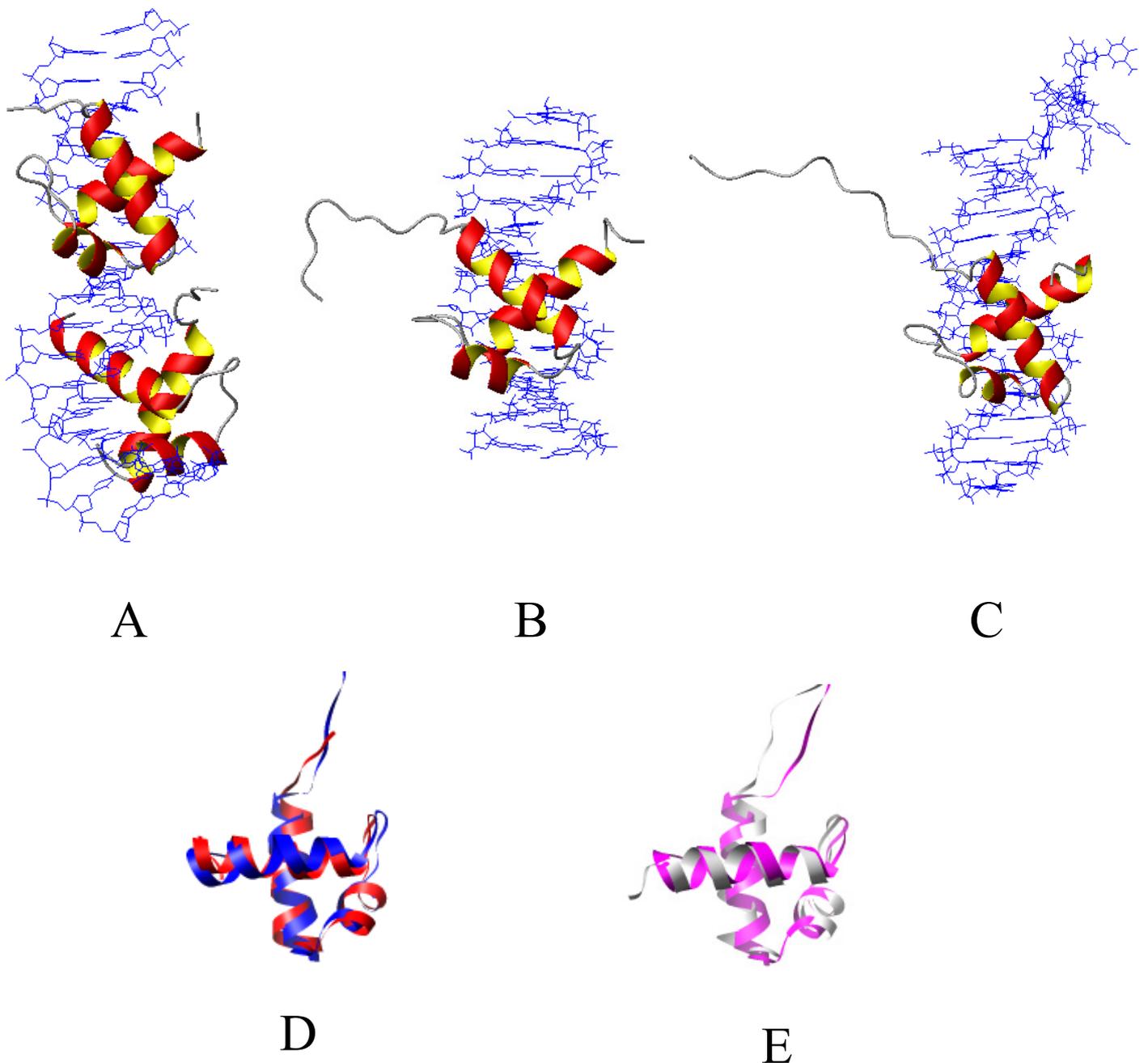


Figure 77 : (A) Structure tridimensionnelle du complexe entre deux domaines Myb et l'ADN télomérique obtenue par radiocristallographie (Court, Chapman et al. 2005)(code PDB :1w0u) (B) Structure tridimensionnelle du complexe obtenue par RMN entre le domaine Myb de TRF2 et l'ADN télomérique obtenue par l'équipe de Nishimura (Hanaoka, Nagadoi et al. 2005) (code PDB : 1vfc) (C) Structure tridimensionnelle choisie arbitrairement de notre complexe (D) Superposition de la structure du domaine Myb complexé obtenue par radiocristallographie (rouge) et par RMN (bleu). Le RMSD est de 1,18 Å sur 39 paires d'atomes du squelette peptidique (E) Superposition des structures du domaine Myb de TRF2 obtenues par notre équipe (rose) et par le groupe de Nishimura (grise). Le RMSD entre les deux structures est de 1,12 Å entre 42 paires d'atomes du squelette peptidique.

## 8 Raffinement de la structure du domaine Myb de TRF2 en utilisant les couplages dipolaires résiduels

Nous avons choisi comme milieu orienté le gel de polyacrylamide pour son faible coût et la possibilité de récupérer l'échantillon par diffusion.

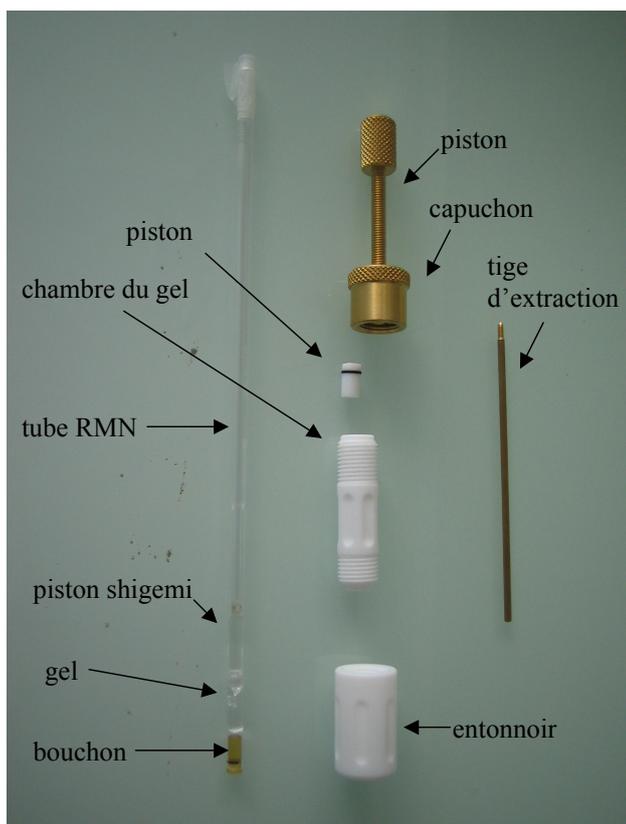
### 8.1 Synthèse du gel et préparation de l'échantillon

Nous avons préparé un gel de polyacrylamide à 7%. Pour cela nous avons utilisé les réactifs suivants (Chou, Gaemers et al. 2001) :

| Pour préparer 1000 $\mu$ l de gel :                  |             |
|--|-------------|
| Acrylamide (ou 2-propénamide) 4%                     | 168 $\mu$ l |
| Bisacrylamide 2%                                     | 175 $\mu$ l |
| Persulfate d'ammonium 5%                             | 5 $\mu$ l   |
| EDTA (dégradation des protéases) 500mM               | 20 $\mu$ l  |
| Eau  | 627 $\mu$ l |
| TEMED 10% (permet l'initiation de la polymérisation) | 5 $\mu$ l   |

*Tableau 5: Réactif utilisés pour préparer 1 ml de gel de polyacrylamide*

Pour la préparation du gel, nous avons utilisé l'appareillage présenté dans la figure ci-dessous :



*Figure78 : Appareil utilisé pour la préparation des milieux orientés*

Cette appareil est disponible à l'achat sur le site internet suivant :

[www.newera-spectro.com](http://www.newera-spectro.com)

La procédure pour la réalisation du gel est détaillée ci-dessous. De nombreux essais ont été nécessaires afin d'obtenir des résultats convenables. Nous avons aussi essayé les cristaux liquides comme milieu orienté mais sans grande réussite.

Protocole de préparation du tube RMN :

Le gel doit polymériser pendant plusieurs heures (au minimum 2h). On le laisse sécher pendant trois heures à l'étuve (40°C) ou bien pendant une nuit à température ambiante. Le gel est alors réhydraté avec la solution contenant la macromolécule.

Pour orienter la molécule dans le gel de polyacrylamide une contrainte mécanique est nécessaire soit par compression soit par élongation horizontale ou verticale du gel. Le sens de l'alignement ne s'effectue pas dans le sens du champ magnétique mais dans la direction de la contrainte mécanique appliquée. Dans notre cas, l'alignement de la molécule est réalisé en poussant le gel à travers l'entonnoir (figure ci-dessous), celui-ci subit alors une compression radiale.

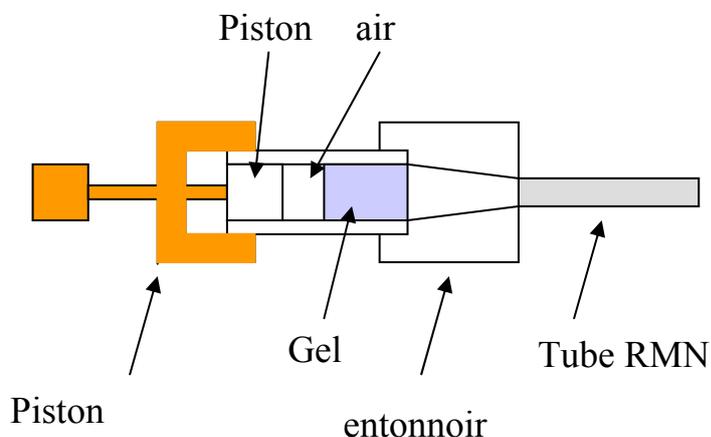


Figure 79 : dispositif permettant l'orientation de la molécule dans le gel

La concentration en complexe du gel était d'environ 0,5 mM.

## 8.2 Etude RMN en milieu orienté

Tout d'abord des spectres de la protéine complexée en solution, donc en milieu isotrope, ont été enregistrés : un spectre  $^{15}\text{N}$ -HSQC ainsi qu'un spectre  $^{15}\text{N}$ -HSQC non découplé pendant le temps incrémentable  $t_1$  ont été enregistrés. Ces spectres permettent de mesurer les constantes de couplage  $^{15}\text{N}$ - $^1\text{H}$ , qui peuvent alors évoluer librement pendant le temps incrémentable  $t_1$  (figure ci-dessous).

Dans le gel, c'est-à-dire en milieu faiblement anisotrope, on refait les mêmes spectres. Cette fois, on mesure de nouveau les couplages  $^{15}\text{N}$ - $^1\text{H}$ , correspondant alors aux couplages scalaires ainsi qu'aux couplages dipolaires résiduels ( $^1J_{^{15}\text{N-H}} + ^1D_{^{15}\text{N-H}}$ ). En soustrayant ces valeurs aux constantes de couplage mesurées en milieu isotrope ( $^1J_{^{15}\text{N-H}}$ ), on obtient les valeurs des couplages dipolaires résiduels ( $^1D_{^{15}\text{N-H}}$ ).

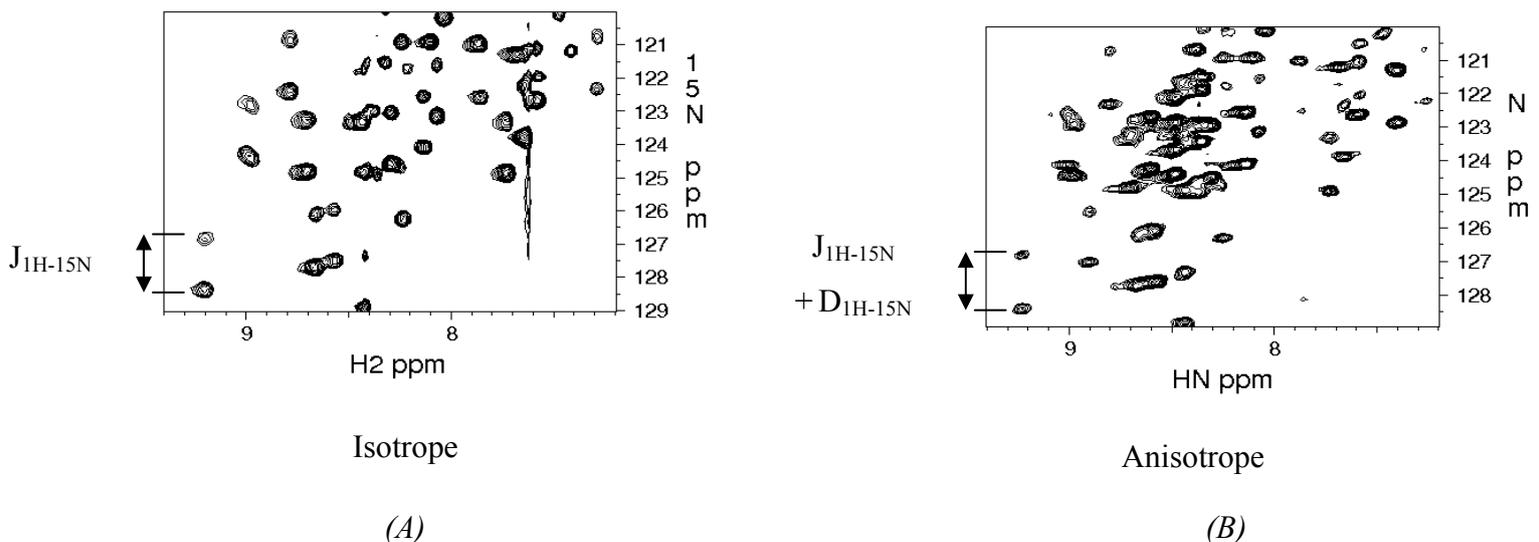


Figure 80 : Spectres  $^1\text{H}$ - $^{15}\text{N}$  HSQC du domaine Myb de TRF2 complexé à l'ADN en milieu isotrope (A) et légèrement anisotrope (B).

Plusieurs types de séquences ont été implémentées. Un point très important pris en compte était les temps de relaxation transversaux très courts de la protéine impliquée dans son complexe. La séquence utilisée devait donc être très sensible et avoir le meilleur rapport signal sur bruit. La séquence qui a donné les meilleurs résultats est la séquence IPAP-SE-NHSQC (Ding and Gronenborn 2003), version améliorée de la séquence IPAP-NHSQC classique mise au point par Ottiger (Ottiger, Delaglio et al. 1998). Elle permet d'enregistrer

simultanément deux sous-spectres en dimension indirecte, avec dans chaque sous-spectre les composantes du doublet  $^{15}\text{N}-^1\text{H}$ . Dans l'un des sous-spectres, nous obtenons le doublet  $^{15}\text{N}-^1\text{H}$  en antiphase alors que dans l'autre sous-spectre, nous obtenons le doublet en phase. La somme ou la soustraction des deux sous-spectres permet d'obtenir respectivement les deux composantes  $^{15}\text{N}-^1\text{H}\beta$  et  $^{15}\text{N}-^1\text{H}\alpha$  du doublet  $^{15}\text{N}$  en dimension indirecte. L'utilisation de cette séquence permet donc d'éviter certains problèmes de recouvrements des raies de résonances.

Pour caractériser l'orientation induite dans le gel, on observe et on mesure le couplage quadrupolaire du deutérium. En milieu isotrope le spectre du deutérium apparaît sous la forme d'un singulet alors qu'en milieu anisotrope on observe un doublet correspondant au couplage quadrupolaire.

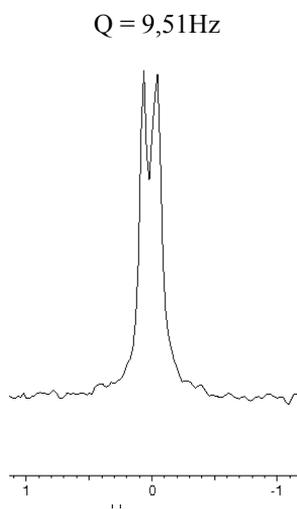


Figure 81 : Spectre deutérium de  $\text{D}_2\text{O}$  du gel de polyacrylamide dans un mélange  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (90%/10%)  $Q = 9,51 \text{ Hz}$

La mesure des couplages est relativement délicate. La précision de ces mesures a été estimée à  $\pm 0,5 \text{ Hz}$ . 46 couplages ont ainsi pu être mesurés. Ils correspondent aux résidus suivants:

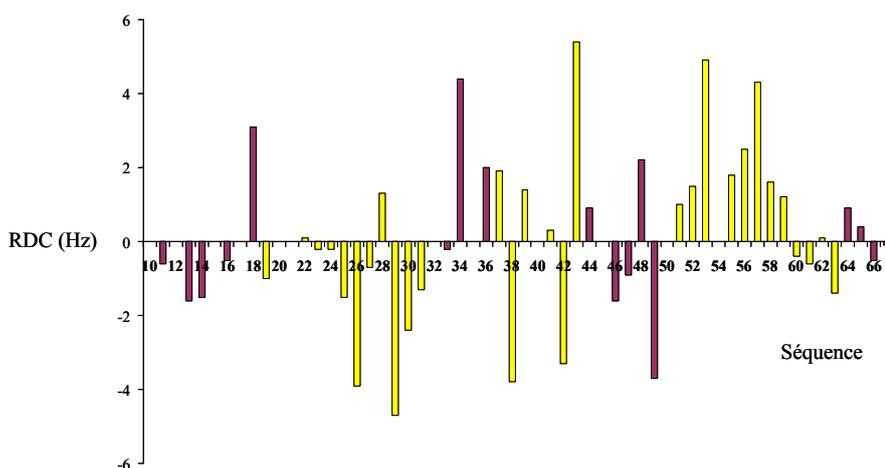


Figure 82: Valeurs des couplages dipolaires résiduels mesurés à l'aide de la séquence IPAP-15NHSQC

Les couplages dipolaires mesurés sont bien répartis sur l'ensemble de la séquence. Les valeurs des RDCs mesurés vont de -5,7Hz à +5,4Hz. Les couplages non mesurés correspondent aux premiers résidus dont les résonances  $^{15}\text{N}$  et  $^1\text{HN}$  n'ont pu être attribuées, ainsi qu'à des couplages non-mesurables en raison de recouvrements.

### 8.3 Détermination du tenseur d'alignement et des angles d'Euler

Le tenseur d'alignement a été déterminé en utilisant le logiciel Module. Celui-ci permet, grâce à une structure PDB ainsi qu'aux valeurs expérimentales de RDCs de déterminer les composantes du tenseur d'alignement ainsi que les angles d'Euler entre le repère moléculaire et l'axe principal du tenseur d'alignement. La définition du tenseur d'alignement est un point crucial pour l'étude structurale sous contraintes RDCs. Pour le déterminer, on utilise le plus généralement une structure PDB qui doit souvent être la plus précise possible. Deux paramètres importants sont à considérer :

- La précision et l'exactitude de la structure PDB
- La précision de la mesure des RDCs

Pour la détermination du tenseur, nous avons sélectionné des couplages dipolaires appartenant aux structures secondaires de la protéine et ayant été mesurés avec la plus grande précision possible. Ainsi les couplages dipolaires résiduels des résidus K26, V29, A37, I39, R49, T50 et I54 ont été retenus. Ces résidus appartiennent aux trois hélices. Les angles d'Euler  $\alpha$ ,  $\beta$  et  $\gamma$  ainsi que les composantes axiale et rhombique du tenseur d'alignement ont pu être déterminées à l'aide du logiciel Module :  $\alpha = 82,11^\circ$ ,  $\beta = 40,77^\circ$  et  $\gamma = -119,22^\circ$ ,  $A_a = 4,44 \cdot 10^{-4}$  et  $A_r = 1,578 \cdot 10^{-4}$

La rhombicité correspond au rapport entre la composante radiale et la composante axiale du tenseur d'alignement. On trouve alors une rhombicité de 0,348. Le calcul de l'amplitude du tenseur nous donne une valeur de 4,38 Hz, ce qui est en accord avec les valeurs expérimentales mesurées.

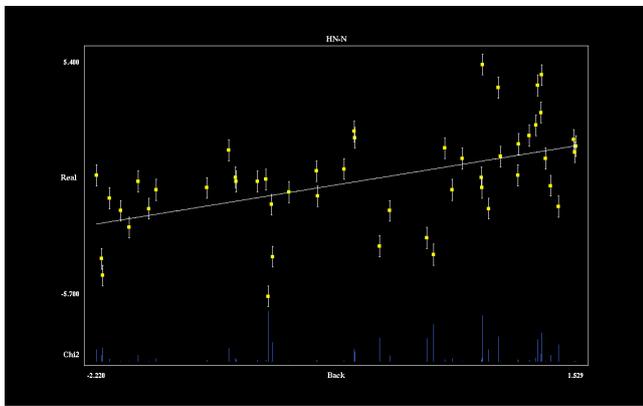
## 8.4 Protocole de raffinement de la structure du domaine Myb de TRF2 à l'aide de RDCs

Les structures de la protéine complexée obtenues précédemment ont été utilisées pour cette étape de raffinement à l'aide des RDCs. L'affinement des structures a été réalisé à l'aide du logiciel ARIA 1.2. L'utilisation d'ARIA a été obligatoire, le logiciel Cyana ne permettant pas de prendre en compte les contraintes RDCs.

Trois fichiers ont été nécessaires :

- ⇒ Les contraintes de distances déterminées à l'aide du logiciels Cyana ont été converties au format CNS
- ⇒ les contraintes d'angles dièdres
- ⇒ les RDCs

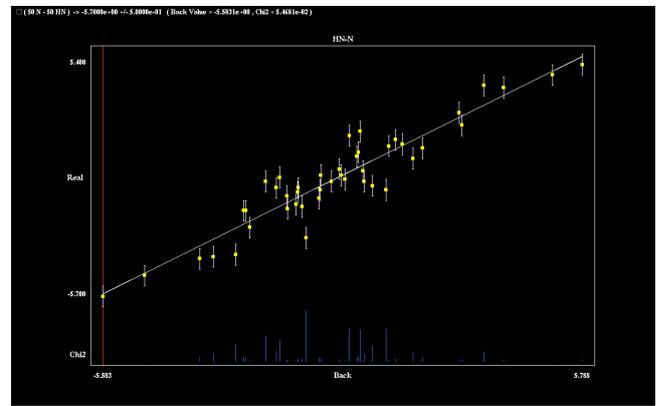
La modélisation moléculaire s'est alors faite en deux étapes, une première minimisation dans l'espace cartésien à une température de 2000K suivie d'une minimisation dans une boîte d'eau. Les constantes de forces appliquées aux contraintes RDCs  $K_{\text{sani}}$  ont des valeurs relativement faibles ( $1.5 \text{ kcal.mol}^{-1}.\text{Hz}^{-1}$  pour le premier refroidissement et  $3.0 \text{ kcal.mol}^{-1}.\text{Hz}^{-1}$  pour le second) par rapport aux contraintes NOE ( $50 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$ ) et dièdres ( $150 \text{ kcal.mol}^{-1}.\text{deg}^{-1}$ ). Le calcul a été effectué sur 200 structures et les 10 structures présentant les énergies totales les plus faibles ont été sélectionnées. Ces 10 structures ont une énergie totale moyenne de  $-2005,5 \text{ kcal.mol}^{-1}$ . Sur la figure ci-dessous, on peut observer la corrélation entre les valeurs RDCs expérimentales et les valeurs recalculées à partir de la structure PDB obtenue après modélisation moléculaire. On observe que l'utilisation des contraintes RDCs permet d'améliorer cette corrélation caractérisée par un  $\chi^2$  plus faible.



RDCs calculés

Structure modélisée sans contraintes RDCs

(A)



RDCs calculés

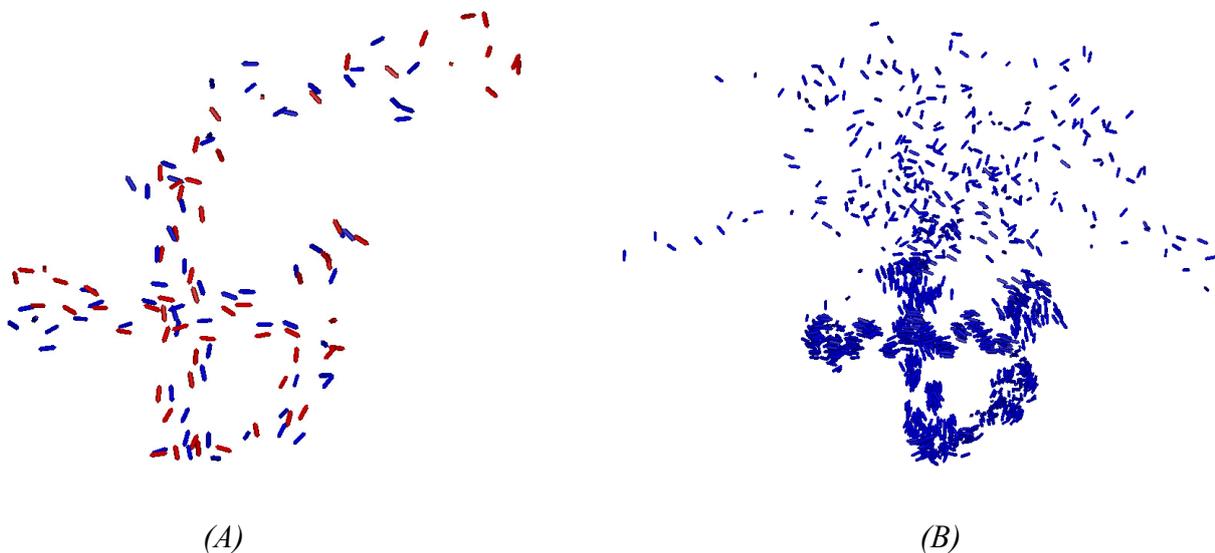
Structure modélisée avec contraintes RDCs

(B)

Figure 83. : Courbes représentant le calcul de RDCs expérimentaux (abscisses) en fonction des RDCs calculés à l'aide de la structure du domaine Myb de TRF2. (A) En milieu anisotrope après raffinement ( $\chi^2=98$ ) ; (B) En milieu isotrope ( $\chi^2=827$ )

## 8.5 Structure du domaine Myb de TRF2 obtenue par raffinement

Les structures du domaine Myb en milieu isotrope et après un raffinement en utilisant les RDCs sont représentées ci-dessous :



(A)

(B)

Figure 84: (A) Représentation de l'orientation des NH pour les deux structures (rouge avec contraintes RDCs, bleu sans contraintes RDCs); (B) Représentation des vecteurs NH de l'ensemble de 10 structures affinées en utilisant les contraintes RDCs

Lorsqu'on compare les structures obtenues avec ou sans contraintes RDCs, on n'observe que de légères différences, notamment au niveau du second coude et de la troisième hélice.

## 9 Conclusion et discussion

Au cours de cette étude plusieurs points très intéressants ont pu être abordés. Tout d'abord nous n'observons la fixation que d'un seul domaine Myb de TRF2 sur l'ADN télomérique, et non de deux domaines Myb, comme cela aurait pu être le cas. Ensuite, nous observons que la fixation se réalise exclusivement sur la partie double brin de l'ADN télomérique, et non au niveau de la jonction double brin-simple brin, montrant la spécificité de la fixation de la protéine sur l'ADN télomérique. Au niveau de la fixation, celle-ci est très proche de celle observée lors des études précédentes. En effet les résidus et les interactions impliquées sont les mêmes que dans les autres complexes étudiés ou publiés.

La modélisation de ce complexe en intégrant les contraintes RDCs est en cours de réalisation et devrait être rapidement achevée.

Un article sur ces travaux est en cours de rédaction et sera soumis dans les mois à venir.

# CHAPITRE 10 : ETUDE DE MODELES DE BOUCLE-D PAR RMN

## 1 Introduction

L'un des rôles essentiels des télomères consiste à protéger les chromosomes contre différentes enzymes de notre corps comme les ligases ou les nucléases. Pour expliquer cette protection, plusieurs modèles ont été proposés dont celui de la boucle-T/boucle-D.

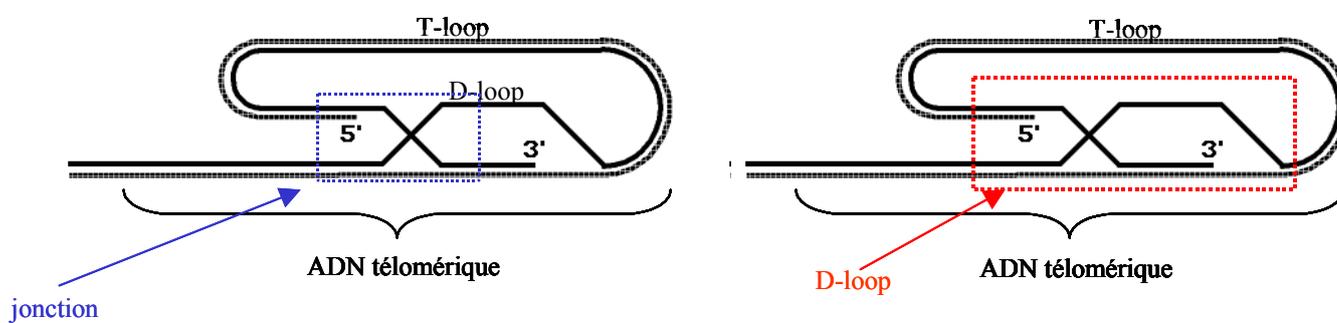


Figure 85: Présentation des boucles télomériques : Le modèle de la boucle-T/boucle-D. La région encadrée en bleu correspond à la zone d'insertion du simple brin. La région encadrée en rouge correspond au modèle que nous avons choisi d'étudier par RMN.

Dans ce modèle on observe le repliement de l'ADN télomérique double brin et l'insertion du simple brin dans le double brin permettant ainsi la formation de deux boucles, la boucle-T (boucle télomérique) et la boucle-D (boucle de déplacement). L'observation attentive de la figure ci-dessus montre la présence d'une jonction où le simple brin s'insère dans le double brin (encadrée en bleu sur la figure ci-dessus). Cette jonction n'est pas sans rappeler les jonctions Holliday, structures bien connues des biologistes étudiant les phénomènes de recombinaison entre chromosomes homologues ou de réparation de l'ADN. Historiquement, c'est en 1964 que Robin Holliday proposa l'implication des jonctions quatre voies (jonctions ADN ou ARN comportant quatre extrémités double brin) dans les phénomènes de recombinaison entre chromosomes homologues (figure ci-dessus). Ces jonctions furent par la suite dénommées « jonctions Holliday ». Les premières structures tridimensionnelles de jonctions Holliday ont été résolues par radiocristallographie en 1998 (Hargreaves, Rice et al.

1998). En 2000, les premiers modèles structuraux de tels motifs fondés sur des études RMN étaient proposés (van Buuren, Schleucher et al. 2000).

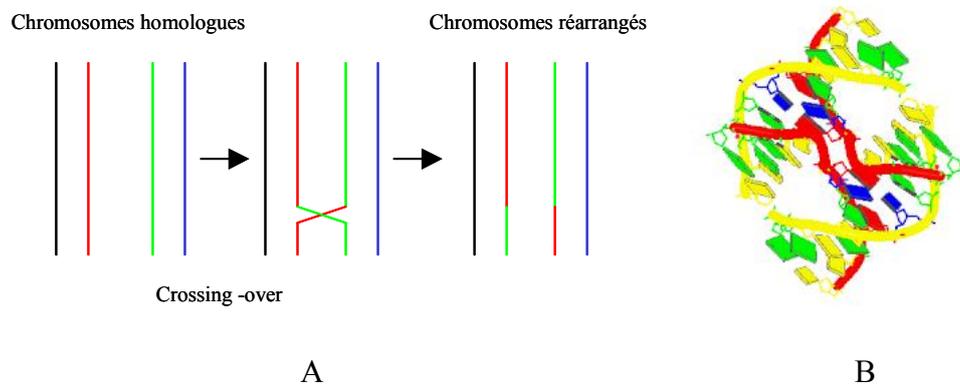


Figure 86: A : Modèle simplifié du phénomène de recombinaison entre chromosomes homologues B : Structure radiocristallographique d'une jonction Holliday (Vargason and Ho 2002).

Dans le même registre, les jonctions trois voies sont présentes dans les ARNs et représentent des motifs structuraux et fonctionnels essentiels. Elles sont constituées de trois portions double brin connectées entre elles et sont classées selon le nombre de paires de bases non-appariées à la jonction. Ainsi un motif 3HS<sub>n</sub> indique la présence de n bases non-appariées sur un brin. Les structures de quelques jonctions trois voies ont été résolues par RMN (Wu, Girard et al. 2004).

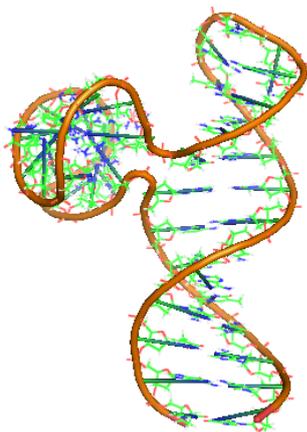


Figure 87: Structure tridimensionnelle d'une jonction trois voies obtenue par RMN (Wu, Girard et al. 2004).

Nous nous sommes particulièrement intéressés à la boucle-D (encadrée en rouge sur la figure 85 page précédente). Aujourd'hui, aucune structure tridimensionnelle de modèle de boucle-D n'est disponible. Afin de mieux comprendre la formation des boucles télomériques, il nous est paru et nous paraît toujours essentiel de connaître la structure tridimensionnelle de tels motifs.

C'est pourquoi nous nous sommes attachés à étudier des modèles de ces structures si particulières.

Notre objectif a consisté tout d'abord à imaginer un modèle de boucle pouvant mimer la boucle-D et dont l'étude par RMN serait réalisable. Une fois l'étude du modèle de boucle par RMN réalisée, notre objectif a été d'étudier la fixation du domaine Myb de la protéine TRF2 sur ce modèle.

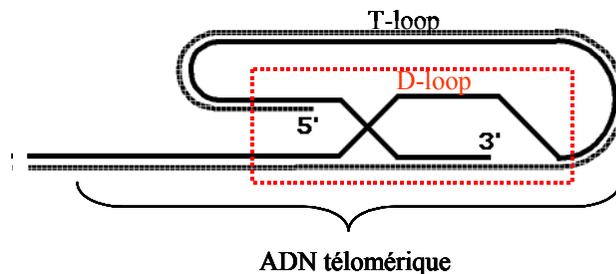
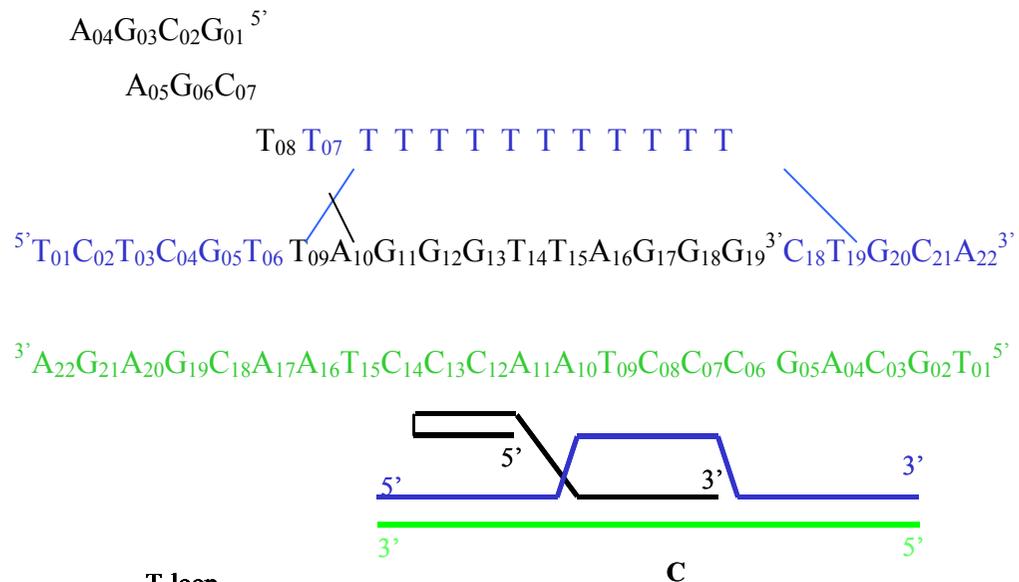
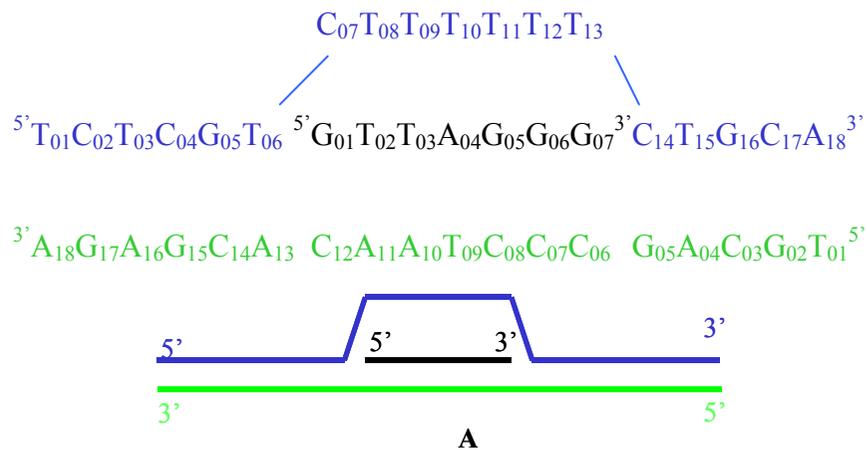
## 2 Présentation des études réalisées

Afin de comprendre les phénomènes conduisant à la formation de la boucle-D, nous avons décidé de réaliser et d'étudier par RMN différents modèles de boucle-D. Plusieurs questions se sont posées : Observe-t-on des irrégularités au niveau de l'ADN ? Que se passe-t-il au niveau des jonctions ? Les bases formant les jonctions sont-elles appariées ou non ? Pour tenter de répondre à ces questions, nous avons réalisé et étudié par RMN quatre différents modèles de boucles.

Les modèles comportent 3 ou 4 brins et contiennent 1 ou 2 répétitions télomériques. Les séquences des quatre modèles de boucle-D étudiées sont présentées sur la figure ci-dessous.

Le brin T, en bleu, représente la partie supérieure de la boucle avec au centre du brin un enchaînement de thymines non appariées et aux extrémités quelques bases complémentaires au brin C. Le brin C correspond au brin central des modèles sur lequel se fixent les brins T et G. Celui-ci comporte la ou les répétitions télomériques A2TC3. Enfin, le brin G est celui qui possède à son extrémité 3' une extension simple brin venant s'insérer à l'intérieur du double brin télomérique.

Le tableau 6 contient un certain nombre d'informations concernant la provenance des oligonucléotides, la préparation des échantillons et les conditions d'études RMN.



Brin T  
 Brin C  
 Brin G

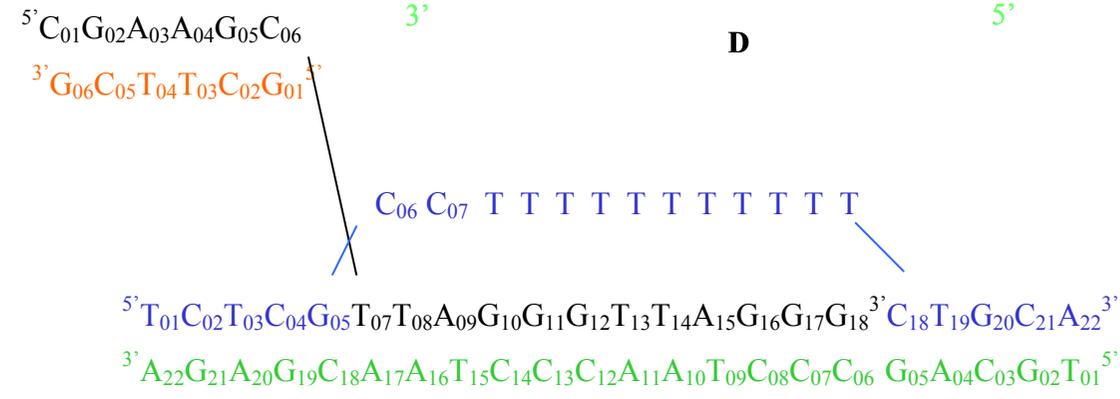
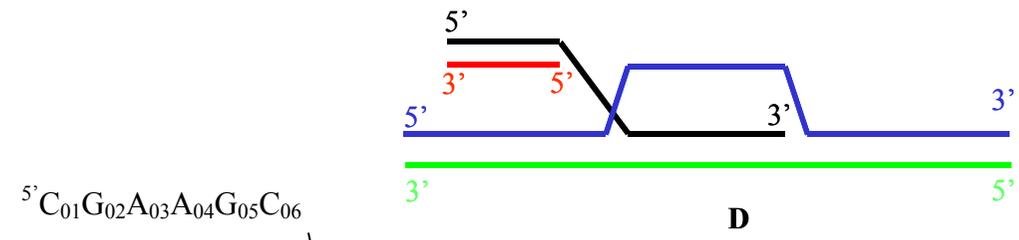
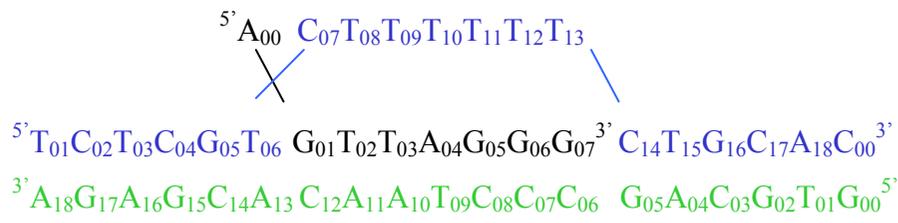
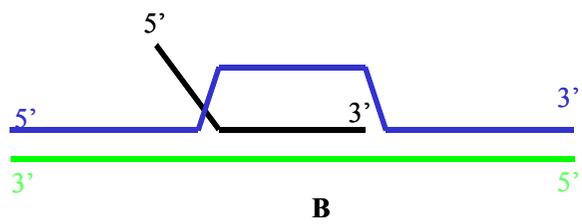


Figure 88: Présentation des différents modèles de Boucle-D

| <i>Modèle de boucle</i> | <i>Fournisseur</i>                                | <i>Nombre de brins/ Noms des brins</i> | <i>Nombre de résidus</i> | <i>Masse moléculaire (kDa)</i> | <i>Concentration Echantillon (mM)</i> | <i>Conditions expérimentales</i> | <i>Températures études RMN (°C)</i> | <i>Expériences réalisées/ Temps de mélange</i> |
|-------------------------|---|--|--------------------------|--------------------------------|---------------------------------------|----------------------------------|-------------------------------------|--|
| <b>A</b>                | Eurogentec  | 3 / CGT                                | 43                       | ~ 14                           | ~ 2                                   | pH6 /<br>100mM<br>NaCl           | 10, 15 et 20°C                      | NOESY/ 100, 150, 200 et 300 ms<br>TOCSY/ 50 ms |
| <b>B</b>                | Eurogentec  | 3 / CGT                                | 48                       | ~ 16                           | ~ 2                                   | pH6 /<br>100mM<br>NaCl           | 10, 15 et 20°C                      | NOESY/ 120 et 300 ms<br>TOCSY/ 50 ms           |
| <b>C</b>                | Eurogentec  | 3 / CGT                                | 63                       | ~ 21                           | ~ 0,5                                 | pH6/<br>100mM<br>NaCl            | 5, 10 et 20°C                       | NOESY/ 80, 150 et 300ms<br>TOCSY/ 50 et 80 ms  |
| <b>D</b>                | Synthèse au laboratoire (méthode phosphoramidite) | 4 / CGTH                               | 68                       | ~ 22                           | ~ 1,5                                 | pH6 /<br>100mM<br>NaCl           | 5, 10 et 20°C                       | NOESY/ 120 et 300 ms<br>TOCSY/ 50 ms           |

*Tableau 6 : Informations concernant les quatre différents modèles de boucles*

### 3 Boucle-D : modèle A

#### 3.1 Présentation

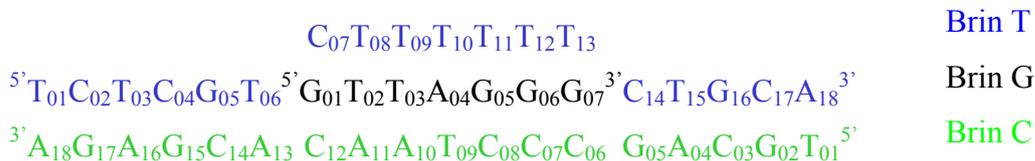


Figure 89 : Séquence du modèle A de boucle étudiée

Cet ADN est constitué des trois brins différents C, T et G. Il comporte 43 résidus. Cette boucle ne contient qu'une seule répétition télomérique et tous les nucléotides appartenant au brin G peuvent s'apparier au brin C.

#### 3.2 Etude RMN

##### 3.2.1 Etude du duplexe entre les brins C et T

Le brin T a été additionné sur le brin C en quatre fois à une température de 10°C. L'attribution séquentielle a pu être réalisée et montre que les deux extrémités du duplexe forment une double hélice droite. La figure 90 ci-dessous montre un agrandissement de la région des protons iminos du duplexe CT d'un spectre NOESY enregistré à 10°C avec un temps de mélange de 300 ms. L'étude des protons iminos montrent que les appariements sont de type Watson et Crick et l'ensemble des bases appariées peut être observé sur la figure 91.

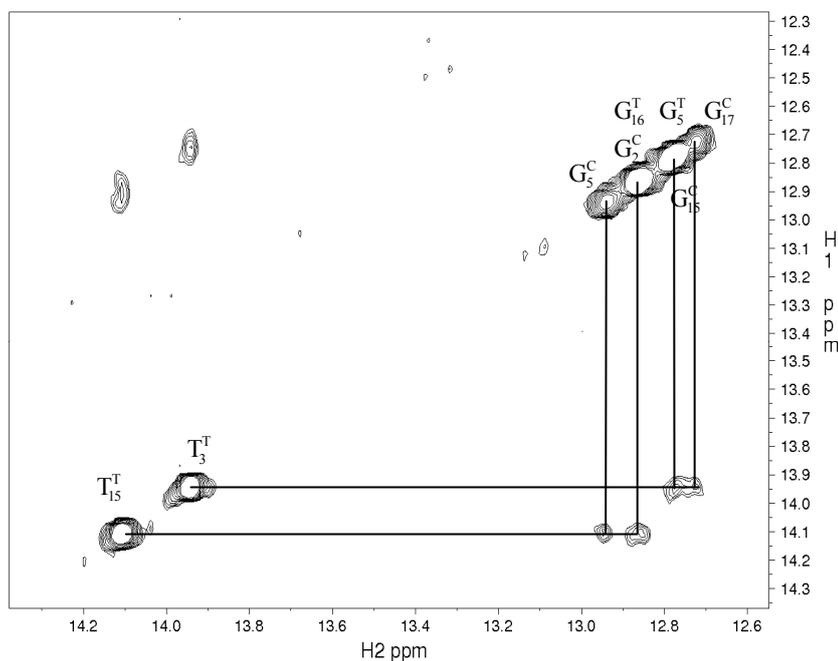


Figure 90 : Duplexe entre les brins C et T : agrandissement de la région des protons iminos d'un spectre NOESY (10°C et tm=300ms) montrant les interactions dipolaires entre les protons iminos des thymines et des guanines.

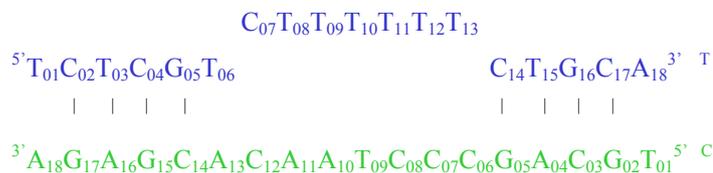


Figure 91 : Récapitulatif des liaisons hydrogène observées sur les différents spectres RMN.  
Les traits représentent les résidus appariés.

### 3.2.2 Etude du duplexe entre les brins C et G

Le brin G a été additionné sur le brin C en quatre fois. La figure 92 montre un agrandissement de la région aromatique H<sub>6</sub>/H<sub>8</sub> –H<sub>1</sub>’/H<sub>5</sub> d’un spectre NOESY enregistré à 10°C avec un temps de mélange de 300ms. Des corrélations entre les protons H<sub>5</sub> et NH<sub>2</sub> des cytosines peuvent être observées. Le proton amino NH<sub>2</sub> résonnant vers les champs faibles (partie gauche du spectre) correspond au proton amino impliqué dans la liaison hydrogène avec le carbonyle de la guanine correspondante alors que le proton amino résonnant vers les champs forts correspond au proton amino non-impliqué dans une liaison hydrogène.

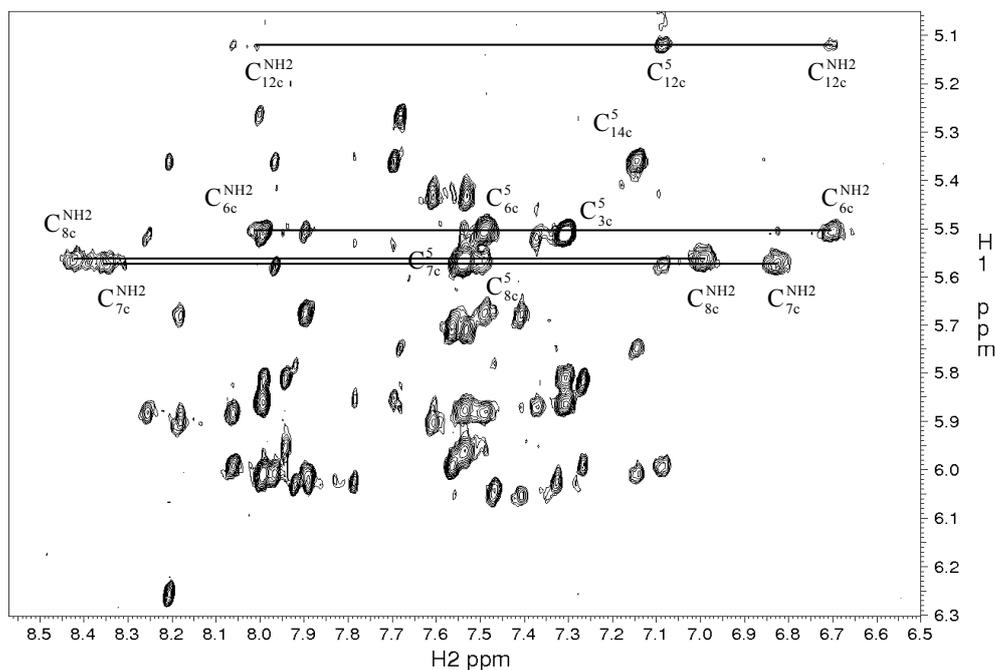


Figure 92: Duplexe entre les brins C et T: agrandissement de la région des protons aromatiques/H1' d'un spectre NOESY (10°C et tm=300ms) montrant les interactions dipolaires entre les protons H6-H5 et aminos(NH2)-H5.



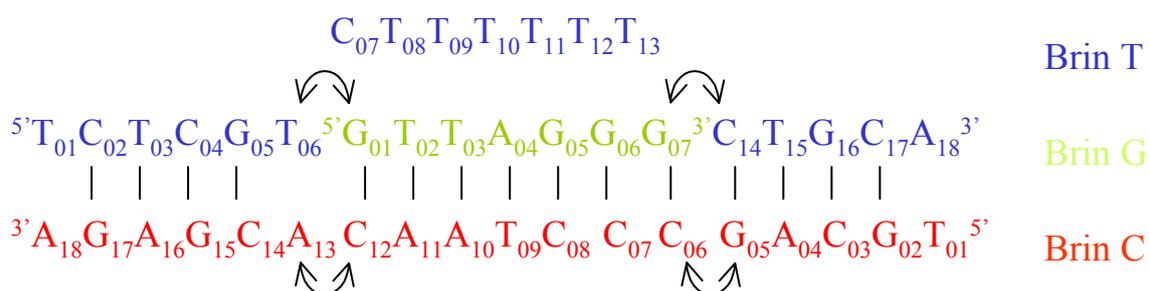
### 3.3 Conclusion

L'ensemble des raies de résonances des protons aromatiques H<sub>6</sub>/H<sub>8</sub> et H<sub>1'</sub> de l'ADN a pu être attribué. Les variations de déplacements chimiques observées entre les duplexes CG, CT et le triplexe CGT, notamment pour les résidus C12C et C14C du brin C montrent que le brin T est bien fixé sur le brin C et que le système est correctement formé. Un point beaucoup plus intéressant est la mise en évidence de corrélations dipolaires entre le H<sub>5</sub> du résidu C14T et le H<sub>8</sub> du résidu G7G (figure 95). De plus, on observe la continuité de l'attribution séquentielle entre les résidus G7G et C14T prouvant que cette partie de la boucle forme une double hélice droite continue. De l'autre côté, des connectivités dipolaires aromatique-H<sub>1'</sub> entre les résidus G1G et T6T montrent que cette partie forme aussi une double hélice droite. Ces connectivités entre les brins C et T ont pu être confirmées par l'étude de la région des protons aromatiques où des connectivités ont été mises en évidence, notamment entre le H<sub>6</sub> de C14T et le H<sub>8</sub> de G7G.

Nous observons que l'attribution séquentielle du brin C a été réalisée en entier, sans aucune anomalie ou irrégularité montrant que le brin C tend à adopter une conformation en double hélice droite.

Toutes ces observations tendent à montrer que l'ensemble du duplexe forme une double hélice droite de type B classique. L'étude des protons iminos montre que l'ensemble des paires de bases complémentaires sont bien appariées, notamment les paires de bases G01G-C12C, G07G-C06C et C14T-G5C.

*Figure 95: Modèle de la boucle A. Les traits représentent les résidus appariés. Les flèches*



*représentent les connectivités dipolaires observées sur les spectres RMN. Les trois parties de la double hélice semblent coaxiales. Cela sera vérifié par la mesure de RDCs.*

## 4 Boucle-D : modèle B

### 4.1 Présentation

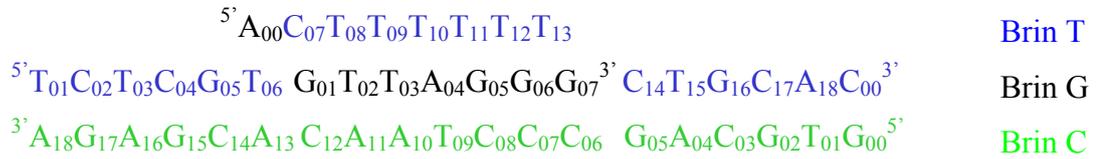


Figure 96 : Séquence de l'assemblage nucléique étudié

Notre objectif, lors de l'étude de ce nouveau modèle, était d'observer les modifications induites au niveau de la jonction par la présence de cette base supplémentaire A0G en 5' du brin G.

### 4.2 Etude RMN et Conclusion

L'analyse des déplacements chimiques montre que les environnements chimiques des cytosines C12C (résonance proton H6 déplacée de 0,21 ppm vers les champs faibles) et C14C (résonance proton H5 déplacée de 0,20 ppm vers les champs faibles) sont perturbés, ces variations de déplacements chimiques peuvent s'expliquer par la proximité de l'adénine 00 du brin G. Néanmoins, l'analyse des spectres NOESY de la région des protons iminos montre que les paires de bases G5T/C14C et C12C/G1G sont appariées. De plus, nous retrouvons les connectivités dipolaires entre les résidus G7G et C14T, que ce soit la connectivité H<sub>5</sub>C14T - H<sub>8</sub>G7G, H<sub>6</sub>C14T - H<sub>1</sub>G7G ou bien la connectivité H<sub>6</sub>C14T - H<sub>8</sub>G7G.

La jonction est juste affectée par la présence de A00G. Notre idée était alors de compliquer le système en ajoutant une extension du brin G comportant plusieurs résidus ainsi que deux répétitions télomériques afin de pouvoir confirmer ces observations et réaliser l'étude d'un complexe entre un modèle de boucle et le domaine Myb de TRF2.

## 5 Boucle-D : modèle C

### 5.1 Présentation

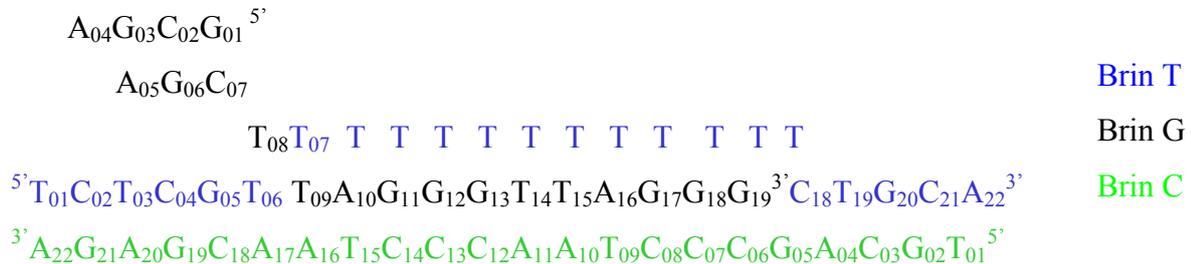


Figure 97: Séquence du fragment d'ADN étudié

Ce fragment d'ADN est constitué de trois brins différents dénommés T, C et G. Il comporte 63 résidus soit une masse moléculaire d'environ 21 kDa (équivalent à une protéine d'environ 200 résidus). Il est totalement non marqué.

Le brin G comporte à son extrémité 5' une succession de 7 résidus formant la séquence d(GCGAAGC). Cette séquence est connue pour se replier sur elle-même en formant une tige-boucle très résistante (même aux nucléases) et très stable. L'étude structurale par RMN de cette tige-boucle a été réalisée (Hirao, Kawai et al. 1994) et montre que les protons  $H_4'$ ,  $H_5'$  et  $H_5''$  appartenant à l'adénine 4 subissent des courants de cycle importants dus à l'adénine 5 (figure 98). Ces protons résonnent à des fréquences caractéristiques et sont facilement identifiables sur les spectres RMN, notamment NOESY.

A4  
A5 G3  
G6 C2  
C7 G1

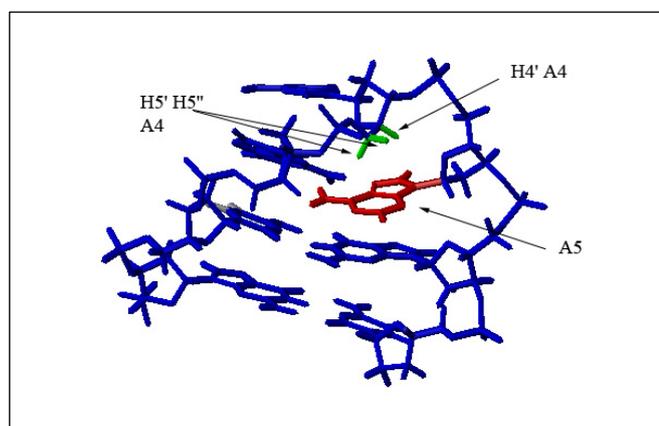


Figure 98 : Séquence et structure tridimensionnelle obtenue par RMN de la tige-boucle formée par la séquence 5'GCGAAGC3' (Hirao, Kawai et al. 1994)

## 5.2 Etude RMN

### 5.2.1 Etude du brin G seul contenant la tige-boucle

Le brin G a d'abord été étudié seul. Celui-ci a été dissous dans un mélange H<sub>2</sub>O/D<sub>2</sub>O (90%/10%), le pH a été ajusté à une valeur d'environ 6. Des spectres NOESY et TOCSY ont alors été enregistrés à différents temps de mélange et à différentes températures.

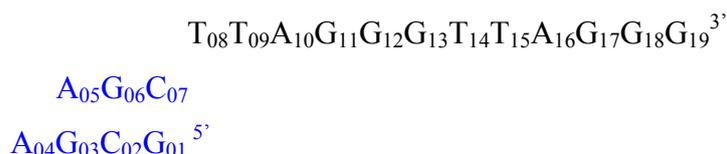


Figure 99 : Séquence du brin G. En bleu apparaissent les résidus constituant la tige-boucle.

L'attribution séquentielle de la région des protons aromatiques H<sub>6</sub>/H<sub>8</sub> et H<sub>1</sub>, ainsi que de la région des connectivités H<sub>6</sub>/H<sub>8</sub>-H<sub>2</sub>'/H<sub>2</sub>'' ont été réalisées. Les résonances caractéristiques correspondant aux protons H5' et H5'' de l'adénine 4 ont pu être identifiées et attribuées. L'étude de la région des résonances des protons iminos, représentée sur la figure ci-dessous, montre la présence de deux raies de résonances correspondant aux protons iminos des guanines G1 et G6. Ceux-ci sont impliqués dans des liaisons hydrogène avec les cytosines C7 et C2. On remarquera la présence d'une déformation de la ligne de base vers 11-11,5 ppm correspondant à la présence de structures quadruplexes.

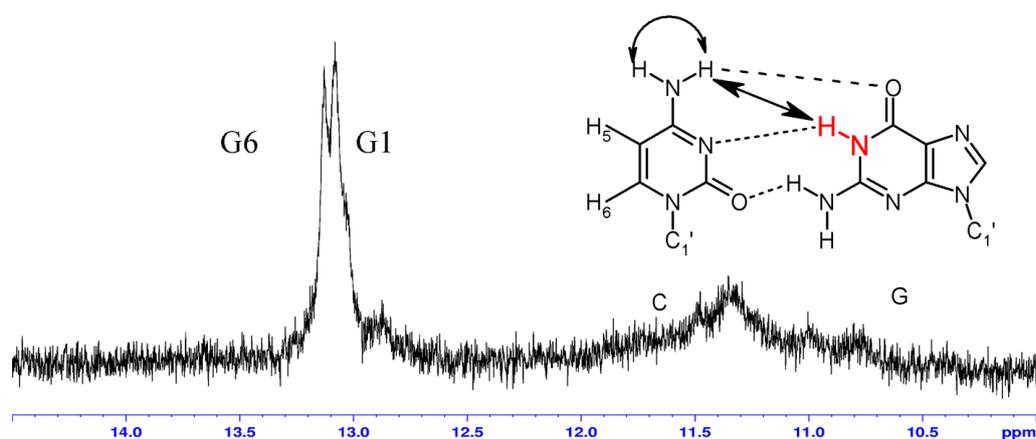


Figure 100 : Agrandissement de la région de résonances des protons iminos des résidus guanines de l'hairpin impliqués dans des liaisons hydrogène.

### 5.2.2 Etude du duplexe entre les brins G et C

La première étape a consisté à ajouter le brin G sur le brin C. Quatre ajouts successifs de brin G ont été nécessaires afin d'obtenir un mélange équimolaire 1 : 1 entre les deux brins.

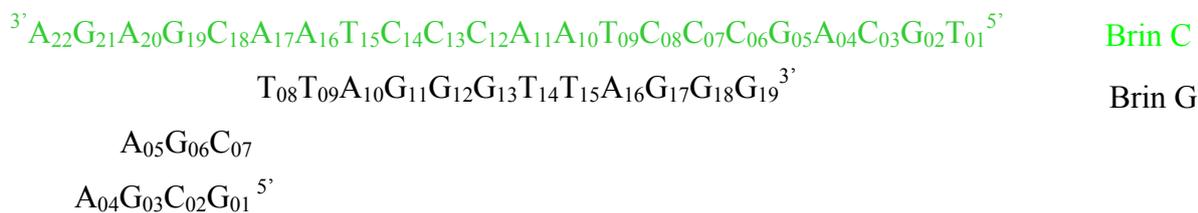
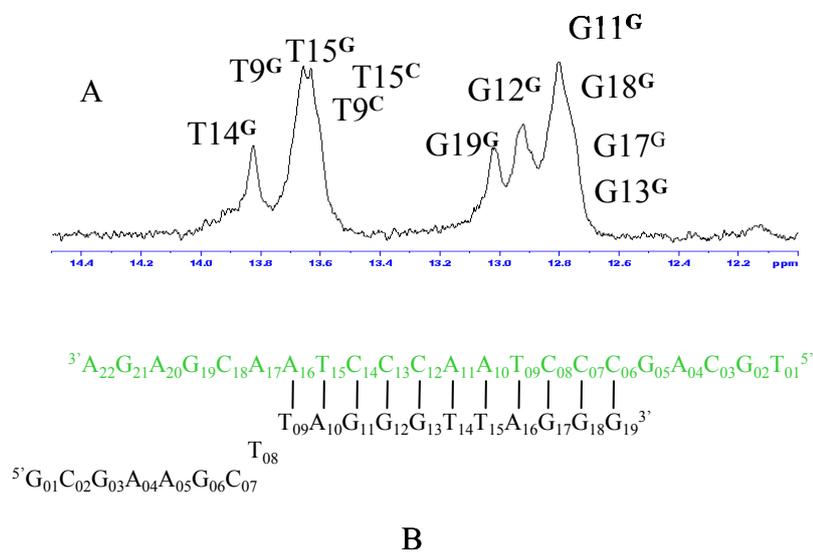


Figure 101: Séquence du duplexe brin G + brin C

Un ensemble de spectres NOESY à différentes températures (5, 10, 15 et 25°C) et pour différents temps de mélange (80, 150 et 300ms) ont été enregistrés. L'attribution séquentielle des résidus formant la partie double brin du télomère a été réalisée et correspond à une double hélice droite. L'étude des protons iminos a montré que l'ensemble des bases du duplexe contenant les deux répétitions télomériques sont appariées et que ces appariements sont de type Watson et Crick. Notons que si les protons iminos de T15C et T9G sont observables, celui de T8G n'est pas observable, cela pouvant s'expliquer par une dynamique importante causée par la présence de la tige-boucle et rendant ce proton beaucoup plus exposé à l'échange avec les protons de l'eau.

Néanmoins l'attribution séquentielle des raies de résonances de la tige-boucle n'était plus réalisable et les raies de résonances caractéristiques des protons H<sub>4</sub>, H<sub>5</sub> et H<sub>5'</sub> de l'adénine 4 n'étaient plus observables. De même l'observation des résonances des protons iminos montre l'absence des raies caractéristiques de l'hairpin. Pour expliquer le désappariement de la tige-boucle, on peut imaginer une gêne stérique déstabilisant l'appariement entre les résidus G1G et C7G et cette déstabilisation entraînant le désappariement de la tige-boucle.

Les raies de résonances du duplexe avec les attributions des protons iminos sont rassemblées dans la figure suivante :



(A) Agrandissement de la région de résonance des protons iminos du duplexe CG d'un spectre 1D réalisé à 10°C. (B) Représentation du duplexe CG montrant les liaisons hydrogène observées.

L'étape suivante a consisté en l'ajout du brin T sur le duplexe.

### 5.2.3 Etude RMN du triplex entre les brins C, G et T

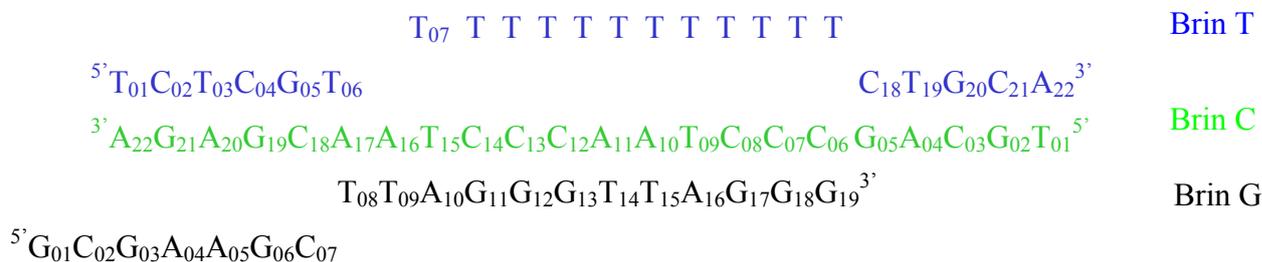


Figure 102: Séquence d'acides nucléiques formant le triplex CGT

L'étude du triplex CGT a été réalisée à 10°C, température à laquelle les spectres se sont révélés les plus exploitables. A cette température, l'attribution séquentielle a pu être réalisée et a montré que les trois brins sont appariés.

L'étude des protons iminos a de plus montré que les appariements C02T-G21C, T03T-A20C, C04T-G19C, G5T-C18C mais aussi T06T-A17C étaient formés. Concernant l'extrémité 3' du triplex les appariements C18T-G05C, T19T-A04C, G20T-C03C et C21T-G02C sont formés. Les protons iminos des thymines T01C et T01T respectivement appariés aux adénines A22T et A22C n'ont pu être mis en évidence (figure 103). L'ensemble des appariements observés correspond à des appariements de type Watson et Crick.

Sur la figure 103 on peut observer les raies de résonances correspondant aux protons iminos des guanines et thymines attribués dans le triplex. En bleu apparaissent les raies de résonances des protons iminos appartenant aux extrémités du brin T et en noir certaines des raies de résonances appartenant au duplexe télomérique GC prouvant que celui-ci est encore formé.

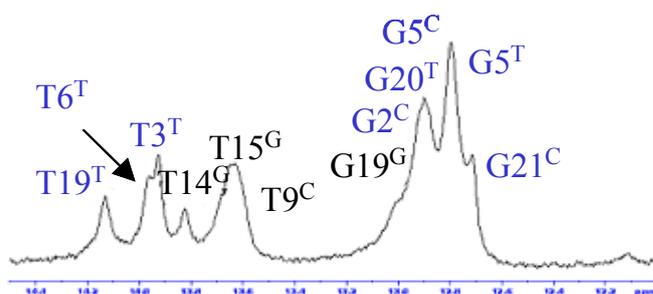


Figure 103 : Agrandissement de la région iminos d'un spectre RMN 1D montrant les nouvelles raies de résonances dues à l'appariement du brin T sur le duplexe GC (en bleu) ainsi que certaines raies de résonances appartenant au duplexe GC.

### 5.3 Conclusion

L'étude des spectres RMN nous a montré que les résidus suivants étaient appariés :

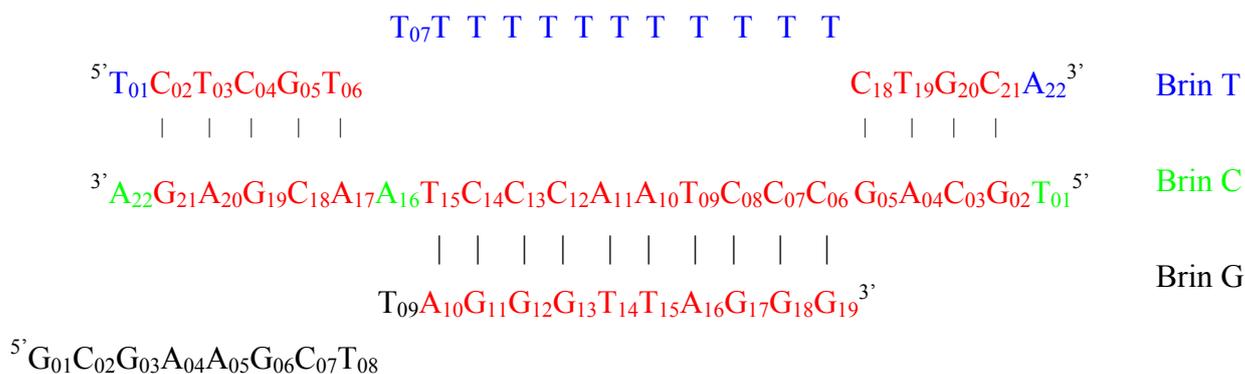


Figure 104 : Séquence montrant en couleur rouge les résidus appariés

Ayant démontré l'appariement des différentes bases azotées, nous nous sommes alors attachés, pendant très longtemps, à chercher des interactions inter brin notamment au niveau des jonctions entre les trois brins. Malheureusement aucune interaction dipolaire entre les résidus T06T et T09G et entre les résidus C18T et G19G n'a pu être observée, que ce soit dans la région des protons échangeables ou non-échangeables.

L'étude RMN de ce modèle s'est révélée très difficile de part la taille de l'oligonucléotide, de part la concentration de l'échantillon (~0,5mM) et de part les conditions expérimentales nécessaires à la formation de la boucle nous obligeant à travailler à basse température (<10°C). Ainsi même si l'attribution séquentielle a pu être partiellement réalisée, aucune interaction dipolaire n'a pu être observée aux jonctions entre les trois brins. Ces corrélations dipolaires n'existent - elles pas dans ce modèle ou ne sont elles pas observables dans nos conditions ? Pour essayer de répondre à ces questions, nous avons décidé de réaliser l'étude d'un quatrième modèle de boucle.

## 6 Boucle-D : modèle D

### 6.1 Présentation

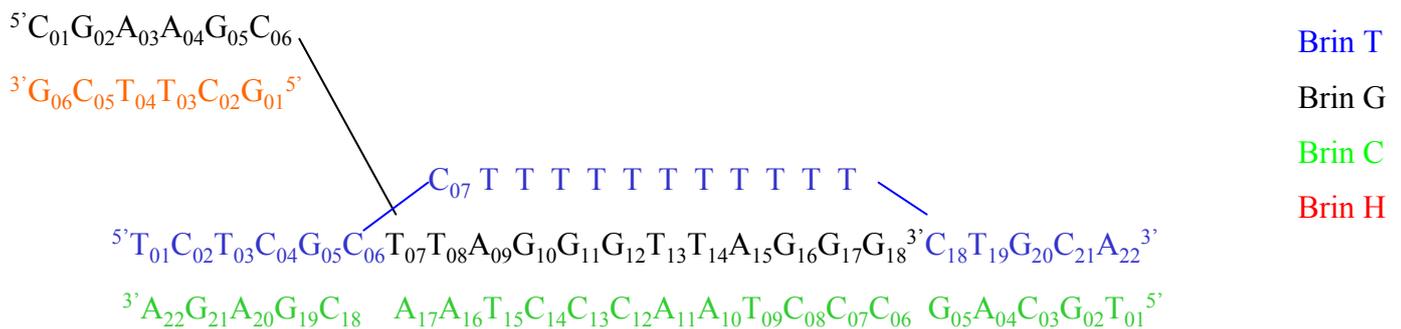


Figure 105: Séquence ADN du modèle D

Ce modèle est constitué de quatre brins différents. Les thymine T6T et T7T du modèle précédent ont été remplacées par deux cytosines C6T et C7T afin d'éviter les problèmes de compétition entre les brins T6T, T7T et T7G, T8G ayant été observés lors de l'étude précédente.

La synthèse des différents brins a été réalisée au sein du laboratoire ainsi qu'à l'ICSN dans l'équipe du Dr Jean-Louis Leroy suivant la méthode aux phosphoramidites. Le protocole est décrit dans l'annexe 1.



## 7 Perspectives : Etude des modèles A et B en milieu orienté

Des études de la boucle en milieu orienté sont en cours de réalisation. Ces études devraient nous permettre d'obtenir des informations sur la conformation globale de nos modèles de boucle. Nous espérons obtenir l'orientation relative des trois fragments de double hélice des modèles A et B (respectivement figures 95 et 96) en mesurant les RDCs présentant une bonne résolution sur les spectres  $^{13}\text{C}$ -HSQC. L'utilisation des couplages dipolaires résiduels pour l'étude des acides nucléiques est très intéressante. La mesure des interactions dipôle - dipôle directes entre le carbone et son proton apporte de nombreuses informations. Les couplages dipolaires résiduels les plus couramment mesurés sont présentés sur la figure ci-dessous.

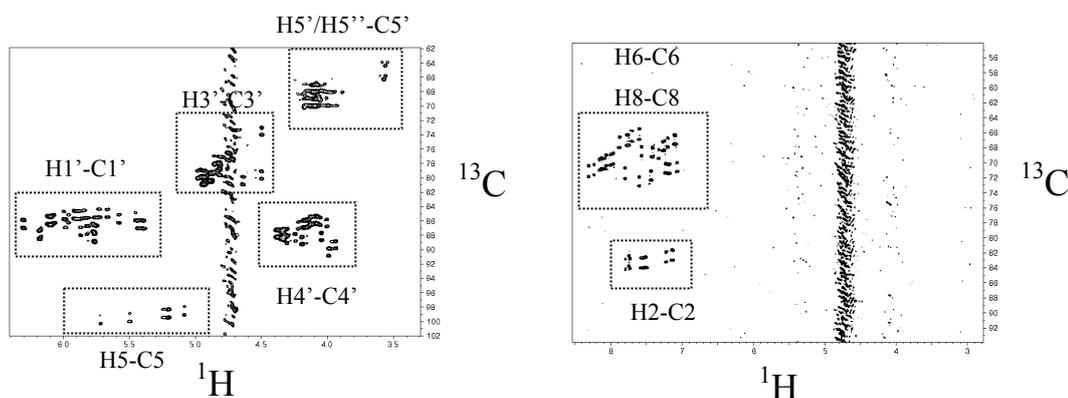


Figure 107 : Spectre  $^1\text{H}$ - $^{13}\text{C}$  HSQC montrant les différentes interactions dipôle-dipôle étudiables dans un fragment d'ADN.

Par exemple la mesure des couplages dipolaires résiduels entre le  $\text{C}1'$  et le  $\text{H}1'$  permet très rapidement d'obtenir la conformation globale de l'ADN. Dans une hélice classique, l'orientation des vecteurs  $\text{C}1'\text{-H}1'$  « tourne » autour de l'axe de l'ADN. Les valeurs des RDCs suivent alors une loi sinusoïdale.

Des spectres  $^1\text{H}$ - $^{13}\text{C}$  HSQC ont été enregistrés en abondance naturelle. La qualité médiocre des spectres n'a, pour l'instant, pas permis d'obtenir des informations exploitables. Deux problèmes sont à surmonter, le temps de relaxation transversale très efficace et le manque de sensibilité, l'échantillon étant non-marqué. En ce moment des essais d'optimisation de séquences RMN sont en cours.

# CONCLUSIONS ET PERSPECTIVES

L'objectif premier de ce travail de thèse a consisté en l'étude par RMN d'un complexe entre le domaine Myb de la protéine TRF2 et un fragment d'ADN télomérique double brin. Le but était de mieux comprendre l'implication de cette protéine dans la formation de la boucle-D et de mieux connaître la structure de cette dernière.

Nous avons déterminé la structure du domaine Myb de TRF2. Le domaine Myb adopte une conformation hélice-tour-hélice de type Myb constitué de trois hélices  $\alpha$  reliées entre elles par deux coudes. L'ensemble de la protéine est stabilisé par un cœur hydrophobe. La dynamique interne a été étudiée grâce à la relaxation nucléaire de l'azote 15. Celle-ci a permis de montrer que le domaine Myb adopte en solution une forme globulaire et qu'il existe sous forme de monomère.

L'étude du premier complexe impliquant le domaine Myb de TRF2, non marqué, et un ADN double brin comportant deux répétitions télomériques a été réalisée. Ce premier complexe a permis d'étudier l'ADN complexé à la protéine. L'étude a montré que le domaine Myb de TRF2 est spécifique de la séquence TAGGGTTA et que deux répétitions télomériques sont nécessaires pour la fixation de celui-ci. Plusieurs résidus impliqués dans la complexation ont pu être identifiés. Ces résidus se situent en général dans la troisième hélice. L'étude par RMN de ce premier complexe a été limitée par l'absence de marquage de la protéine. L'étude de l'ADN complexé montre que le domaine Myb n'est pas, seul, responsable de la formation de la boucle. De plus, la comparaison entre les complexes Myb-TRF2/ADN et Myb-TRF1/ADN montre que ceux-ci sont peu différents et n'explique nullement les fonctions différentes des protéines TRF2 et TRF1 dans la formation des boucles télomériques.

Notre objectif consistait ensuite à étudier le même complexe mais cette fois-ci en utilisant le domaine Myb de TRF2 uniformément marqué à l'azote 15 et au carbone 13. La parution d'un article présentant ce même complexe nous a amené à modifier notre projet. Nous avons alors décidé de réaliser et d'étudier des complexes entre le domaine Myb uniformément marqué  $^{15}\text{N}$  et  $^{13}\text{C}$  et différents types d'ADN, en particulier ceux qui possèdent une jonction double brin-simple brin.

Ainsi, un second complexe a été étudié par RMN. Celui-ci est constitué du domaine Myb de TRF2 doublement marqué et d'un fragment d'ADN possédant une partie double brin et une extension simple brin. L'objectif était de déterminer si la protéine allait se fixer sur la partie double brin ou à la jonction entre la partie double et simple brin de l'ADN. L'étude RMN de ce complexe montre que le domaine Myb de TRF2 ne se fixe que sur la première répétition et le début de la seconde appartenant à la partie double brin. La structure tridimensionnelle de ce complexe a été déterminée. Celle-ci montre que la troisième hélice s'insère dans le grand sillon de l'ADN télomérique et qu'une partie de la protéine (résidus 15 à 17) réalise des contacts avec le petit sillon de l'ADN télomérique. La comparaison des deux complexes montre que la fixation de la protéine sur l'ADN est quasiment identique dans les deux complexes. Le complexe a alors été étudié en milieu orienté et des couplages dipolaires résiduels  $1H-15N$  ont pu être mesurés pour la protéine. L'obtention de ces nouvelles contraintes a permis le raffinement de la structure du domaine Myb de TRF2 complexée à la protéine. Les contraintes RDCs sont couramment utilisées pour la validation ou le raffinement de structures, même si celles-ci sont beaucoup plus intéressantes dans l'étude de complexes protéine-protéine en permettant ainsi de palier au faible nombre de contraintes NOEs intermoléculaires et en permettant d'orienter les différents domaines les uns par rapport aux autres.

Mon travail de thèse a ensuite consisté à réaliser l'étude RMN de modèles de boucle-D. L'idée générale était de mieux connaître la structure de telles boucles. Nous avons pour cela réalisé l'étude de quatre modèles de boucles différents contenant un nombre plus ou moins important de résidus.

Nous avons commencé par des modèles plutôt simples, comprenant une seule répétition télomérique. Ainsi, le premier modèle de boucle (modèle A; figure 88) a pu être étudié par RMN, et notamment des connectivités entre les trois brins au niveau des jonctions ont pu être mises en évidence. Nous avons ainsi pu démontrer que notre fragment d'ADN adopte une conformation en double hélice de type B sur toute la longueur de la partie double brin.

L'ajout dans le modèle suivant d'une base libre (modèle B; figure 88) a modifié l'environnement chimique des résidus situés au niveau de la jonction, mais l'ensemble de la partie double brin semble aussi adopter une conformation en double hélice.

Nous souhaitons approfondir ces études et obtenir notamment plus d'informations sur l'orientation des trois fragments de double-hélice les uns par rapport aux autres en étudiant

ces boucles en milieu orienté. L'optimisation des séquences RMN ( $^{13}\text{C}$ -HSQC en abondance naturelle) est en cours de réalisation et nous restons confiants quant à la suite de ces travaux. En effet, les premières investigations sur les boucles-D A et B (figure 88) montrent que ces mesures sont faisables à une concentration d'environ 2mM, à 15°C et 600MHz.

La mesure des RDCs réalisée, nous pourrions procéder à la modélisation moléculaire des modèles de boucle. Celle-ci pourra se faire de deux façons différentes, soit en utilisant un raffinement et la méthode du recuit simulé à basse température, soit en utilisant le concept de fragments rigides, méthode reposant sur une augmentation des constantes de forces associées aux angles dièdres et aux NOEs. L'étape suivante de ce travail consistera en la réalisation de simulations dynamiques (de quelques ns) de nos différents systèmes en les intégrant dans une boîte d'eau. Ces simulations permettront d'obtenir des renseignements sur la stabilité et les mouvements de ces différentes structures.

Pour terminer, la réalisation de complexes entre le domaine Myb de TRF2 et nos modèles de boucle est aussi envisagée.

Au début de la réalisation de cette thèse, les connaissances concernant la protéine TRF2 étaient relativement faibles. En particulier la manière dont cette protéine peut induire la formation de la boucle-T/boucle-D était incomprise. Trois ans plus tard, de nombreuses avancées ont été réalisées et les mécanismes pouvant permettre la formation de ces boucles commencent à être élucidés. Concernant la protéine TRF2, le domaine Myb de TRF2 seul n'explique pas la formation des boucles. Ces résultats, que nous avons pu observer lors de nos études, ont été confirmés en 2007 par des biologistes. Ils ont créé une protéine mutante en remplaçant le domaine Myb de TRF2 par le domaine Myb de TRF1. Ils ont observé que cette protéine mutante induisait l'insertion du simple brin dans le double brin, alors que la protéine TRF1 seule ne l'induit pas. Cette étude prouve, que même si le domaine Myb de TRF2 permet la fixation à l'ADN, il n'est en aucun cas responsable de la formation de la boucle.

D'autres études ont montré l'importance du domaine de dimérisation TRFH dans le phénomène d'insertion du simple brin. Il a été envisagé que si le domaine Myb permet de se fixer spécifiquement à l'ADN télomérique, la partie N-terminale pourrait reconnaître certains motifs structuraux particuliers. La protéine TRF2 est connue pour se lier sous forme d'homodimère à l'ADN télomérique. Néanmoins, il semblerait que la formation des boucles fasse intervenir la protéine TRF2 sous forme de tétramère. Ce tétramère en se liant à l'ADN pourrait induire, à proximité, un surenroulement négatif du double brin créant une tension au

niveau de l'ADN télomérique double-brin et permettant l'ouverture et l'insertion de l'extrémité simple brin de l'ADN télomérique.

# ANNEXE 1 : Synthèse chimique des différents brins d'ADN

La synthèse par voie chimique des oligonucléotides a été réalisée sur un appareil de la marque Pharmacia. La synthèse s'est effectuée sur support solide en utilisant la méthode Phosphoramidite.

Les réactifs utilisés sont les suivants :

| <i>Réactif</i>                                  | <i>Volume à utiliser pour<br/>chaque base constituant<br/>l'oligo à synthétiser (ml)</i> | <i>Fonction du réactif</i>                                       |
|---|--|--|
| Acétonitrile                                    | 6,00   | Solvant  |
| H <sup>+</sup> (généralement HCl)               | 6,00   | Détritylation  |
| Tétrazole                                       | 0,40   | Couplage   |
| I <sub>2</sub> , Pyridine                       | 0,75   | Oxydation  |
| Ac <sub>2</sub> O                               | 0,4  | Réaction de "capping":<br>piégeage des séquences trop<br>courtes |
| Dichlorométhane CH <sub>2</sub> Cl <sub>2</sub> | 3,25   | Solvant  |
| p-amidite (A,T,G ou C)                          | 0,12   | Nucléotides  |

Le synthétiseur automatique utilisé est contrôlé par ordinateur à l'aide du programme Gène Assembleur. Une fois les réactifs et la résine mise en place, toutes les manipulations s'effectuent à partir de l'ordinateur.

Les réactifs sont placés sur tamis moléculaires 3 ou 4 Å pendant 24h. On prépare alors la résine sur laquelle est greffé le premier nucléotide. Les synthèses de nos différents brins ont été réalisées à l'échelle de 10 µmol d'oligonucléotides (environ 400mg de résine). La résine a été pesée et insérée dans la colonne du synthétiseur. La colonne est alors équilibrée (débit

d'environ 3,4ml par 1min30, indiquant que la colonne n'est ni trop ni pas assez tassée). Le système est purgé avec de l'acétonitrile. Le taux de couplage minimum (sinon la synthèse s'arrête) est fixé à 70%, la détritylation du dernier résidu est sélectionnée.

Après chaque cycle de couplage, la détritylation est observée par l'apparition d'une coloration rouge-orangé, correspondant aux radicaux trityls. Le rendement de chaque étape de couplage doit être compris entre 99 et 100%. Une fois la synthèse terminée, la résine est séchée sous azote. On obtient finalement une poudre blanchâtre (ou légèrement jaunâtre) contenant l'oligonucléotide greffé sur la résine.

#### Purification des différents brins d'ADN

Les différents oligonucléotides fixés au support solide ont été placés pendant une nuit à 65°C dans de l'ammoniaque à 28%, permettant ainsi le clivage des oligonucléotides de la résine ainsi que la déprotection des bases azotées. Le surnageant est prélevé, et la résine est lavée avec de l'eau distillée. L'ammoniaque est évaporé au Büchi. Les différents brins sont dissous dans environ 1,5ml d'eau et sont élués sur une colonne d'exclusion Shephadex G10. Les groupements protecteurs libres sont retenus par la colonne alors que les oligonucléotides, plus gros, migrent dans le volume mort. La purification des oligonucléotides a ensuite été réalisée par HPLC en phase inverse en utilisant comme solvant un mélange ACN/AcO<sup>-</sup>/NH<sub>4</sub><sup>+</sup>. Les masses ont été contrôlées par Spectrométrie de Masse et les Densités Optiques ont été mesurées. Cette synthèse nous a permis d'obtenir un échantillon avec une concentration finale en oligonucléotides d'environ 1,5mM dans un volume de 280µl.

## ANNEXE 2 : Tableau de déplacements chimiques des différentes macromolécules étudiées

Télomère 2 répétitions (25°C)

| residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''  | NH/NH2             |
|------------|-------|-----------|------|-----------|--------------------|
| <i>G1</i>  | 7,92  |           | 6,07 | 2,64/2,77 | 13,06              |
| <i>T2</i>  | 7,49  | 1,32      | 6,18 | 2,18/2,55 | 13,98              |
| <i>T3</i>  | 7,43  | 1,64      | 5,75 | 2,06/2,33 | 13,59              |
| <i>A4</i>  | 8,20  | 7,26      | 6,00 | 2,02/2,35 |                    |
| <i>G5</i>  | 7,56  |           | 5,52 | 2,42      | 12,84              |
| <i>G6</i>  | 7,51  |           | 5,78 | 2,44/2,66 | 12,92              |
| <i>G7</i>  | 7,47  |           | 5,92 | 2,41/2,48 | 12,84              |
| <i>T8</i>  | 7,24  | 1,16      | 6,01 | 2,49/2,67 | 13,88              |
| <i>T9</i>  | 7,37  | 1,57      | 5,74 | 2,03/2,44 | 13,68              |
| <i>A10</i> | 8,19  | 7,26      | 5,97 | 2,62/2,78 |                    |
| <i>G11</i> | 7,60  |           | 5,53 | 2,45/2,53 | 12,88              |
| <i>G12</i> | 7,56  |           | 5,74 | 2,45/2,61 | 13,10              |
| <i>G13</i> | 7,65  |           | 6,03 | 2,45/2,62 | 13,06              |
| <i>T14</i> | 7,41  | 1,46      | 6,26 | 2,17/2,49 | 13,69              |
| <i>A28</i> | 8,27  | 8,03      | 6,28 | 2,67/2,75 |                    |
| <i>C27</i> | 7,50  | 5,51      | 5,90 | 2,43      | 4f 6,79<br>4b 8,18 |
| <i>C26</i> | 7,54  | 5,58      | 6,01 | 2,01/2,18 | 4f 6,85<br>4b 8,44 |
| <i>C25</i> | 7,54  | 5,58      | 5,90 | 1,99/2,17 | 4f 6,98<br>4b 8,50 |
| <i>T24</i> | 7,36  |           | 5,56 | 1,98/2,30 | 13,69              |
| <i>A23</i> | 8,23  | 6,90      | 5,90 | 2,65/2,77 |                    |
| <i>A22</i> | 8,09  | 7,57      | 6,07 | 2,51/2,73 |                    |
| <i>C21</i> | 7,20  | 5,15      | 5,71 | 1,97/2,33 | 4f 6,54<br>4b 7,95 |
| <i>C20</i> | 7,44  | 5,38      | 5,89 | 1,96/2,35 | 4f 6,64<br>4b 8,35 |
| <i>C19</i> | 7,50  | 5,44      | 5,87 | 1,92/2,38 | 4f 6,93<br>4b 8,40 |
| <i>T18</i> | 7,38  | 1,60      | 5,58 | 2,00/2,35 | 13,69              |
| <i>A17</i> | 8,25  | 6,89      | 5,94 | 2,66/2,78 |                    |
| <i>A16</i> | 8,11  | 7,69      | 6,15 | 2,47/2,73 |                    |
| <i>C15</i> | 7,31  | 5,32      | 6,01 | 1,98/2,03 | 4f 6,77<br>4b 8,12 |

## ADN télomérique contenant trois répétitions (23°C)

| residue    | H6/H8 | H5/H2/<br>CH3 | H1'  | H2'/H2''  | NH/N<br>H2               | residue    | H6/H8 | H5/H2/<br>CH3 | H1'  | H2'/H2''  | NH/NH2             |
|------------|-------|---------------|------|-----------|--------------------------|------------|-------|---------------|------|-----------|--------------------|
| <b>G1</b>  | 7,94  |               | 6,04 | 2,69/2,97 | <i>non<br/>déterminé</i> | <b>C21</b> | 7,32  | 5,46          | 5,99 | 1,75/1,84 | 4f 6,77<br>4b 8,11 |
| <b>T2</b>  | 7,47  | 1,43          | 6,14 | 2,20/2,60 | 14,05                    | <b>A22</b> | 8,09  | 7,75          | 6,12 | 2,55/2,79 |                    |
| <b>T3</b>  | 7,41  | 1,69          | 5,71 | 2,12/2,43 | 13,72                    | <b>A23</b> | 8,23  | 7,05          | 5,88 | 2,69/2,85 |                    |
| <b>A4</b>  | 8,18  | 7,35          | 4,95 | 2,60/2,84 |                          | <b>T24</b> | 7,33  | 1,65          | 5,55 | 2,03/2,37 | 13,65              |
| <b>G5</b>  | 7,53  |               | 5,49 | 2,48/2,65 | 12,8                     | <b>C25</b> | 7,47  | 5,48          | 5,86 | 2,05/2,46 | 4f 6,90<br>4b 8,37 |
| <b>G6</b>  | 7,48  |               | 5,72 | 2,46/2,70 | 12,8                     | <b>C26</b> | 7,42  | 5,37          | 5,87 | 2,03/2,40 | 4f 6,65<br>4b 8,23 |
| <b>G7</b>  | 7,43  |               | 5,86 | 2,44/2,70 | 12,8                     | <b>C27</b> | 7,19  | 5,15          | 5,86 | 2,01/2,39 | 4f 6,50<br>4b 7,85 |
| <b>T8</b>  | 7,21  | 1,62          | 5,95 | 2,04/2,54 | 13,88                    | <b>A28</b> | 8,06  | 7,57          | 6,00 | 2,60/2,77 |                    |
| <b>T9</b>  | 7,31  | 1,80          | 5,67 | 2,07/2,42 | 13,62                    | <b>A29</b> | 8,18  | 6,97          | 5,84 | 2,72/2,85 |                    |
| <b>A10</b> | 8,18  | 7,32          | 5,93 | 2,72/2,84 |                          | <b>T30</b> | 7,29  | 1,60          | 5,46 | 2,01/2,32 | 13,65              |
| <b>G11</b> | 7,53  |               | 5,49 | 2,48/2,65 | 12,8                     | <b>C31</b> | 7,42  | 5,48          | 5,81 | 2,01/2,70 | 4f 6,90<br>4b 8,37 |
| <b>G12</b> | 7,47  |               | 5,72 | 2,46/2,70 | 12,8                     | <b>C32</b> | 7,38  | 5,37          | 5,85 | 2,01/2,40 | 4f 6,65<br>4b 8,23 |
| <b>G13</b> | 7,43  |               | 5,86 | 2,44/2,70 | 12,8                     | <b>C33</b> | 7,17  | 5,13          | 5,69 | 2,01/2,39 | 4f 6,50<br>4b 7,85 |
| <b>T14</b> | 7,21  | 1,59          | 5,96 | 2,03/2,53 | 13,88                    | <b>A34</b> | 8,06  | 7,57          | 6,00 | 2,60/2,77 |                    |
| <b>T15</b> | 7,33  | 1,78          | 5,68 | 2,05/2,39 | 13,62                    | <b>A35</b> | 8,19  | 6,97          | 5,85 | 2,72/2,85 |                    |
| <b>A16</b> | 8,16  | 7,32          | 5,94 | 2,68/2,73 |                          | <b>T36</b> | 7,32  | 1,65          | 5,50 | 2,01/2,37 | 13,65              |
| <b>G17</b> | 7,58  |               | 5,48 | 2,49/2,61 | 12,85                    | <b>C37</b> | 7,51  | 5,59          | 5,87 | 2,05/2,45 | 4f 6,95<br>4b 8,43 |
| <b>G18</b> | 7,52  |               | 5,71 | 2,52/2,68 | 12,85                    | <b>C38</b> | 7,51  | 5,62          | 5,97 | 2,22/2,52 | 4f 6,95<br>4b 8,43 |
| <b>G19</b> | 7,64  |               | 5,99 | 2,52/2,67 | 12,98                    | <b>C39</b> | 7,48  | 5,56          | 5,87 | 2,22/2,53 | 4f 6,72<br>4b 8,08 |
| <b>T20</b> | 7,41  | 1,55          | 6,21 | 2,16/2,25 | <i>non<br/>déterminé</i> | <b>A40</b> | 8,24  |               | 6,24 | 2,72/2,85 |                    |

Téломère 2 répétitions complexé au domaine Myb de TRF2 (20°C)

| residue    | H6/H8 | H5/H2/CH3 | H1'  | NH/NH2                        |
|------------|-------|-----------|------|-------------------------------|
| <i>G1</i>  | 7,98  |           | 6,06 |                               |
| <i>T2</i>  | 7,48  | 1,48      | 6,10 | 14,22                         |
| <i>T3</i>  | 7,69  | 1,73      | 6,01 | 13,53                         |
| <i>A4</i>  | 8,22  | 7,08      | 6,16 |                               |
| <i>G5</i>  | 7,61  |           | 5,21 | 13,02                         |
| <i>G6</i>  | 7,72  |           | 6,00 | 13,16                         |
| <i>G7</i>  | 7,38  |           | 5,82 | 12,62                         |
| <i>T8</i>  | 7,22  | 1,17      | 6,05 | 13,94                         |
| <i>T9</i>  | 7,32  | 1,58      | 5,49 | 13,32                         |
| <i>A10</i> | 8,27  | 7,63      | 6,02 |                               |
| <i>G11</i> | 7,66  |           | 5,42 | 12,80                         |
| <i>G12</i> | 7,58  |           | 5,77 | 12,98                         |
| <i>G13</i> | 7,65  |           | 6,02 | 12,79                         |
| <i>T14</i> | 7,41  | 1,57      | 6,25 |                               |
| <i>A28</i> | 8,27  | 8,12      | 6,25 |                               |
| <i>C27</i> | 7,55  | 5,45      | 5,90 | 4f 6,98<br>4b 8,42<br>4f 6,75 |
| <i>C26</i> | 7,53  | 5,56      | 6,02 | 4b 8,32<br>4f 6,98            |
| <i>C25</i> | 7,39  | 5,59      | 5,89 | 4b 8,42                       |
| <i>T24</i> | 7,38  | 1,65      | 5,60 | 13,92                         |
| <i>A23</i> | 8,23  | 7,16      | 6,15 |                               |
| <i>A22</i> | 8,33  | 7,53      | 6,11 |                               |
| <i>C21</i> | 7,50  | 5,56      | 5,75 | 4f 6,90<br>4b 8,37            |
| <i>C20</i> | 7,31  | 5,22      | 6,05 |                               |
| <i>C19</i> | 7,61  | 5,78      | 6,05 | 4f 7,03<br>4b 8,45            |
| <i>T18</i> | 7,39  | 1,73      | 5,20 | 13,94                         |
| <i>A17</i> | 8,28  | 7,05      | 6,11 |                               |
| <i>A16</i> | 8,14  | 7,77      | 6,10 |                               |
| <i>C15</i> | 7,28  | 5,26      | 6,01 | 4f 6,90<br>4b 8,37            |

Télomère 3R/2R (20°C)

| residue | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''  | NH/NH2             | residue | H6/H8 | H5/H2/CH3 | H1'   | H2'/H2''  | NH/NH2             |
|---------|-------|-----------|------|-----------|--------------------|---------|-------|-----------|-------|-----------|--------------------|
| C1      | 7,82  | 5,95      | 5,92 | 2,09/2,52 | 4f 7,10<br>4b 7,81 | A38     | 8,24  |           | 6,25  | 2,70/2,80 |                    |
| T2      | 7,51  | 1,65      | 5,88 | 2,21/2,51 | 13,87              | C37     | 7,53  | 5,52      | 5,96  | 2,22/2,48 | 4f 6,82            |
| G3      | 7,92  |           | 6,05 | 2,61/2,77 | 12,59              | C36     | 7,53  | 5,57      | 5,96  | 2,22/2,48 | 4b 8,35<br>4f 6,88 |
| T4      | 7,25  | 1,35      | 5,99 | 2,02/2,49 | 13,82              | C35     | 7,53  | 5,52      | 5,92  | 2,21/2,48 | 4b 8,48<br>4f 6,88 |
| T5      | 7,34  | 1,59      | 5,65 | 2,01/2,35 | 13,63              | T34     | 7,35  | 1,58      | 5,628 | 2,00/2,36 | 4b 8,53<br>13,82   |
| A6      | 8,12  | 7,19      | 5,91 | 2,59/2,78 |                    | A33     | 8,18  | 6,68      | 5,82  | 2,61/2,75 |                    |
| G7      | 7,49  |           | 5,48 | 2,52/2,65 | 12,72              | A32     | 8,1   | 7,49      | 6,05  | 2,52/2,73 |                    |
| G8      | 7,48  |           | 5,7  | 2,48/2,64 | 12,79              | C31     | 7,16  | 5,11      | 5,68  | 1,95/2,32 | 4f 6,45            |
| G9      | 7,43  |           | 5,85 | 2,49/2,62 | 12,81              | C30     | 7,38  | 5,34      | 5,85  | 2,09/2,39 | 4b 7,83<br>4f 6,58 |
| T10     | 7,16  | 1,12      | 6,05 | 1,92/2,40 | 13,75              | C29     | 7,42  | 5,49      | 5,72  | 1,99/2,39 | 4b 8,20<br>4f 6,82 |
| T11     | 7,31  | 1,56      | 5,64 | 2,02/2,42 | 13,36              | T28     | 7,35  | 1,58      | 5,52  | 2,00/2,32 | 4b 8,35<br>13,57   |
| A12     | 8,12  | 7,09      | 5,9  | 2,60/2,80 |                    | A27     | 8,24  | 6,82      | 5,85  | 2,72/2,81 |                    |
| G13     | 7,48  |           | 5,71 | 2,66/2,84 | 12,91              | A26     | 8,12  | 7,51      | 6,01  | 2,50/2,72 |                    |
| G14     | 7,49  |           | 5,43 | 2,61      | 12,78              | C25     | 7,15  | 5,22      | 5,34  | 1,72/2,11 | 4f 6,46<br>4b 8,15 |
| G15     | 7,49  |           | 5,72 | 2,51      |                    | A24     | 8,11  | 7,65      | 6,01  | 2,60/2,78 |                    |
| T16     | 7,13  | 1,22      | 5,74 |           |                    | G23     | 7,7   |           | 5,97  | 2,21/2,39 |                    |
| T17     |       | 1,54      |      |           |                    |         |       |           |       |           |                    |
| A18     |       |           |      |           |                    |         |       |           |       |           |                    |
| G19     |       |           |      |           |                    |         |       |           |       |           |                    |
| G20     |       |           |      |           |                    |         |       |           |       |           |                    |
| G21     |       |           |      |           |                    |         |       |           |       |           |                    |
| T22     |       |           |      |           |                    |         |       |           |       |           |                    |

ADN télomérique 3R/2R complexé au domaine Myb de TRF2 (25°C)

| residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''  | NH/NH2 | residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''     | NH/NH2  |
|------------|-------|-----------|------|-----------|--------|------------|-------|-----------|------|--------------|---------|
| <i>C1</i>  | 7,81  | 5,94      | 5,94 | 2,14/2,58 |        | <i>A38</i> | 8,22  |           | 6,24 | 2,75/2,82    |         |
| <i>T2</i>  | 7,49  | 1,71      | 5,84 | 2,26/2,55 |        | <i>C37</i> | 7,52  | 5,46      | 5,90 | recouvrement | 4f 6,82 |
| <i>G3</i>  | 7,94  |           | 6,04 | 2,71/2,82 | 12,65  | <i>C36</i> | 7,51  | 5,55      | 5,92 | recouvrement | 4b 8,38 |
| <i>T4</i>  | 7,20  | 1,48      | 5,91 | 2,10/2,59 | 13,82  | <i>C35</i> | 7,52  | 5,54      | 5,89 | 2,18/2,51    | 4f 6,82 |
| <i>T5</i>  | 7,32  | 1,62      | 5,88 | 2,09/2,56 | 13,51  | <i>T34</i> | 7,34  | 1,61      | 5,53 | 2,10/2,32    | 4b 8,38 |
| <i>A6</i>  | 8,11  | 7,50      | 6,05 | 2,85      |        | <i>A33</i> | 8,11  | 7,09      | 6,00 | 2,73         |         |
| <i>G7</i>  | 7,55  |           | 5,11 | 2,22/2,55 | 12,92  | <i>A32</i> | 8,20  | 7,48      | 6,08 | 2,92         |         |
| <i>G8</i>  | 7,55  |           | 5,75 | 2,22/2,50 | 13,13  | <i>C31</i> | 7,46  | 5,52      | 5,97 | 2,10/2,34    |         |
| <i>G9</i>  | 7,38  |           | 5,77 | 2,11/2,35 | 12,64  | <i>C30</i> | 7,28  | 5,17      | 5,99 | 2,11/2,34    | 4f 6,82 |
| <i>T10</i> | 7,15  | 1,11      | 5,98 | 2,04      | 13,84  | <i>C29</i> | 7,55  | 5,54      | 5,95 | 2,12/2,49    | 4b 8,09 |
| <i>T11</i> | 7,32  |           | 5,71 | 2,12/2,32 | 13,28  | <i>T28</i> | 7,27  |           | 5,18 | 2,10/2,56    | 4f 6,82 |
| <i>A12</i> | 8,14  | 7,52      | 6,02 | 2,85      |        | <i>A27</i> | 8,22  | 6,88      | 6,01 | 2,92         | 4b 8,38 |
| <i>G13</i> | 7,55  |           | 5,70 | 2,22/2,55 | 12,81  | <i>A26</i> | 8,13  | 7,52      | 5,97 | 2,55/2,75    |         |
| <i>G14</i> | 7,53  |           | 5,44 | 2,24/2,54 | 12,91  | <i>C25</i> | 7,14  | 5,21      | 5,38 | 1,75/2,15    | 4f 6,42 |
| <i>G15</i> | 7,55  |           | 5,74 | 2,18/2,49 | 12,80  | <i>A24</i> | 8,12  |           | 6,03 | 2,65/2,82    | 4b 8,10 |
| <i>T16</i> | 7,16  |           | 5,79 | 2,05      |        | <i>G23</i> | 7,72  |           | 6,02 | 2,26/2,45    |         |
| <i>T17</i> |       |           |      |           |        |            |       |           |      |              |         |
| <i>A18</i> |       |           |      |           |        |            |       |           |      |              |         |
| <i>G19</i> |       |           |      |           |        |            |       |           |      |              |         |
| <i>G20</i> |       |           |      |           |        |            |       |           |      |              |         |
| <i>G21</i> |       |           |      |           |        |            |       |           |      |              |         |
| <i>T22</i> |       |           |      |           |        |            |       |           |      |              |         |

### Modèle de loop C (10°C)

| residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''     | NH/NH2             | residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''     | NH/NH2             |
|------------|-------|-----------|------|--------------|--------------------|------------|-------|-----------|------|--------------|--------------------|
| <i>T1</i>  | 7,59  | 1,72      | 6,12 | 2,25/2,54    | non-observé        | <i>A17</i> | 8,18  | 7,72      | 6,13 | 2,61/2,82    |                    |
| <i>C2</i>  | 7,75  | 5,82      | 6,06 | 2,33/2,56    | 4f 7,16<br>4b 8,42 | <i>C18</i> | 7,16  | 5,16      | 5,51 | 1,79/2,21    | 4f 6,48<br>4b 8,09 |
| <i>T3</i>  | 7,48  | 1,66      | 6,05 | 2,20/2,41    | 13,92              | <i>G19</i> | 7,83  |           | 5,39 | 2,61/2,69    | 12,76              |
| <i>C4</i>  | 7,53  | 5,66      | 5,56 | 2,30/2,54    | 4f 6,96<br>4b 8,55 | <i>A20</i> | 8,05  | 7,68      | 5,91 | 2,56/2,74    |                    |
| <i>G5</i>  | 7,94  |           | 5,97 | 2,81         | 12,81              | <i>G21</i> | 7,66  |           | 5,48 | 2,40         | 12,71              |
| <i>C6</i>  | 7,75  | 5,89      | 5,96 | 2,22/2,47    | 4f 6,98<br>4b 8,34 | <i>A22</i> | 8,09  |           | 6,30 | 2,55         |                    |
| <i>T7</i>  |       | 1,44      |      | 2,20/2,41    | 13,92              | <i>G1</i>  |       |           |      |              |                    |
| <i>C18</i> | 7,76  | 5,90      | 5,96 | 2,21/2,44    | 4f 7,28<br>4b 8,35 | <i>C2</i>  |       |           |      |              |                    |
| <i>T19</i> | 7,38  | 1,71      | 5,74 | 2,13/2,42    | 14,12              | <i>G3</i>  |       |           |      |              |                    |
| <i>G20</i> | 7,94  |           | 5,82 | 2,64         | 12,9               | <i>A4</i>  |       |           |      |              |                    |
| <i>C21</i> | 7,39  | 5,47      | 5,76 | 1,99/2,30    | 4f 6,71<br>4b 8,38 | <i>A5</i>  |       |           |      |              |                    |
| <i>A22</i> | 8,25  |           | 6,30 | 2,45/2,66    |                    | <i>G6</i>  |       |           |      |              |                    |
| <i>T1</i>  | 7,38  | 1,56      | 5,88 | 1,86/2,24    | non-observé        | <i>C7</i>  |       |           |      |              |                    |
| <i>G2</i>  | 8,04  |           | 5,97 | 2,74         | 12,87              | <i>T8</i>  |       |           |      |              |                    |
| <i>C3</i>  | 7,42  | 5,46      | 6,04 | 2,01/2,25    | 4f 6,62<br>4b 8,50 | <i>T9</i>  | 7,32  |           | 5,75 |              | 13,52              |
| <i>A4</i>  | 8,12  | 7,78      | 6,04 | 2,68/2,82    |                    | <i>A10</i> | 8,12  |           | 6,11 | 2,65/2,75    |                    |
| <i>G5</i>  | 7,68  |           | 5,78 | 2,42         | 12,79              | <i>G11</i> | 7,55  |           | 5,60 | Recouvrement | 12,82              |
| <i>C6</i>  | 7,54  | 5,55      | 5,91 | 2,20/2,44    | 4f 6,92<br>4b 8,35 | <i>G12</i> | 7,54  |           | 5,74 | 2,20/2,45    | 12,90              |
| <i>C7</i>  | 7,54  | 5,55      | 5,96 | 2,20/2,44    | 4f 6,92<br>4b 8,35 | <i>G13</i> | 7,48  |           | 5,84 | 2,20/2,45    | 12,79              |
| <i>C8</i>  | 7,52  | 5,56      | 5,86 | 2,20/2,44    | 4f 6,96<br>4b 8,38 | <i>T14</i> | 7,19  | 1,19      | 5,96 | 2,04/2,51    | 13,84              |
| <i>T9</i>  | 7,34  | 1,58      | 5,46 | 2,02/2,42    | 13,64              | <i>T15</i> | 7,32  | 1,59      | 5,62 | 2,32/2,52    | 13,63              |
| <i>A10</i> | 8,22  | 6,88      | 5,84 | 2,66/2,82    |                    | <i>A16</i> | 8,16  | 7,28      | 5,86 |              |                    |
| <i>A11</i> | 8,09  | 7,49      | 6,01 | 2,60/2,72    |                    | <i>G17</i> | 7,51  |           | 5,88 | recouvrement | 12,76              |
| <i>C12</i> | 7,16  | 5,09      | 5,66 | 1,78/2,23    | 4f 6,48<br>4b 8,05 | <i>G18</i> | 7,60  |           | 5,72 | recouvrement | 12,78              |
| <i>C13</i> | 7,38  | 5,32      | 5,82 | Recouvrement | 4f 6,68<br>4b 8,50 | <i>G19</i> | 7,92  |           | 6,06 | 2,58/2,71    | 13,06              |
| <i>C14</i> | 7,46  | 5,43      | 5,84 | Recouvrement | 4f 6,88<br>4b 8,32 |            |       |           |      |              |                    |
| <i>T15</i> | 7,34  | 1,58      | 5,46 | 2,02/2,40    | 13,62              |            |       |           |      |              |                    |
| <i>A16</i> | 8,18  | 7,38      | 5,76 | 2,61/2,69    |                    |            |       |           |      |              |                    |

## Hairpin et résonance caractéristiques (20°C)

| residue      | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''  |
|--------------|-------|-----------|------|-----------|
| <i>G1</i>    | 7,89  |           | 5,91 | 2,42/2,65 |
| <i>C2</i>    | 6,95  | 5,01      | 5,93 | 1,59/2,25 |
| <i>G3</i>    | 8,05  |           | 5,38 | 2,51/2,63 |
| <i>A4</i>    | 8,12  |           | 5,95 | 2,25/2,42 |
| <i>A5</i>    | 8,02  |           | 6,31 | 2,85      |
| <i>G6</i>    | 8,00  |           | 5,52 | 2,48/2,61 |
| <i>C7</i>    | 7,48  | 5,40      | 5,53 | 2,15/2,35 |
| <i>T8...</i> |       |           |      |           |

| residue   | H4'  | H5'  | H5'' |
|-----------|------|------|------|
| <i>A4</i> | 2,12 | 3,40 | 3,09 |

### Modèle de loop A (10°C)

| residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''     | NH/NH2             | residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''     | NH/NH2 |
|------------|-------|-----------|------|--------------|--------------------|------------|-------|-----------|------|--------------|--------|
| <i>T1</i>  | 7,56  | 1,72      | 6,07 | 2,26/2,56    | non-observable     | <i>A18</i> | 8,07  |           | 6,27 | 2,38/2,53    |        |
| <i>C2</i>  | 7,72  | 5,76      | 6,04 | 2,21/2,52    | 4f 7,12<br>4b 8,45 | <i>G1</i>  | 7,92  |           | 6,11 | 2,24/2,52    | 12,51  |
| <i>T3</i>  | 7,45  | 1,62      | 5,92 | 2,12/2,47    | 13,91              | <i>T2</i>  | 7,46  | 1,07      | 5,92 | 2,44         | 13,95  |
| <i>C4</i>  | 7,49  | 5,63      | 5,49 | 2,18/2,40    | 4f 6,90<br>4b 8,37 | <i>T3</i>  | 7,38  | 1,55      | 5,71 | 2,45         | 13,63  |
| <i>G5</i>  | 7,90  | 1,45      | 5,93 | recouvrement | 12,71              | <i>A4</i>  | 8,17  | 7,28      | 5,92 | 2,68/2,79    |        |
| <i>T6</i>  | 7,19  | 1,43      | 6,09 | recouvrement | non-observable     | <i>G5</i>  | 7,60  |           | 5,38 | Recouvrement | 12,84  |
| <i>CI4</i> | 7,43  | 5,27      | 5,62 | recouvrement | 4f 6,78<br>4b 8,34 | <i>G6</i>  | 7,58  |           | 5,64 | Recouvrement | 12,98  |
| <i>TI5</i> | 7,48  | 1,55      | 5,84 | 2,09/2,40    | 13,98              | <i>G7</i>  | 7,69  |           | 5,89 | 2,42/2,65    | 12,82  |
| <i>G16</i> | 7,86  |           | 5,75 | 2,26/2,49    | 12,71              |            |       |           |      |              |        |
| <i>CI7</i> | 7,33  | 5,41      | 5,72 | 1,93/2,26    | 4f 6,69<br>4b 8,35 |            |       |           |      |              |        |
| <i>A18</i> | 8,21  |           | 6,28 | 2,42/2,64    |                    |            |       |           |      |              |        |
| <i>T1</i>  | 7,37  | 1,56      | 5,86 | 1,88/2,27    | non-observable     |            |       |           |      |              |        |
| <i>G2</i>  | 8,01  |           | 5,94 | 2,50/2,71    | 12,83              |            |       |           |      |              |        |
| <i>C3</i>  | 7,37  | 5,41      | 5,48 | 2,04/2,39    | 4f 6,58<br>4b 8,43 |            |       |           |      |              |        |
| <i>A4</i>  | 8,15  | 7,55      | 6,07 | 2,69/2,91    |                    |            |       |           |      |              |        |
| <i>G5</i>  | 7,61  |           | 5,95 | 2,25/2,49    | 12,96              |            |       |           |      |              |        |
| <i>C6</i>  | 7,51  | 5,57      | 5,87 | recouvrement | 4f 6,88<br>4b 8,37 |            |       |           |      |              |        |
| <i>C7</i>  | 7,56  | 5,65      | 5,94 | recouvrement | 4f 7,29<br>4b 7,94 |            |       |           |      |              |        |
| <i>C8</i>  | 7,59  | 5,68      | 5,79 | recouvrement | 4f 6,97<br>4b 8,38 |            |       |           |      |              |        |
| <i>T9</i>  | 7,35  | 1,63      | 5,45 | 2,08/2,39    | 13,67              |            |       |           |      |              |        |
| <i>A10</i> | 8,22  | 6,82      | 5,91 | 2,72/2,87    |                    |            |       |           |      |              |        |
| <i>A11</i> | 8,05  | 7,45      | 6,01 | 2,77/2,82    |                    |            |       |           |      |              |        |
| <i>CI2</i> | 7,09  | 5,10      | 5,53 | 1,90/2,34    | 4f 6,90<br>4b 8,37 |            |       |           |      |              |        |
| <i>A13</i> | 7,99  |           | 6,05 | 2,50/2,73    |                    |            |       |           |      |              |        |
| <i>CI4</i> | 7,15  | 5,10      | 5,42 | 1,82/2,20    | 4f 6,45<br>4b 8,12 |            |       |           |      |              |        |
| <i>G15</i> | 7,81  |           | 5,34 | 2,56/2,68    | 12,78              |            |       |           |      |              |        |
| <i>A16</i> | 8,03  | 7,69      | 5,89 | 2,42/2,53    |                    |            |       |           |      |              |        |
| <i>G17</i> | 7,62  |           | 5,45 | 2,42/2,497   | 12,73              |            |       |           |      |              |        |

Modèle de loop B (10°C)

| residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''     | NH/NH2             | residue    | H6/H8 | H5/H2/CH3 | H1'  | H2'/H2''     | NH/NH2 |
|------------|-------|-----------|------|--------------|--------------------|------------|-------|-----------|------|--------------|--------|
| <i>T1</i>  | 7,56  | 1,73      | 6,09 | 2,28/2,52    | non-observable     | <i>A16</i> | 8,04  | 7,62      | 5,92 | 2,72/2,76    |        |
| <i>C2</i>  | 7,75  | 5,81      | 6,07 | 2,25/2,56    | 4f 7,12<br>4b 8,43 | <i>G17</i> | 7,64  |           | 5,48 | 2,62/2,72    | 12,78  |
| <i>T3</i>  | 7,51  | 1,68      | 6,09 | 2,21/2,41    | 13,76              | <i>A18</i> | 8,08  |           | 6,29 | 2,41/2,56    |        |
| <i>C4</i>  | 7,52  | 5,71      | 5,68 | 2,05/2,41    | 4f 7,05<br>4b 8,51 | <i>A0</i>  | 7,86  |           | 5,92 |              |        |
| <i>G5</i>  | 7,53  |           | 5,64 | 2,35/2,52    | 12,97              | <i>G1</i>  | 7,81  |           | 6,03 | 2,71         | 12,39  |
| <i>T6</i>  | 7,35  | 1,62      |      | recouvrement | Non-observable     | <i>T2</i>  | 7,32  | 1,35      | 5,92 | 2,10/2,41    | 13,71  |
| <i>C14</i> | 7,39  | 5,28      | 5,60 | 2,21         | 4f 6,68<br>4b 7,87 | <i>T3</i>  | 7,31  | 1,62      | 5,68 | 2,05/2,38    | 13,71  |
| <i>T15</i> | 7,48  | 1,53      | 5,92 | 2,15/2,51    | 13,84              | <i>A4</i>  | 8,16  | 7,35      | 5,91 | 2,68/2,81    |        |
| <i>G16</i> | 7,86  |           | 5,80 | 2,58/2,65    | 12,67              | <i>G5</i>  | 7,58  |           | 5,36 | recouvrement | 12,82  |
| <i>C17</i> | 7,55  | 5,58      | 5,82 | 2,01/2,38    | 4f 6,47<br>4b 8,34 | <i>G6</i>  | 7,58  |           | 5,64 | 2,54/2,64    | 12,92  |
| <i>A18</i> | 8,24  | 7,82      | 6,23 | 2,68/2,88    |                    | <i>G7</i>  | 7,69  |           | 5,92 | 2,45/2,66    | 12,92  |
| <i>C00</i> | 7,31  | 5,29      | 6,04 | 2,10/2,41    | 4f 6,83<br>4b 8,12 |            |       |           |      |              |        |
| <i>G00</i> | 7,92  |           | 5,86 | 2,61/2,78    |                    |            |       |           |      |              |        |
| <i>T1</i>  | 7,37  | 1,32      | 5,91 | 2,05/2,36    | 13,76              |            |       |           |      |              |        |
| <i>G2</i>  | 7,87  |           | 5,86 | 2,52/2,68    | 12,79              |            |       |           |      |              |        |
| <i>C3</i>  | 7,34  | 5,37      | 5,51 | 2,00/2,39    | 4f 6,45<br>4b 8,33 |            |       |           |      |              |        |
| <i>A4</i>  | 8,12  | 7,49      | 6,08 | 2,68/2,90    |                    |            |       |           |      |              |        |
| <i>G5</i>  | 7,52  |           | 5,64 | 2,41/2,54    | 12,92              |            |       |           |      |              |        |
| <i>C6</i>  | 7,48  | 5,57      | 5,85 | 2,22/2,41    | 4f 6,69<br>4b 8,25 |            |       |           |      |              |        |
| <i>C7</i>  | 7,50  | 5,57      | 5,85 | 2,22/2,41    | 4f 6,77<br>4b 8,15 |            |       |           |      |              |        |
| <i>C8</i>  | 7,51  | 5,49      | 5,81 | 2,22/2,41    | 4f 6,77<br>4b 8,15 |            |       |           |      |              |        |
| <i>T9</i>  | 7,31  | 1,63      | 5,51 |              | 13,58              |            |       |           |      |              |        |
| <i>A10</i> | 8,23  | 6,92      | 5,86 | 2,72/2,82    |                    |            |       |           |      |              |        |
| <i>A11</i> | 8,13  | 7,62      | 6,00 |              |                    |            |       |           |      |              |        |
| <i>C12</i> | 7,32  | 5,11      | 5,53 |              | 4f 6,62<br>4b 8,00 |            |       |           |      |              |        |
| <i>A13</i> | 8,11  | 7,76      | 6,03 | 2,69/2,91    |                    |            |       |           |      |              |        |
| <i>C14</i> | 7,22  | 5,31      | 5,40 |              | 4f 6,68<br>4b 8,22 |            |       |           |      |              |        |
| <i>G15</i> | 7,84  |           | 5,36 |              | 12,79              |            |       |           |      |              |        |

## Domaine Myb de TRF2 seule

Table 1 resonance assignments (proton chemical shifts are given in parentheses).

| residue | N              | CO      | C <sup>α</sup> | C <sup>β</sup>       | Autres  |
|---------|----------------|---------|----------------|----------------------|---|
| S2      | (8,839)        |         | (4.618)        | (3.980)              |   |
| H3      | (8,874)        |         | (4.860)        | (3.438, 3.326)       | C <sup>δ2</sup> , * (7.416); C <sup>ε1</sup> , * (8.684)  |
| M4      | (8,643)        | 179,793 | (4.574)        | (2.198, 2.107)       | C <sup>γ</sup> , * (2.643)  |
| E5      | (8,758)        | 180,756 | 55,358(4.425)  | 28,966(2.202, 2.080) | C <sup>γ</sup> , * (2.420)  |
| D6      | 123,365(8,519) | 181,210 | 53,031(4.779)  | 39,905(2.877, 2.829) |   |
| S7      | 118,254(8,532) | 179,901 | 57,558(4.631)  | 62,406(4.111, 4.030) |   |
| T8      | 117,110(8,412) | 179,870 | 61,243(4.508)  | 68,191(4.436)        | C <sup>γ2</sup> , * (1.364)   |
| T9      | 117,334(8,207) |         | 61,158(4.437)  | 68,194(4.359)        | C <sup>γ2</sup> , * (1.335)   |
| N10     | 117,440(8,051) | 179,910 | 52,302(4.845)  | 999(2.974, 2.914)    | N <sup>δ2</sup> , * (7.775, 7.065)  |
| I11     | 122,624(8,060) | 180,301 | 60,416(4.314)  | 37,405(2.056)        | C <sup>γ1</sup> , * (1.348); C <sup>γ2</sup> , * (1.047); C <sup>δ1</sup> , * (0.998)   |
| T12     | 123,788(8,170) | 179,388 | 61,158(4.408)  | 68,347(4.347)        | C <sup>γ2</sup> , * (1.351)   |
| K13     | 125,421(8,186) | 180,550 | 58,652(4.417)  | 31,623(1.978, 1.908) | C <sup>γ</sup> , * (1.540); C <sup>δ</sup> , * (1.820); C <sup>ε</sup> , * (3.121)  |
| K14     | 127,189(8,406) | 180,756 | 56,625(4.410)  | 29,435(1.968, 1.898) | C <sup>γ</sup> , * (1.539); C <sup>δ</sup> , * (1.846); C <sup>ε</sup> , * (3.152)  |
| Q15     | 123,408(8,367) | 180,901 | 54,169(4.462)  | 28,498(2.155, 2.077) | C <sup>γ</sup> , * (2.450, *); N <sup>ε2</sup> , * (7.664, 7.074)   |
| K16     | 119,345(8,166) | 181,210 | 55,329(4.483)  | 31,779(1.947, 1.947) | C <sup>γ</sup> , * (1.585); C <sup>δ</sup> , * (1.838); C <sup>ε</sup> , * (3.145)  |
| W17     | 126,102(8,715) | 182,189 | 56,410(4.629)  | 27,873(3.161, 3.058) | C <sup>δ1</sup> , * (7.570); C <sup>n2</sup> , * (6.687); C <sup>ε2</sup> , * (7.228); C <sup>ε3</sup> , * (7.123);<br>N <sup>ε1</sup> , * (10.262)       |
| T18     | 116,982(8,760) | 181,210 | 58,823(4.795)  | 70,226(4.898)        | C <sup>γ2</sup> , * (1.446)   |
| V19     | 125,958(8,841) | 182,820 | 65,189(3.883)  | 30,373(2.149)        | C <sup>γ1</sup> , * (1.223); C <sup>γ2</sup> , * (1.073)  |
| E20     | 123,049(8,485) | 182,005 | 58,313(3.761)  | 27,638(1.798, 1.548) | C <sup>γ</sup> , * (1.681)  |
| E21     | 117,046(7,429) | 184,385 | 57,382(3.837)  | 30,842(2.674, 2.674) | C <sup>γ</sup> , * (2.674)  |
| S22     | 117,043(8,105) |         | 62,021(4.450)  | 999 (4.354, 4.181)   |   |
| E23     | 123,982(8,212) | 184,753 | 56,153(4.189)  | 31,623(2.249, 2.144) | C <sup>γ</sup> , * (2.471)  |
| W24     | 123,984(7,739) | 182,901 | 57,092(4.364)  | 27,873(3.592, 2.455) | C <sup>δ1</sup> , * (7.175); C <sup>n2</sup> , * (7.150); C <sup>ε2</sup> , * (7.459); C <sup>ε3</sup> , * (6.607);<br>N <sup>ε1</sup> , * (10.075)       |
| V25     | 120,185(8,095) | 181,500 | 66,948(3.685)  | 30,373(2.684)        | C <sup>γ1</sup> , * (1.522, 1.522, 1.522); C <sup>γ2</sup> , * (1.249, 1.249, 1.249)  |
| K26     | 119,273(7,702) | 183,510 | 58,817(3.978)  | 31,936(2.191, 2.082) | C <sup>γ</sup> , * (1.681); C <sup>δ</sup> , * (1.951); C <sup>ε</sup> , * (3.144)  |
| A27     | 123,869(8,605) | 185,557 | 53,501(4.240)  | 16,621(1.717)        |   |
| G28     | 111,068(8,923) | 180,157 | 46,469(3.974)  |                      |   |
| V29     | 126,491(8,512) | 183,894 | 65,544(3.269)  | 29,592(1.436)        | C <sup>γ1</sup> , * (0.453, 0.453, 0.453); C <sup>γ2</sup> , * (0.082, 0.082, 0.082)  |
| Q30     | 120,896(7,585) | 181,721 | 57,576(4.052)  | 26,935(2.335, 2.271) | C <sup>γ</sup> , * (2.525, *); N <sup>ε2</sup> , * (7.594, 7.032)   |
| K31     | 119,88(8,016)  | 182,295 | 45,074(4.104)  | 32,561(1.782, 1.707) | C <sup>γ</sup> , * (0.208, *); C <sup>δ</sup> , * (1.564, *); C <sup>ε</sup> , * (2.876, *)   |
| Y32     | 116,681(9,056) | 181,402 | 58,045(4.574)  | 37,718(3.255, 2.886) | C <sup>δ1</sup> , * (7.471); C <sup>δ2</sup> , * (7.471); C <sup>ε1</sup> , * (6.953); C <sup>ε2</sup> , * (6.953)  |
| G33     | 112,263(7,663) | 177,628 | 43,063(3.886)  |                      |   |
| E34     | 122,428(8,407) | 178,954 | 56,640(3.229)  | 28,254(2.214, 1.980) | C <sup>γ</sup> , * (2.393, *)   |
| G35     | 121,003(5,386) | 178,299 | 43,621(3.582)  |                      |   |
| N36     | 123,497(7,188) | 179,793 | 49,967(5.111)  | 35,130(2.905, 2.473) | N <sup>δ2</sup> , 114,931 (7.969, 7.090)  |
| W37     | 120,301(6,734) | 183,100 | 56,939(4.440)  | 28,654(3.757, 3.243) | C <sup>δ1</sup> , * (7.608); C <sup>n2</sup> , * (7.269); C <sup>ε2</sup> , * (7.929); C <sup>ε3</sup> , * (6.673);<br>N <sup>ε1</sup> , 130,859 (10.929) |
| A38     | 122,538(7,345) | 185,055 | 53,735(4.261)  | 16,465(1.606)        |   |
| A39     | 122,651(7,510) | 185,055 | 53,579(3.980)  | 16,465(1.003)        |   |
| I40     | 119,957(7,469) | 181,999 | 65,541(3.758)  | 37,561(2.183)        | C <sup>γ1</sup> , * (1.001, *); C <sup>γ2</sup> , * (1.478, *, *); C <sup>δ1</sup> , * (0.792, *, *)  |
| S41     | 113,317(8,192) | 179,581 | 60,048(4.365)  | 61,936(4.092, 4.043) |   |
| K42     | 117,579(7,192) | 182,550 | 55,523(4.655)  | 32,717(2.004, 1.888) | C <sup>γ</sup> , * (1.642, *); C <sup>δ</sup> , * (1.761, *); C <sup>ε</sup> , * (3.077, *)   |

|     |                |         |               |                      |   |
|-----|----------------|---------|---------------|----------------------|---|
| N43 | 117,537(7,515) | 176,951 | 48,856(4.734) | 39,593(2.754, 2.485) | $N^{\delta 2}, *$ (7.321, 7.106)  |
| Y44 | 118,244(7,327) | 183,395 | 54,282(4.364) | 37,569(2.397, 1.744) | $C^{\delta 1}, *$ (7.385); $C^{\delta 2}, *$ (7.385); $C^{e1}, *$ (6.900); $C^{e2}, *$ (6.900)  |
| P45 |                | 180,505 | 61,360(4.649) | 26,154(2.155, 1.571) | $C^{\gamma}, *$ (1.242, *)  |
| F46 | 122,834(8,206) | 181,825 | 58,547(4.912) | 39,920(3.273, 3.134) | $C^{\delta 1}, *$ (7.918); $C^{\delta 2}, *$ (7.918); $C^{e1}, *$ (8.072); $C^{e2}, *$ (8.073);<br>$C^{\zeta}, *$ (7.282)               |
| V47 | 126,698(9,814) | 179,901 | 60,852(4.288) | 31,016(2.204)        | $C^{\gamma 1}, *$ (1.158, 1.158, 1.158); $C^{\gamma 2}, *$ (1.046, 1.046, 1.046)  |
| N48 | 123,851(8,920) | 179,109 | 52,900(4.444) | 36,155(3.124, 2.939) | $N^{\delta 2}, *$ (7.742, 7.031)  |
| R49 | 115,416(8,055) | 180,204 | 51,272(5.275) | 30,373(1.795, 1.438) | $C^{\gamma}, *$ (1.910, *); $C^{\delta}, *$ (3.130, *)  |
| T50 | 114,226(7,433) | 179,650 | 58,199(4.689) | 69,939(4.900)        | $C^{\gamma 2}, *$ (1.431)   |
| A51 | 124,287(8,923) | 183,700 | 54,898(4.045) | 16,699(1.620)        |   |
| V52 | 116,642(7,744) | 181,721 | 64,464(3.684) | 30,373(2.092)        | $C^{\gamma 1}, *$ (1.157, 1.157, 1.157); $C^{\gamma 2}, *$ (1.046, 1.046, 1.046)  |
| M53 | 120,793(7,424) | 184,103 | 57,720(4.424) | 31,936(2.495, 2.456) | $C^{\gamma}, *$ (2.987, *)  |
| I54 | 122,782(7,861) | 180,388 | 63,971(3.580) | 36,467(2.110)        | $C^{\gamma 1}, *$ (0.879, *); $C^{\gamma 2}, *$ (1.251, *, *); $C^{\delta 1}, *$ (0.327, *, *)  |
| K55 | 123,386(7,429) | 183,100 | 57,527(1.912) | 29,592(1.477, 0.069) | $C^{\gamma}, *$ (0.947, *); $C^{\delta}, *$ (1.583, *); $C^{\epsilon}, *$ (3.044, *)  |
| D56 | 119,836(7,933) | 183,510 | 55,636(4.355) | 38,968(2.762, 2.683) |   |
| R57 | 123,876(8,020) | 182,626 | 55,491(3.922) | 27,873(1.004, 0.941) | $C^{\gamma}, *$ (-0.156, *); $C^{\delta}, *$ (2.905, *)   |
| W58 | 122,838(8,438) | 184,156 | 999(4.752)    | 999(3.320, 3.224)    | $C^{\delta 1}, *$ (7.561); $C^{\eta 2}, *$ (7.432); $C^{\zeta 2}, *$ (7.692); $C^{\zeta 3}, *$ (7.024);<br>$N^{\epsilon 1}, *$ (10.915) |
| R59 | 118,538(7,782) | 181,420 | 58,658(3.764) | 28,185(2.078)        | $C^{\gamma}, *$ (1.714, *)  |
| T60 | 118,538(7,782) | 179,003 | 65,378(3.912) | 67,409(4.302)        | $C^{\gamma 2}, *$ (1.247, *, *)   |
| M61 | 122,838(8,438) | 182,901 | 58,560(3.806) | 30,373(2.049, 1.987) | $C^{\gamma}, *$ (2.495, *)  |
| K62 | 120,171(8,009) | 183,811 | 57,420(3.740) | 30,998(1.497, 1.335) | $C^{\gamma}, *$ (0.864, *); $C^{\delta}, *$ (1.602, *); $C^{\epsilon}, *$ (2.869, *)  |
| R63 | 121,865(7,518) | 182,550 | 57,230(4.180) | 28,654(2.063, 1.986) | $C^{\gamma}, *$ (1.728, *); $C^{\delta}, *$ (3.267, *); $N^{\epsilon}, *$ (7.375)   |
| L64 | 119,348(7,822) | 182,295 | 53,768(4.404) | 40,976(1.840, 1.787) | $C^{\gamma}, *$ (1.728); $C^{\delta 1}, *$ (0.954, 0.954, 0.954); $C^{\delta 2}, *$ (0.954, 0.954, 0.954)                               |
| G65 | 110,174(7,778) | 179,432 | 44,756(4.003) |                      |   |
| M66 | 120,434(8,211) | 179,101 | 54,219(4.540) | 32,404(2.336, 1.853) | $C^{\gamma}, *$ (2.562, *)  |
| N67 | 124,714(8,113) | 180,880 | 53,579(4.494) | 38,655(2.846, 2.400) | $N^{\delta 2}, *$ 114,873 (7.551, 6.877)  |

## Domaine Myb de TRF2 complexé à l'ADN télomérique

Table 2 resonance assignments (proton chemical shifts are given in parentheses).

| residue | N               | C <sup>α</sup>        | C <sup>β</sup>        | other  |
|---------|-----------------|-----------------------|-----------------------|--|
| G1      |                 |                       |                       |  |
| S2      |                 |                       |                       |  |
| H3      |                 |                       |                       |  |
| M4      |                 | * (4.428)             |                       |  |
| E5      |                 |                       |                       |  |
| D6      |                 | 53,045 (4.270)        | * (1.944, 2.063)      |  |
| S7      |                 | 57,224(4.512)         | 62,209 (3.867, 3.964) |  |
| T8      |                 | 61,212(4.409)         | * (*)                 |  |
| T9      | * (8.122)       | 60.651 (4.415)        | 41,201 (4.300)        | C <sup>γ2</sup> , 20,581 (*)   |
| N10     | 117,998(8.134)  | 51.921 (4.785)        | 37.416 (2.836, 2.756) | N <sup>δ2</sup> , 114.671 (7.626, 6.938)   |
| I11     | 123.348 (8.136) | 59.594 (4.295)        | 37.567 (1.880)        | C <sup>γ1</sup> , 22,000 (1.343, 0.919); C <sup>γ2</sup> , 16,406 (0,934); C <sup>δ1</sup> , 11,990 (0,887)  |
| T12     | 123.619 (8.473) | 60.513 (4.376)        | 68.131 (4.206)        | C <sup>γ2</sup> , 20,968 (1,283)   |
| K13     | 126.786 (8.575) | 54.855 (4.284)        | 31,632 (1.888, 1.783) | C <sup>γ</sup> , 23,948 (1,439); C <sup>δ</sup> , 28,212 (*); C <sup>ε</sup> , 41,052 (*)  |
| K14     | 128.143 (8.415) | 52.948 (4.073)        | 29,435 (2.089, 2.031) | C <sup>γ</sup> , 23,800 (1,517); C <sup>δ</sup> , 27,985 (*); C <sup>ε</sup> , 40,998 (*)  |
| Q15     | 123.076 (8.395) | 55.438 (4.655)        | 27.838 (*)            | C <sup>γ</sup> , 31.861 (2.754, 2.703); N <sup>ε2</sup> , 114.244 (7.578, 6.866)   |
| K16     | 122.557 (8.403) | 56.383 (4,362)        | 30.945 (1,830)        | C <sup>γ</sup> , 23,845 (1,472, 1,491); C <sup>δ</sup> , 28,320 (2,421); C <sup>ε</sup> , 41,000 (*)   |
| W17     | 127.600 (9.229) | 56.588 (4.501)        | 27,861 (3.134, 3.028) | C <sup>δ1</sup> , * (7.570); C <sup>η2</sup> , * (6.687); C <sup>ζ2</sup> , * (7.228); C <sup>ζ3</sup> , * (7.123); N <sup>ε1</sup> , 131,589 (10.268) |
| T18     | 118.331 (9.132) | 59.005 (4.805)        | 70,215 (?)            | C <sup>γ2</sup> , 20,954 (1,328)   |
| V19     | 126,386 (8.389) | 65.345 (3.783)        | 30,145 (2.034)        | C <sup>γ1</sup> , 21,005 (0,965); C <sup>γ2</sup> , 19,647 (1,118)   |
| E20     | 123.800 (8.388) | 58.245 (3.643)        | 27,548 (1.711, 1.534) | C <sup>γ</sup> , 35,105 (1.889, 1.643)   |
| E21     | 117.010 (7.551) | 57.736 (3,710)        | 30,823 (2,559)        | C <sup>γ</sup> , 35,254 (2,461)  |
| S22     | 116.548 (8.224) | 62.018 (4,238)        | 65.202 (*)            |  |
| E23     | 124.162 (8.382) | 57.839 (4.098)        | 27.116 (*)            | C <sup>γ</sup> , 35,120 (1.748, 2.351)   |
| W24     | 126.424 (7.732) | 57.257 (4.323)        | 27,273 (2,351, 3,497) | C <sup>δ1</sup> , 126,05 (7.689); N <sup>ε1</sup> , 129,105 (9.930)  |
| V25     | 120.112 (8.247) | 67.108 (3.618)        | 30,313 (2.585)        | C <sup>γ1</sup> , * (1,137); C <sup>γ2</sup> , 19,125 (1,127)  |
| K26     | 119.373 (8.038) | 58.958 (3.934)        | 31.801 (1.982, 2.066) | C <sup>γ</sup> , 24.126 (1.859, 1.500); C <sup>δ</sup> , 27,992 (1,841); C <sup>ε</sup> , 40,989 (3.036, 3.040)  |
| A27     | 124.071 (8.712) | 53.512 (4.143)        | 16.567 (1,621)        |  |
| G28     | 111.304 (8.955) | 46.580 (3,868,4,519)  |                       |  |
| V29     | 126.967 (8.666) | 65.728 (3.070)        | 29,159 (1.517)        | C <sup>γ1</sup> , 21,503 (-0,038); C <sup>γ2</sup> , 19 ,504 (0,435)   |
| Q30     | 120.509 (7.681) | 57.568 (3.959)        | 26.693 (2.172, 2.206) | C <sup>γ</sup> , 32.455 (2.575, 2.436); N <sup>ε2</sup> , 113.550 (7.511, 6.935)   |
| K31     | 120.124 (8.113) | 57.596 (3.988)        | 32.359 (1.575, 1.607) | C <sup>γ</sup> , 23,890 (0.913, 0.096); C <sup>δ</sup> , 28,175 (1.447, 1.452); C <sup>ε</sup> , 41,204 (2.738, 2.735)                                 |
| Y32     | 117.037 (9.158) | 58.034 (4.474)        | 37,708 (2,794,3,132)  |  |
| G33     | 112.484 (7.592) | 42.912 (3,795, 4,291) |                       |  |
| E34     | 124.071 (8.451) | 57.121 (4,749)        | 28,054 (*)            | C <sup>γ</sup> , 35,987 (2.298, 2.417)   |
| G35     | 122.629 (5.436) | 43.916 (4.177, 3.311) |                       |  |
| N36     | 122.360 (6.694) | 49.243 (4.987)        | 34.655 (2.763, 2.311) | N <sup>δ2</sup> , 114.781 (7.630, 6.871)   |

|     |                     |                       |                       |   |
|-----|---------------------|-----------------------|-----------------------|---|
| W37 | 122.736 (6.881)     | 56.719 (4.330)        | 28,654 (3.653, 3.132) | $C^{\delta 1}$ , * (7.608); $C^{\eta 2}$ , * (7.269); $C^{\zeta 2}$ , * (7.929); $C^{\zeta 3}$ , * (6.673); $N^{\epsilon 1}$ , 130,859 (10.790) |
| A38 | 122.538 (8.757)     | 54.101 (4.148)        | 16.424 (1,485)        |   |
| A39 | 121.881 (7.586)     | 53.658 (3.858)        | 16,565 (0,880)        |   |
| I40 | 121.210 (7.578)     | 65.760 (3.639)        | 37,561 (2,090)        | $C^{\gamma 1}$ , 21.810 (1.417, 0.889); $C^{\gamma 2}$ , 16,350 (1,371); $C^{\delta 1}$ , 12,560 (0,790)  |
| S41 | 114.068 (8.778)     | 60.035 (4.233)        | 61.541 (3,927, 3,983) |   |
| K42 | 118.131 (7.162)     | 55.630 (4.512)        | 32,717 (1.882, 1.775) | $C^{\gamma}$ , 23,649 (1.677, 1.646); $C^{\delta}$ , 30,501 (*); $C^{\epsilon}$ , 40,458 (*)  |
| N43 | 117.725 (7.555)     | 53.400 (4.681)        | 39.987 (2.621, 2.342) | $N^{\delta 2}$ , * (7,385, *)   |
| Y44 | 118.146 (7.371)     | 54.426 (4.276)        | 37,569 (1,648, 2,299) |   |
| P45 |                     | 61.327 (*)            | 32,515 (1,475)        | $C^{\gamma}$ , 26,120 (1,286,1,126); $C^{\delta}$ , ? (2,985)   |
| F46 | 122.359 (8.085)     | 58.469 (*)            | 40,999 (3.170, 3.000) | $C^{\delta 1}$ , 130,125 (7.918); $C^{\epsilon 1}$ , 131,122 (8.072); $C^{\epsilon 2}$ , * (8.073); $C^{\zeta}$ , 131,312 (7.282)               |
| V47 | 127.238<br>(10.050) | 61.108 (4.141)        | 35.357 (2.109)        | $C^{\gamma 1}$ , (0,962); $C^{\gamma 2}$ , 19,888(0,944)  |
| N48 | 123.709 (8.967)     | 52.792 (4.366)        | 35.921 (3.026, 2.877) | $N^{\delta 2}$ , 115.217 (7.667, 6.896)   |
| R49 | 115.298 (8.013)     | 51.388 (5.172)        | 30,373(1,504, 1,662)  | $C^{\gamma}$ , 28,952 (2,221); $C^{\delta}$ , 41,358 (3.024, 3.025)   |
| T50 | 114.364 (7.467)     | 58.173 (4,591)        | 69,939 (*)            | $C^{\gamma 2}$ , 20,125 (*)   |
| A51 | 123.438 (9.009)     | 54.442 (3.967)        | 17,005 (0,528)        |   |
| V52 | 119.312 (7.488)     | 64.470 (3.487)        | 30.671 (1.773)        | $C^{\gamma 1}$ , 20,004(1,075); $C^{\gamma 2}$ , 19,106(0,799)  |
| M53 | 120.414 (7.424)     | 57.696 (4.262)        | 31,936 (2,354)        | $C^{\gamma}$ , 33,000 (2,377, 2,884); $C^{\epsilon}$ , 16,300 (1,492)   |
| I54 | 121.807 (7.850)     | 64.308 (3.464)        | 35,368 (1.799)        | $C^{\gamma 1}$ , * (1.546, 0.787); $C^{\gamma 2}$ , 21,564 (1,138); $C^{\delta 1}$ , 13,002 (0,216)   |
| K55 | 124.162 (7.122)     | 57.560 (1,877)        | * (-0,044,1,342)      | $C^{\gamma}$ , 24,005 (0,886, 1,387); $C^{\delta}$ , 31,502 (1,483, 1,804); $C^{\epsilon}$ , 41,256 (2,965, 3,040)                              |
| D56 | 120.911 (8.229)     | 56.406 (4.251)        | 39,988 (0,886, 2,637) |   |
| R57 | 125.429 (8.254)     | 54.663 (*)            | 27,873 (0,854, 0,885) | $C^{\gamma}$ , 28,619 (1,361, 1,491); $C^{\delta}$ , 41,402 (2,780)   |
| W58 | 122.675 (8.438)     | 58.099 (4.638)        | * (3.146, 3.094)      | $C^{\delta 1}$ , * (7.661); $C^{\eta 2}$ , * (7.387); $C^{\zeta 2}$ , * (7.833); $C^{\zeta 3}$ , * (*); $N^{\epsilon 1}$ , 133,002 (11.057)     |
| R59 | 118.920 (7.692)     | 58.922 (3.774)        | 28,185 (1,961)        | $C^{\gamma}$ , 26,562 (1,593, 1,947); $C^{\delta}$ , 42,312 (3,242)   |
| T60 | 120.151 (7.875)     | 65.508 (3.809)        | 67,409 (4,200)        | $C^{\gamma 2}$ , 22,001 (1,127)   |
| M61 | 121.616 (8.793)     | 58.879 (3,677)        | * (1,958, 1,847)      | $C^{\gamma}$ , 31,212 (2,379); $C^{\epsilon}$ , 16,406 (2,346)  |
| K62 | 120.736 (8.338)     | 58.319 (3.644)        | 30.237 (1,408)        | $C^{\gamma}$ , 23,865 (0.750, 0.708); $C^{\delta}$ , 28,002 (1.595, 1.560); $C^{\epsilon}$ , 40,982 (2.931, 2.776)                              |
| R63 | 123.076 (7.658)     | 57.938 (3.993)        | 28.500 (1,621, 1,700) | $C^{\gamma}$ , 28,924 (1.608, 1.620); $C^{\delta}$ , 42,312 (3.243, 3.178)  |
| L64 | 118.370 (7.808)     | 53.597 (4.294)        | 40.994 (1,648, 1,684) | $C^{\gamma}$ , 25,958 (1.661); $C^{\delta 1}$ , 24,458 (0,836); $C^{\delta 2}$ , 24,458 (0,865)   |
| G65 | 110.302 (7.754)     | 45.048 (4.044, 3.929) |                       |   |
| M66 | 119.945 (8.375)     | 54.339 (4.435)        | 32.426 (1,569, 1,726) | $C^{\gamma}$ , 32,585 (2.513, 2.422); $C^{\epsilon}$ , 16,400 (2,500)   |
| N67 | 123.800 (8.301)     | 53.465 (4,382)        | 37.810 (2.767, 2.291) | $N^{\delta 2}$ , 115.362 (7.365, 6.761)   |



# ANNEXE 3 : L'ARTICLE



Available online at [www.elsevier.com/locate/CRCI](http://www.elsevier.com/locate/CRCI)



C. R. Chimie 9 (2006) 452–458



<http://www.elsevier.com/locate/CRCI>

## NMR studies of telomeric nucleoprotein complexes involving the Myb-like domain of the human telomeric protein TRF2

Yann Bilbille, Françoise Paquet \*, Hervé Meudal,  
Marie-Josèphe Giraud-Panis, Gérard Lancelot

Centre de Biophysique Moléculaire, UPR du CNRS 4301, rue Charles-Sadron, 45071 Orléans cedex 2, France

Received 21 March 2005; accepted 6 June 2005

Available online 25 August 2005

### Abstract

In order to study the binding of the Myb-like domain of the human telomeric protein TRF2 (Myb-TRF2) with different structural components of the t-loop model, we report NMR studies of the binding of Myb-TRF2 protein with two repeats human telomeric DNA under three conformations. Our results showed that Myb-TRF2 binds to the duplex and even to the quadruplex and the random coil G-rich strand. The solution structure of Myb-TRF2 reported here looks like Myb-TRF1 suggesting similar DNA binding mode. As a matter of fact, we have shown that its binding to the human telomeric duplex presents great similarities with this of Myb-TRF1. To cite this article: Y. Bilbille et al., C. R. Chimie 9 (2006).

© 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.

### Résumé

Afin d'étudier les interactions du domaine de fixation Myb de la protéine télomérique TRF2 (Myb-TRF2) avec les différentes composantes structurales du modèle de la t-loop, nous présentons ici la structure en solution de la protéine et une étude RMN de ses interactions avec une séquence à deux répétitions de l'ADN télomérique sous trois conformations différentes. Nous montrons que cette protéine interagit avec le duplexe mais également avec le quadruplexe et la conformation aléatoire du brin riche en guanines. Une étude détaillée du duplexe et de son complexe est présentée. La grande similarité de structure de Myb-TRF1 et Myb-TRF2 suggère fortement un mode de liaison à l'ADN très comparable. Pour cet article : Y. Bilbille et al., C. R. Chimie 9 (2006).

© 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.

**Keywords:** NMR structure; Telomere; DNA; Quadruplex; Myb-TRF2 DNA binding protein

**Mots clés:** Structure par RMN; Télomère; ADN; Quadruplexe; Myb-TRF2

\* Corresponding author.

E-mail address: [paquet@cns-orleans.fr](mailto:paquet@cns-orleans.fr) (F. Paquet).

## 1. Introduction

Telomeres are special nucleoprotein complexes at the ends of eukaryotic chromosomes that can switch stochastically between two states: capped and uncapped [1]. Capping is functionally defined as preserving the physical integrity of the telomere. Griffith and de Lange have shown that the 3' single stranded *TTAGGG* overhang is tucked back inside the double stranded *TTAGGG/CCCTAA* telomere DNA, resulting in a large t-loop that is stabilized by an intratelomeric D-loop [2]. A complex of proteins including TRF1 and TRF2 is involved in the stabilization of the telomeric structure [3]. This t-loop/D-loop model could explain how TRF2 protects DNA from the enzymatic attack and end-to-end fusion [4]. However, the molecular mechanism of such t-loop/D-loop is speculative, especially the processes that involve TRF2 [5]. The structures of the TRF proteins themselves and in complex with double stranded DNA have been partially solved by X-ray crystallography [6,7] and NMR [8–10].

In this report, we describe the structure of the Myb-TRF2 and its binding with G-rich strand in random coil conformation or/and in quadruplexes. In order to compare the affinity of Myb-TRF2 with the different components of the t-loop, we also present two repeats telomeric DNA duplex structure itself and in complex with Myb-TRF2.

## 2. Material and methods

The oligonucleotides d[G(TTAGGG)<sub>2</sub>T] and d[A(CCCCTAA)<sub>2</sub>C] were purchased from Eurogentec and were passed through Chelex 100 in order to remove the paramagnetic ions.

The proteins were prepared and purified in our laboratory (M.-J. Giraud-Panis, unpublished results).

### 2.1. NMR experiments

The free oligonucleotides were lyophilized and dissolved in D<sub>2</sub>O or in H<sub>2</sub>O/D<sub>2</sub>O 9:1 (v/v) argon degassed solvent. The protein Myb + 7-TRF2, containing seven N-terminal residues that do not belong to the Myb-like domain, and the protein Myb-TRF2 were used for NMR experiments. These proteins were dissolved in a 50 mM NaCl, 2 mM sodium phosphate buffer (pH 5.3) to which

NaN<sub>3</sub> was added in order to prevent microbial growth. The NMR samples were degassed then sealed under argon. The DNA duplex titration was made by adding amounts of protein (2.3 mM) dissolved in a 50 mM NaCl, 2 mM sodium phosphate buffer (pH 6.0) to the oligonucleotide duplex (1.25 mM in double strand) dissolved in a 10 mM NaCl buffer (pH 6.0).

NMR experiments were carried out on a VARIAN UNITY-INOVA spectrometer operating at 14.08 Tesla and equipped with a z-axis pulsed field gradient. DQF-COSY, TOCSY with MLEV17 and isotropic period of 40 and 70 ms, and NOESY spectra with mixing time of 50, 100, 120 and 150 ms were recorded at 30 °C. Chemical shifts were referenced to DSS. Sequence-specific <sup>1</sup>H NMR assignments were obtained following the conventional strategy [11,12]. Spin systems of the individual amino acids were identified in a D<sub>2</sub>O solution at 30 °C using DQF-COSY, TOCSY and NOESY spectra. Corresponding spectra in 90% H<sub>2</sub>O, 10% D<sub>2</sub>O solution were used to assign the exchangeable protons.

### 2.2. Structure calculations

The cross-peak intensities on the NOESY spectrum recorded with a 100 ms mixing time were integrated and partially assigned within NMRView [13]. The Myb + 7-TRF2 related NOEs were used as input to ARIA 1.1 [14] implemented in CNS 1.1 [15]. The calculations were performed starting from 150 random templates, using the standard parameters of ARIA (Ambiguous Restraints for Iterative Assignment). ARIA is a method combining an iterative NOE interpretation scheme with a dynamical assignment of ambiguous NOE cross-peaks treated as the sum of contributions from all possible assignments. A large number of NOEs were calibrated and assigned automatically during the structure calculation by ARIA. The procedure of assignment/refinement was repeated iteratively. At each step, the new assignments proposed by ARIA were carefully checked manually and introduced (or not) into the subsequent run. Rejected restraints and residual NOE violations were analyzed, and the assignments were corrected if needed. In the final iteration, the 100 structures with the lowest energy were further refined by molecular dynamics calculations in explicit solvent to remove artefacts [16]. Structures were visualized and aligned with MOLMOL 2.6 [17] and analyzed with



the high definition of the structure from residue 18–67 and the mobility found in the N-terminal arm. The folding of the protein can explain some unusual shifts observed for several protons in the NOESY spectra. The amide proton of Val47 appears to be in the plane of the cycle of Trp24, in agreement with the observed downfield shift of its resonance at 9.85 ppm [ $H\zeta_3(\text{Trp24})-\text{NH}(\text{Val47}) = 3.0 \text{ \AA}$ ]. The amide proton of Gly35 and  $C_\alpha$  proton of Lys55 are located on both sides and near to the cycles of the tryptophan Trp37, in accordance with the strong up field shift observed for their resonance: 5.45 and 1.78 ppm respectively. One  $H_\alpha$  of Gly28 due to its proximity with the Tyr32 and Tyr44 rings is low field shifted at 4.62 ppm. The hydroxylic protons of Thr18 and Thr60 are hydrogen bonded to the carboxylate group of Glu21 and to the carbonyl group of Arg57 respectively, this explains that their resonance was observed and located at 5.66 and 5.12 ppm.

The overall structure of the folded domain is very similar to that of TRF1 with the presence of three  $\alpha$ -helices (residues 19–31, 37–43 and 51–64) linked by two turns (residues 32–36 and 44–50). A hydrophobic core is formed by the residues Trp17, Val25, Val29, Trp37, Ile40, Phe46, Ala51, Ile54, Trp58, and Met61. Most of these residues are conserved between Myb-TRF1 and Myb-TRF2 (Val25 and Ile54 in Myb-TRF2 instead of Leucines in Myb-TRF1) and in both structures they are involved in maintaining the three helices. The first helix resembles that of Myb-TRF1 in length and relative orientation but presents a different distribution of the residues along the surface. In Myb-TRF1, two faces are clearly defined on the structure, one hydrophobic and one charged. In Myb-TRF2 residues of a different nature are mixed along the helix.

The first turn, classified as a  $\beta$ -turn IV, is stabilized by two hydrogen bonds between NH(Asn36) and CO(Gly35), and between NH(Trp37) and CO(Gly35). Helices 2 and 3 form the variant of the helix–turn–helix motif responsible for most of the interactions between Myb-TRF1 and DNA [8]. Although being very similar to Myb-TRF1, helix 2 has a slightly different position in Myb-TRF2 due to the presence of Pro45, which induces the formation of a  $\gamma$ -turn (44–46). This second turn is stabilized by the interaction of Phe46 ring (a conserved residue) with the side chain of a hydrophobic residue in each helix: Val25 (a leucine in Myb-TRF1), Ile40 (conserved) and Ile54 (another leucine in Myb-TRF1). Two hydrogen bonds between HN(Phe46)

and CO(Tyr44), and between NeH(Arg49) and CO(Val47) participate also to the stability of this structural element. Following the  $\gamma$ -turn (44–46) that is missing in Myb-TRF1, the end of the turn (46–49) recovers a similar position between both domains as a classical  $\beta$ -turn IV. Helix 3 is undistinguishable from that of Myb-TRF1 owing to the remarkable homology between both sequences (11 residues are conserved in the 14 forming the helix). Could the differences that we observed between Myb-TRF2 and Myb-TRF1, be responsible for differences in DNA binding behavior? The most likely answer to that question is no. Indeed, the overall shape of the molecule is uncannily similar and the small differences observed occur in regions of the molecule that are not involved in DNA interactions in Myb-TRF1. Finally the vast majority of the residues that are implicated in DNA binding are conserved between the two proteins and in both cases are exposed to the solvent (Trp17, Arg49, Met53, Lys55, Asp56, Arg57, Arg59, Thr60 and Arg63).

### 3.3. Single stranded telomeric human DNA structure

It is known that the C-rich human telomeric strand  $d[(\text{CCCTAA})_3\text{CCC}]$  adopts an *i*-motif structure at acidic pH [20] and that  $d[\text{AGGG}(\text{TTAGGG})_3]$  folds on itself as a G tetrad [21].

At DNA concentrations  $\leq 0.2 \text{ mM}$  and in absence of salt, the  $^1\text{H NMR}$  spectra of  $d[\text{G}(\text{T TAGGG})_2\text{T}]$  were almost independent of the temperature between 5 and 70 °C, demonstrating a random coil conformation in agreement with the absence of imino protons resonances. Adding KCl led to the appearance of several sharp imino resonance in the range of 10–11.5 ppm. These resonances were assigned to quadruplexe structures since it has been shown that the four-repeats G-rich strand folded in a single strand with quartets of guanine [21] and that the two repeats human telomeric DNA showed parallel or antiparallel quadruplexes in presence of salt and after substitution of thymines in uracil or bromouracil [22].

The NMR spectra of  $10^{-4}$ – $10^{-3} \text{ M}$   $d[\text{A}(\text{CCCTAA})_2\text{C}]$  showed sharp proton resonance without resonance in the range of 10–16 ppm, even at pH 5.0 demonstrating a random coil single stranded conformation.

### 3.4. Titration of the two repeats G-rich telomeric strand with the protein Myb-TRF2

Adding successive amounts of Myb-TRF2 (0.0–0.2 mM) to d[G(TTAGGG)<sub>2</sub>T] (0.2 mM) in random coil conformation led to a broadening of the resonance lines and slight shifts of the oligonucleotide resonances. It is worth noting that inter-residue NOE connectivities (H6/8–H1', –H2'', –H6/8, –H5, and –CH3) were well distinguishable on the 150 ms NOESY spectrum at 25 °C, whereas they were absent with the single strand. This corresponds to the fact that the random coil conformation exhibited large internal dynamics, which led to an effective correlation time so that  $\omega \tau \approx 1.1$  and reduced dramatically the NOE efficiency. The presence of well-definite NOE connectivities corresponding to the increase of the effective correlation time strongly suggested the binding of Myb-TRF2 to the single-stranded d[G(TTAGGG)<sub>2</sub>T] in random coil conformation.

Adding Myb-TRF2 (0.2 mM) to the quadruplex form (0.2 mM) in presence of KCl (50 mM) led to several changes in the 1-D proton spectrum and to the appearance of several intermolecular NOE connectivities between the quadruplex and the protein. It is well known that unspecific binding of protein to DNA leads to the formation of several types of complexes, each of them owing different contact points with low populations, generating undetectable NOEs. On the contrary, specific interaction only produces one type of complex where NOEs observation is expected. Such situations were observed in NOESY spectra of specific and unspecific complexes between the Lac Repressor headpiece and DNA for example [23]. Consequently, presence of intermolecular NOEs is indicative of specific binding. Our data strongly suggest a tight binding of Myb-TRF2 with the quadruplex.

### 3.5. Titration of the telomeric DNA duplex by the protein Myb-TRF2

The 1:1 mixing of the strands d[G(TTAGGG)<sub>2</sub>T] and d[A(CCCCTAA)<sub>2</sub>C] in absence of salts, showed NMR spectra with exchangeable and non-exchangeable proton resonances corresponding to a right-handed duplex conformation, even at pH 5.5, despite the possible folding of the separate strands. The N1H imino protons of

the guanines were found in the range of 12–12.5 ppm and the N3H imino protons of the thymines were located in the range of 13–14 ppm. No imino resonances corresponding to quadruplexes species or protonated cytosines were observed. All the exchangeable and non-exchangeable proton resonances were sequentially assigned to a B-type right-handed helix except the 5' and 5'' ones that were found to overlap.

Adding successive amounts of protein to a 0.25 mM solution of DNA at 20 °C led to the increase of new set of bound imino protons resonances and to a decrease of the free-form resonances until the ratio 1/1 between protein and DNA was reached (Fig. 2). These spectra gave evidence of a slow exchange process on the NMR time scale between the free form and the bound form of the DNA, suggesting a tight binding of the protein to the telomeric DNA.

It can be pointed out that the binding of Myb-TRF2 without the seven supplementary residues led to very similar spectra with the same imino proton resonances in slow chemical exchange. We conclude that these seven supplementary amino acids did not belong to the specific protein fixation domain.

### 3.6. Contribution of individual base pairs to the sequence-specific binding

The binding specificity of Myb-TRF2 with the two repeats telomeric DNA was tested by running NOESY spectra at 50, 150 and 300 ms mixing time (Fig. 3). Several intermolecular NOESY connectivities were observed, indicating a specific interaction. Although a complete assignment of these connectivities requires the use of NOESY experiments with half filter on a mixture of labeled and unlabeled molecule, we were able to assign some of them: NH(W58) with H8(A22), H3'(A22), H1'(A22) and H1'(C21); H<sub>β</sub>(W58) with H8(A22); CH<sub>3</sub>(V19) with H5(C19) and H5(C20); CH<sub>3</sub>(T60) with H1'(A22) and H3'(A23).

In order to estimate the contribution of each base pair in the telomeric binding site, the assignment of the imino, aromatic and H1' resonances of the DNA was made for the free and bound DNA. The lack of protein resonances in the range of 12–15 ppm, except the N<sub>ε</sub>H of the tryptophan 37 at 12.27 ppm, and the small number of NOESY protein–protein connectivities between non-exchangeable protein protons in the range of 7–9 ppm allow us to assign without difficulty almost

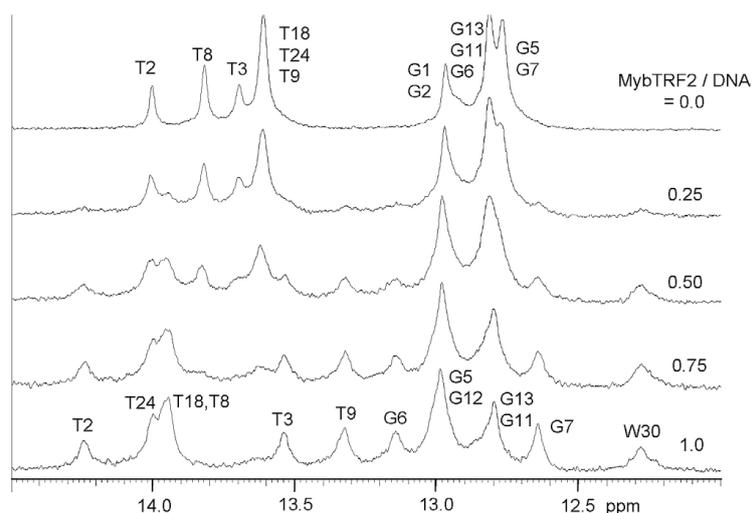


Fig. 2. Imino proton resonance spectra of two repeats duplex in the presence of increasing amounts of Myb-TRF2 showing slow chemical exchange.

all the imino, aromatic and  $H1'$  resonances of the bound DNA. Important chemical shift variations were observed for these protons located in the minor groove and in the major groove of the DNA. The most important variations were observed for the paired nucleotides T2-A16 until A10-T24. The imino proton resonances of the residues T2, G5, G6, T8, T14, T18 and T24 were downfield shifted, indicating a strengthening of their hydrogen bonds. The imino resonance of T9 were upfield shifted as well as the H2 proton of its paired adenine A23, suggesting an increase of the stacking of the base pair T9-A23 with its surrounding T8-A22 and A10-T24 base pairs. It is worth noting that the H2 resonance of the base pair T3-A17 did not exhibit such a trend. These chemical shift variations indicated that the binding domain of Myb-TRF2 overlapped parts of each of the two motifs TTAGGG, in agreement with the formation of a 1–1 complex. The same observation can be made with the  $H1'$  protons located in the minor groove like the H2 protons. Several important shifts were also observed for some aromatic resonances (H5, H6 and H8) corresponding to protons of the residues T3, G6, C19, C20, C21 and A22 located in the major groove of the double helix.

The  $N\epsilon H$  of the tryptophan 37 was downfield shifted by 1.2 ppm upon complexation suggesting the formation of a hydrogen bond or of an electrostatic bond. All these data are in agreement with the crystallographic structure reported for the complex with TRF2 [7].

#### 4. Conclusion

We demonstrated here, that Myb-TRF2 binds to the duplex, to the random coil G-rich strand and to the quadruplex of the human telomeric DNA, contrary to already reported results [24]. The great similarity of the structures of Myb-TRF1 and Myb-TRF2 as well as their NMR features suggests a similar binding mode to the double stranded telomere. It is therefore clear that TRF2 functional singularity cannot originate from the DNA binding domain per se. Nevertheless these data are important to understand the formation of the t-loop since it requires the binding of TRF2 and the coexistence of single strand, double strand and presumably of quadruplex [25].

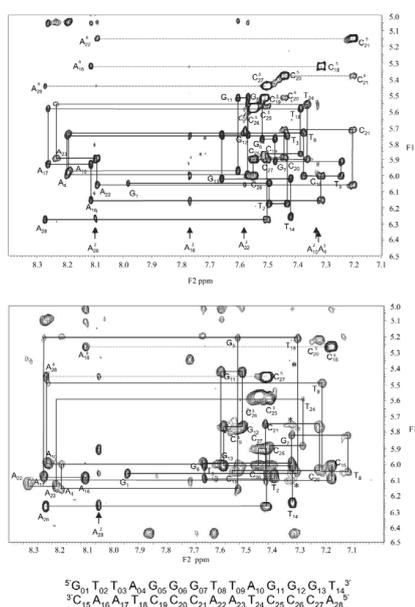


Fig. 3. Expansion of the H6/8-H1' NOE connectivities of the duplex free (top) and complexes to Myb-TRF2 (bottom). Intermolecular cross-peaks corresponding to NH(W58) (7.32 ppm) with H1'(C21), H1'(A22), and H3'(A22) (5.55 ppm) are showed by a star.

## References

- [1] E.H. Blackburn, *Nature* 408 (2000) 53.
- [2] J.D. Griffith, L. Comeau, S. Rosenfield, R.M. Stansel, A. Bianchi, H. Moss, T. de Lange, *Cell* 97 (1999) 503.
- [3] A. Smogorzewska, T. de Lange, *Annu. Rev. Biochem.* 73 (2004) 177.
- [4] T. de Lange, *Oncogene* 21 (2002) 532.
- [5] R.M. Stansel, T. de Lange, J.D. Griffith, *EMBO J.* 20 (2001) 5532.
- [6] P. Konig, L. Fairall, D. Rhodes, *Nucleic Acids Res.* 26 (1998) 1731.
- [7] R. Court, L. Chapman, L. Fairall, D. Rhodes, *EMBO Rep.* 6 (2005) 39.
- [8] T. Nishikawa, H. Okamura, A. Nagadoi, P. Konig, D. Rhodes, Y. Nishimura, *Structure (Camb.)* 9 (2001) 1237.
- [9] T. Nishikawa, A. Nagadoi, S. Yoshimura, S. Aimoto, Y. Nishimura, *Structure* 6 (1998) 1057.
- [10] S. Hanaoka, A. Nagadoi, Y. Nishimura, *Protein Sci.* 14 (2005) 119.
- [11] S.W. Englander, A.J. Wand, *Biochemistry* 26 (1987) 5953.
- [12] K. Wüthrich, *NMR of proteins and nucleic acids*, Wiley-Interscience Publication, New York, 1986.
- [13] B.A. Johnson, R.A. Blevins, *J. Biomol. NMR* 4 (1994) 603.
- [14] J.P. Linge, S.I. O'Donoghue, M. Nilges, *Methods Enzymol.* 339 (2001) 71.
- [15] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. De Lano, P. Gros, R.W. Grosse-Kuntze, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, *Acta Crystallogr. D. Biol. Crystallogr.* 54 (1998) 905.
- [16] J.P. Linge, M.A. Williams, C.A. Spronk, A.M. Bonvin, M. Nilges, *Proteins* 50 (2003) 496.
- [17] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graph.* 14 (1996) 51.
- [18] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, *J. Appl. Crystallogr.* 26 (1993) 283.
- [19] R.A. Laskowski, J.A. Rullmann, M.W. MacArthur, R. Kaptein, J.M. Thornton, *J. Biomol. NMR* 8 (1996) 477.
- [20] J.L. Leroy, M. Gueron, J.L. Mergny, C. Helene, *Nucleic Acids Res.* 22 (1994) 1600.
- [21] Y. Wang, D.J. Patel, *Structure* 1 (1993) 263.
- [22] A.T. Phan, D.J. Patel, *J. Am. Chem. Soc.* 125 (2003) 15021.
- [23] G. Lancelot, F. Paquet, in: G.A. Webb (Ed.), *Annual Reports on NMR Spectroscopy*, Elsevier Science Ltd., 2003, p. 170.
- [24] C.W. Greider, *Cell* 97 (1999) 419.
- [25] S.H. Yoshimura, H. Maruyama, F. Ishikawa, R. Ohki, K. Takeyasu, *Genes Cells* 9 (2004) 205.

## Références bibliographiques

- Abragam, A. (1961). Principles of nuclear magnetism. Oxford, Clarendon.
- Ackerman, M. S. and D. Shortle (2002). "Molecular alignment of denatured states of staphylococcal nuclease with strained polyacrylamide gels and surfactant liquid crystalline phases." Biochemistry **41**(9): 3089-95.
- Ahmed, S., A. Kintanar, et al. (1994). "Human telomeric C-strand tetraplexes." Nat Struct Biol **1**(2): 83-8.
- Amiard, S., M. Doudeau, et al. (2007). "A topological mechanism for TRF2-enhanced strand invasion." Nat Struct Mol Biol **14**(2): 147-54.
- Ancelin, K., M. Brunori, et al. (2002). "Targeting assay to study the cis functions of human telomeric proteins: evidence for inhibition of telomerase by TRF1 and for activation of telomere degradation by TRF2." Mol Cell Biol **22**(10): 3474-87.
- Azzalin, C. M. and J. Lingner (2004). "Cell biology. Telomere wedding ends in divorce." Science **304**(5667): 60-2.
- Barbato, G., M. Ikura, et al. (1992). "Backbone dynamics of calmodulin studied by <sup>15</sup>N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible." Biochemistry **31**(23): 5269-78.
- Baumann, P. and T. R. Cech (2001). "Pot1, the putative telomere end-binding protein in fission yeast and humans." Science **292**(5519): 1171-5.
- Bax, A., G. M. Clore, et al. (1990). "Practical Aspects of Proton-Carbon-Carbon-Proton Three Dimensional Correlation Spectroscopy of <sup>13</sup>C-Labeled Proteins." Journal of Magnetic Resonance **87**: 620-627.
- Bax, A., G. M. Clore, et al. (1990). "<sup>1</sup>H-<sup>1</sup>H Correlation via Isotropic Mixing of <sup>13</sup>C Magnetization, a new Three Dimensional Approach for assigning <sup>1</sup>H and <sup>13</sup>C Spectra of <sup>13</sup>C-Enriched Proteins." Journal of Magnetic Resonance **88**: 425-431.
- Bax, A. and M. Ikura (1991). "An efficient 3D NMR technique for correlating the proton and <sup>15</sup>N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly <sup>15</sup>N/<sup>13</sup>C enriched proteins." J Biomol NMR **1**(1): 99-104.
- Bax, A., G. Kontaxis, et al. (2001). "Dipolar couplings in macromolecular structure determination." Methods Enzymol **339**: 127-74.
- Beattie, T. L., W. Zhou, et al. (2001). "Functional multimerization of the human telomerase reverse transcriptase." Mol Cell Biol **21**(18): 6151-60.
- Bhattacharya, A., R. Tejero, et al. (2007). "Evaluating protein structures determined by structural genomics consortia." Proteins **66**(4): 778-95.
- Bianchi, A. and T. de Lange (1999). "Ku binds telomeric DNA in vitro." J Biol Chem **274**(30): 21223-7.
- Bianchi, A., S. Smith, et al. (1997). "TRF1 is a dimer and bends telomeric DNA." Embo J **16**(7): 1785-94.
- Bianchi, A., R. M. Stansel, et al. (1999). "TRF1 binds a bipartite telomeric site with extreme spatial flexibility." Embo J **18**(20): 5735-44.
- Bilaud, T., C. E. Koering, et al. (1996). "The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human." Nucleic Acids Res **24**(7): 1294-303.
- Blackburn, E. H. (1991). "Structure and function of telomeres." Nature **350**(6319): 569-73.
- Blackburn, E. H. (2001). "Switching and signaling at the telomere." Cell **106**(6): 661-73.
- Blasco, M. A. (2002). "Mouse models to study the role of telomeres in cancer, aging and DNA repair." Eur J Cancer **38**(17): 2222-8.
- Bodnar, A. G., M. Ouellette, et al. (1998). "Extension of life-span by introduction of telomerase into normal human cells." Science **279**(5349): 349-52.

- Borer, P. N., L. Pappalardo, et al. (1997). "NMR-based structure determination for unlabelled RNA and DNA." Adv Biophys Chem **6**: 173-216.
- Broccoli, D., L. Chong, et al. (1997). "Comparison of the human and mouse genes encoding the telomeric protein, TRF1: chromosomal localization, expression and conserved protein domains." Hum Mol Genet **6**(1): 69-76.
- Brünger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system : A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-21.
- Bryan, T. M. and T. R. Cech (1999). "Telomerase and the maintenance of chromosome ends." Curr Opin Cell Biol **11**(3): 318-24.
- Cavanagh, J., W. J. Fairbrother, et al. (1996). Protein NMR spectroscopy - Principles and Practice.
- Cech, T. R. (2000). "Life at the End of the Chromosome: Telomeres and Telomerase." Angew Chem Int Ed Engl **39**(1): 34-43.
- Chai, W., L. P. Ford, et al. (2002). "Human Ku70/80 associates physically with telomerase through interaction with hTERT." J Biol Chem **277**(49): 47242-7.
- Chang, W., J. N. Dynek, et al. (2003). "TRF1 is degraded by ubiquitin-mediated proteolysis after release from telomeres." Genes Dev **17**(11): 1328-33.
- Chong, L., B. van Steensel, et al. (1995). "A human telomeric protein." Science **270**(5242): 1663-7.
- Chou, J. J., S. Gaemers, et al. (2001). "A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles." J Biomol NMR **21**(4): 377-82.
- Cimino-Reale, G., E. Pascale, et al. (2001). "The length of telomeric G-rich strand 3'-overhang measured by oligonucleotide ligation assay." Nucleic Acids Res **29**(7): E35.
- Clore, G. M., A. Szabo, et al. (1990). "Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins." J Am Chem Soc **112**: 4989-4991.
- Clubb, R., T., V. Thanabal, et al. (1992). "A Constant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  Chemicals Shifts in  $^{15}\text{N}$ - $^{13}\text{C}$ -Labeled Proteins." Journal of Magnetic Resonance **97**: 213-217.
- Cole, R. and J. P. Loria (2003). "FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data." J Biomol NMR **26**(3): 203-13.
- Colgin, L. M., K. Baran, et al. (2003). "Human POT1 facilitates telomere elongation by telomerase." Curr Biol **13**(11): 942-6.
- Collins, K. (1996). "Structure and function of telomerase." Curr Opin Cell Biol **8**(3): 374-80.
- Cook, B. D., J. N. Dynek, et al. (2002). "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2 at human telomeres." Mol Cell Biol **22**(1): 332-42.
- Cornilescu, G., F. Delaglio, et al. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." J Biomol NMR **13**(3): 289-302.
- Court, R., L. Chapman, et al. (2005). "How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures." EMBO Rep **6**(1): 39-45.
- de Alba, E. and N. Tjandra (2002). "NMR dipolar couplings for the structure determination of biopolymers in solution." Prog NMR Spectrosc **40**: 175-197.
- Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-93.
- Ding, K. and A. M. Gronenborn (2003). "Sensitivity-enhanced 2D IPAP, TROSY-anti-TROSY, and E.COSY experiments: alternatives for measuring dipolar  $^{15}\text{N}$ - $^1\text{H}$ N couplings." J Magn Reson **163**(2): 208-14.

- Dominguez, C., R. Boelens, et al. (2003). "HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information." *J Am Chem Soc* **125**: 1731-1737.
- Dosset, P., J. C. Hus, et al. (2001). "A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings." *J Biomol NMR* **20**(3): 223-31.
- Fairall, L., L. Chapman, et al. (2001). "Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2." *Mol Cell* **8**(2): 351-61.
- Feng, J., W. D. Funk, et al. (1995). "The RNA component of human telomerase." *Science* **269**(5228): 1236-41.
- Fischer, M. W. F., A. Majumdar, et al. (1998). "Protein NMR relaxation: theory, applications and outlook." *Prog Nucl Magn Reson Spectrosc* **33**: 207-272.
- Fouché, N., A. J. Cesare, et al. (2006). "The Basic Domain of TRF2 Directs Binding to DNA Junctions Irrespective of the Presence of TTAGGG Repeats." *The Journal of Biological Chemistry* **281**(49): 37486-37495.
- Gavathiotis, E. and M. S. Searle (2003). "Structure of the parallel-stranded DNA quadruplex d(TTAGGGT)<sub>4</sub> containing the human telomeric repeat: evidence for A-tetrad formation from NMR and molecular dynamics simulations." *Org Biomol Chem* **1**(10): 1650-6.
- Gehring, K., J. L. Leroy, et al. (1993). "A tetrameric DNA structure with protonated cytosine-cytosine base pairs." *Nature* **363**(6429): 561-5.
- Greider, C. W. (1991). "Telomerase is processive." *Mol Cell Biol* **11**(9): 4572-80.
- Greider, C. W. (1996). "Telomere length regulation." *Annu Rev Biochem* **65**: 337-65.
- Greider, C. W. (1999). "Telomeres do D-loop-T-loop." *Cell* **97**(4): 419-22.
- Greider, C. W. and E. H. Blackburn (1985). "Identification of a specific telomere terminal transferase activity in Tetrahymena extracts." *Cell* **43**(2 Pt 1): 405-13.
- Griffith, J. D., L. Comeau, et al. (1999). "Mammalian telomeres end in a large duplex loop." *Cell* **97**(4): 503-14.
- Grzesiek, S., J. Anglister, et al. (1993). "Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in <sup>13</sup>C/<sup>15</sup>N - Enriched Proteins by isotropic Mixing of <sup>13</sup>C Magnetization." *Journal of Magnetic Resonance Ser B* **101**: 114-119.
- Grzesiek, S. and A. Bax (1992). "Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR." *J Am Chem Soc* **114**(16): 6291-6293.
- Grzesiek, S. and A. Bax (1992). "An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins." *Journal of Magnetic Resonance* **99**: 201-207.
- Grzesiek, S. and A. Bax (1993). "Amino acid type determination in the sequential assignment procedure of uniformly <sup>13</sup>C/<sup>15</sup>N-enriched proteins." *J Biomol NMR* **3**(2): 185-204.
- Grzesiek, S. and A. Bax (1993). "The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements." *J Am Chem Soc* **115**: 12593-12594.
- Gueron, M. and J. L. Leroy (2000). "The i-motif in nucleic acids." *Curr Opin Struct Biol* **10**(3): 326-31.
- Guntert, P. (2004). "Automated NMR structure calculation with CYANA." *Methods Mol Biol* **278**: 353-78.
- Guntert, P., C. Mumenthaler, et al. (1997). "Torsion angle dynamics for NMR structure calculation with the new program DYANA." *J Mol Biol* **273**(1): 283-98.
- Guntert, P. and K. Wuthrich (1991). "Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints." *J Biomol NMR* **1**(4): 447-56.

- Halvorsen, T. L., G. Leibowitz, et al. (1999). "Telomerase activity is sufficient to allow transformed cells to escape from crisis." *Mol Cell Biol* **19**(3): 1864-70.
- Hanaoka, S., A. Nagadoi, et al. (2005). "Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities." *Protein Sci* **14**(1): 119-30.
- Hanaoka, S., A. Nagadoi, et al. (2001). "NMR structure of the hRap1 Myb motif reveals a canonical three-helix bundle lacking the positive surface charge typical of Myb DNA-binding domains." *J Mol Biol* **312**(1): 167-75.
- Hargreaves, D., W. D. Rice, et al. (1998). "Crystal structure of E.coli RuvA with bound DNA Holliday junction at 6Å resolution." *Nat Struct Biol* **5**(8): 441-446.
- Hayflick, L. (1997). "Mortality and immortality at the cellular level. A review." *Biochemistry (Mosc)* **62**(11): 1180-90.
- Hemann, M. T., M. A. Strong, et al. (2001). "The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability." *Cell* **107**(1): 67-77.
- Hirao, I., G. Kawai, et al. (1994). "Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat." *Nucleic Acids Res* **22**(4): 576-82.
- Hsu, H. L., D. Gilley, et al. (2000). "Ku acts in a unique way at the mammalian telomere to prevent end joining." *Genes Dev* **14**(22): 2807-12.
- Huffman, K. E., S. D. Levene, et al. (2000). "Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang." *J Biol Chem* **275**(26): 19719-22.
- Hutchinson, E. G. and J. M. Thornton (1996). "PROMOTIF a program to identify and analyze structural motifs in proteins." *Protein Sci* **5**(2): 212-20.
- Johnson, B. A. (2004). "Using NMRView to visualize and analyze the NMR spectra of macromolecules." *Methods Mol Biol* **278**: 313-52.
- Johnson, B. A. and R. A. Blevins (1994). "NMRview: A computer program for the visualization and analysis of NMR data." *J Biomol NMR* **4**: 603-14.
- Karlseder, J., A. Smogorzewska, et al. (2002). "Senescence induced by altered telomere state, not telomere loss." *Science* **295**(5564): 2446-9.
- Kay, L. E., M. Ikura, et al. (1990). "Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Protein." *Journal of Magnetic Resonance* **89**: 496-514.
- Kay, L. E., D. A. Torchia, et al. (1989). "Backbone dynamics of proteins as studied by <sup>15</sup>N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease." *Biochemistry* **28**(23): 8972-9.
- Kilian, A., D. D. Bowtell, et al. (1997). "Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types." *Hum Mol Genet* **6**(12): 2011-9.
- Kim, N. W., M. A. Piatyszek, et al. (1994). "Specific association of human telomerase activity with immortal cells and cancer." *Science* **266**(5193): 2011-5.
- Kim, S. H., C. Beausejour, et al. (2004). "TIN2 mediates functions of TRF2 at human telomeres." *J Biol Chem* **279**(42): 43799-804.
- Kim, S. H., S. Han, et al. (2003). "The human telomere-associated protein TIN2 stimulates interactions between telomeric DNA tracts in vitro." *EMBO Rep* **4**(7): 685-91.
- Kim, S. H., P. Kaminker, et al. (1999). "TIN2, a new regulator of telomere length in human cells." *Nat Genet* **23**(4): 405-12.
- Konig, P., L. Fairall, et al. (1998). "Sequence-specific DNA recognition by the myb-like domain of the human telomere binding protein TRF1: a model for the protein-DNA complex." *Nucleic Acids Res* **26**(7): 1731-40.
- Larson, D. D., E. A. Spangler, et al. (1987). "Dynamics of telomere length variation in *Tetrahymena thermophila*." *Cell* **50**(3): 477-83.

- Laskowski, R. A., J. A. Rullmann, et al. (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-86.
- Lei, M., E. R. Podell, et al. (2004). "Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection." Nat Struct Mol Biol **11**(12): 1223-9.
- Leroy, J. L., M. Gueron, et al. (1994). "Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif." Nucleic Acids Res **22**(9): 1600-6.
- Levitt, M. H. (2001). Spin dynamics Basics of Nuclear Magnetic Resonance.
- Li, B., S. Oestreich, et al. (2000). "Identification of human Rap1: implications for telomere evolution." Cell **101**(5): 471-83.
- Li, W., D. Miyoshi, et al. (2003). "Structural competition involving G-quadruplex DNA and its complement." Biochemistry **42**(40): 11736-44.
- Linge, J. P., M. Habeck, et al. (2003). "ARIA: automated NOE assignment and NMR structure calculation." Bioinformatics **19**(2): 315-6.
- Lingner, J., T. R. Hughes, et al. (1997). "Reverse transcriptase motifs in the catalytic subunit of telomerase." Science **276**(5312): 561-7.
- Lipari, G. and A. Szabo (1982a). "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. Theory and range of validity." J. Am. Chem. Soc **104**: 4546-4559.
- Lipari, G. and A. Szabo (1982b). "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. Analysis of experimental result." J. Am. Chem. Soc. **104**: 4559-4570.
- Logan, T. M., E. T. Olejniczak, et al. (1993). "A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments." J Biomol NMR **3**(2): 225-31.
- Luthy, R., J. U. Bowie, et al. (1992). "Assessment of protein models with three-dimensional profiles." Nature **356**(6364): 83-5.
- Luu, K. N., A. T. Phan, et al. (2006). "Structure of the human telomere in K<sup>+</sup> solution: an intramolecular (3 + 1) G-quadruplex scaffold." J Am Chem Soc **128**(30): 9963-70.
- Lyons, B. A. and G. T. Montelione (1993). "An HCCNH Triple-Resonance Experiment Using Carbon-13 Isotropic Mixing for Correlating Backbone Amide and Side-Chain Aliphatic Resonances in Isotopically Enriched Proteins." Journal of Magnetic Resonance Ser B **101**: 206-209.
- Makarov, V. L., Y. Hirose, et al. (1997). "Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening." Cell **88**(5): 657-66.
- Mandel, A. M., M. Akke, et al. (1995). "Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme." J Mol Biol **246**(1): 144-63.
- McClintock, B. (1941). "The Stability of Broken Ends of Chromosomes in Zea Mays." Genetics **26**(2): 234-82.
- McIntosh, L. (1999). "Investigating Protein-Ligand Interactions: Isotope Editing and Filtering." Embo Course.
- Meyerson, M., C. M. Counter, et al. (1997). "hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization." Cell **90**(4): 785-95.
- Montelione, G. T., B. A. Lyons, et al. (1992). "An Efficient Triple - Resonance Experiment Using Carbon-13 Isotropic Mixing for Determining Sequence-Specific Resonance Assignments of Isotopically-Enriched Proteins." J Am Chem Soc **114**: 10974-10975.

- Morin, G. B. (1989). "The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats." *Cell* **59**(3): 521-9.
- Moyzis, R. K., J. M. Buckingham, et al. (1988). "A highly conserved repetitive DNA sequence, (TTAGGG)<sub>n</sub>, present at the telomeres of human chromosomes." *Proc Natl Acad Sci U S A* **85**(18): 6622-6.
- Muller, J. (1938). "The remaking of chromosomes." *The collecting Net* **13**(187-95).
- Muniyappa, K. and K. M. Kironmai (1998). "Telomere structure, replication and length maintenance." *Crit Rev Biochem Mol Biol* **33**(4): 297-336.
- Neidle, S. and G. N. Parkinson (2003). "The structure of telomeric DNA." *Curr Opin Struct Biol* **13**(3): 275-83.
- Nishikawa, T., A. Nagadoi, et al. (1998). "Solution structure of the DNA-binding domain of human telomeric protein, hTRF1." *Structure* **6**(8): 1057-65.
- Nishikawa, T., H. Okamura, et al. (2001). "Solution structure of a telomeric DNA complex of human TRF1." *Structure* **9**(12): 1237-51.
- O'Connor, M. S., A. Safari, et al. (2004). "The human Rap1 protein complex and modulation of telomere length." *J Biol Chem* **279**(27): 28585-91.
- Ogata, K., S. Morikawa, et al. (1995). "Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb." *Nature* **2**: 309-320.
- Ottiger, M., F. Delaglio, et al. (1998). "Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra." *J Magn Reson* **131**(2): 373-8.
- Otting, G. and K. Wüthrich (1989). "Extended Heteronuclear Editing of 2D <sup>1</sup>H NMR Spectra of Isotope-Labeled Proteins, Using the X(ω<sub>1</sub>,ω<sub>2</sub>) Double Half filter." *Journal of Magnetic Resonance* **85**: 586.
- Palmer III, A. G., M. Rance, et al. (1991). "Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterised by proton-detected natural abundance <sup>13</sup>C heteronuclear NMR spectroscopy." *J Am Chem Soc* **113**: 4371-4380.
- Parkinson, G. N., M. P. Lee, et al. (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA." *Nature* **417**(6891): 876-80.
- Patel, P. K. and R. V. Hosur (1999). "NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats." *Nucleic Acids Res* **27**(12): 2457-64.
- Pervushin, K., R. Riek, et al. (1997). "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." *Proc Natl Acad Sci U S A* **94**(23): 12366-71.
- Phan, A. T., M. Gueron, et al. (2000). "The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere." *J Mol Biol* **299**(1): 123-44.
- Phan, A. T. and J. L. Mergny (2002). "Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix." *Nucleic Acids Res* **30**(21): 4618-25.
- Piotto, M., V. Saudek, et al. (1992). "Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions." *J Biomol NMR* **2**(6): 661-5.
- Plateau, P. and M. Gueron (1982). "Exchangeable Proton NMR without Base-Line Distortion, Using New Strong-Pulse Sequences." *J Am Chem Soc* **104**: 7310-7311.
- Prestegard, J. H., H. M. al-Hashimi, et al. (2000). "NMR structures of biomolecules using field oriented media and residual dipolar couplings." *Q Rev Biophys* **33**(4): 371-424.
- Prestegard, J. H. and A. I. Kishore (2001). "Partial alignment of biomolecules: an aid to NMR characterization." *Curr Opin Chem Biol* **5**(5): 584-90.
- Rhodes, D. and R. Giraldo (1995). "Telomere structure and function." *Curr Opin Struct Biol* **5**(3): 311-22.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*. New York, Springer-Verlag.

- Sass, H. J., G. Musco, et al. (2000). "Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes." *J Biomol NMR* **18**(4): 303-9.
- Sattler, M., J. Schleucher, et al. (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradient." *Progress in NMR spectroscopy* **34**: 93-158.
- Shay, J. W. (1999). "At the end of the millennium, a view of the end." *Nat Genet* **23**(4): 382-3.
- Simonsson, T. (2001). "G-quadruplex DNA structures--variations on a theme." *Biol Chem* **382**(4): 621-8.
- Sippl, M. J. (1993). "Recognition of errors in three-dimensional structures of proteins." *Proteins* **17**(4): 355-62.
- Sklenar, V., M. Piotto, et al. (1993). "Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity." *Journal of Magnetic Resonance Series A* **102**: 241-245.
- Slijepcevic, P. (1998). "Telomeres and mechanisms of Robertsonian fusion." *Chromosoma* **107**(2): 136-40.
- Smith, S. and T. de Lange (2000). "Tankyrase promotes telomere elongation in human cells." *Curr Biol* **10**(20): 1299-302.
- Smith, S., I. Gariat, et al. (1998). "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres." *Science* **282**(5393): 1484-7.
- Smogorzewska, A., B. van Steensel, et al. (2000). "Control of human telomere length by TRF1 and TRF2." *Mol Cell Biol* **20**(5): 1659-68.
- Song, K., D. Jung, et al. (2000). "Interaction of human Ku70 with TRF2." *FEBS Lett* **481**(1): 81-5.
- Stansel, R. M., T. de Lange, et al. (2001). "T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang." *EMBO Journal* **20**(19): 5532-5540.
- Stewart, S. A., I. Ben-Porath, et al. (2003). "Erosion of the telomeric single-strand overhang at replicative senescence." *Nat Genet* **33**(4): 492-6.
- Sundquist, W. I. and A. Klug (1989). "Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops." *Nature* **342**(6251): 825-9.
- Tjandra, N., J. G. Omichinski, et al. (1997). "Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution." *Nat Struct Biol* **4**(9): 732-8.
- Tommerup, H., A. Dousmanis, et al. (1994). "Unusual chromatin in human telomeres." *Mol Cell Biol* **14**(9): 5777-85.
- Tugarinov, V., P. M. Hwang, et al. (2004). "Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins." *Annu Rev Biochem* **73**: 107-46.
- van Buuren, B. N. M., J. Schleucher, et al. (2000). "NMR Structural Studies on a DNA Four-Way Junction: Stacking Preference and Localization of the Metal-ion Binding Site." *J. Biomol. Struct. Dyn.* **11**(2): 237-243.
- van Steensel, B. and T. de Lange (1997). "Control of telomere length by the human telomeric protein TRF1." *Nature* **385**(6618): 740-3.
- van Steensel, B., A. Smogorzewska, et al. (1998). "TRF2 protects human telomeres from end-to-end fusions." *Cell* **92**(3): 401-13.
- Vargason, J. M. and P. S. Ho (2002). "The effect of cytosine methylation on the structure and geometry of the Holliday junction : the structure of d(CCGGTACm5CGG) at 1.5 Å resolution." *J.Biol.Chem* **277**: 21041-21049.
- Walker, J. R., R. A. Corpina, et al. (2001). "Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair." *Nature* **412**(6847): 607-14.

- Wang, A. C., P. J. Lodi, et al. (1994). "An Efficient Triple-Resonance Experiment for Proton-Directed Sequential Backbone Assignment of Medium-Sized Proteins." Journal of Magnetic Resonance Ser B **105**: 196-198.
- Wang, Y. and D. J. Patel (1993). "Solution structure of a parallel-stranded G-quadruplex DNA." J Mol Biol **234**(4): 1171-83.
- Wei, C. and M. Price (2003). "Protecting the terminus: t-loops and telomere end-binding proteins." Cell Mol Life Sci **60**(11): 2283-94.
- Williamson, M. P., T. F. Havel, et al. (1985). "Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry." J Mol Biol **182**(2): 295-315.
- Wittekind, M. and L. Mueller (1993). "HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins." Journal of Magnetic Resonance Ser B **101**: 201-205.
- Wright, W. E., V. M. Tesmer, et al. (1997). "Normal human chromosomes have long G-rich telomeric overhangs at one end." Genes Dev **11**(21): 2801-9.
- Wu, B., F. Girard, et al. (2004). "Global structure of a DNA three-way junction by solution NMR: towards prediction of 3H fold." Nucleic Acids Res **32**(10): 3228-3239.
- Wu, G., W. H. Lee, et al. (2000). "NBS1 and TRF1 colocalize at promyelocytic leukemia bodies during late S/G2 phases in immortalized telomerase-negative cells. Implication of NBS1 in alternative lengthening of telomeres." J Biol Chem **275**(39): 30618-22.
- Wüthrich, K. (1986). NMR of proteins and nucleic acids. New York.
- Yamazaki, T., W. Lee, et al. (1994). "A suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity." J Am Chem Soc **116**: 11655.
- Ye, J. Z. and T. de Lange (2004). "TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex." Nat Genet **36**(6): 618-23.
- Ye, J. Z., D. Hockemeyer, et al. (2004). "POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex." Genes Dev **18**: 1649-1654.
- Zahler, A. M., J. R. Williamson, et al. (1991). "Inhibition of telomerase by G-quartet DNA structures." Nature **350**(6320): 718-20.
- Zakian, V. A. (1995). "Telomeres: beginning to understand the end." Science **270**(5242): 1601-7.
- Zhou, X. Z. and K. P. Lu (2001). "The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor." Cell **107**(3): 347-59.
- Zhu, X. D., B. Kuster, et al. (2000). "Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres." Nat Genet **25**(3): 347-52.
- Zhu, X. D., L. Niedernhofer, et al. (2003). "ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes." Mol Cell **12**(6): 1489-98.

