



**HAL**  
open science

# Étude structurale de l'assemblage du complexe télomérique humain TRF2/RAP1

Guillaume Gaullier

► **To cite this version:**

Guillaume Gaullier. Étude structurale de l'assemblage du complexe télomérique humain TRF2/RAP1. Biologie structurale [q-bio.BM]. Université Paris Sud - Paris XI, 2015. Français. NNT : 2015PA114831 . tel-01281310

**HAL Id: tel-01281310**

**<https://theses.hal.science/tel-01281310v1>**

Submitted on 2 Mar 2016

**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.



Comprendre le monde,  
construire l'avenir®



## UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 425  
INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL À L'APPLIQUÉ

Équipe Enveloppe nucléaire, télomères et réparation de l'ADN  
Département de Biochimie, Biophysique et Biologie Structurale (B3S)  
Institut de Biologie Intégrative de la Cellule (I2BC)  
UMR 9198 CNRS, CEA, Université Paris-Sud

# THÈSE DE DOCTORAT

en  
Biochimie et Biologie Structurale

par

**Guillaume GAULLIER**

Étude structurale de l'assemblage  
du complexe télomérique humain TRF2/RAP1

**Date de soutenance : 22/09/2015**

**Composition du jury :**

Directrice de thèse :	Marie-Hélène LE DU	Chercheur CEA	(CEA, Saclay)
Rapporteurs :	Vincent GÉLI Dominique HOUSSET	DR CNRS Chercheur CEA	(CRCM, Marseille) (IBS, Grenoble)
Examineurs :	Laure CRABBÉ Claudine MAYER Herman VAN TILBEURGH	CR CNRS Professeur Professeur	(CGM, Gif-sur-Yvette) (Institut Pasteur, Paris) (Univ. Paris-Sud, Orsay)

## Étude structurale de l'assemblage du complexe télomérique humain TRF2/RAP1

### Résumé

Les télomères sont les extrémités des chromosomes linéaires des eucaryotes. Ils sont constitués de répétitions en tandem d'un motif court riche en guanine, et liés par des protéines spécifiques. Chez les vertébrés ces protéines forment un complexe appelé le *shelterin* et dont l'intégrité est critique pour assurer la réplication correcte des extrémités des chromosomes, et pour les protéger contre une prise en charge illicite par les voies de réparation des cassures double-brin de l'ADN. Des dysfonctions des télomères engendrent une instabilité du génome qui peut conduire à la sénescence ou au cancer. Les télomères représentent une région subnucléaire où les protéines du *shelterin* sont fortement enrichies, ce qui permet l'implication dans les fonctions biologiques d'interactions de basse affinité. Parmi les protéines du *shelterin*, la protéine de liaison aux répétitions télomériques TRF2 et son partenaire constitutif RAP1 sont les facteurs majeurs responsables de la protection des extrémités. Nous avons étudié en détails l'assemblage du complexe TRF2/RAP1 par des approches intégrées de biologie structurale, de biophysique et de biochimie. Nous avons montré que cet assemblage s'accompagne d'importants ajustements de conformation des deux protéines, et implique une interaction de basse affinité qui engage de grandes régions des deux protéines et affecte leurs propriétés d'interactions.

**Mots-clés :** télomères, TRF2, RAP1, SAXS, ITC, cristallographie, empreinte protéique.

### Structural study of the assembly of human TRF2/RAP1 telomeric complex

### Abstract

Telomeres are the ends of eukaryotic linear chromosomes. They are made of tandem repeats of a short guanine-rich motif and bound by specific proteins. In vertebrates, these proteins form a complex called shelterin, the integrity of which is critical to ensure proper replication of chromosome ends and to protect them against illicit targeting by DNA double-strand break repair pathways. Telomere dysfunctions lead to genome instability, which can ultimately cause senescence or cancer. Telomeres are a subnuclear region in which shelterin proteins are highly enriched, enhancing low affinity interactions of important biological function. Among shelterin proteins, telomeric repeat-binding protein TRF2 and its constitutive partner RAP1 are the main factors responsible for end protection. We studied in details the assembly of TRF2/RAP1 complex by means of integrated structural, biophysical and biochemical approaches. We showed that this assembly displays important conformational adjustments of both proteins, and involves a low affinity interaction engaging large regions in both proteins which affects their interaction properties.

**Keywords:** telomeres, TRF2, RAP1, SAXS, ITC, crystallography, protein footprinting.

## Table des matières

<b>Table des matières</b> .....	<b>1</b>
<b>Table des figures et tableaux</b> .....	<b>5</b>
<b>Glossaire</b> .....	<b>9</b>
<b>Remerciements</b> .....	<b>13</b>
<b>0 Introduction</b> .....	<b>19</b>
<b>0.1 Fonctions des télomères</b> .....	<b>19</b>
0.1.1 Le support de l'information génétique.....	19
0.1.2 Les télomères.....	19
0.1.3 Origine des chromosomes linéaires.....	20
0.1.4 Le problème de la réplication des extrémités.....	21
0.1.4.1 Aperçu du problème de la réplication.....	21
0.1.4.2 L'addition de répétitions télomériques par la télomérase.....	23
0.1.4.3 La voie alternative d'élongation des télomères.....	26
0.1.5 Autres solutions évolutives au problème de la réplication.....	29
0.1.5.1 Les rétro-transposons de <i>Drosophila</i> .....	29
0.1.5.2 Les chromosomes fermés coalementment de <i>Borrelia</i> .....	30
0.1.5.3 La protéine terminale de <i>Streptomyces</i> .....	32
0.1.6 Le problème de la protection des extrémités.....	33
0.1.6.1 Jonction d'extrémités non-homologues (NHEJ).....	35
0.1.6.2 Recombinaison homologue.....	37
0.1.7 Solutions moléculaires au problème de la protection apparues au cours de l'évolution.....	39
0.1.7.1 Chez les ciliés.....	39
0.1.7.2 Chez la drosophile.....	41
0.1.7.3 Chez les levures.....	42
<b>0.2 Les télomères humains</b> .....	<b>45</b>
0.2.1 L'ADN télomérique et la <i>t-loop</i> .....	45
0.2.2 Les protéines télomériques forment le complexe <i>shelterin</i> .....	46
0.2.3 Les facteurs télomériques accessoires.....	48
0.2.4 Régulation de la télomérase par les protéines télomériques.....	49
0.2.5 Protection de l'extrémité du chromosome par le complexe <i>shelterin</i> .....	50
<b>0.3 TRF2 et RAP1 : deux protéines essentielles</b> .....	<b>51</b>
0.3.1 Fonctions de TRF2 et RAP1.....	51
0.3.2 TRF2 et RAP1 sont des protéines multi-domaines.....	52
0.3.3 Partenaires de TRF2 et RAP1.....	55
<b>0.4 Objectifs de ces travaux</b> .....	<b>57</b>
<b>1 Partie 1 : Étude détaillée de l'interaction entre TRF2 et RAP1</b> .....	<b>59</b>
<b>1.1 Résultats</b> .....	<b>62</b>
1.1.1 L'interaction entre TRF2 et RAP1 entières est complexe.....	62
1.1.2 L'interaction est associée à des changements de conformation.....	65
1.1.3 L'interaction entre TRF2 et RAP1 implique un motif TBM.....	75
1.1.4 Les surfaces d'interaction entre TRF2 et RAP1 ne sont pas limitées au motif TBM.....	79
<b>1.2 Discussion</b> .....	<b>88</b>
1.2.1 Mise en évidence de la transition mineure dans la courbe de titration.....	88
1.2.2 Convergence des résultats et corrélation avec la conservation des protéines.....	89

1.2.3	Importance de l'étude des protéines entières.....	91
1.2.4	Effets de RAP1 sur les propriétés d'interaction de TRF2 avec l'ADN.....	92
1.2.5	Rôle du domaine BRCT de RAP1 .....	93
1.2.6	Étude en présence de RAP1 de l'interaction entre TRF2 et ses partenaires connus.....	94
<b>2</b>	<b>Partie 2 : Partenaires de TRF2 et du complexe TRF2/RAP1 .....</b>	<b>95</b>
<b>2.1</b>	<b>Résultats.....</b>	<b>96</b>
2.1.1	Apollo .....	96
2.1.2	WRN .....	98
2.1.3	Ku .....	103
2.1.4	<i>Pull down</i> streptavidine-ADN.....	105
<b>2.2</b>	<b>Discussion.....</b>	<b>108</b>
2.2.1	Apollo .....	108
2.2.2	WRN .....	109
2.2.3	Ku .....	111
2.2.4	<i>Pull down</i> streptavidine-ADN.....	112
2.2.5	Perspectives sur l'étude des partenaires de TRF2 et RAP1 .....	113
<b>3</b>	<b>Partie 3 : Mode d'interaction de TRF2 avec l'ADN .....</b>	<b>115</b>
<b>3.1</b>	<b>Résultats.....</b>	<b>117</b>
3.1.1	Calorimétrie.....	117
3.1.2	SAXS .....	120
3.1.3	Cristallographie.....	124
3.1.3.1	Cristallisation.....	124
3.1.3.2	Collecte et traitement des données.....	125
3.1.3.3	Résolution de la structure et affinement.....	127
<b>3.2</b>	<b>Discussion.....</b>	<b>130</b>
<b>4</b>	<b>Conclusion.....</b>	<b>131</b>
<b>5</b>	<b>Matériel et méthodes.....</b>	<b>133</b>
<b>5.1</b>	<b>Constructions utilisées.....</b>	<b>133</b>
5.1.1	Protéines.....	133
5.1.1.1	TRF2.....	133
5.1.1.2	RAP1.....	134
5.1.1.3	WRN.....	135
5.1.1.4	Ku.....	135
5.1.2	Oligonucléotides.....	135
5.1.2.1	Première génération.....	135
5.1.2.2	Seconde génération.....	136
5.1.3	Peptides.....	137
<b>5.2</b>	<b>Expression des protéines.....</b>	<b>138</b>
5.2.1	Transformation.....	138
5.2.2	Culture des bactéries et induction de l'expression.....	138
5.2.3	Détails spécifiques aux protéines.....	139
5.2.3.1	TRF2, TRFH, B-TRFH, WRN_949-1079, Myb-court, Myb-long.....	139
5.2.3.2	RAP1, RAP1-ARAGA, RAP1_1-208.....	139
5.2.4	Résultats des expressions.....	139
5.2.4.1	TRF2.....	140
5.2.4.2	RAP1.....	140
5.2.4.3	RAP1-ARAGA .....	141
5.2.4.4	RAP1_1-208.....	141
5.2.4.5	B-TRFH.....	142

5.2.4.6	WRN_949-1079 .....	143
5.2.4.7	Myb-court.....	144
<b>5.3</b>	<b>Purification des protéines recombinantes.....</b>	<b>145</b>
5.3.1	Protocole .....	145
5.3.1.1	Tampons .....	145
5.3.1.2	Lyse des bactéries.....	145
5.3.1.3	Purification des protéines.....	145
5.3.2	Détails spécifiques à RAP1 et RAP1-ARAGA.....	146
5.3.2.1	Tampons .....	146
5.3.2.2	Protocole .....	147
5.3.3	Détails spécifiques à TRF2, WRN_979-1049 et le domaine Myb-court de TRF2.....	147
5.3.3.1	Tampons .....	147
5.3.3.2	Protocole .....	148
5.3.4	Résultats des purifications.....	148
5.3.4.1	TRF2 .....	148
5.3.4.2	RAP1.....	149
5.3.4.3	RAP1-ARAGA .....	150
5.3.4.4	RAP1_1-208.....	151
5.3.4.5	TRFH et B-TRFH.....	152
5.3.4.6	WRN_949-1079 .....	153
5.3.4.7	Myb-court.....	154
5.3.4.8	Myb-long.....	155
<b>5.4</b>	<b>Dichroïsme circulaire .....</b>	<b>157</b>
5.4.1	Protocole .....	157
5.4.2	Résultats.....	157
<b>5.5</b>	<b>Titration calorimétriques .....</b>	<b>159</b>
<b>5.6</b>	<b>SAXS .....</b>	<b>161</b>
5.6.1	Préparation des échantillons .....	161
5.6.2	Acquisition des courbes de diffusion.....	161
5.6.3	Analyse des données.....	161
5.6.3.1	Traitement des données brutes.....	161
5.6.3.2	Calcul des enveloppes.....	162
<b>5.7</b>	<b>Empreinte protéique.....</b>	<b>164</b>
5.7.1	Principe et mise au point .....	164
5.7.2	Protocole.....	165
5.7.2.1	Réaction .....	165
5.7.2.2	Spectrométrie de masse.....	166
5.7.2.3	Analyse .....	167
<b>5.8</b>	<b>Tests d'interaction par chromatographie d'exclusion stérique.....</b>	<b>168</b>
<b>5.9</b>	<b>Test d'interaction entre TRFH et Ku par rétention sur résine de nickel.....</b>	<b>170</b>
<b>5.10</b>	<b>Culture de cellules HeLa .....</b>	<b>171</b>
5.10.1	Matériel.....	171
5.10.2	Protocoles .....	171
5.10.2.1	Démarrer une nouvelle culture.....	171
5.10.2.2	Passage des cellules .....	172
5.10.2.3	Stockage d'aliqots de cellules.....	172
<b>5.11</b>	<b>Préparation des extraits nucléaires .....</b>	<b>173</b>
<b>5.12</b>	<b>Pull down streptavidine-ADN.....</b>	<b>174</b>
5.12.1	Principe.....	174
5.12.2	Protocole .....	174
<b>5.13</b>	<b>Hybridation des oligonucléotides.....</b>	<b>176</b>
5.13.1	Matériel.....	176

5.13.2	Protocole .....	176
5.13.3	Remarques.....	177
<b>5.14</b>	<b>Résolution de la structure du complexe Myb-long/ADN.....</b>	<b>179</b>
5.14.1	Préparation du complexe.....	179
5.14.1.1	Propriétés du complexe et de ses composants .....	180
5.14.2	Cristallogenèse .....	180
5.14.3	Collecte et traitement des données de diffraction .....	181
5.14.4	Détermination des phases .....	181
5.14.5	Affinement des modèles .....	182
<b>5.15</b>	<b>Alignements de séquences .....</b>	<b>183</b>
<b>6</b>	<b>Références.....</b>	<b>185</b>
<b>7</b>	<b>Annexes.....</b>	<b>197</b>
7.1	Alignement des séquences de RAP1 .....	197
7.2	Alignement des séquences de TRF2.....	198
7.3	Publication.....	199

## Table des figures et tableaux

Figure 1 : Le problème de la réplication des extrémités.....	22
Figure 2 : Représentation schématique de la télomérase humaine .....	24
Figure 3 : Synthèse des répétitions télomériques par la télomérase .....	25
Figure 4 : Longueur des télomères maintenus par la voie ALT et par la télomérase .....	27
Figure 5 : Modèles actuels de l'élongation des télomères par la voie ALT.....	28
Figure 6 : Principe de la réplication des ADN à télomères en épingle à cheveux.....	30
Figure 7 : Mécanisme des résolvases de télomères.....	31
Figure 8 : Mécanisme de synthèse de la fin du brin tardif chez <i>Streptomyces</i> .....	33
Figure 9 : La voie NHEJ de réparation des cassures double-brin de l'ADN.....	36
Figure 10 : La recombinaison homologe .....	38
Figure 11 : Complexe télomérique d' <i>Oxytricha nova</i> .....	40
Figure 12 : Complexes télomériques d' <i>O. nova</i> et <i>T. thermophila</i> .....	41
Figure 13 : Complexe <i>terminin</i> de <i>D. melanogaster</i> .....	42
Figure 14 : Complexe télomérique de <i>S. cerevisiae</i> .....	43
Figure 15 : Complexe télomérique de <i>S. pombe</i> .....	44
Figure 16 : L'ADN télomérique des mammifères .....	45
Figure 17 : La <i>t-loop</i> .....	45
Figure 18 : Domaines structuraux des protéines du <i>shelterin</i> .....	47
Figure 19 : Organisation des domaines de TRF2 et RAP1 .....	53
Figure 20 : Représentation schématique du complexe TRF2/RAP1 .....	54
Figure 21 : Conservation des résidus en surface du domaine TRFH de TRF2.....	56
Figure 22 : Les motifs TBM actuellement connus et validés expérimentalement .....	56
Figure 23 : Effets de RAP1 sur les propriétés d'interaction de TRF2 avec l'ADN.....	60
Figure 24 : Représentation schématique d'un calorimètre de titration isotherme.....	62
Figure 25 : Titration calorimétrique de TRF2 par RAP1 et par RAP1-RCT.....	64



Figure 26 : Courbes de SAXS de nos protéines et complexes.....	66
Figure 27 : Profils de $R_g$ et $I_0$ le long des chromatogrammes.....	68
Figure 28 : Représentation de Kratky .....	69
Figure 29 : Représentations de Kratky de nos protéines.....	70
Figure 30 : Formes caractéristiques de la fonction $P(r)$ .....	71
Figure 31 : Fonctions de distribution des distances de nos protéines.....	72
Figure 32 : Enveloppes de nos protéines.....	74
Figure 33 : Alignement du motif YRLGP de RAP1 avec les motifs TBM connus. ....	76
Figure 34 : Titration calorimétrique du domaine TRFH par le peptide RAP1-TBM .....	77
Figure 35 : Structure cristallographique du complexe TRFH/RAP1-TBM.....	78
Figure 36 : Titration calorimétrique de TRF2 par RAP1-ARAGA.....	79
Figure 37 : Principe de l'expérience d'empreinte protéique.....	80
Figure 38 : Positions des lysines dans les séquences et structures de TRF2 et RAP1 .....	81
Figure 39 : Interprétation des spectres de masse.....	82
Figure 40 : Empreintes de TRF2 sur RAP1 et de RAP1 sur TRF2 .....	83
Figure 41 : Protection des lysines K327 et K329 de TRF2 dans le complexe TRF2/RAP1 ....	84
Figure 42 : Empreinte de RAP1 sur le domaine TRFH de TRF2 .....	85
Figure 43 : Empreinte de TRF2 sur la région N-terminale de RAP1.....	86
Figure 44 : Modèle du complexe TRF2/RAP1.....	87
Figure 45 : Titration calorimétrique de TRFH par RAP1-TBM.....	89
Figure 46 : Région du motif TBM dans l'alignement des séquences de RAP1 .....	90
Figure 47 : Titration calorimétrique de B-TRFH par RAP1_1-208 .....	91
Figure 48 : Superposition des structures TRFH/RAP1-TBM et TRFH/Apollo-TBM.....	96
Figure 49 : Titrations calorimétriques d'Apollo-TBM.....	97
Figure 50 : Alignement du motif KLLLP de WRN avec les motifs TBM connus.....	99
Figure 51 : Titrations calorimétriques du domaine TRFH par WRN-RQC .....	100
Figure 52 : Titration calorimétrique de TRF2 par WRN-RQC.....	101

Figure 53 : Titration calorimétrique de TRF2/RAP1 et de RAP1 par WRN-RQC.....	102
Figure 54 : Titration calorimétrique de WRN-RQC par l'oligonucléotide T2-S5S6.....	103
Figure 55 : Tests d'interaction entre Ku et TRFH.....	104
Figure 56 : Titrations calorimétriques de Ku par TRFH et par TRF2.....	105
Figure 57 : Chromatogrammes d'élué des <i>pull down</i> .....	106
Figure 58 : Analyse SDS-PAGE des éluats des <i>pull down</i> .....	107
Figure 59 : Titration de TRF2 et du complexe TRF2/RAP1 préformé par TIN2-TBM.....	109
Figure 60 : Structure tridimensionnelle de l'hétérodimère Ku.....	111
Figure 61 : Titrations calorimétriques de Myb-court et Myb-long par l'ADN.....	117
Figure 62 : Contributions enthalpique et entropique dans les interactions Myb/ADN .....	119
Figure 63 : Variation des paramètres $R_g$ et $I_0$ au cours de l'élué des Myb/ADN .....	121
Figure 64 : Courbes de SAXS des complexes Myb/ADN .....	122
Figure 65 : Représentations de Kratky des complexes Myb/ADN.....	122
Figure 66 : Fonctions de distribution des distances des complexes Myb/ADN.....	123
Figure 67 : Cristaux du complexe Myb-long/T2-S5S7 .....	125
Figure 68 : Expression de TRF2 .....	140
Figure 69 : Expression de RAP1.....	140
Figure 70 : Expression de RAP1-ARAGA .....	141
Figure 71 : Expression de RAP1_1-208.....	141
Figure 72 : Expression de B-TRFH.....	142
Figure 73 : Expression de WRN_949-1079 .....	143
Figure 74 : Expression du Myb-court.....	144
Figure 75 : Purification de TRF2.....	149
Figure 76 : Purification de RAP1 .....	150
Figure 77 : Purification de RAP1-ARAGA.....	151
Figure 78 : Purification de RAP1_1-208 .....	152
Figure 79 : Purification de B-TRFH .....	153

Figure 80 : Purification de WRN_949-1079.....	154
Figure 81 : Purification du Myb-court.....	155
Figure 82 : Purification du Myb-long.....	156
Figure 83 : Spectres de dichroïsme circulaire de TRF2 et de RAP1 et RAP1-ARAGA.....	158
Figure 84 : Dénaturation thermique de RAP1 et RAP1-ARAGA.....	158
Figure 85 : Principe de l'expérience d'empreinte protéique.....	164
Figure 86 : Structure du sulfo-NHS-acetate.....	165
Figure 87 : Chromatogrammes des tests d'interaction entre TRF2 et RAP1. ....	169
Figure 88 : Contrôle de l'hybridation des oligonucléotides T2-S5 et T2-S7.....	177
Figure 89 : Chromatogramme d'exclusion stérique du complexe Myb-long/ADN.....	179
Figure 90 : Modèle de remplacement moléculaire.....	182
Tableau 1 : Séquences des répétitions télomériques d'organismes modèles.....	20
Tableau 2 : Structures tridimensionnelles des domaines de TRF2 et RAP1.....	54
Tableau 3 : Comparaison des paramètres thermodynamiques des interactions TRF2/RAP1- RCT et TRF2/RAP1.....	64
Tableau 4 : $R_g$ , $D_{max}$ et masses moléculaires des protéines et des complexes.....	73
Tableau 5 : Conditions expérimentales de la titration calorimétrique de TRF2 par RAP1....	88
Tableau 6 : Comparaison des paramètres thermodynamiques des interactions Myb- court/ADN et Myb-long/ADN.....	118
Tableau 7 : $R_g$ et $D_{max}$ des complexes Myb-court/ADN et Myb-long/ADN.....	123
Tableau 8 : Jeux de données cristallographiques et statistiques.....	126
Tableau 9 : Résultats du remplacement moléculaire et du premier affinement.....	128
Tableau 10 : Conditions expérimentales des titrations calorimétriques.....	160
Tableau 11 : Tests d'interaction par chromatographie d'exclusion stérique. ....	168

## Glossaire

- Å : unité de mesure de distance,  $1 \text{ \AA} = 10^{-10} \text{ m}$
- A, T, G, C : adénine, thymine, guanine, cytosine
- ADN : acide désoxyribonucléique
- AFM : *atomic force microscopy*, microscopie à force atomique, technique consistant à sonder une surface avec une pointe microscopique, permet de mesurer la hauteur en chaque point de la surface
- ALT : *alternative lengthening of telomeres*, voie d'élongation des télomères indépendante de la télomérase
- ARN : acide ribonucléique
- ATM : *ataxia telangiectasia mutated*, protéine kinase initiant la voie DDR
- ATP : adénosine triphosphate
- ATR : *ataxia telangiectasia and Rad3 related*, protéine kinase initiant la voie DDR
- BRCT : *BRCA1 C-terminus*, domaine C-terminal de la protéine BRCA1, désigne par extension la famille de domaines homologues
- CCE1 : résolvase mitochondriale de levure
- Cdc13 : protéine télomérique de *S. cerevisiae* liant l'ADN simple brin et appartenant au complexe CST
- ChIP : *chromatin immunoprecipitation*, technique de biologie moléculaire permettant de déterminer les régions du génome auxquelles se lie une protéine donnée
- CO-FISH : *chromosome-orientation fluorescence in situ hybridization*, technique de marquage fluorescent spécifique d'un brin
- CST : complexe formé par Cdc13, Stn1 et Ten1
- CV : *column volume* soit en français volumes de colonne
- D-loop : *displacement loop*, boucle d'ADN simple-brin formée suite à l'invasion d'un duplex par un simple-brin
- DAPI : 4',6-diamidino-2-phenylindole, intercalant de l'ADN utilisé comme marqueur fluorescent non spécifique d'une séquence
- db : double-brin
- dCMP : désoxycytidine monophosphate
- DDR : *DNA damage response*, voie de signalisation des dommages à l'ADN
- DNA-PK<sub>cs</sub> : sous-unité catalytique de la protéine kinase dépendante de l'ADN, initiant la voie DDR
- DNaseI : endonucléase I
- DSBR : *double-strand break repair*, réparation des cassures double-brin de l'ADN
- EDTA : éthylène diamine tétra-acétique (chélateur des cations divalents)
- FISH : *fluorescence in situ hybridization*, technique de détection de séquence d'ADN *in situ*
- GEN1 : résolvase humaine
- GFP : *green fluorescent protein*
- GST : *glutathione S-transférase*

- H2AX : variant de l'histone H2A, phosphorylé par ATM lors de l'activation de la voie DDR
- HeT-A, TART, TAHRE : rétro-transposons télomériques de *Drosophila melanogaster*
- HipHop : protéine télomérique de drosophile
- HOAP : protéine télomérique de drosophile
- IPTG : isopropyl  $\beta$ -D-1-thiogalactopyranoside (inducteur synthétique non hydrolysable de l'opéron lactose)
- ITC : *isothermal titration calorimetry*, titration calorimétrique isotherme, technique de biophysique permettant de mesurer les paramètres thermodynamiques des interactions entre molécules purifiées
- K : unité de mesure de température (degré Kelvin), 273 K = 0 °C
- $K_A$  : constante d'équilibre d'association (unité  $M^{-1}$ )
- kb : 1000 pb
- $K_D$  : constante d'équilibre de dissociation (unité : M)
- MCPH1 : *microcephalin 1*
- Moi : protéine télomérique de drosophile
- MRE11 : *meiotic recombination 11*
- MRN : complexe formé par MRE11, RAD50 et NBS1
- Myb : Domaine structural à trois hélices
- NBS1 : *nibrin*
- NHEJ : *non-homologous end-joining*, voie de réparation des cassures double-brin de l'ADN
- NLS : *nuclear localization signal*, signal de localisation nucléaire
- OB-fold : *oligonucleotide-binding fold*, domaine de liaison à l'ADN simple-brin
- PARP1 : *poly-ADP ribosylase protein 1*
- Pat1 : protéine télomérique de *Tetrahymena thermophila*
- Pat2 : protéine télomérique de *Tetrahymena thermophila*
- pb : paires de bases
- PBS : *phosphate buffered saline*
- PCA : *protein complementation assay*, technique de biologie moléculaire permettant de détecter des interactions protéine-protéine dans des cellules en culture
- PDB : *protein data bank*, dépôt centralisé de structures tridimensionnelles de macromolécules
- PMSF : fluorure de phénylméthylsulfonyl (inhibiteur des protéases à sérine)
- PNUTS : *Phosphatase nuclear targeting subunit*, protéine partenaire du *shelterin*
- Pot1 : *protection of telomeres 1*, protéine télomérique de liaison aux répétitions simple-brin
- POT1 : *protection of telomeres 1*, protéine télomérique de liaison aux répétitions simple-brin
- Pot1a : *protection of telomeres 1*, protéine télomérique de liaison aux répétitions simple-brin
- Poz1 : protéine télomérique de *S. pombe*
- qsp : quantité suffisante pour
- RAD50 : *radiation sensitive mutant 50*
- Rap1 : *repressor/activator protein 1*
- RAP1 : *repressor/activator protein 1*, protéine du complexe *shelterin*

- RAP1-ARAGA : RAP1 humaine dont les résidus du motif TBM sont mutés en alanine
- RBM : *RAP1-binding motif*, région de TRF2 liant le domaine C-terminal de RAP1
- RCT : *RAP1 C-terminus*, domaine C-terminal de RAP1
- ResT : résolvasse de télomères de *Borrelia burgdorferi*
- Rif : *Rap1 interacting factor*, protéine télomérique de levure (1 ou 2), partenaire de Rap1
- RMN : résonance magnétique nucléaire, technique de biologie structurale permettant d'étudier des molécules en solution
- RPA : *replication protein A*, protéine de liaison à l'ADN simple-brin
- SAXS, SEC-SAXS : *small angle X-ray scattering*, diffusion des rayons X aux petits angles, et *size-exclusion chromatography - SAXS*, chromatographie d'exclusion stérique couplée à la mesure de diffusion des rayons X aux petits angles
- sb : simple-brin
- SPR : *surface plasmon resonance*, résonance plasmonique de surface, technique de biophysique permettant de mesurer les vitesses d'association et de dissociation de complexes entre molécules purifiées (l'une des deux étant immobilisée)
- SDS-PAGE : *sodium dodecyl sulfate - polyacrylamide gel electrophoresis*, électrophorèse sur gel de polyacrylamide en conditions dénaturantes
- SDSA : *synthesis-dependant strand annealing*
- shRNA : *short hairpin RNA*, outil de biologie moléculaire permettant de réduire l'expression d'un gène en dégradant l'ARN messenger correspondant
- Sir : *silent information regulator*, protéines de *saccharomyces cerevisiae* responsables de la répression transcriptionnelle
- SLX4 : *Fanconi anemia protein*, protéine humaine partenaire de TRF2
- Stn1 : protéine télomérique de *S. cerevisiae* appartenant au complexe CST
- STORM : *stochastic optical reconstruction microscopy*, technique de microscopie dite à "super-résolution"
- T : température
- T-SCE : *telomere sister-chromatid exchange*, échange de télomères entre chromatides soeurs
- TAE : Tris-acétate/EDTA
- Taz1 : orthologue de TRF1/2 chez *S. pombe*
- Tbf1 : *TTAGGG repeat-binding factor 1*, orthologue de TRF1/2 chez *S. cerevisiae*
- TBM : *TRFH-binding motif*, motif de liaison au domaine TRFH, de séquence consensus YxLxP ou YRL pour la liaison à TRF2
- TEBP  $\alpha$ ,  $\beta$  : *telomere end-binding proteins  $\alpha$ ,  $\beta$* , protéines télomériques d'*Oxytricha nova*
- TelA : résolvasse de télomères d'*Agrobacterium tumefaciens*
- TelK : résolvasse de télomères d'un bactériophage
- TeloPIN : *Telomeric Proteins Interaction Network*, base de données recensant les partenaires des protéines du *shelterin*
- Ten1 : protéine télomérique de *S. cerevisiae* appartenant au complexe CST
- TERT : *telomerase reverse transcriptase*, sous-unité rétro-transcriptase de la télomérase
- TIF : *telomere dysfunction-induced foci*, foyers de dommages télomériques
- TIN2 : *TRF1-interacting nuclear factor 2*, protéine du complexe *shelterin*
- TP : *terminal protein*, protéine terminale des *Streptomyces*

- TPP1 : protéine du *shelterin* reliant TIN2 et POT1
- Tpt1 : protéine télomérique de *Tetrahymena thermophila*
- Tpz1 : protéine télomérique de *S. pombe*
- TR, TERC : *telomerase RNA component*, sous-unité ARN de la télomérase
- TRF1 et TRF2 : *telomere repeat-binding factor 1 & 2*, protéines télomériques de liaison aux répétitions double-brin chez les eucaryotes supérieurs, formant le complexe *shelterin*
- TRFH : *TRF homology domain*, domaine de dimérisation des protéines de la famille des TRF
- Ver : protéine télomérique de drosophile
- VIH : virus de l'immunodéficience humaine
- WRN : *Werner syndrome protein*, hélicase dont les déficiences provoquent le syndrome de Werner
- XRCC4 : *X-ray repair cross-complementing protein 4*, protéine de la voie NHEJ
- YFP : *yellow fluorescent protein*
- YxLxP : Motif consensus d'interaction avec le domaine TRFH de TRF2
- $\Delta G$  : variation d'enthalpie libre
- $\Delta H$  : variation d'enthalpie (chaleur échangée au cours d'une réaction ou de la formation d'un complexe)
- $\Delta S$  : variation d'entropie

## Remerciements

Ces travaux ont été réalisés au laboratoire de biologie structurale et radiobiologie (LBSR) du service de bioénergétique, biologie structurale et mécanismes (SB2SM) du CEA. Je souhaite tout d'abord remercier le programme Irtélis de la direction des sciences de la vie du CEA qui a bien voulu m'attribuer un contrat doctoral pour travailler dans ce laboratoire pendant ces trois dernières années. Merci à Bruno Robert, directeur du SB2SM, qui a soutenu ma candidature au concours. Je remercie également tous les chercheurs du LBSR qui m'y ont fait un très bon accueil.

Je remercie Dominique Housset et Vincent Géli qui ont accepté d'être rapporteurs de ma thèse, ainsi que Claudine Mayer, Laure Crabbé et Herman van Tilbeurgh qui ont accepté de constituer le jury de la soutenance avec eux. Merci également Claudine d'avoir accepté le rôle de rapportrice de mi-parcours.

Une thèse serait irréalisable sans un directeur, en l'occurrence une directrice : merci Marie-Hélène, merci pour tout. Merci tout d'abord d'avoir donné sa chance à un petit étudiant (avec ce petit stage de licence) ; qui aurait deviné à l'époque que cela nous mènerait jusqu'ici ? Les remerciements pourraient être longs si je tentais de les détailler complètement. Merci pour ta rigueur scientifique, et pour ta rigueur et ton souci du détail en général : je pensais être maniaque, je me suis aperçu que tu l'es au moins autant, mais sur d'autres choses (alors c'était plutôt bien complémentaire :-). Surtout merci pour ton optimisme et ton enthousiasme inébranlables et toujours communicatifs, qui ont fait que reproduire des manips un certain nombre de fois n'était finalement pas insurmontable (je m'en suis aperçu avec les expériences d'acétylation, qu'il a fallu refaire tellement de fois...). C'est aussi cet état d'esprit qui a confirmé mon goût pour la recherche.

Marie-Hélène et moi ne nous serions pas rencontrés, et cette thèse aurait donc été tout autre, si je n'avais pas fait ce stage hors cursus à la plateforme de cristallogenèse de l'Institut Pasteur en 2009. J'adresse un grand merci à Ahmed Haouz, directeur de la plateforme, qui m'y avait alors accueilli et m'avait très bien encadré. Merci également Ahmed pour ton soutien au début de ma thèse, et pour ton aide qui a permis l'obtention des cristaux mentionnés dans la troisième partie du manuscrit. Je regrette de ne pas avoir eu plus souvent l'occasion de travailler en collaboration avec la plateforme, mais le projet a pris une tournure telle que nous avons eu besoin d'autres techniques que la cristallographie, c'est ainsi.

Les expériences réalisées pendant cette thèse ont pu être menées à bien grâce à la participation de nombreuses personnes. J'adresse mes plus grands remerciements à Yann-Vaï Le Bihan : tes conseils m'ont été précieux pendant toute ma thèse, et c'est en grande



partie grâce à toi que je suis devenu autonome pour les manipulations. Merci également Simona Miron pour m'avoir enseigné la calorimétrie et pour ta participation à tous nos travaux. Merci à Carine Tellier-Lebègue pour son aide avec la biologie moléculaire et les purifications de protéines, ainsi que divers tests d'expression et d'interaction, et surtout pour l'assistance quotidienne au laboratoire. Un immense merci à Bob Thai pour sa grande implication dans les expériences d'empreinte par acétylation : tu m'as enseigné les bases de la spectrométrie de masse, tu as été super réactif quand il fallait qu'on reproduise rapidement des expériences, j'en suis très reconnaissant. Merci Clément pour avoir fourni le dimère Ku purifié, et pour avoir réalisé quelques expériences, et merci pour ta motivation de doctorant fraîchement arrivé (ça aide quand, en troisième année, on a épuisé une bonne partie de cette motivation des débuts). Merci Wala pour ta participation à quelques-unes de ces expériences, c'était une aide bienvenue. Je te souhaite bon courage pour la fin de ta thèse, et sois certaine que je viendrai à ta soutenance. :-)

Je remercie également les scientifiques des lignes de lumière du synchrotron SOLEIL pour leur assistance durant les collectes et le traitement initial des données : Beatriz Guimarães et Pierre Legrand (PROXIMA 1), Martin Savko (PROXIMA 2), et Aurélien Thureau (SWING). Ils n'hésitent jamais à sacrifier des nuits et des week-ends pour s'assurer du bon fonctionnement de l'instrumentation complexe des lignes de lumière, et méritent donc des remerciements particuliers.

Au cours de la deuxième année, j'ai eu l'opportunité de passer une semaine au laboratoire de l'équipe d'Eric Gilson à l'IRCAN à Nice. Je remercie chaleureusement Eric, Sabrina, Delphine et Marie-Jo pour leur accueil. Plus particulièrement, merci Sabrina et Marie-Jo pour toutes ces expériences qui constituent une grande contribution à l'article qui sera j'espère bientôt publié. Merci également Delphine pour notre discussion très constructive et pour avoir partagé ton expérience sur les manipulations d'empreinte protéique. Je n'ai malheureusement pas eu le temps de poursuivre le projet qui m'avait initialement fait venir apprendre une technique dans votre laboratoire, mais ce court séjour m'a appris beaucoup de choses (expérimentales et théoriques) et j'en garde un excellent souvenir.

Je remercie toutes les personnes qui ont installé une bonne ambiance dans le bureau que nous avons partagé pendant que je l'ai occupé ces trois dernières années : Benjamin, Isslam, Wala, Philippe, Jingqi et Jonathan. C'était un plaisir d'être votre voisin pendant ces moments de rigolade et parfois même de discussions scientifiques sérieuses (si, si). Merci aux voisins des bureaux du rdc, en particulier Michel pour ton *coaching* actif pendant les derniers mois. Merci aussi à ceux "d'en haut" : le labo et les bureaux voisins. L'étage, cette autre dimension où on s'aperçoit toujours trop tard qu'on a oublié quelque chose au rez-de-chaussée : stylo, cahier, timer, marqueur, info nécessaire pour la manip qu'on a oublié de noter, etc. chaque chose utile imaginable a été oubliée en bas au moins une fois. Merci aux habitants de cette dimension d'avoir supporté les emprunts récurrents de stylos, marqueurs, etc. C'est aussi

une dimension où on prend conscience de la relativité du temps : lorsque l'on s'y trouve, le temps semble passer très vite tandis que les occupants du rez-de-chaussée ont l'impression que l'attente pour partir à la cantine dure des siècles. Occupant un bureau au rdc et travaillant dans le labo à l'étage, ma perception du temps a été pour le moins étrange pendant ces trois ans.

Enfin, je tiens à remercier tout particulièrement toutes les *autres* personnes sans qui cette thèse ne serait pas aboutie. Celles qui n'étaient pas au labo, et qui ont justement réussi à m'en faire sortir suffisamment régulièrement. Merci à mes amis du lycée (c'est super d'avoir pu rester en contact avec vous depuis tout ce temps !) et aux personnes rencontrées par votre intermédiaire : Anne-Sophie, Béatrice & Kévin, Claire, Guillaume, Adrien, Mikaël, Quentin, Virgile & Lenka (merci pour ce magnifique séjour en Slovaquie), Elsa. Merci à tous de m'avoir aidé à me changer les idées régulièrement, c'était très important. Merci aussi Diyendo et Maud, je suis content d'avoir pu garder le contact avec vous après l'université alors que nous avons tous commencé une thèse au même moment. Merci à mes amis musiciens, tous présents pendant la thèse mais surtout dès mon arrivée à Paris il y a 5 ans pour la plupart. Merci Léo, Sonia, Sarah, Isabelle, Jean-Yves (ton humour décapant !!), Manu violoniste (si curieux de la biologie, c'est toujours un plaisir de te raconter mes recherches), Patricia & Jean-Michel, Claire (promis je ne te parlerai plus de passer ton HDR), Aline & Pierre, Béatrice (ta bonne humeur constante, et ton grog...), Lea (merci pour cette jolie mélodie que tu as composée), Alexis, Julien, Patricia (Mapaúlina), Robin, Sylvain, Camille, Erwan, Pablo et les autres musiciens du Kwaga band. Merci en particulier Robin et Sylvain pour avoir bien pimenté les derniers mois de ma thèse :- ) au sens figuré comme au sens propre (vos plats végétaliens sont formidables).

Merci à toute ma famille, en premier mes parents Marie-Bé et Jacques, ma sœur Juliette (pardon, Juju) et mon frère Arthur. Merci pour votre grand soutien, vous avez toujours été là malgré la distance. Merci à tous les autres : tantes, oncles, cousines et cousins. Merci à mes grands-parents Nicole et René. Vous m'avez tous toujours encouragé, et vous avez supporté mes difficultés à vulgariser un sujet complexe. Merci pour tout. Enfin, merci infiniment Rebecca. Merci pour ton soutien de chaque instant, pour ta bonne humeur, pour ta musique, pour tous ces moments merveilleux déjà passés ensemble (les souterrains de Paris, la côte Ouest des États-Unis, les sentiers de Belle-Île, la Bogue d'Or, les Alpes en plein été, le festival Interceltique, et j'en oublie certainement...), et pour tous ceux à venir.



C'est que malheureusement on ne se doute pas que le livre le plus précieux du plus savant serait celui où il dirait tout ce qu'il ne sait pas, c'est qu'on ne se doute pas qu'un auteur ne nuit jamais tant à ses lecteurs que quand il dissimule une difficulté. Quand la concurrence, c'est-à-dire l'égoïsme, ne règnera plus dans la science, quand on s'associera pour étudier, au lieu d'envoyer aux Académies des paquets cachetés, on s'empressera de publier ses moindres observations pour peu qu'elles soient nouvelles et on ajoutera : « Je ne sais pas le reste. »

– Évariste Galois, Préface aux *Deux mémoires d'Analyse pure*, décembre 1831

The quiet beginnings of telomerase research emphasize the importance of basic, curiosity-driven research. At the time that it is conducted, such research has no apparent practical applications. Our understanding of the way the world works is fragmentary and incomplete, which means that progress does not occur in a simple, direct and linear manner. It is important to connect the unconnected, to make leaps and to take risks, and to have fun talking and playing with ideas that might at first seem outlandish.

– Elizabeth Blackburn, Carol Greider & Jack Szostak (2006) Telomeres and telomerase: the path from maize, *Tetrahymena* and yeast to human cancer and aging. *Nature Medicine* **12**: 1133–1138



## 0 Introduction

### 0.1 Fonctions des télomères

#### 0.1.1 Le support de l'information génétique

Tous les organismes cellulaires utilisent l'ADN comme support de leur information génétique. Grâce aux propriétés de complémentarité des bases nucléiques, cette molécule permet en effet la réplication et donc la transmission de l'information génétique d'une génération à la suivante. Au delà de ce point commun fondamental, on peut observer une très vaste diversité dans l'organisation des génomes (Budd, 2012) : la taille des molécules d'ADN, leur nombre, leur géométrie, leur condensation, le degré de ploïdie, etc. sont autant de propriétés propres à chaque espèce.

Parmi cette diversité d'organisation des génomes, deux groupes principaux se distinguent : d'une part la majorité des procaryotes possède un génome composé d'une ou plusieurs molécules d'ADN circulaires (chromosomes et plasmides), et d'autre part le génome nucléaire de tous les eucaryotes est constitué exclusivement de chromosomes linéaires. Nous nous intéresserons ici aux extrémités des chromosomes linéaires des eucaryotes.

#### 0.1.2 Les télomères

Les extrémités des chromosomes linéaires sont appelées les télomères, du grec τέλος signifiant "fin" et μέρος signifiant "partie". Chez la majorité des organismes qui possèdent des ADN linéaires, les télomères partagent les caractéristiques suivantes :

- ils sont composés de courtes séquences répétées en tandem et riches en guanine (appelées "répétitions télomériques") sur le brin sens 5'-3' (appelé le "brin G"),
- ce brin G forme une extrémité simple-brin sortante en 3',
- ces répétitions télomériques sont liées par des protéines spécifiques,
- enfin, une enzyme appelée télomérase est capable d'allonger le brin G à l'aide d'une matrice d'ARN et de son activité de type rétro-transcriptase.

Ces caractéristiques déterminent ce qui est appelé un "télomère canonique". Quelques organismes possèdent des télomères dits "non canoniques" qui ne partagent pas toutes ces caractéristiques ; le caractère non canonique sera indiqué lorsque nous mentionnerons ces télomères dans la suite de cette introduction.

Les séquences des répétitions télomériques de quelques organismes modèles courants en biologie des télomères sont listées dans le tableau suivant.

**Tableau 1 : Séquences des répétitions télomériques d'organismes modèles, d'après la base de données <http://telomerase.asu.edu> (Podlevsky *et al*, 2008)**

Organisme	Répétition télomérique
<i>Tetrahymena thermophila</i>	TTGGGG
<i>Oxytricha nova</i>	TTTTGGGG
<i>Saccharomyces cerevisiae</i>	T(G) <sub>2-3</sub> (TG) <sub>1-6</sub>
<i>Schizosaccharomyces pombe</i>	G <sub>2-8</sub> TTAC(A)
<i>Drosophila melanogaster</i>	rétrotransposons
vertébrés	TTAGGG

### 0.1.3 Origine des chromosomes linéaires

L'origine des chromosomes linéaires est liée à celle des eucaryotes, puisque le génome nucléaire de tous les eucaryotes est exclusivement constitué de chromosomes linéaires. Néanmoins, la linéarité a aussi émergé indépendamment chez quelques groupes de procaryotes. Les bactéries du genre *Borrelia* ont en effet une majorité de chromosomes et plasmides linéaires, aux télomères très différents des télomères canoniques : leurs molécules d'ADN linéaires sont fermées covalamment par une structure en épingle à cheveux liant les deux brins (Chaconas & Kobryn, 2010). On trouve aussi ce type de télomères dans un plasmide linéaire de la bactérie *Agrobacterium tumefaciens* (Huang *et al*, 2012). Le genre *Streptomyces* possède également des plasmides linéaires, avec des télomères non canoniques encore différents : ils ne sont pas fermés par une épingle à cheveux mais sont associés à une protéine terminale liée covalamment à l'extrémité 5' et permettant d'initier la réplication des extrémités (Chen, 2007).

Les chromosomes linéaires semblent au premier abord défavorables car ils nécessitent des protéines spécialisées pour leur maintenance (les protéines télomériques et la télomérase), et la synthèse de ces protéines représente un coût énergétique important. Il a été proposé que les chromosomes linéaires ont été conservés au cours de l'évolution des eucaryotes car ils rendent possible la méiose, et car la diversité génétique qu'elle engendre confère un avantage qui compense le coût de la maintenance des télomères (Ishikawa & Naito, 1999). La méiose est en effet physiquement plus difficile à réaliser avec des chromosomes circulaires : ces chromosomes ne peuvent pas regrouper leurs extrémités pour faciliter leur appariement (puisque'ils n'ont pas d'extrémités), et dans le cas où un appariement aurait lieu par simple rencontre aléatoire, ils doivent nécessairement subir un nombre pair d'événements de recombinaison pour ségréger correctement dans les cellules filles (Ishikawa & Naito, 1999). Ces deux événements ont des probabilités très faibles, et pour qu'une cellule possédant plusieurs chromosomes circulaires réalise une méiose avec succès, il faut que ces événements se produisent pour chaque paire de chromosomes : la

probabilité totale qu'une méiose termine correctement avec des chromosomes circulaires est donc infime (Ishikawa & Naito, 1999). D'autre part, le rôle important et conservé des télomères pour le bon appariement des chromosomes homologues lors de la méiose est bien reconnu (Hiraoka, 1998; Scherthan, 2007; Siderakis & Tarsounas, 2007).

La conservation des chromosomes linéaires chez les eucaryotes est une évidence et trouve une explication convaincante avec leur rôle dans la méiose et l'importance de cette dernière, puisqu'elle génère une diversité génétique favorable. En revanche expliquer leur apparition est plus difficile. Une hypothèse proposée (Garavís *et al*, 2013) est que les chromosomes linéaires pourraient être apparus lors du premier événement d'endosymbiose (celui qui a conduit à l'apparition des mitochondries par endocytose d'un organisme d'origine bactérienne) : le génome de l'ancêtre des eucaryotes aurait été envahi par des éléments mobiles venant du génome du symbiote, et ces éléments auraient été "capturés" à plusieurs reprises aux extrémités générées par des cassures double-brin dans l'ADN (auparavant circulaire). Ils auraient ainsi permis de stabiliser le génome sous une forme linéaire suite à ces cassures non réparées, en engendrant des répétitions télomériques primitives. Ces répétitions télomériques auraient favorisé le maintien de la longueur des télomères par un mode de réplication dépendant de la recombinaison homologue (la machinerie de recombinaison homologue existait déjà chez les bactéries avant l'apparition des eucaryotes). La télomérase aurait pu émerger par la suite en évoluant à partir de rétro-transposons (Garavís *et al*, 2013). Cette hypothèse de "domestication" d'éléments mobiles pour survivre suite à une linéarisation accidentelle est aussi formulée dans l'autre sens, à savoir qu'un élément mobile "égoïste" aurait forcé la linéarisation d'un génome ancestral circulaire en s'y intégrant (Nosek *et al*, 2006). Cet élément aurait permis la maintenance des extrémités de l'ADN par la réplication en cercle roulant, jusqu'à être supplanté par la télomérase ou son ancêtre (Nosek *et al*, 2006).

La linéarité des chromosomes confère un avantage en rendant possible la méiose, mais elle pose aussi deux problèmes principaux : la réplication des extrémités, et la protection des extrémités contre la dégradation par des nucléases et contre les fusions. Dans les paragraphes suivants nous expliquerons les principales causes de ces deux problèmes.

#### **0.1.4 Le problème de la réplication des extrémités**

##### **0.1.4.1 Aperçu du problème de la réplication**

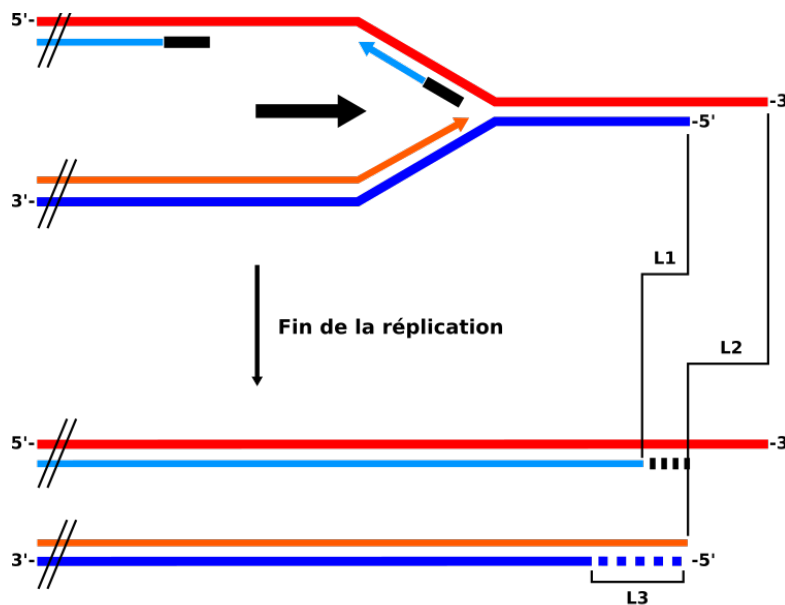
La réplication nécessite une machinerie complexe que nous ne détaillerons pas ici. Pour comprendre le problème de la réplication des extrémités, il suffit de connaître les propriétés générales des ADN polymérases. Ces enzymes utilisent un brin d'ADN matrice qu'elles lisent dans le sens 3'-5' pour synthétiser un nouveau brin complémentaire de 5' vers 3'. En plus du brin matrice, elles nécessitent une amorce exposant une extrémité 3'-OH



libre. Ces amorces sont de petits brins d'ARN d'une dizaine de nucléotides, synthétisés par une primase, et appelés fragments d'Okazaki.

Lors de la réplication, le brin orienté 3'-5' de la molécule d'ADN initiale sert de matrice pour la synthèse du nouveau brin dit "précoce" (aussi appelé le "brin sens", car sa synthèse ne nécessite qu'une seule amorce (l'ADN polymérase peut en effet parcourir tout le brin une fois la synthèse initiée), et elle est en avance par rapport à celle de l'autre nouveau brin qui est appelé par conséquent le brin "tardif" (aussi appelé le "brin anti-sens"). La synthèse du brin tardif nécessite plusieurs amorces car le brin servant de matrice est orienté de 5' vers 3' et la polymérase chargée de la synthèse de ce brin doit progresser dans le sens opposé au sens de progression de la fourche de réplication.

La réplication des extrémités d'un chromosome est présentée dans la figure suivante :



**Figure 1 : Le problème de la réplication des extrémités (inspiré de Gilson & Géli, 2007; Matot, 2010). Les brins d'ADN d'origine sont représentés en traits épais rouge et bleu foncé et leur orientation 5'-3' est indiquée. Les brins néo-synthétisés sont représentés en traits plus fins orange (brin précoce) et bleu clair (brin tardif), avec une flèche indiquant le sens de progression de la polymérase. Les fragments d'Okazaki sont représentés en traits noirs épais. La flèche noire épaisse indique le sens de progression de la fourche de réplication. L'emplacement du dernier fragment d'Okazaki est indiqué en pointillés noirs, la résection de l'extrémité 5' est représentée en pointillés bleus.**

Une première cause de raccourcissement vient du fait que le dernier fragment d'Okazaki est dégradé à la fin de la réplication, mais n'est pas remplacé par de l'ADN. Ceci a pour conséquence le raccourcissement du brin tardif (longueur L1 dans la figure).

Une deuxième cause du raccourcissement est due à l'extrémité simple-brin sortante en 3' : puisqu'il n'existe pas de matrice pour cette extension simple-brin, la molécule d'ADN issue de la synthèse du brin précoce se trouve raccourcie de la longueur de l'extension simple-brin (longueur L2 dans la figure).

Une troisième cause du raccourcissement vient de la résection par des exonucléases de l'extrémité 5' (longueur L3 dans la figure) du télomère issu de la synthèse du brin précoce (Gilson & Géli, 2007). Cette résection permet de rétablir une extrémité simple-brin sortante en 3'.

Ce raccourcissement des extrémités à chaque réplication est à l'origine de la sénescence observée dans une culture *in vitro* de cellules somatiques humaines après une cinquantaine de doublements (Hayflick, 1965). Cette limitation du nombre de mitoses avant la sénescence est appelée la "limite de Hayflick", et constitue un mécanisme suppresseur de tumeur important. Les seules cellules qui y échappent sont les lignées germinales et certaines cellules tumorales hautement prolifératives, ainsi que les organismes unicellulaires.

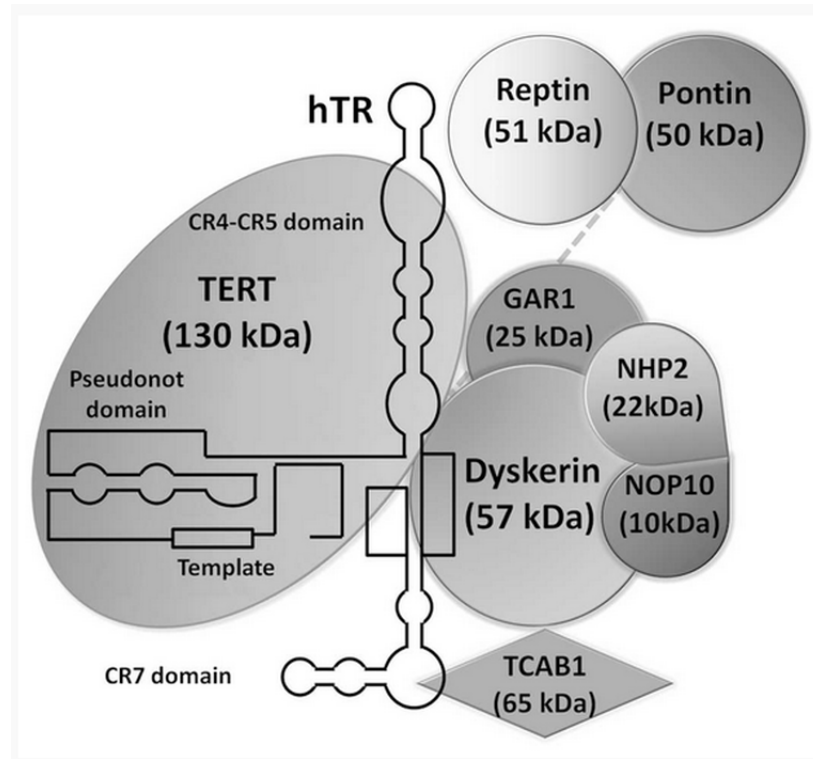
Deux mécanismes principaux compensent l'érosion des télomères et permettent donc de contourner le problème de la réplication : l'activité de la télomérase, et une voie alternative appelée ALT (*alternative lengthening of telomeres*) utilisant la recombinaison homologue. Il est intéressant de remarquer ici la diversité des solutions au problème de la réplication qui ont émergé au cours de l'évolution, comme chez la drosophile qui a perdu la télomérase ou chez les quelques procaryotes qui possèdent des chromosomes linéaires mais ne possèdent pas de télomérase. Ces systèmes de réplication des télomères sont présentés dans les paragraphes suivants.

#### **0.1.4.2 L'addition de répétitions télomériques par la télomérase**

La télomérase est l'enzyme qui allonge le brin G des télomères. Sa découverte chez l'organisme modèle *Tetrahymena thermophila* (Greider & Blackburn, 1985) ainsi que les autres travaux initiaux sur la biologie des télomères ont valu un prix Nobel en 2009 à Elizabeth Blackburn, Carol Greider et Jack Szostak.

La télomérase fonctionnelle est un complexe ribonucléoprotéique (Greider & Blackburn, 1987) composé d'une sous-unité protéique appelée TERT (*telomerase reverse transcriptase*), d'une sous-unité d'ARN appelée TR (*telomerase RNA*) ou TERC (*telomerase*

RNA component) et de protéines accessoires. La télomérase humaine est représentée schématiquement dans la figure suivante :



**Figure 2 : Représentation schématique de la télomérase humaine (Gomez, 2012)**

TERT sert de matrice pour la synthèse des répétitions télomériques par TERC. La première structure tridimensionnelle d'une sous-unité TERT entière a montré que cette enzyme a un repliement similaire à celui de rétro-transcriptases virales telle que celle du VIH (Gillis *et al*, 2008). La structure de la même enzyme en complexe avec un hybride ADN/ARN mimant le produit de synthèse a fourni des informations sur le mécanisme de synthèse des répétitions télomériques (Mitchell *et al*, 2010). Ce mécanisme de synthèse est représenté schématiquement dans la figure suivante :

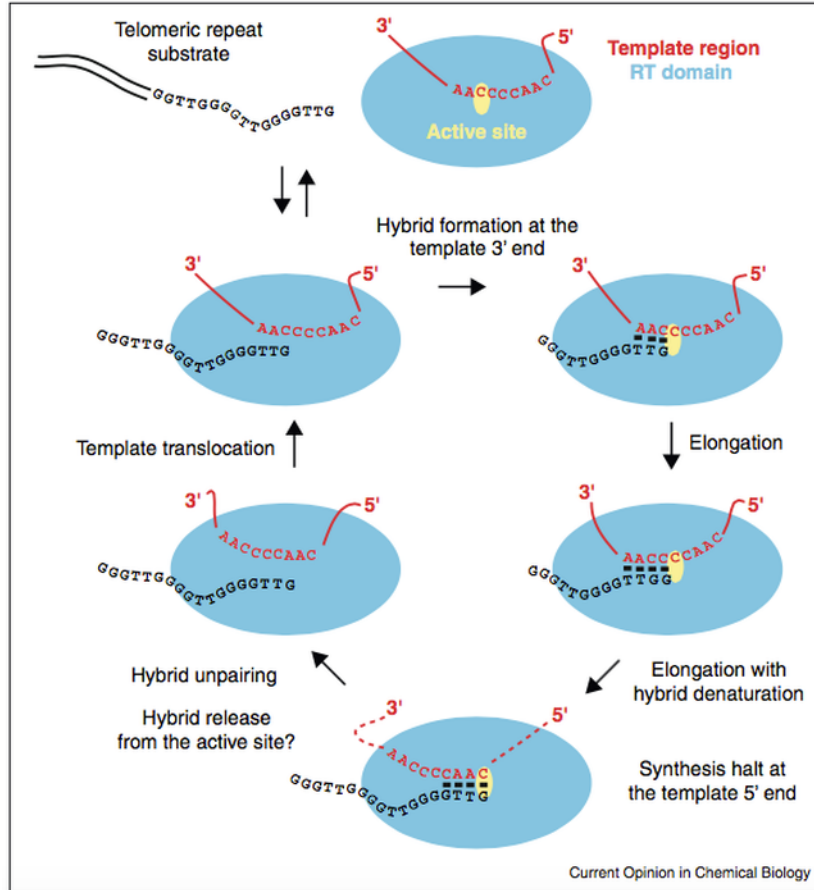


Figure 3 : Synthèse des répétitions télomériques par la télomérase (Collins, 2011)

L'enzyme apparie la matrice d'ARN à l'extrémité simple-brin du télomère et synthétise les répétitions télomériques jusqu'à la fin de la séquence matrice, tout en dissociant les deux brins au fil de la progression de la synthèse. Lorsque la fin de la séquence matrice est atteinte, les brins se dissocient et l'extrémité nouvellement synthétisée est éventuellement réappariée au début de la matrice pour démarrer la synthèse d'une nouvelle répétition. Contrairement aux autres rétro-transcriptases qui circulent le long de leur ARN matrice, TERT n'utilise qu'une petite région de TERC comme matrice, et cette dernière est maintenue à proximité du site actif. Ce mécanisme de synthèse répétition par répétition avec une étape de translocation du brin synthétisé et de la matrice entre chaque synthèse peut être modulé par des partenaires de la télomérase, la rendant plus ou moins processive en termes de répétitions (cette propriété unique de la télomérase est appelée *repeat addition processivity*).

L'activité de la télomérase est régulée globalement par le niveau d'expression de TERT (Cong *et al*, 2002, et références associées). L'expression de cette sous-unité s'arrête dans les cellules somatiques au cours du développement embryonnaire, et sa ré-expression suffit à conférer aux cellules un potentiel de prolifération illimité. L'expression de la télomérase est détectée dans environ 90 % des cancers (Cong *et al*, 2002, et références associées). Les cellules cancéreuses chez lesquelles l'expression de TERT n'est pas détectée utilisent une voie alternative d'élongation des télomères.

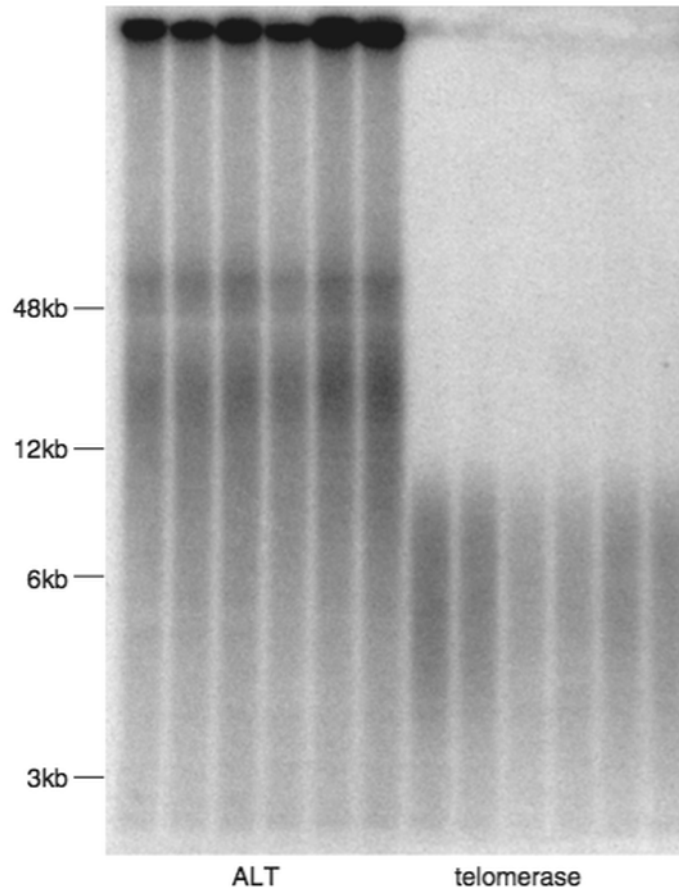
#### 0.1.4.3 La voie alternative d'élongation des télomères

Cette voie alternative d'élongation des télomères (appelée ALT pour *alternative lengthening of telomeres*) a été découverte pour la première fois dans des cellules tumorales qui n'exprimaient pas la télomérase (Bryan *et al*, 1995, 1997). On estime que 10 à 15 % des tumeurs maintiennent leurs télomères par cette voie indépendante de la télomérase (Pickett & Reddel, 2009). Chez les organismes pluricellulaires la voie ALT semble être une caractéristique unique de ces tumeurs, même si un phénomène similaire peut être détecté chez la levure *Saccharomyces cerevisiae*. Dans des souches chez lesquelles la télomérase est inactivée la majorité des levures meurent, mais en prolongeant la culture on peut tout de même isoler des "survivants", et leur survie dépend du bon fonctionnement de la recombinaison homologue (Lundblad & Blackburn, 1993).

Les télomères maintenus par la voie ALT ont une longueur moyenne plus grande que ceux maintenus par la télomérase (respectivement entre 12 et 48 kb, et autour de 6 kb), et présentent une plus grande hétérogénéité de longueur, comme le montre la méthode de *terminal restriction fragment length analysis* (Kimura *et al*, 2010). Cette méthode de mesure de la longueur des télomères est la plus couramment utilisée dans la littérature, et s'exécute avec les étapes suivantes : extraction et purification de l'ADN génomique, digestion par des enzymes de restriction, séparation des fragments de restriction par électrophorèse, transfert sur une membrane, hybridation d'une sonde complémentaire des répétitions

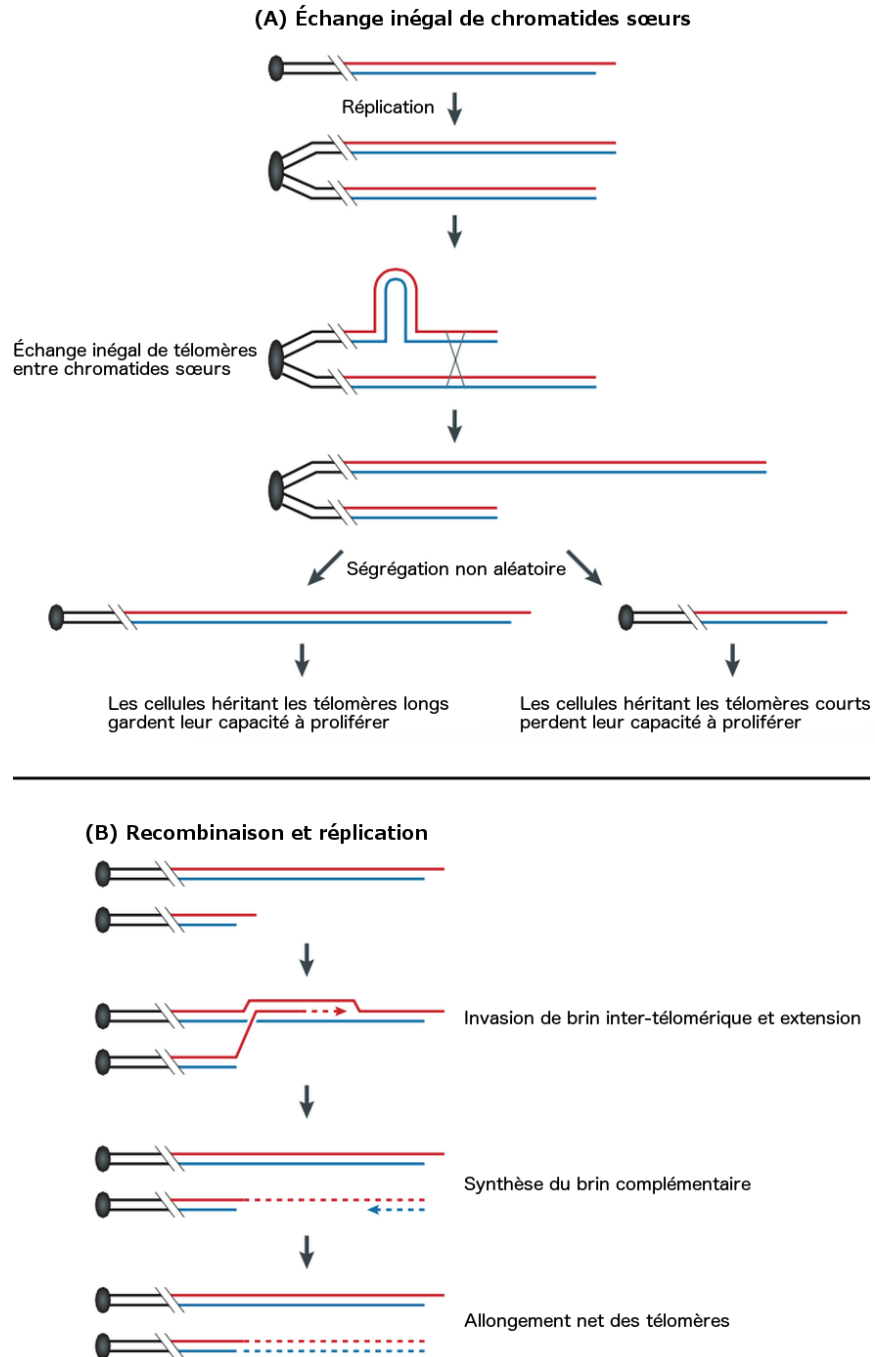
téломériques et révélation. Elle permet de mesurer la longueur moyenne des télomères dans un échantillon de cellules et la dispersion autour de cette longueur moyenne (ce qui correspond à l'hétérogénéité de longueur des télomères au sein de l'échantillon).

**Fig.** Telomere length heterogeneity in ALT cells. Terminal restriction fragment (TRF) length analysis of an ALT cell line (six lanes derived from the same individual) and a telomerase-positive cell line (six lanes derived from the same individual), demonstrating the characteristic telomere lengths of the two telomere maintenance mechanisms. TRF analysis was carried out by digesting genomic DNA with restriction enzymes that do not recognise the telomeric sequence, separating by pulsed field gel electrophoresis and hybridising with a radioactively labelled probe complementary to the telomeric sequence



**Figure 4 : Longueur moyenne des télomères maintenus par la voie ALT, à gauche, et par la télomérase, à droite (Pickett & Reddel, 2009). Les télomères maintenus par la télomérase ont une longueur plus homogène que ceux maintenus par la voie ALT.**

Les modèles d'élongation des télomères par la voie ALT actuellement proposés sont schématisés dans la figure suivante :



**Figure 5 : Modèles actuels de l'élongation des télomères par la voie ALT (adapté de Cesare & Reddel, 2010)**

Le premier modèle fait intervenir un échange inégal de chromatides sœurs (A) : en phase G2 les chromosomes sont entièrement répliqués et des événements de recombinaison peuvent se produire entre chromatides sœurs (indiqués par la croix entre les deux chromatides). Des jonctions de Holliday sont formées entre les deux chromatides, et leur résolution conduit à des échanges de portions d'ADN de longueurs différentes. La cellule fille qui reçoit les chromosomes aux télomères longs conserve donc une grande capacité à proliférer, contrairement à celle qui hérite les télomères courts.

Dans le deuxième modèle, dit de réplication dépendante de la recombinaison (B), l'extrémité simple-brin sortante en 3' d'un télomère envahit la région double-brin d'un autre télomère en formant une jonction de Holliday, et utilise cet autre télomère comme matrice pour la machinerie de réplication. Le brin complémentaire est ensuite complété pour reconstituer la région double-brin du télomère. Cette voie conduit à un allongement net des télomères.

### 0.1.5 Autres solutions évolutives au problème de la réplication

Outre l'élongation par la télomérase et par la voie ALT, d'autres mécanismes de réplication des télomères et de maintien de leur longueur sont apparus au cours de l'évolution. Dans les paragraphes suivants, nous présentons trois mécanismes originaux associés à des télomères non canoniques.

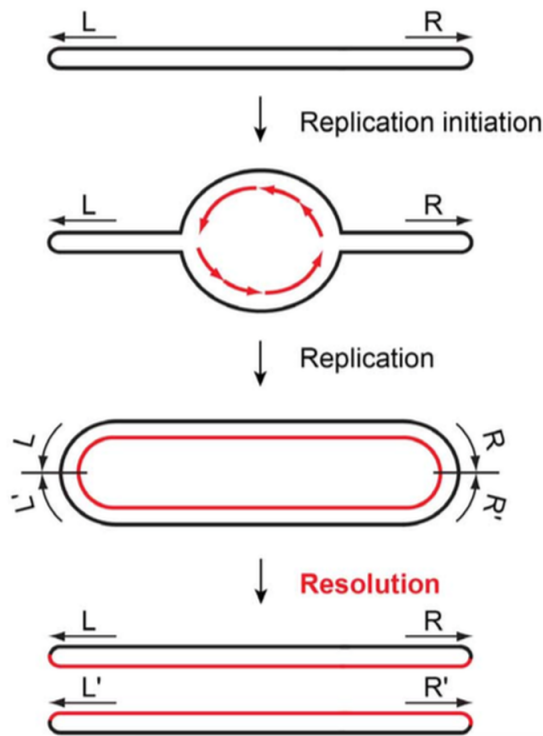
#### 0.1.5.1 Les rétro-transposons de *Drosophila*

*Drosophila melanogaster* ne possède pas de télomérase. Chez cet organisme, les télomères sont maintenus par l'action de rétrotransposons (Raffa *et al*, 2013). Trois rétrotransposons sont trouvés aux télomères : HeT-A, TART et TAHRE. Ils contiennent tous un gène *gag* codant pour des protéines prenant en charge le transcrit. TAHRE et TART contiennent aussi un gène *pol* codant pour une rétro-transcriptase. Grâce à cette enzyme, ces rétro-transposons synthétisent au site d'insertion (à l'extrémité du chromosome) l'ADN complémentaire de leur transcrit. La régulation de la longueur des télomères est ainsi principalement contrôlée par le taux de transcription de ces rétrotransposons. Elle est aussi en partie régulée par l'hétérodimère Ku70/Ku80 qui occupe potentiellement l'extrémité du chromosome et la rend inaccessible à la machinerie d'insertion des transposons. Bien qu'il ne code pas sa propre rétro-transcriptase, ce qui le rend dépendant des autres rétrotransposons, HeT-A est le plus abondant aux télomères en termes de nombre de copies. Ceci peut être dû au fait que sa protéine Gag est la seule capable de s'associer spécifiquement aux télomères, et permet de recruter les protéines Gag de TART et TAHRE. Ces trois rétrotransposons se transposent à une fréquence de  $10^{-2}$  à  $10^{-4}$  par génération, tandis que l'érosion due au problème de la réplication est d'environ 70 pb par génération en moyenne. La stabilité de la longueur des télomères résulte donc d'un équilibre dynamique entre ces deux phénomènes.



### 0.1.5.2 Les chromosomes fermés coalementment de *Borrelia*

Les bactéries du genre *Borrelia* (dont *Borrelia burgdorferi*, l'agent pathogène de la maladie de Lyme) possèdent des ADN linéaires et contournent le problème de la réplication des extrémités d'une façon bien plus simple. Les extrémités sont fermées coalementment par une épingle à cheveux (*closed hairpin telomeres*), et une enzyme appelée résolvasse de télomères (parfois appelée protélomérase bien qu'elle n'ait rien en commun avec la télomérase des eucaryotes ; voir Chaconas & Kobryn, 2010) permet de séparer les deux molécules filles, concaténées en un dimère circulaire suite à la réplication (Chaconas & Kobryn, 2010). La figure suivante illustre le mode de réplication des ADN linéaires à télomères en épingles à cheveux avec une initiation centrale et une étape de résolution des télomères qui reconstitue deux ADN linéaires fermés coalementment par des épingles à cheveux.

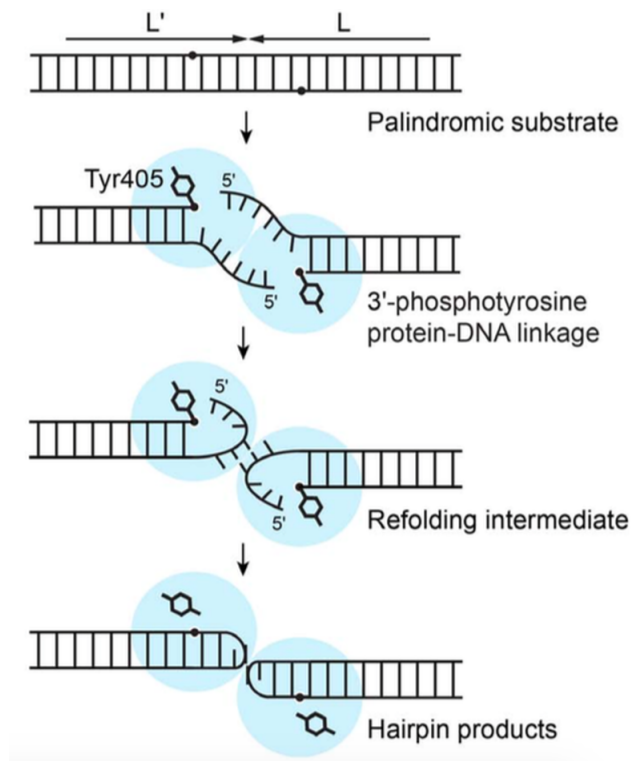


**Figure 6 : Principe de la réplication des ADN linéaires à télomères en épingle à cheveux (Shi *et al*, 2013)**

L'enzyme catalysant la résolution du dimère, appelée ResT (résolvase de télomères), a été purifiée et son activité a été caractérisée *in vitro* (Kobryn & Chaconas, 2002). L'enzyme seule, sans partenaire protéique ni cations divalents ni ATP, est capable de reconstituer *in vitro* la réaction de résolution. Les auteurs de cette étude ont pu montrer que l'enzyme agit selon un mécanisme similaire à celui des topoisomérases de type IB et des recombinaisons

site-spécifiques : une tyrosine catalytique clive un des brins d'ADN et forme un intermédiaire covalent avec l'extrémité 3'-phosphate, et l'enzyme est finalement libérée par la formation d'une nouvelle liaison 3'-5' phosphodiester.

La structure tridimensionnelle d'une résolvasse de télomères a été publiée pour la première fois en 2007 avec l'enzyme TelK d'un bactériophage (Aihara *et al*, 2007). Cette structure a permis de confirmer le mécanisme proposé d'après les expériences *in vitro*. La seule autre résolvasse de télomères dont une structure tridimensionnelle est publiée vient de la bactérie *Agrobacterium tumefaciens* (Shi *et al*, 2013). Cette enzyme, appelée TelaA, a été mise en évidence plus récemment (Huang *et al*, 2012). Dans les deux cas, l'enzyme se fixe sous la forme d'un dimère à la séquence télomérique répliquée, et chaque monomère clive un brin de l'ADN en attaquant un groupement phosphate avec sa tyrosine catalytique. La conformation de la protéine force ensuite la formation d'une épingle à cheveux par la courte extrémité 5'-sortante générée par le clivage. Cette extrémité 5'-OH attaque le groupement phosphate lié à la tyrosine catalytique pour reformer une liaison phosphodiester et libérer l'enzyme. La figure suivante illustre le mécanisme des résolvases de télomères.



**Figure 7 : Mécanisme des résolvases de télomères (Shi *et al*, 2013)**

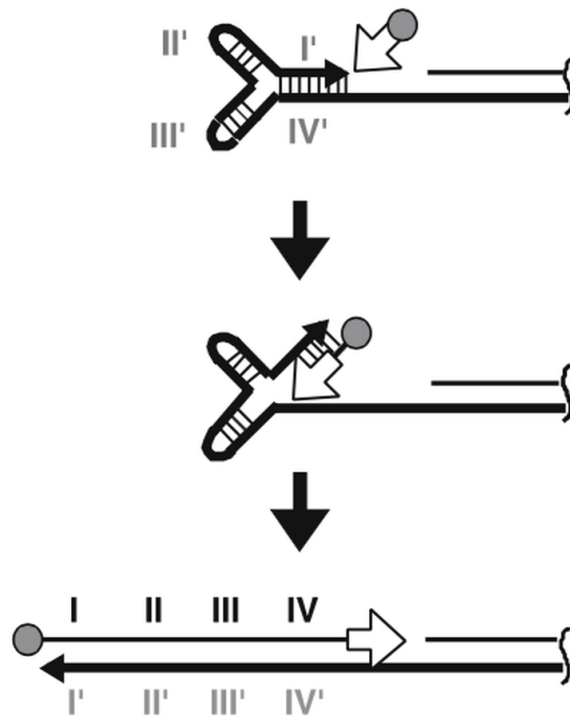
Il a par la suite été démontré que la seule acquisition d'une résolvasse de télomères suffit à linéariser un génome normalement circulaire comme celui d'*E. coli* sans altérer la viabilité de la bactérie (Cui *et al*, 2007). Ceci indique que la résolvasse de télomères est un facteur nécessaire et suffisant pour contourner le problème de la réplication des extrémités des ADN linéaires.

### 0.1.5.3 La protéine terminale de *Streptomyces*

Les bactéries du genre *Streptomyces* (qui produisent de nombreuses molécules utilisées comme antibiotiques) possèdent des ADN linéaires portant une protéine appelée TP (*terminal protein*) liée covalamment à chaque extrémité 5' et servant d'amorce pour terminer la synthèse du brin tardif à la fin de la réplication, un processus appelé *end patching*, soit "colmatage" des extrémités (Chen, 2007).

Les séquences télomériques de ces bactéries sont composées de six palindromes couvrant environ 140 pb. Lors de la réplication, la synthèse incomplète du brin tardif laisse une extrémité simple-brin sortante en 3' pouvant atteindre jusqu'à environ 300 nucléotides de longueur. Les palindromes permettent alors le repliement de cette extrémité simple-brin en structures secondaires, formant une dizaine de tiges-boucles. L'extrémité 3' se trouve appariée et peut servir d'amorce pour la synthèse de la fin du brin tardif. Cette synthèse est initiée par la coopération entre une ADN polymérase et TP activée par un nucléotide (un dCMP lié covalamment à une thréonine de la protéine). La protéine TP reste associée à l'extrémité 5' une fois la réplication terminée, et permet de protéger l'extrémité (Chen, 2007).

Aucune structure tridimensionnelle de protéine TP n'a encore été publiée à ce jour, ainsi le mécanisme de synthèse de la fin du brin tardif n'est pas encore complètement élucidé. Le mécanisme le plus probable est présenté dans la figure suivante.



**Figure 8 : Mécanisme de synthèse de la fin du brin tardif chez *Streptomyces* (Chen, 2007). La protéine TP est représentée par le disque gris.**

Dans ce système, comme dans celui des *Borrelia*, la longueur des chromosomes et des plasmides linéaires reste constante au fil des réplifications car ces organismes utilisent à la place des mécanismes qui évitent totalement (chez les *Borrelia*) ou compensent exactement (chez les *Streptomyces*) l'érosion des extrémités. Il est intéressant de remarquer que les systèmes de réplication des télomères des eucaryotes génèrent nécessairement une certaine hétérogénéité de longueur, contrairement à ces systèmes évolutivement distants.

### 0.1.6 Le problème de la protection des extrémités

Les extrémités des ADN linéaires sont vulnérables à différents processus cellulaires : la dégradation par des exonucléases, l'activation illicite de la voie de signalisation des dommages à l'ADN, les fusions d'extrémités et les événements de recombinaison.

La digestion par des exonucléases est un phénomène nécessaire pour établir un télomère fonctionnel chez la plupart des organismes, pour générer l'extrémité simple-brin sortante en 3'. Elle doit par conséquent être parfaitement régulée par les protéines télomériques.

L'activation de la voie de signalisation des dommages à l'ADN (DDR pour *DNA damage response*) est initiée par deux kinases appelées ATM (*ataxia telangiectasia mutated*) et ATR

(*ataxia telangiectasia and Rad3 related*). Le complexe MRN (MRE11/RAD50/NBS1) détecte les cassures double-brin et recrute ATM qui initie la voie DDR. De façon complémentaire, c'est la présence d'ADN simple-brin exposé qui active la signalisation par ATR, recrutée par la protéine RPA. Ces deux kinases phosphorylent de nombreuses protéines, ce qui a pour conséquences principales l'arrêt du cycle cellulaire et l'activation des voies de réparation appropriées (voir les revues Palm & de Lange, 2008; Lamarche *et al*, 2010; Gobbin *et al*, 2013). L'activation de la voie DDR par ces deux kinases est détectable par la formation de foyers enrichis en protéines cibles de ces kinases (comme par exemple l'histone H2AX phosphorylé sur la sérine 139), visibles en microscopie après immunomarquage fluorescent. Le double marquage fluorescent de ces foyers de dommages à l'ADN et des télomères permet de mettre en évidence l'activation illicite de la voie DDR aux extrémités des chromosomes. Ces foyers de dommages à l'ADN localisés aux télomères sont alors appelés des TIFs (*telomere dysfunction induced foci* ; voir Takai *et al*, 2003).

Les fusions d'extrémités sont particulièrement délétères car elles conduisent à des défauts de ségrégation des chromosomes lors de la mitose ou de la méiose. Par conséquent elles sont à l'origine de caryotypes aberrants chez les cellules issues de la division, si celle-ci se termine, ou bien provoquent un arrêt du cycle cellulaire et potentiellement l'apoptose. Les fusions de chromosomes sont détectables dans des préparations de chromosomes métaphasiques dans lesquelles l'ADN est coloré par le DAPI (ou un autre intercalant), et les séquences télomériques sont marquées par la technique de FISH (*fluorescence in situ hybridization*) avec une sonde fluorescente complémentaire des répétitions télomériques : un signal télomérique situé dans une région interne du chromosome trahit une fusion (voir Celli & de Lange, 2005). Les fusions de chromosomes indiquent que les télomères ont été pris en charge "illégalement" par la voie de réparation des cassures double-brin NHEJ (*non-homologous end joining* : jonction d'extrémités non homologues), normalement réprimée aux télomères.

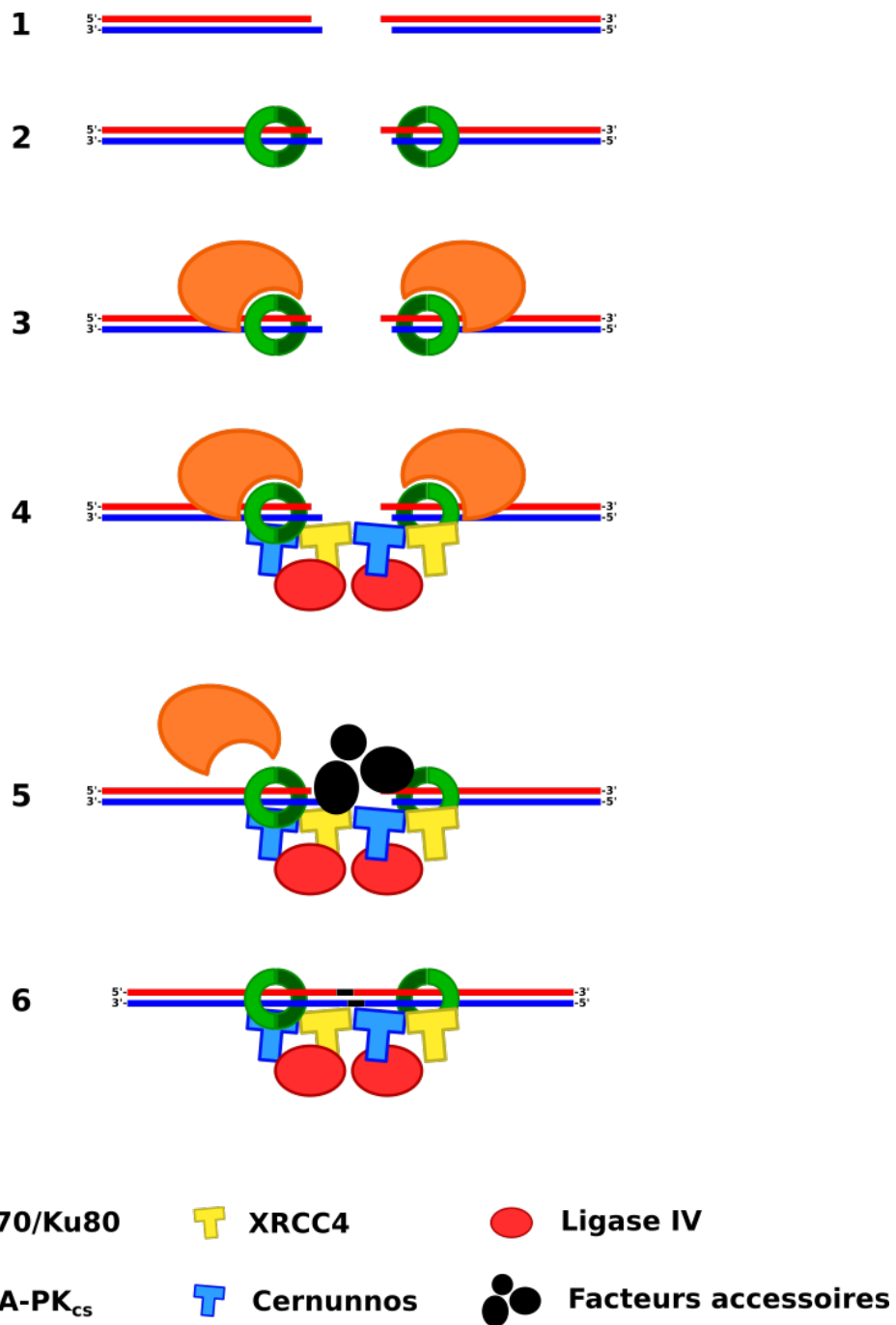
Enfin, les événements de recombinaison sont aussi délétères car ils conduisent à des variations brutales de la longueur des télomères, qui sont associées à une prolifération tumorale facilitée dans le cas d'un allongement (voie ALT) ou à une sénescence précoce dans le cas d'un raccourcissement. Ces événements peuvent être mis en évidence par une technique appelée CO-FISH (*chromosome orientation - fluorescence in situ hybridization* ; Williams & Bailey, 2009). Les cellules sont cultivées le temps d'un doublement en présence d'un nucléotide modifié, incorporé dans l'ADN par la réplication. Une préparation de chromosomes métaphasiques est réalisée, et les brins néo-synthétisés sont dégradés sélectivement grâce à la présence du nucléotide modifié. Les télomères sont enfin révélés avec deux sondes fluorescentes (de couleurs différentes) complémentaires des répétitions télomériques de chacun des deux brins. Les échanges de télomères entre chromatides sœurs (T-SCE pour *telomere sister chromatid exchange*) indiquent que les télomères subissent des événements de recombinaison.

La voie de signalisation DDR, le NHEJ et la recombinaison homologue sont réprimés par les télomères fonctionnels, et se manifestent aux télomères présentant un défaut de protection.

Les mécanismes du NHEJ et de la recombinaison homologue sont présentés brièvement dans les paragraphes suivants.

#### **0.1.6.1 Jonction d'extrémités non-homologues (NHEJ)**

Le déroulement de la réparation d'une cassure double-brin par la voie NHEJ (chez les mammifères) est représenté schématiquement sur la figure suivante :



**Figure 9 : La voie NHEJ de réparation des cassures double-brin de l'ADN (inspiré de Grundy *et al*, 2014)**

Une cassure double-brin (1) est reconnue par l'hétérodimère Ku70/Ku80 (2), qui recrute la kinase DNA-PK<sub>cs</sub> (3) permettant d'initier la voie de signalisation de réponse aux dommages de l'ADN (DDR, pour *DNA damage response*). Le recrutement des protéines XRCC4,

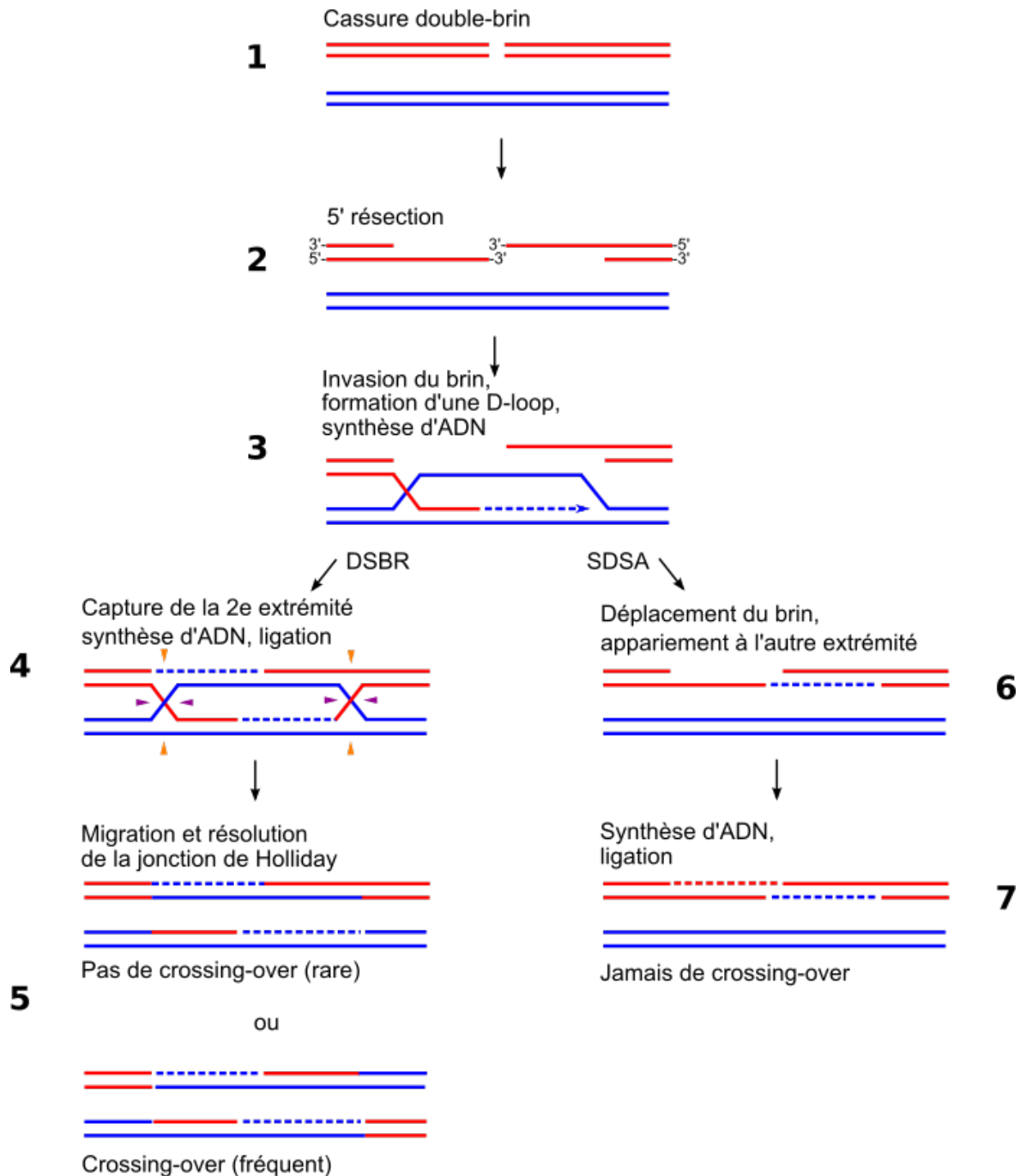
Cernunnos et Ligase 4 forme le complexe basal du NHEJ et permet la synapse des deux extrémités générées par la cassure (4). Les extrémités sont alors “nettoyées” par divers facteurs accessoires dont des polymérases et des exonucléases (5) jusqu’à ce que des extrémités cohésives soient générées. La ligature s’opère enfin sur ces extrémités cohésives (6) (Grundy *et al*, 2014).

La machinerie du NHEJ est capable de lier deux extrémités d’ADN sans nécessiter une molécule de séquence similaire pour servir de matrice : pour cette raison cette voie est utilisée pendant la phase G1, au cours de laquelle les chromosomes ne sont pas encore répliqués (il n’existe donc pas encore de chromatide sœur pouvant servir de matrice pour la réparation). Du fait de l’absence de matrice pour guider la réparation, cette voie est potentiellement mutagène.

#### **0.1.6.2 Recombinaison homologue**

L’autre voie de réparation des cassures double-brin est la recombinaison homologue (San Filippo *et al*, 2008). Cette voie est utilisée principalement pendant les phases S et G2, quand une chromatide sœur existe et peut donc servir de matrice pour guider la réparation. Contrairement au NHEJ, la recombinaison homologue produit des réparations très fidèles à la séquence d’origine. Le principe de la recombinaison homologue est illustré dans la figure suivante :





**Figure 10 : La recombinaison homologue (inspiré de San Filippo *et al*, 2008, et adapté à partir d'une figure originale de Ewm sur Wikimedia Commons, CC BY-SA 3.0 : [http://commons.wikimedia.org/wiki/File%3AHR\\_schematic\\_diagram.svg](http://commons.wikimedia.org/wiki/File%3AHR_schematic_diagram.svg))**

Suite à une cassure double-brin (1), les extrémités 5' sont digérées par des exonucléases qui génèrent une extrémité simple-brin sortante en 3' (2). Cette extrémité est liée par la protéine RAD51 qui polymérise et forme un filament, permettant la recherche d'une région homologe dans une autre molécule d'ADN. Lorsque cette région est trouvée, l'extrémité

simple-brin envahit la molécule d'ADN homologue en formant une jonction de Holliday (structure branchée à 4 brins), s'apparie au brin qui lui est complémentaire, et son groupement 3'-OH sert d'amorce pour la synthèse d'ADN (3). À partir de cette étape, deux voies principales sont possibles :

- La voie DSBR (*double-strand break repair* : réparation de la cassure double-brin) implique la capture de la deuxième extrémité générée par la cassure, son appariement au niveau de la *D-loop* (*displacement loop* : le brin d'ADN déplacé suite à l'invasion de l'étape précédente) et la synthèse d'ADN. Lorsque la synthèse a terminé de compléter les deux brins, une ligature est opérée (4) et produit un intermédiaire à deux jonctions de Holliday qui est finalement résolu (5), le plus souvent avec un échange de chromatides (*crossover product*).
- La voie SDSA (*synthesis-dependant strand annealing* : appariement de brins après synthèse d'ADN) débute avec le désappariement du brin qui avait envahi la molécule homologue. Ce brin, étendu après la synthèse, est capable de s'apparier à l'autre extrémité de la cassure double-brin (6). La réparation est finalement complétée par la synthèse du brin complémentaire et une ligature des deux brins (7). Cette voie ne provoque jamais d'échange de chromatides.

La recombinaison homologue fait intervenir la protéine RAD51 et de nombreux facteurs associés comme le complexe MRN. Comme dans la voie ALT, la jonction de Holliday est un intermédiaire omniprésent dans les mécanismes de la recombinaison homologue.

### 0.1.7 Solutions moléculaires au problème de la protection apparues au cours de l'évolution

Comme nous l'avons vu dans les paragraphes précédents, les molécules d'ADN linéaires sont confrontées aux problèmes de la réplication et de la protection des extrémités. Plusieurs stratégies pour contourner ces problèmes ont émergé au cours de l'évolution, et nous avons déjà présenté quelques solutions au problème de la réplication.

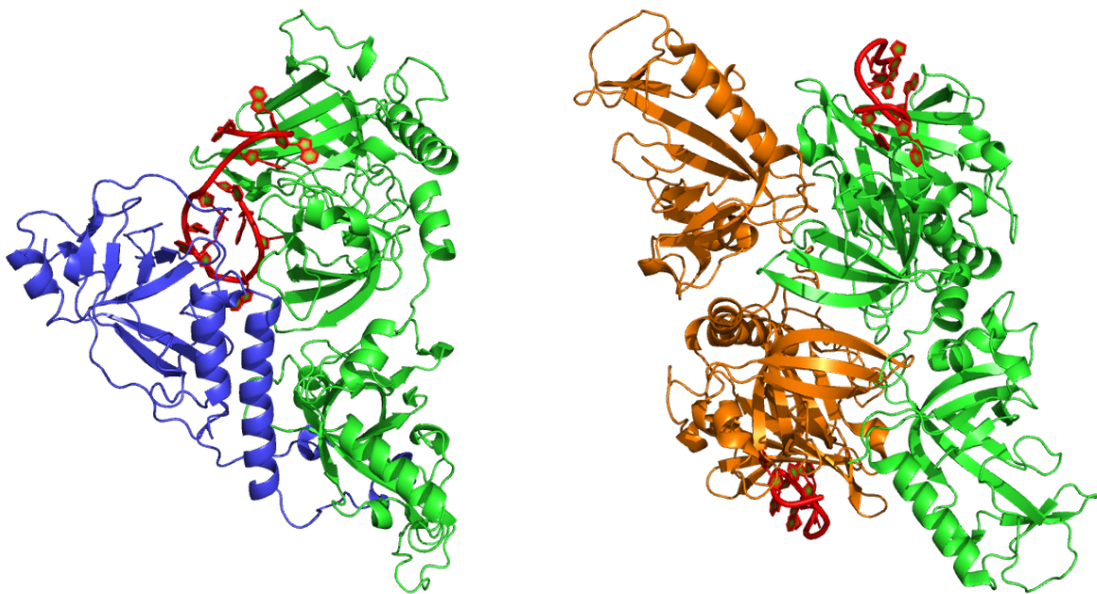
Le problème de la protection des extrémités est résolu par l'existence de protéines spécifiques des télomères, qui s'associent aux répétitions télomériques et assurent la répression de la voie DDR, du NHEJ et de la recombinaison homologue. Dans les paragraphes qui suivent nous présenterons brièvement l'organisation des complexes télomériques de quelques organismes modèles, avec un exemple de télomères non canoniques, et en précisant comment l'étude de ces complexes a contribué à établir les connaissances actuelles sur les télomères en général.

#### 0.1.7.1 Chez les ciliés

Les premières protéines télomériques ont été mises en évidence chez l'unicellulaire cilié *Oxytricha nova* (Gottschling & Zakian, 1986). Chez cet organisme les répétitions

télomériques simple-brin de l'extrémité sortante en 3' sont liées spécifiquement par deux protéines appelées TEBP- $\alpha$  et TEBP- $\beta$  (*telomere end binding proteins*), comprenant chacune deux domaines *OB-fold* (*oligonucleotide-binding fold*) et qui s'assemblent en un hétérodimère (Fang & Cech, 1993). La structure cristallographique du complexe TEBP- $\alpha$ /TEBP- $\beta$ /ADNsb (Horvath *et al*, 1998) montre que l'extrémité 3' de l'ADN est profondément enfouie dans le complexe (figure suivante), ce qui explique la forte affinité de ces protéines pour l'ADNsb et la protection efficace de cet ADNsb contre les nucléases *in vitro*. TEBP- $\alpha$  est aussi capable de former des homodimères (Peersen *et al*, 2002), ce qui suggère qu'elle pourrait être responsable du regroupement spatial des télomères observé dans la chromatine native extraite du noyau d'*O. nova* (Murti & Prescott, 2002).

Les structures tridimensionnelles des complexes TEBP- $\alpha$ /TEBP- $\beta$ /ADNsb (entrée PDB 1OTC) et (TEBP- $\alpha$ /ADNsb) $_2$  (entrée PDB 1KIX) sont représentées dans la figure suivante :



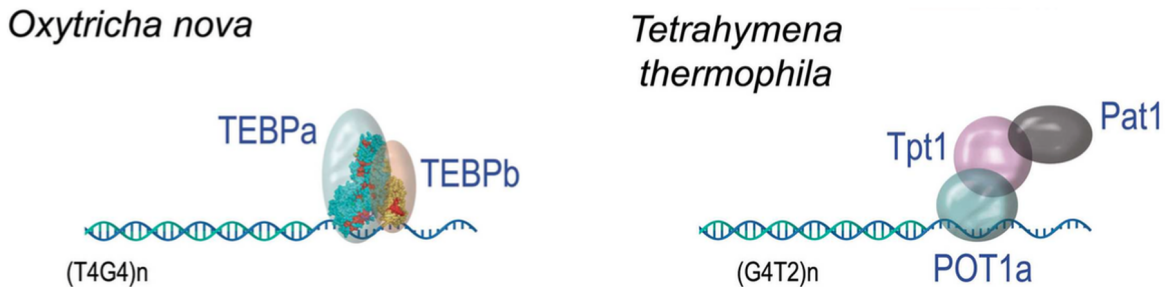
**Figure 11 : Complexe télomérique d'*Oxytricha nova*. L'ADNsb est coloré en rouge. À gauche, l'hétérodimère TEBP- $\alpha$  (en vert) / TEBP- $\beta$  (en bleu). À droite, l'homodimère (TEBP- $\alpha$ /ADNsb) $_2$  avec le monomère TEBP- $\alpha$  vert dans la même orientation qu'à gauche.**

Ce complexe télomérique simple a conduit au premier modèle de protection des extrémités des chromosomes par la séquestration physique de l'extrémité 3'-OH. Ce modèle a par la suite servi de base conceptuelle pour l'étude des télomères d'autres organismes, en effet la séquestration de l'extrémité 3' est un thème récurrent en biologie des télomères.

*Tetrahymena thermophila*, un autre unicellulaire cilié, est aussi un modèle très utilisé en biologie des télomères. Cet organisme a surtout été étudié pour comprendre le fonctionnement de la télomérase, puisque cette enzyme a été découverte chez lui (Greider

& Blackburn, 1985, 1987). En revanche son complexe télomérique a été moins étudié que celui d'*O. nova* (par exemple aucune structure tridimensionnelle n'est disponible à ce jour), et semble plus élaboré puisqu'il contient au moins quatre protéines (Premkumar *et al*, 2014) : Pot1a, Tpt1, Pat1 et Pat2.

L'organisation des complexes télomériques de ces deux ciliés est schématisée dans la figure suivante :

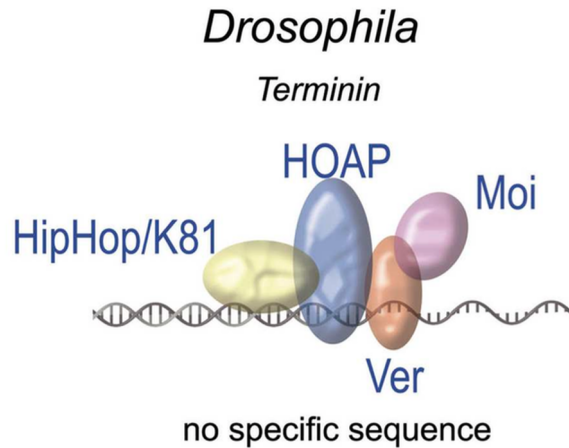


**Figure 12 : Complexes télomériques d'*O. nova* et *T. thermophila* (Giraud-Panis *et al*, 2013)**

#### 0.1.7.2 Chez la drosophile

Comme nous l'avons vu précédemment, chez la drosophile la longueur des télomères est régulée par l'activité de trois rétrotransposons, et leur séquence n'est donc pas faite de répétitions d'un motif court mais des gènes de ces rétrotransposons. À cet égard, les télomères de la drosophile sont un exemple de télomères non canoniques, et le complexe télomérique est également atypique.

La protection des télomères de la drosophile est assurée par un complexe protéique appelé *terminin* (Raffa *et al*, 2013). Les protéines de ce complexe ont été identifiées par mutagenèse aléatoire et recherche de phénotypes de dysfonction des télomères. Ces expériences ont permis d'identifier les protéines HOAP, HipHop, Moi et Ver dont la localisation télomérique a été démontrée par des expériences d'immunofluorescence et par la localisation de fusions avec la GFP. Ces protéines sont localisées exclusivement aux télomères pendant tout le cycle cellulaire, et les phénotypes associés à des mutations dans les gènes codant pour ces protéines sont exclusivement télomériques. Les protéines du *terminin* protègent les télomères en recrutant des facteurs accessoires dont la plupart sont impliqués dans les voies de réparation de l'ADN, comme MRN, ATM et ATR. Ces caractéristiques font de ce complexe l'équivalent fonctionnel des complexes télomériques trouvés aux télomères canoniques d'autres organismes. Le complexe *terminin* de la drosophile est représenté schématiquement dans la figure suivante :



**Figure 13 : : Complexe *terminin* de *D. melanogaster* (Giraud-Panis *et al*, 2013)**

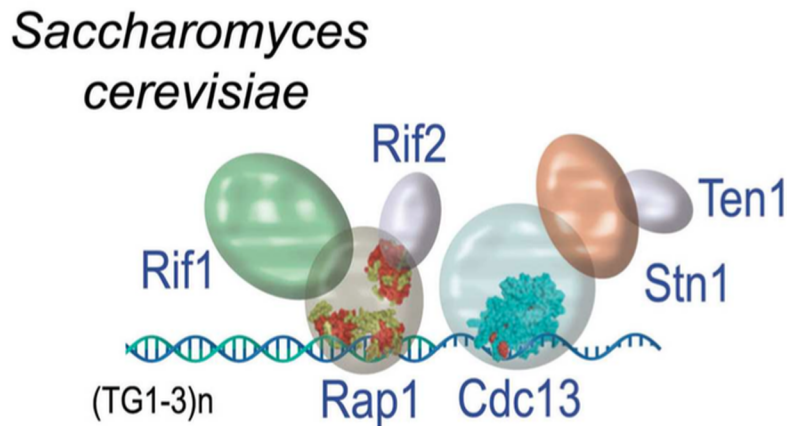
Cependant les protéines du *terminin* n'ont pas d'orthologues connus en dehors du groupe des drosophilidae, et leur séquence mute rapidement (Raffa *et al*, 2013). L'hypothèse évolutive suivante est actuellement proposée : les drosophiles auraient perdu la télomérase (présente chez l'ancêtre commun de tous les eucaryotes), ce qui aurait conduit à une "domestication" rapide des rétrotransposons (en quelques générations) permettant de compenser l'érosion répliquative des extrémités. Ces changements ont provoqué la perte des répétitions télomériques riches en G, et donc la perte des protéines télomériques ancestrales, et ont favorisé l'émergence rapide des protéines du *terminin* liant les extrémités de l'ADN sans spécificité de séquence.

### 0.1.7.3 Chez les levures

Les deux levures les plus utilisées comme organismes modèles en biologie des télomères sont *Saccharomyces cerevisiae* (levure bougeonnante) et *Schizosaccharomyces pombe* (levure fissipare). Ces levures sont de bons modèles car elles possèdent des télomères canoniques évolutivement proches de ceux des vertébrés (notamment, la protéine RAP1 est conservée), car leur culture est facile, et car les techniques de biologie moléculaire et de modification génétique de ces levures sont bien établies. Leurs complexes télomériques sont décrits dans les paragraphes suivants.

Chez *S. cerevisiae*, le complexe télomérique est organisé autour de la protéine Rap1 d'une part et autour du complexe CST d'autre part (Kupiec, 2014). La protéine Rap1 est capable de lier directement les répétitions télomériques de l'ADN double-brin par un domaine composé de deux repliements de type Myb répétés en tandem (König *et al*, 1996), à partir desquels les extrémités N- et C-terminales s'étendent de part et d'autre de l'axe de l'ADN (Matot *et al*, 2012). Par son domaine C-terminal, Rap1 recrute les protéines Rif1 et Rif2 impliquées dans la régulation de la longueur des télomères (Wotton & Shore, 1997), ainsi que les protéines Sir3 et Sir4 responsables de l'établissement et du maintien de la

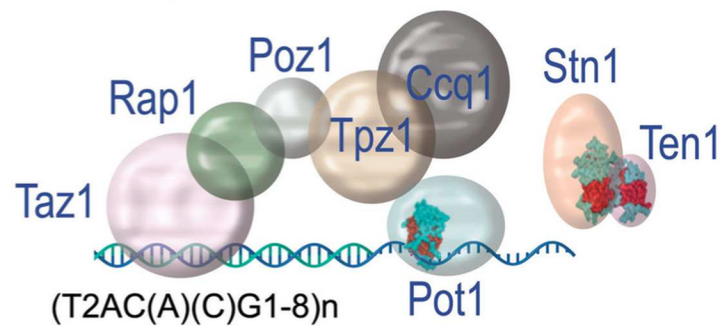
répression de la transcription des gènes situés dans les régions proches des télomères (*telomere position effect* ; Kyrion *et al*, 1993). Le complexe CST (pour Cdc13, Stn1, Ten1) se lie à l'extrémité simple-brin sortante en 3' du brin G et est essentiel pour sa protection contre les nucléases et également pour le recrutement de la télomérase lors de la réplication (Kupiec, 2014). L'organisation du complexe télomérique de *S. cerevisiae* est présentée schématiquement dans la figure suivante :



**Figure 14 : Complexe télomérique de *S. cerevisiae* (Giraud-Panis *et al*, 2013)**

Chez *S. pombe*, le complexe télomérique est organisé autour de la protéine Taz1 qui lie l'ADN télomérique double-brin, et la protéine Pot1 qui lie l'ADN télomérique simple-brin (Pot1 est l'orthologue de la protéine TEBP- $\alpha$  d'*O. nova*). Ces deux protéines sont reliées par un réseau d'interactions protéine-protéine faisant intervenir Rap1, Poz1, et Tpz1, et permettant au complexe de lier à la fois les répétitions télomériques double-brin et simple-brin. Rap1 est conservée chez cette levure, mais n'interagit plus directement avec l'ADN et est recrutée aux télomères par son interaction avec Taz1. *S. pombe* possède aussi les protéines Stn1 et Ten1, mais à ce jour aucun orthologue de Cdc13 n'est connu chez cette levure et les fonctions de son complexe (C)ST sont encore très peu décrites (Giraud-Panis *et al*, 2010). L'organisation du complexe télomérique de *S. pombe* est présentée schématiquement dans la figure suivante :

## *Schizosaccharomyces pombe*



**Figure 15 : Complexe télomérique de *S. pombe* (Giraud-Panis *et al*, 2013)**

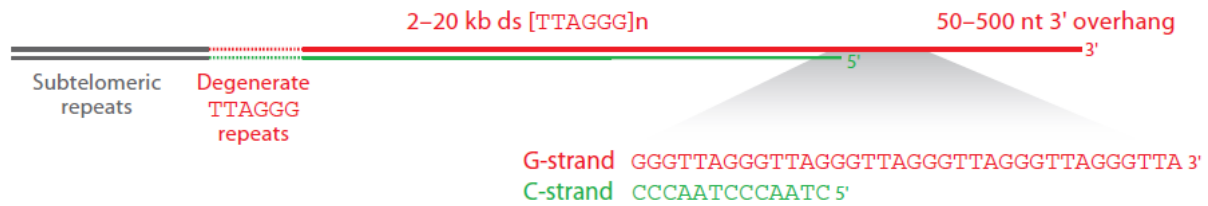
Comme nous le verrons dans le chapitre suivant, le complexe télomérique de *S. pombe* ressemble beaucoup à celui des vertébrés : la protéine Taz1 est en effet l'orthologue de TRF2, et elle recrute une protéine Rap1 également conservée ; ce complexe TRF2/RAP1 constitue notre principal sujet d'étude.

À ce stade nous pouvons remarquer que l'architecture globale du complexe télomérique est bien conservée, même si les protéines impliquées sont différentes d'un organisme à l'autre. Le complexe télomérique comporte toujours au minimum une ou plusieurs protéines de liaison aux répétitions simple-brin, et les protéines de liaison aux répétitions double-brin sont présentes chez la majorité des organismes.

## 0.2 Les télomères humains

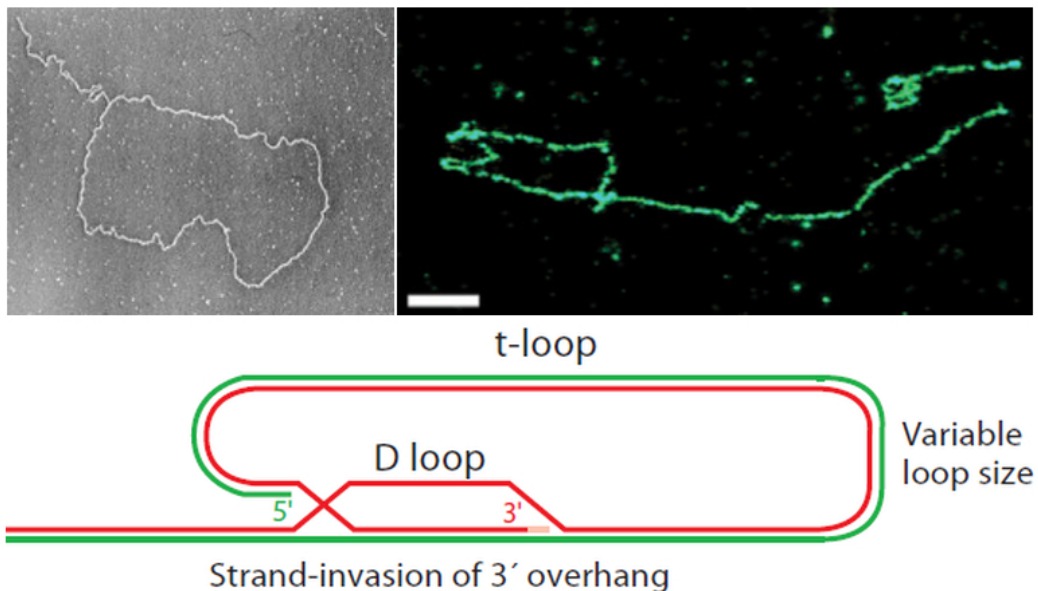
### 0.2.1 L'ADN télomérique et la *t-loop*

Chez l'humain, l'ADN télomérique est constitué de répétitions en tandem du motif TTAGGG (Moyzis *et al*, 1988) qui s'étendent dans une région double-brin d'une longueur typique de 10 à 15 kb et se prolongent par une extrémité simple-brin sortante en 3' longue de 50 à 500 nucléotides (Makarov *et al*, 1997). L'organisation de l'ADN télomérique est représentée schématiquement sur la figure suivante :



**Figure 16 : L'ADN télomérique des mammifères (Palm & de Lange, 2008)**

L'extrémité simple-brin et la nature répétée des séquences télomériques rend possible la formation d'une structure en boucle appelée la *t-loop*, représentée dans la figure suivante :



**Figure 17 : La *t-loop*. En haut à gauche, cliché de microscopie électronique (Griffith *et al*, 1999). En haut à droite, cliché de microscopie STORM (*stochastic optical reconstruction microscopy*; Doksani *et al*, 2013), la barre d'échelle équivaut à 400 nm. En bas, le modèle de la *t-loop* (Palm & de Lange, 2008).**



Cette structure a été mise en évidence pour la première fois par microscopie électronique (Griffith *et al*, 1999), purifiée à partir de cellules HeLa fixées par un traitement au psoralène et aux UV (le psoralène est un intercalant de l'ADN qui induit des pontages covalents inter-brins lorsqu'il est irradié aux UV). Selon le modèle actuel de formation de la *t-loop*, la molécule d'ADN se replie vers l'arrière et l'extrémité simple-brin envahit la région double-brin et s'apparie au brin C (panneau du bas dans la figure). Ce modèle suggère également que la *t-loop* est une solution pour la protection des extrémités contre les systèmes de réparation des cassures double-brin : elle bloque physiquement l'accès à l'extrémité de la molécule d'ADN (Griffith *et al*, 1999).

Plus récemment, la *t-loop* a pu être visualisée dans des cellules de souris sans extraction ni purification, par une technique de microscopie de fluorescence dite à "super-résolution" avec laquelle le mode d'acquisition et le traitement des images résultantes permet d'observer des objets plus petits que la limite due à la diffraction (Doksani *et al*, 2013). Il est donc clair que la *t-loop* existe *in vivo*, en revanche son temps de vie et sa dynamique d'ouverture et de fermeture au cours du cycle cellulaire ne sont pas encore caractérisés. Ces questions sont importantes pour comprendre comment la *t-loop* peut réguler l'accès à l'extrémité de la molécule d'ADN pour permettre l'élongation par la télomérase.

### 0.2.2 Les protéines télomériques forment le complexe *shelterin*

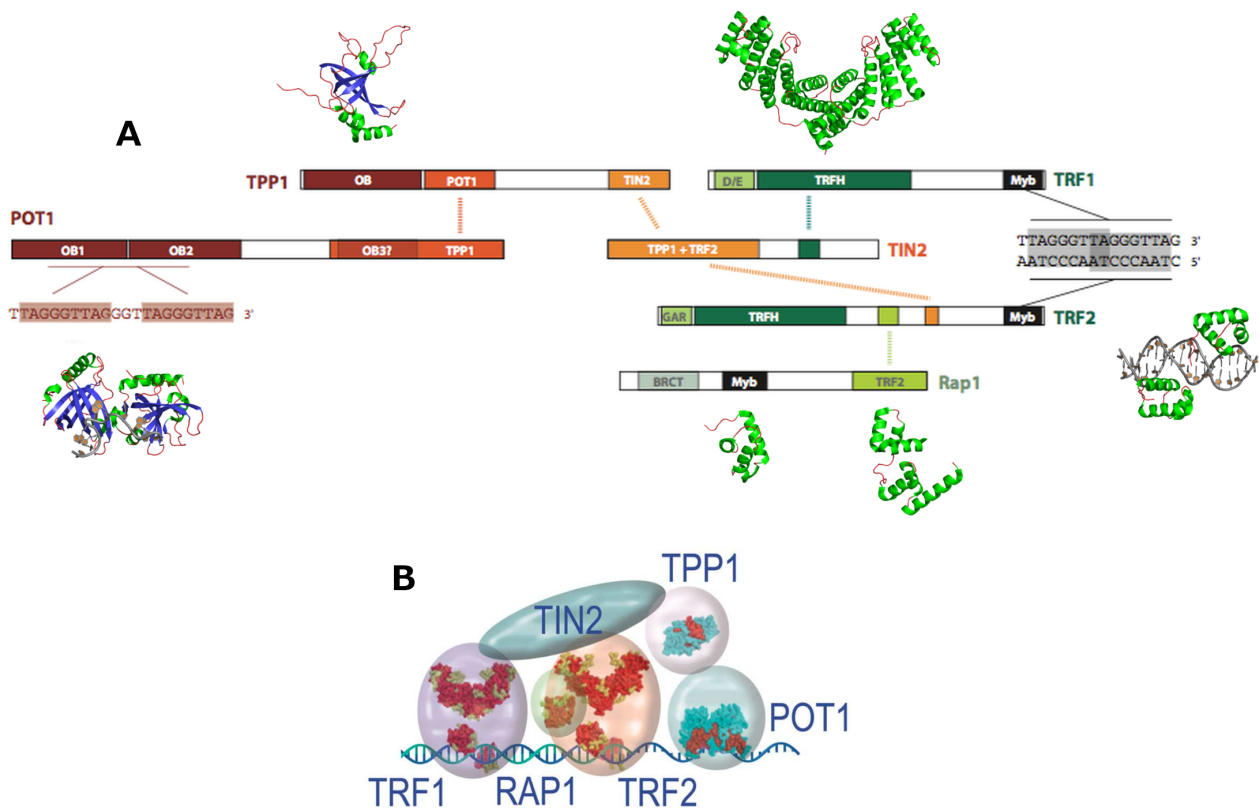
Chez les vertébrés, en plus des nucléosomes, l'ADN télomérique est lié par un complexe de protéines spécifiques appelé le complexe *shelterin* (de Lange, 2005) ou télosome (Liu *et al*, 2004). Ces protéines sont associées aux télomères pendant toutes les phases du cycle cellulaire, elles sont absentes ou très peu abondantes dans les autres régions du noyau, et leurs fonctions principales connues sont télomériques (c'est-à-dire que le phénotype observé suite à leur délétion ou mutation est essentiellement télomérique).

Les 6 protéines qui forment ce complexe sont (dans l'ordre de leur découverte) :

- TRF1 : *telomere repeat-binding factor 1*, identifiée initialement par son affinité et sa spécificité *in vitro* pour les répétitions télomériques double-brin (Zhong *et al*, 1992), et dont la localisation télomérique a été confirmée peu de temps après par immunofluorescence (Chong *et al*, 1995) ;
- TRF2 : *telomere repeat-binding factor 2*, identifiée par sa similarité de séquence avec TRF1 et ses propriétés similaires de liaison à l'ADN (Bilaud *et al*, 1996), et dont la localisation télomérique a été confirmée de la même façon (Bilaud *et al*, 1997) ;
- TIN2 : *TRF1-interacting nuclear factor 2*, identifiée comme un partenaire de TRF1 dans un crible de double-hybride, localisée aux télomères et dont les mutations provoquent des phénotypes télomériques (Kim *et al*, 1999) ;
- RAP1 : *repressor/activator protein 1*, identifiée comme un partenaire de TRF2 dans un crible de double-hybride, et par sa similarité de séquence avec scRap1 (Li *et al*, 2000) ;

- POT1 : *protection of telomeres 1*, identifiée par sa similarité de séquence avec les protéines TEBP des ciliés (Baumann & Cech, 2001) ;
- TPP1 : *POT1-binding partner*, identifiée indépendamment dans trois études comme un partenaire de TIN2 et POT1, et comme nécessaire pour le recrutement de POT1 aux télomères (Liu *et al*, 2004; Ye *et al*, 2004; Houghtaling *et al*, 2004).

Toutes ces protéines sont organisées en domaines multiples connectés par des régions flexibles. La figure suivante donne un aperçu des domaines dont une structure tridimensionnelle est disponible à ce jour :



**Figure 18 : Domaines structuraux des protéines du *shelterin* (A, adapté de Palm & de Lange, 2008). Dans les représentations des structures tridimensionnelles, les hélices sont colorées en vert, les feuillets en bleu et les boucles en rouge, et l'ADN est coloré en gris. Ces représentations ont été produites à partir des entrées PDB 3BUA, 1W0U, 3K6G, 1FEX, 2I46 et 1XJV. (B) Représentation schématique du complexe *shelterin* d'après les connaissances actuelles (Giraud-Panis *et al*, 2013).**

TRF1 et TRF2 (les orthologues de Taz1 de *S. pombe*) s'assemblent en homodimères par leur domaine TRFH (*TRF homology*) et lient les répétitions télomériques double-brin par leur domaine Myb. RAP1 interagit spécifiquement avec TRF2 par son domaine RCT (*RAP1 C-terminus*). POT1 lie les répétitions télomériques simple-brin avec son tandem de deux domaines OB-fold. TPP1 interagit avec POT1 et TIN2, et TIN2 interagit avec TRF1, TRF2 et

TPP1.

Les 6 protéines du *shelterin* peuvent être co-purifiées (Liu *et al*, 2004), ce qui démontre les interactions multiples qui existent entre elles.

La stoechiométrie des protéines du *shelterin* a été analysée dans plusieurs lignées de cellules HeLa à télomères longs et courts et dans des fibroblastes primaires humains, par une méthode de Western blot quantitatif (Takai *et al*, 2010). Un dosage précis par cette méthode est délicat, et cette étude ne prétend fournir que des estimations. Les résultats les plus clairs de cette analyse montrent que TRF2 et RAP1 sont présents aux télomères à un ratio de 1:1 l'un par rapport à l'autre. Il en est de même pour TPP1 et POT1. TRF1 est deux à quatre fois moins abondant que TRF2 aux télomères. TIN2 est présent en quantité suffisante pour lier l'ensemble de TRF1 et TRF2. TPP1 et POT1 sont environ 10 fois moins abondants que TIN2. De façon notable, un réservoir de RAP1 est également présent dans le cytosol. La déplétion de TRF2 d'un facteur 10 par un shRNA est sans impact sur les quantités de TRF1, TIN2, TPP1 et POT1 aux télomères, ce qui suggère l'existence de sous-complexes plus ou moins indépendants du *shelterin*. Cette déplétion de TRF2 diminue notablement la quantité de RAP1 télomérique, mais n'affecte que peu la quantité de RAP1 cytosolique, ce qui suggère pour RAP1 des fonctions extra-télomériques indépendantes de son interaction avec TRF2 (voir Kabir *et al*, 2010; Teo *et al*, 2010; Martínez & Blasco, 2011; Ghosh & Tergaonkar, 2012).

Il a été montré que le complexe *shelterin* complet se lie préférentiellement aux régions contenant une jonction double-brin/simple-brin, c'est-à-dire avec un site de liaison de POT1 et au moins un site de liaison du domaine Myb de TRF2 ou TRF1 (Choi *et al*, 2011). Cette étude a également montré que deux complexes majoritaires possèdent cette spécificité de liaison : un complexe contenant les six protéines du *shelterin*, et un autre qui ne contient pas TRF1. Puisque les télomères contiennent peu de ces sites de haute affinité, et bien plus de sites de plus faible affinité (répétitions télomériques double-brin éloignées des sites de liaison de POT1), les auteurs supposent que les sites de haute affinité permettent de positionner le complexe aux endroits critiques du télomère où il peut ainsi lier et rapprocher physiquement l'extrémité simple-brin et la région double-brin et ainsi favoriser la formation de la *t-loop* (Choi *et al*, 2011). La région double-brin serait occupée par des sous-complexes ne contenant pas forcément TPP1 et POT1.

### 0.2.3 Les facteurs télomériques accessoires

En plus des protéines du *shelterin*, présentes pendant toutes les phases du cycle cellulaire, et dont les fonctions connues sont essentiellement télomériques, on trouve aux télomères un grand nombre de facteurs accessoires. La base de données TeloPIN (*telomeric proteins interaction network* ; Luo *et al*, 2015) recense un très grand nombre de partenaires des protéines du *shelterin* : 300 partenaires pour TRF1, 248 pour TRF2, 131 pour RAP1, 167

pour TIN2, 65 pour TPP1 et 184 pour POT1 (certaines de ces protéines étant probablement des partenaires communs à plusieurs des protéines du *shelterin*). Ces facteurs accessoires, contrairement aux protéines télomériques, ne sont pas présents uniquement aux télomères, et n'y sont pas présents pendant toutes les phases du cycle cellulaire. Leurs fonctions sont essentiellement non télomériques : ces facteurs sont impliqués dans les voies de réparation de l'ADN, dans la signalisation des dommages à l'ADN, dans la réplication, dans le remodelage de la chromatine, etc. (Lee *et al*, 2011).

Les détails moléculaires du mode de recrutement des facteurs accessoires sont encore peu connus. Un mode de recrutement a déjà été mis en évidence et implique le domaine TRFH des protéines TRF1 et TRF2 : ce domaine n'est pas seulement un domaine d'homodimérisation mais aussi d'interactions protéine-protéine. Le domaine TRFH est en effet capable de recruter des partenaires contenant un motif appelé TBM (*TRFH-binding motif*) de séquence consensus FxLxP pour le domaine TRFH de TRF1 (Chen *et al*, 2008) et YxLxP (Chen *et al*, 2008) ou YRL (Kim *et al*, 2009) pour le domaine TRFH de TRF2. D'autres modes de recrutement existent puisque de nombreux facteurs accessoires sont recrutés par les autres protéines télomériques qui ne possèdent pas de domaine TRFH, et au niveau des *linkers* de TRF1 et TRF2 qui sont également des régions d'interactions protéine-protéine. En revanche ces autres modes de recrutement ne sont pas encore décrits au niveau moléculaire.

#### 0.2.4 Régulation de la télomérase par les protéines télomériques

La télomérase agit préférentiellement sur les télomères courts. Cette propriété a été mise en évidence pour la première fois chez la levure *Saccharomyces cerevisiae* et a conduit au modèle du "comptage protéique" (*protein counting model of telomere length regulation* ; Marcand *et al*, 1997). Selon ce modèle, le télomère est enrichi en répresseurs de la télomérase et un télomère long possède ainsi un fort potentiel d'inhibition de la télomérase. Le raccourcissement dû au problème de la réplication provoque la diminution du nombre de sites télomériques et donc l'appauvrissement en répresseurs de la télomérase. Lorsque le nombre de répresseurs chute en dessous d'une valeur critique, c'est-à-dire lorsque le télomère devient suffisamment court, la répression est levée et la télomérase peut allonger le télomère. Chez *S. cerevisiae*, ces répresseurs de la télomérase sont les protéines Rif1 et Rif2 recrutées par scRap1 (Bianchi & Shore, 2008). Chez l'humain, TRF1 est principalement responsable de l'inhibition de la télomérase (van Steensel & de Lange, 1997; Ancelin *et al*, 2002), et agit de concert avec POT1 (Loayza & de Lange, 2003; Ye *et al*, 2004). POT1 semble avoir un double rôle dans le contrôle de l'activité de la télomérase : en masquant l'extrémité 3', cette protéine empêche physiquement l'élongation du brin G (Lei *et al*, 2005) ; mais lorsqu'elle glisse le long du brin d'ADN en s'éloignant de l'extrémité 3' elle augmente la processivité de la télomérase (Lei *et al*, 2005) probablement en maintenant son association avec l'ADN avec l'aide de TPP1 (Wang *et al*, 2007; Xin *et al*, 2007). Le mécanisme précis de

passage de l'état extensible à l'état non extensible n'est pas encore complètement compris, de même que le rôle possible de la *t-loop* dans la régulation de l'accessibilité de l'extrémité 3'.

### 0.2.5 Protection de l'extrémité du chromosome par le complexe *shelterin*

La fonction du *shelterin* à l'origine de son nom est celle de protection des extrémités contre les voies de réparation des cassures double-brin de l'ADN et contre la dégradation par des exonucléases.

Ces fonctions ont été récemment étudiées chez la souris par des méthodes génétiques de délétion conditionnelle sélective d'un ou plusieurs gènes codant pour les protéines du complexe (Sfeir & de Lange, 2012). La double délétion de TRF1 et TRF2 suffit à réduire à un niveau indétectable (par CHIP et immunofluorescence) la présence aux télomères de toutes les protéines du *shelterin*. Les télomères "nus" ainsi générés sont très rapidement la cible de la machinerie du NHEJ classique (c-NHEJ, dépendant de Ku et de la ligase 4) : des fusions de chromosomes se produisent à une fréquence très élevée. Pour pouvoir observer les autres phénotypes provoqués par l'absence du *shelterin*, les auteurs ont étudié des cellules chez lesquelles la ligase 4 est inactivée (*Lig4<sup>-/-</sup>*) : dans ce contexte génétique les fusions de chromosomes sont plus rares mais tout de même possibles, ce qui suggère que la voie NHEJ alternative est aussi active (a-NHEJ, dépendant de PARP1 et la ligase 3). L'ensemble de leurs résultats leur permet d'identifier quelles protéines du *shelterin* répriment chaque voie menaçant l'intégrité des télomères :

- la voie DDR initiée par ATM est réprimée par TRF2,
- la voie DDR initiée par ATR est réprimée par POT1,
- la voie NHEJ classique est réprimée par TRF2 et POT1,
- la recombinaison homologue est réprimée par RAP1 et POT1,
- la voie NHEJ alternative et la dégradation de l'extrémité 5' par des exonucléases sont réprimées de façon redondante par plusieurs des protéines du *shelterin*.

TRF2 et RAP1 sont les protéines les plus abondantes aux télomères, et sont les principales responsables de la protection des extrémités. Elles sont en effet critiques pour la répression de la signalisation DDR par ATM, la répression de la voie NHEJ et la répression de la recombinaison homologue. Dans le chapitre suivant nous décrirons plus en détails ces deux protéines au niveau moléculaire.

## 0.3 TRF2 et RAP1 : deux protéines essentielles

### 0.3.1 Fonctions de TRF2 et RAP1

TRF2 a fait l'objet de nombreuses études *in vitro*, et ses propriétés d'interaction avec l'ADN sont maintenant bien décrites. Par son domaine Myb, TRF2 se lie spécifiquement aux répétitions télomériques TTAGGG (Bilaud *et al*, 1997). Son extrémité N-terminale basique lui permet de lier des structures d'ADN branchées sans spécificité de séquence mais avec une préférence pour les structures à quatre branches comme la jonction de Holliday (Fouché *et al*, 2006). La liaison de TRF2 à ces jonctions accélère leur formation et les protège contre le clivage par des résolvasés (Poulet *et al*, 2009), et contre la séparation de brins par l'hélicase WRN (Nora *et al*, 2010). TRF2 favorise aussi l'invasion d'un duplex télomérique par un simple-brin télomérique, en induisant des super-tours négatifs dans l'ADN qui facilitent la dissociation locale des deux brins du duplex et donc facilitent l'invasion du duplex par le simple-brin (Amiard *et al*, 2007). Cette propriété topologique de TRF2 semble s'expliquer par le fait que TRF2 condense l'ADN en l'enroulant autour de lui en tours positifs, ce qui est compensé dans le reste de la molécule d'ADN par des tours négatifs. La condensation est dépendante du domaine B de TRF2, en effet des protéines chimériques dans lesquelles les domaines A et B de TRF1 et TRF2 ont été inversés ont des propriétés de condensation de l'ADN elles aussi inversées (Poulet *et al*, 2012). Ces propriétés de TRF2 lui permettent de favoriser la formation de la *t-loop* comme cela a été démontré *in vitro* (TRF2 forme des boucles à partir d'un ADN télomérique modèle, voir Griffith *et al*, 1999) et *in vivo* (les *t-loops* ne sont pas détectées dans des cellules délétées de TRF2, voir Doksani *et al*, 2013).

TRF2 et RAP1 sont responsables de la protection des extrémités contre le NHEJ. Elles inhibent *in vitro* la machinerie du NHEJ présente dans des extraits nucléaires de cellules humaines (Bae & Baumann, 2007). Dans ces expériences les deux protéines sont nécessaires pour réprimer le NHEJ. Dans des cellules en culture, la délétion de TRF2 suffit à induire des fusions de chromosomes (Celli & de Lange, 2005), tandis que le rétablissement de RAP1 aux télomères indépendamment de TRF2 (sous forme d'une protéine chimérique comportant un domaine de liaison aux répétitions TTAGGG) est suffisant pour réprimer les fusions de chromosomes (Sarthý *et al*, 2009).

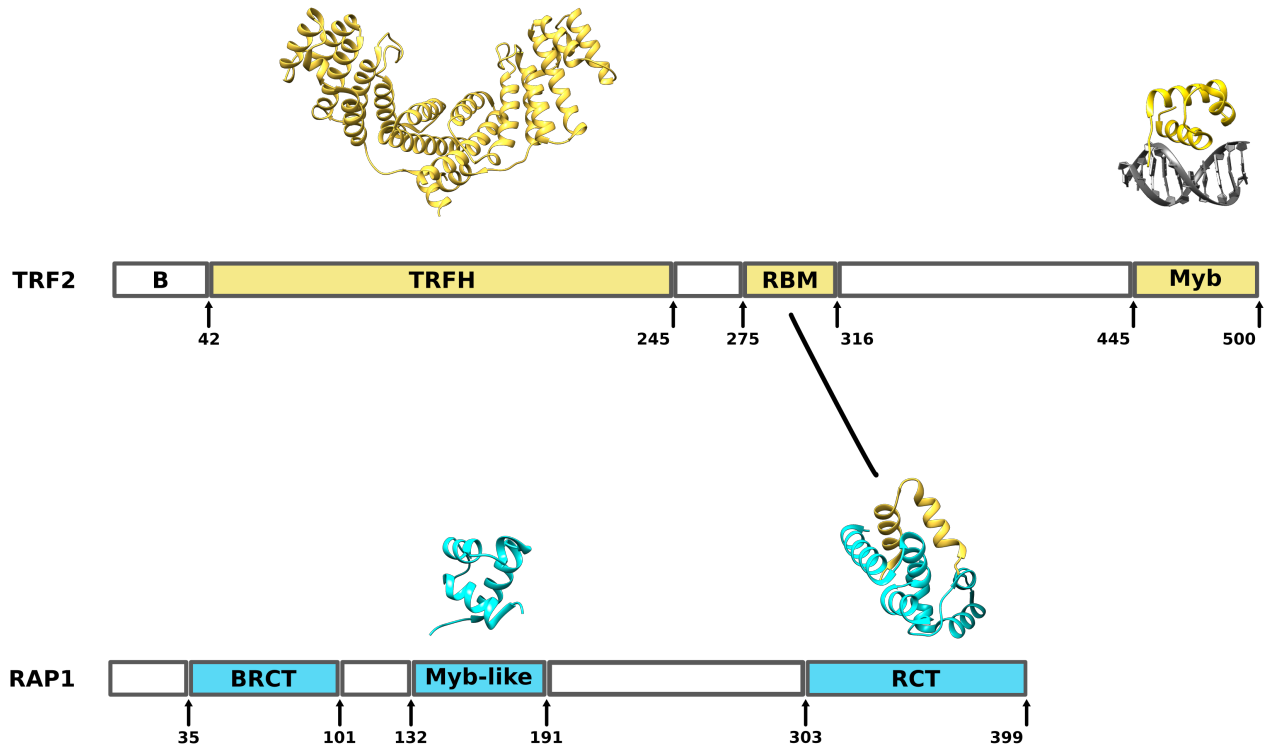
RAP1 semble également avoir un rôle dans la régulation de la longueur des télomères, puisque sa déplétion par un shRNA (O'Connor *et al*, 2004) ou la surexpression de variants tronqués (Li & de Lange, 2003) dans des cellules humaines en culture provoquent un allongement des télomères. Une étude récente des fonctions de RAP1 par sa délétion par une technique d'édition de génome (Kabir *et al*, 2014) montre cependant que son absence complète dans des cellules humaines en culture ne provoque aucun phénotype télomérique. Cette étude montre que les résultats des expériences de surexpression et de déplétion sont

difficiles à interpréter, car les effets de ces manipulations peuvent être variés. Les auteurs s'interrogent sur la grande conservation de RAP1 malgré son rôle télomérique apparemment limité. Ils suggèrent que les rôles extra-télomériques de RAP1 (voir Crabbé & Karlseder, 2010, et références associées) expliqueraient sa conservation chez les eucaryotes supérieurs. Une autre possibilité pourrait être que les fonctions télomériques de RAP1 sont critiques et donc assurées de façon redondante par plusieurs protéines, ce qui expliquerait l'absence de phénotype télomérique dans les cellules où RAP1 est délétée.

### 0.3.2 TRF2 et RAP1 sont des protéines multi-domaines

TRF2 et RAP1 contiennent chacune plusieurs domaines connectés par des régions peu structurées relativement longues. TRF2 comporte une région N-terminale basique appelée B et prédite non structurée, un domaine de dimérisation appelé TRFH, une longue région également prédite non structurée (*linker*) contenant le motif de liaison à RAP1 (RBM pour *RAP1-binding motif*), le motif de liaison à TIN2 et un signal de localisation nucléaire (NLS), et en C-terminal un domaine Myb de liaison à l'ADN. RAP1 comporte un domaine BRCT en N-terminal, un domaine pseudo-Myb (au repliement en trois hélices  $\alpha$  similaire à celui du domaine Myb de TRF2, mais dont les résidus de surface ne permettent pas la liaison à l'ADN ; voir Hanaoka *et al*, 2001), une longue région prédite non structurée (*linker*) riche en résidus acides, et un domaine C-terminal par lequel elle interagit avec TRF2 (Li *et al*, 2000; Chen *et al*, 2011). Comme chez *S. pombe*, RAP1 n'interagit pas avec l'ADN et dépend de son interaction avec une protéine de la famille des TRF pour son recrutement aux télomères.

L'organisation des domaines de TRF2 et RAP1 est présentée dans la figure suivante, avec une représentation des structures tridimensionnelles disponibles dans la PDB :



**Figure 19 : Organisation des domaines de TRF2 et RAP1**

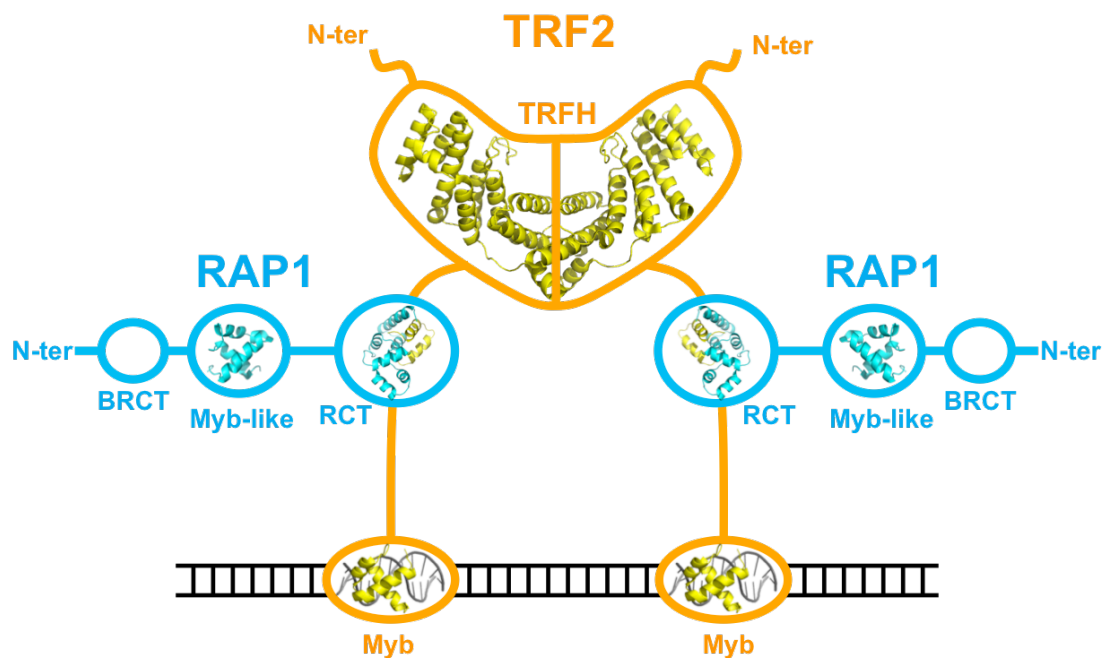
Cette organisation en domaines connectés par des régions peu ou pas structurées confère à ces protéines une plasticité conformationnelle probablement importante pour leurs fonctions, mais les rend particulièrement difficiles à étudier. En effet, aucune donnée structurale n'a été publiée à ce jour sur les protéines entières ni sur le complexe, seules des structures tridimensionnelles de domaines isolés sont actuellement disponibles dans la PDB. Ces structures sont listées dans le tableau suivant :



**Tableau 2 : Structures tridimensionnelles des domaines de TRF2 et RAP1**

Domaine	Entrée PDB	Référence
RAP1-Myb	1FEX	Hanaoka <i>et al</i> (2001)
RAP1-RCT/TRF2-RBM	3K6G	Chen <i>et al</i> (2011)
TRF2-TRFH	1H6P	Fairall <i>et al</i> (2001)
TRF2-TRFH/Apollo-TBM	3BUA	Chen <i>et al</i> (2008)
TRF2-TRFH/TIN2-TBM	3BU8	Chen <i>et al</i> (2008)
TRF2-TRFH/SLX4-TBM	4M7C	Wan <i>et al</i> (2013)
TRF2-Myb	1VF9	Hanaoka <i>et al</i> (2005)
TRF2-Myb/ADN	1VFC	Hanaoka <i>et al</i> (2005)
(TRF2-Myb) <sub>2</sub> /ADN	1WOU	Court <i>et al</i> (2005)
(TRF2-Myb) <sub>2</sub> /ADN	3SJM	Nair <i>et al</i> (2011)

D'après la stoechiométrie connue du complexe (deux monomères de RAP1 s'associent à un dimère de TRF2) et les structures des domaines isolés, les connaissances structurales actuelles sur le complexe TRF2/RAP1 peuvent se résumer très schématiquement comme suit :



**Figure 20 : Représentation schématique du complexe TRF2/RAP1 d'après les connaissances actuelles**

Remarquons que la conformation globale des deux protéines et du complexe est complètement hypothétique, car les informations disponibles ne nous renseignent pas sur ces aspects.

### 0.3.3 Partenaires de TRF2 et RAP1

TRF2 et RAP1 recrutent de nombreux partenaires aux télomères. La base de données TeloPIN (*telomeric proteins interaction network* ; Luo *et al*, 2015) liste 248 partenaires pour TRF2 et 131 pour RAP1.

Les partenaires de RAP1 ont été principalement identifiés par co-immunoprécipitation et spectrométrie de masse (O'Connor *et al*, 2004), et par PCA en cellules vivantes et validation par GST *pull down* en extraits cellulaires (Lee *et al*, 2011). De façon intéressante, RAP1 recrute Ku80, MRE11 et RAD50 (toutes des protéines impliquées dans la réparation des cassures double-brin de l'ADN) indépendamment de son interaction avec TRF2 (O'Connor *et al*, 2004). Cependant, à ce jour aucune étude ciblée de partenaires de RAP1 n'est publiée et nous n'avons pas d'informations sur la nature directe ou indirecte des interactions mises en évidence par les cribles à haut débit, ni sur les détails de ces interactions à l'échelle moléculaire.

Concernant TRF2, un des modes de recrutement de ses partenaires a été bien caractérisé au niveau moléculaire. Son domaine TRFH contient un site conservé de liaison pour un motif appelé TBM (*TRFH-binding motif*) de séquence consensus YxLxP, initialement mis en évidence par une étude ciblée de son interaction avec la protéine Apollo (Chen *et al*, 2008), et également découvert dans la protéine NBS1 (mais pour laquelle l'interaction a seulement été vérifiée *in vitro* avec le motif isolé ; Chen *et al*, 2008). Une autre étude a précisé la spécificité de séquence de ce site en criblant une banque de peptides pour leur liaison au domaine TRFH, et a déterminé le motif consensus YRL (Kim *et al*, 2009). Les auteurs de cette étude ont aussi découvert ce motif dans les protéines PNUTS et MCPH1, et ont confirmé leur interaction avec TRF2 *in vitro* avec un peptide contenant le motif, et *in vivo* en montrant leur localisation télomérique dépendante de TRF2 et en observant des phénotypes télomériques lorsque l'interaction est rompue en mutant le motif TBM (Kim *et al*, 2009). Enfin, une étude ciblée de la protéine SLX4 a démontré son interaction avec TRF2 par un motif de séquence HxLxP (Wan *et al*, 2013), ce qui élargit la séquence consensus des motifs TBM.

La figure suivante montre le site conservé de liaison des motifs TBM du domaine TRFH de TRF2 :

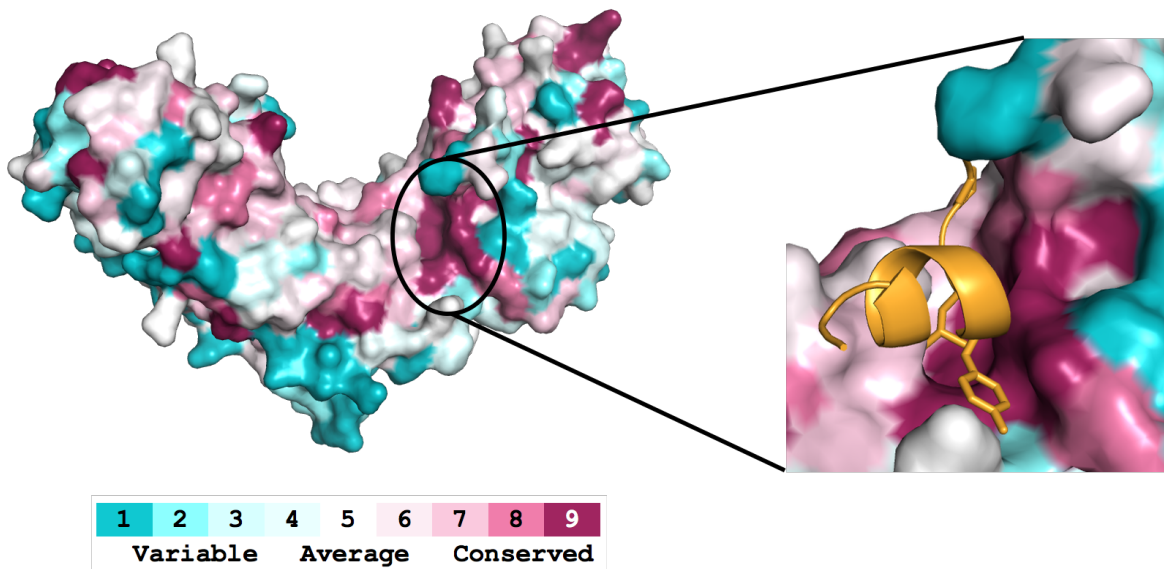


Figure 21 : Conservation des résidus en surface du domaine TRFH de TRF2 (à gauche). La figure a été réalisée avec l'entrée PDB 3BUA et un alignement des séquences de TRF2 chez 14 espèces de vertébrés (visible en annexe), à l'aide du serveur ConSurf (<http://consurf.tau.ac.il> ; Glaser *et al*, 2003). Le site conservé de liaison des motifs TBM est entouré. Vue détaillée du site de liaison occupé par le motif TBM d'Apollo (à droite). Le peptide d'Apollo est représenté en orange, avec les résidus Y, L et P représentés en bâtons.

Les motifs TBM validés expérimentalement à ce jour sont présentés dans la figure suivante :

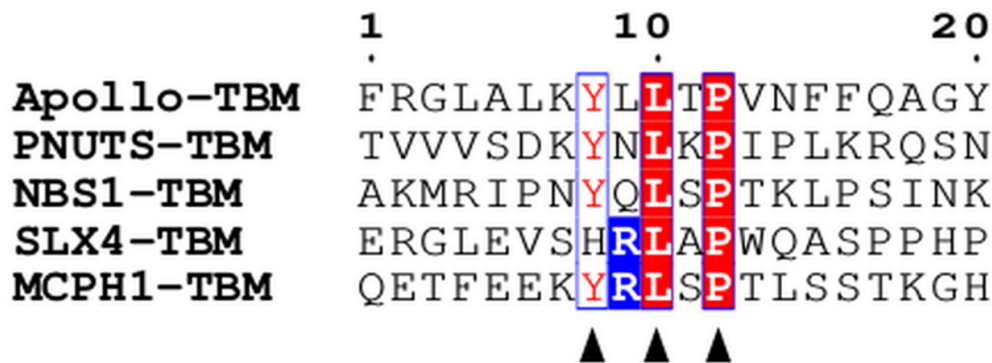


Figure 22 : Les motifs TBM actuellement connus et validés expérimentalement. Les résidus identiques sont surlignés en rouge, les résidus similaires sont encadrés en bleu. Les résidus du motif sont indiqués par des triangles noirs, le résidu R du motif consensus défini par Kim *et al* (2009) est surligné en bleu.

Parmi les résidus du motif, la leucine centrale est critique pour la liaison. Sa mutation en alanine abolit complètement l'interaction *in vitro* et *in vivo* (Chen *et al*, 2008; Kim *et al*, 2009; Wan *et al*, 2013). Dans le crible de la banque de peptides de Kim *et al* (2009),

l'isoleucine et la valine ne sont jamais sélectionnées à cette position du motif alors que ce sont des résidus similaires à la leucine (aliphatiques et branchés), ce qui confirme son importance dans le motif. La mutation de la tyrosine en alanine réduit l'affinité *in vitro* d'un facteur variable selon le motif, d'environ 30 pour PNUTS (Kim *et al*, 2009) à environ 10 000 pour Apollo (Chen *et al*, 2008). La mutation de l'histidine en alanine chez SLX4 réduit l'affinité *in vitro* d'un facteur 60 environ (Wan *et al*, 2013). Enfin la mutation de la proline a un effet très variable : elle réduit l'affinité *in vitro* d'un facteur 5 pour PNUTS (Kim *et al*, 2009), d'un facteur 100 pour SLX4 (Wan *et al*, 2013) et abolit complètement l'interaction pour Apollo (Chen *et al*, 2008).

Kim *et al* (2009) proposent que l'affinité variable du domaine TRFH de TRF2 pour les différents motifs TBM de ses partenaires serait un moyen de réguler leur recrutement aux télomères, en plus des niveaux d'expression de ces partenaires. De plus, la présence de nombreux dimères de TRF2 aux télomères (environ 1000 dimères en moyenne sont disponibles pour chaque télomère, voir Takai *et al*, 2010) permet le recrutement simultanément à un même télomère de nombreux facteurs accessoires différents, permettant d'assurer les nombreuses fonctions contrôlées par TRF2.

## 0.4 Objectifs de ces travaux

Le choix de présentation que nous avons fait dans l'introduction de cette thèse met en avant l'apport de la biologie structurale dans la compréhension générale d'un mécanisme ou d'une fonction biologique. Ainsi, l'apport des informations structurales dans la compréhension des fonctions des protéines télomériques est manifeste, notamment avec la visualisation par microscopie électronique de la *t-loop*, et la compréhension grâce à la cristallographie du mode de recrutement des partenaires de TRF2 contenant un motif TBM. Réciproquement, les questions fonctionnelles encore non résolues coïncident souvent avec un manque de données structurales. Citons par exemple le recrutement aux télomères de protéines des voies de réparation des cassures double-brin de l'ADN (par exemple Ku et le complexe MRN), paradoxalement indispensable pour réprimer correctement ces voies aux extrémités des chromosomes, et qui n'est pas documenté d'un point de vue structural. Le même écart entre les informations fonctionnelles et structurales se retrouve pour TRF2 et RAP1.

TRF2 est une protéine essentielle aux télomères, où elle assure des fonctions multiples : elle contrôle la topologie et la condensation de l'ADN, elle permet la formation de la *t-loop*, elle recrute un grand nombre de facteurs accessoires, et elle protège les extrémités contre la voie de signalisation des dommages à l'ADN initiée par ATM et contre les fusions opérées par la machinerie du NHEJ. RAP1 est la protéine télomérique la plus conservée chez les levures et les vertébrés. Chez ces derniers, elle est le partenaire constitutif de la protéine multi-fonctionnelle TRF2. Malgré cela, ses fonctions chez les vertébrés restent mal

comprises. Dans les deux cas, les données structurales disponibles se limitent aux structures tridimensionnelles de domaines isolés.

Ces deux protéines ont été très étudiées individuellement, mais peu d'études s'intéressent de façon approfondie au complexe TRF2/RAP1. Par ailleurs, l'organisation modulaire de ces protéines composées de multiples domaines et de longues régions prédites comme non structurées suggère une plasticité conformationnelle importante, directement reliée aux propriétés multi-fonctionnelles du complexe. D'autre part, les répétitions télomériques constituent une région subnucléaire fortement enrichie en protéines du *shelterin*, et cette élévation locale de concentration favorise les interactions transitoires de basse affinité. Il est maintenant reconnu que ce type d'interaction, auparavant considéré comme non spécifique, a une grande importance fonctionnelle (Qin & Gronenborn, 2014). Ce sont précisément ces deux notions de plasticité conformationnelle et d'interactions transitoires qui définissent la dynamique fonctionnelle du *shelterin*.

L'obstacle majeur pour comprendre ces phénomènes provient des limites techniques des méthodes de biophysique et de biologie structurale. En particulier, pour comprendre comment la plasticité conformationnelle peut influencer les fonctions, et pour détecter des éventuelles interactions de basse affinité entre domaines, il est essentiel d'étudier l'architecture des protéines entières et de leurs complexes. Ainsi, les principales questions que nous nous sommes posées sont les suivantes :

- Comment l'architecture de TRF2 peut-elle expliquer ses fonctions si nombreuses et variées ?
- Quels sont les effets de RAP1 sur TRF2 au niveau conformationnel et fonctionnel ?
- Comment ces deux protéines s'influencent-elles mutuellement concernant le recrutement de leurs nombreux partenaires ?
- Quels déterminants structuraux de TRF2 permettent ses modes d'interaction avec l'ADN ?

Pour tenter de répondre à ces questions, nous avons utilisé une approche de biologie structurale intégrative combinant la calorimétrie, la diffusion des rayons X aux petits angles, la cristallographie et des expériences de biochimie. Dans une première partie nous avons étudié la conformation en solution de TRF2 et RAP1 entières, ainsi que l'assemblage et la conformation du complexe TRF2/RAP1. Dans la deuxième partie nous nous sommes intéressés à l'influence de RAP1 sur l'interaction entre TRF2 et quelques-uns de ses partenaires connus. Enfin dans la troisième et dernière partie, nous avons tenté d'apporter des détails à l'échelle moléculaire sur le mode d'interaction de TRF2 avec l'ADN.

## 1 Partie 1 : Étude détaillée de l'interaction entre TRF2 et RAP1

Comme nous l'avons vu dans l'introduction, les connaissances actuelles sur TRF2 et RAP1 sont très partielles. Du point de vue structural, nous n'avons des informations que sur des domaines isolés mais aucune connaissance sur les protéines entières. Du point de vue fonctionnel, beaucoup d'études sur TRF2 et RAP1 sont publiées, mais la compréhension des effets de RAP1 sur TRF2 dans les expériences de biologie cellulaire est difficile car les conséquences de la surexpression ou de la déplétion de l'une comme de l'autre sont multiples et inter-dépendantes. Enfin, peu d'études *in vitro* s'intéressent spécifiquement à l'interaction entre TRF2 et RAP1 entières, et les études structurales publiées concernent uniquement des domaines isolés des protéines. L'importance de TRF2 au coeur du complexe *shelterin*, sa capacité à interagir avec de multiples partenaires, et la compréhension limitée du rôle de RAP1 malgré son extrême conservation nous ont encouragé à chercher à comprendre comment s'assemble le complexe formé par ces deux protéines.

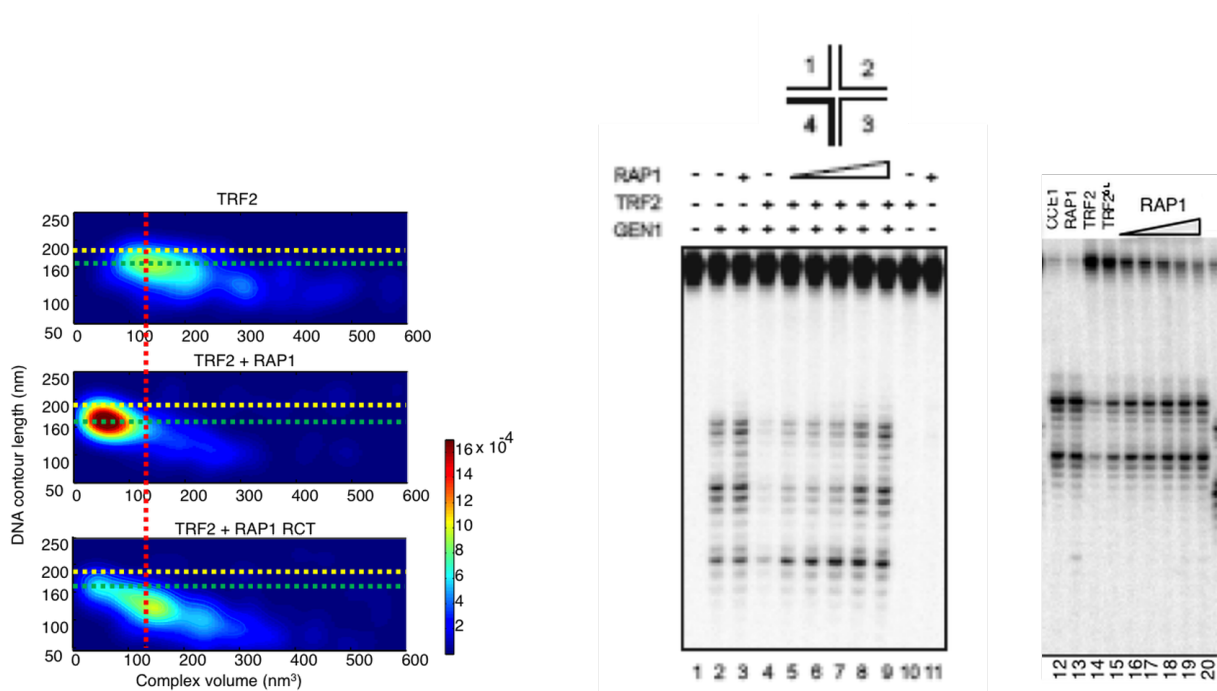
Nos collaborateurs au laboratoire d'Eric Gilson ont commencé à étudier l'effet de RAP1 sur les propriétés de TRF2 d'interaction et de condensation de l'ADN. Ils utilisent l'AFM (*atomic force microscopy*, soit microscopie à force atomique) pour mesurer la capacité de TRF2 à condenser l'ADN, et une expérience de protection de la jonction de Holliday contre le clivage par une résolvasse.

L'AFM leur permet de visualiser individuellement les complexes TRF2/ADN et de mesurer pour chacun la longueur apparente de l'ADN et le volume du complexe. Ils ont tout d'abord observé que TRF2 est capable de former des complexes ayant une longueur apparente d'ADN courte et un grand volume, c'est-à-dire des oligomères hautement condensés (Poulet *et al*, 2012). Plus récemment, ils ont observé que RAP1 diminue fortement cette capacité de TRF2 à condenser l'ADN télomérique et que cet effet de RAP1 dépend de la présence de sa région N-terminale puisque le domaine RCT isolé de RAP1 (pourtant suffisant pour lier TRF2 avec une forte affinité) n'affecte pas cette propriété de TRF2 (figure suivante, panneau de gauche).

L'expérience de protection de la jonction de Holliday consiste à traiter la jonction par une résolvasse (une endonucléase reconnaissant spécifiquement la structure branchée de la jonction de Holliday) et à observer le profil de clivage par électrophorèse (figure suivante, panneau de droite). Les résolvases utilisées dans ces expériences sont GEN1, une résolvasse nucléaire humaine, et CCE1, une résolvasse mitochondriale. En complexe avec TRF2, la jonction de Holliday est nettement protégée contre la résolvasse, tandis que la présence de RAP1 abolit cette protection. Notamment, la présence de RAP1 abolit aussi la protection par un mutant de TRF2 délété de la région RBM, même si dans ce cas des concentrations

plus élevées de RAP1 sont nécessaires pour observer l'effet sur la protection. Ceci signifie que l'interaction entre RAP1-RCT et TRF2-RBM n'est pas le seul déterminant responsable de l'effet de RAP1, et que d'autres régions de RAP1 et de TRF2 sont certainement impliquées.

Par ces deux approches nos collaborateurs ont pu ainsi mettre en évidence que RAP1 modifie les propriétés de TRF2, et que cet effet de RAP1 dépend de la présence de sa région N-terminale (communication personnelle de Marie-Josèphe Giraud-Panis, figure suivante et article en annexe).



**Figure 23 : Effets de RAP1 sur les propriétés d'interaction de TRF2 avec l'ADN. À gauche : la condensation de l'ADN par TRF2 est affectée par RAP1 (comparer le panneau central au panneau du haut) mais pas par RAP1-RCT (comparer le panneau du bas au panneau du haut). L'axe vertical représente la longueur apparente de l'ADN (en nm, la ligne pointillée jaune situe la longueur de l'ADN libre), l'axe horizontal représente le volume du complexe (en nm<sup>3</sup>), et le code couleur représente la fréquence d'observation d'un complexe de longueur apparente d'ADN et de volume donnés. À droite : TRF2 protège la jonction de Holliday contre le clivage par les enzymes GEN1 (comparer les pistes 2 et 4) et CCE1 (comparer les pistes 12 et 14), l'interaction de RAP1 avec TRF2 abolit cette protection (pistes 5 à 9). RAP1 abolit aussi la protection d'un mutant de TRF2 délété du motif RBM (pistes 16 à 20).**

Pour tenter de comprendre à l'échelle moléculaire comment RAP1 interagit avec TRF2 et module ainsi ses propriétés, dans cette première partie nous avons étudié en détails l'interaction entre TRF2 et RAP1 par des méthodes de biochimie, de biophysique et de

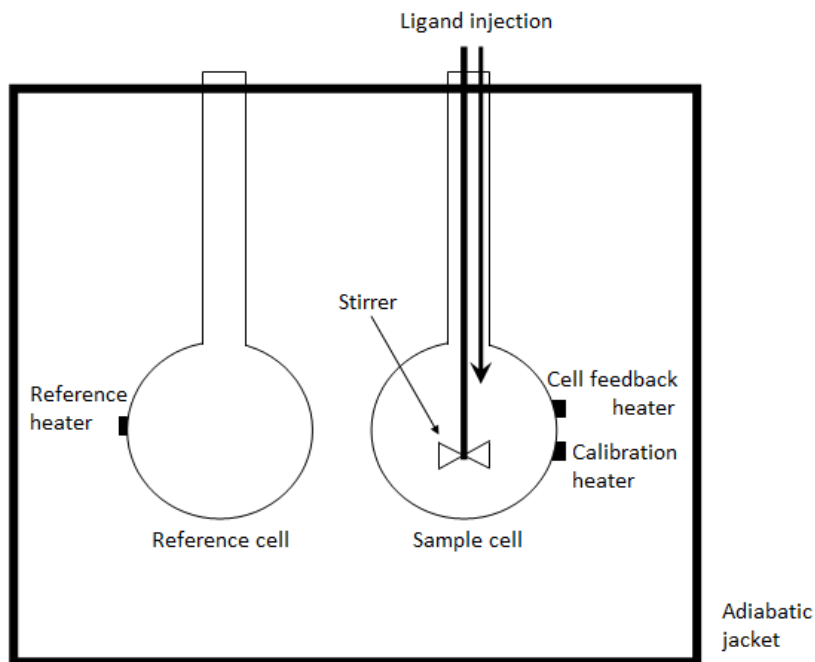
biologie structurale, ce qui nous a permis de mettre en évidence un mode d'interaction entre les deux protéines plus complexe que ce que les données disponibles décrivaient (Chen *et al*, 2011; Janoušková *et al*, 2015).



## 1.1 Résultats

### 1.1.1 L'interaction entre TRF2 et RAP1 entières est complexe

Dans un premier temps, nous avons entrepris de mesurer l'interaction entre TRF2 et RAP1 entières, car si cette interaction est déjà connue (Li *et al*, 2000) elle n'a en revanche été étudiée de façon quantitative qu'avec des domaines isolés (Chen *et al*, 2011). Pour cette mesure nous avons utilisé la titration calorimétrique isotherme (ITC pour *isothermal titration calorimetry*) car cette technique est la seule qui permette d'obtenir une valeur réelle de la constante d'équilibre de dissociation ( $K_D$ ), sans marquage (contrairement aux méthodes de mesure de polarisation ou d'anisotropie de fluorescence) ni immobilisation d'un des partenaires (contrairement à la résonance plasmonique de surface) ni séparation (contrairement aux techniques de centrifugation, de chromatographie et d'électrophorèse). Cette technique en solution nous apparaît pertinente pour l'étude de protéines solubles telles que TRF2 et RAP1 en absence d'ADN. Remarquons que la résonance plasmonique de surface (SPR) serait bien adaptée à l'étude des interactions entre protéines télomériques en présence d'ADN, puisqu'un ADN immobilisé permettrait de mimer le télomère. Le calorimètre est schématisé dans la figure suivante :

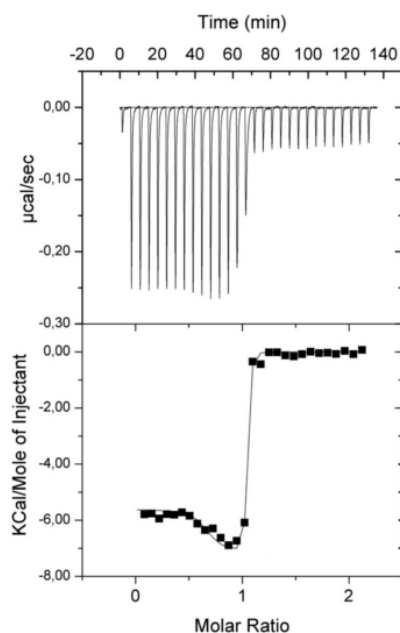


**Figure 24 : Représentation schématique d'un calorimètre de titration isotherme (domaine public : <https://en.wikipedia.org/wiki/File:ITC1.png>)**

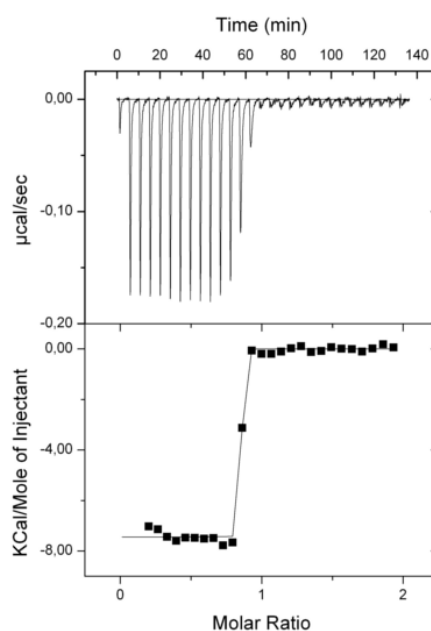
Une des deux molécules est placée dans la cuve et l'autre dans la seringue. Les concentrations doivent être de l'ordre de  $10 \times K_D$  dans la cuve et  $100 \times K_D$  dans la seringue, ainsi lorsque l'on ne connaît pas l'ordre de grandeur du  $K_D$  il faut réaliser plusieurs essais pour mettre au point l'expérience. Cette technique a donc le défaut d'être très consommatrice en matériel biologique : elle est d'autant plus consommatrice que l'affinité entre les deux partenaires est faible (c'est-à-dire que le  $K_D$  est élevé). Une fois la cuve et la seringue remplies (respectivement 2 mL et 500  $\mu$ L sur un appareil VP-ITC), l'appareil stabilise la température dans l'enceinte adiabatique et injecte le contenu de la seringue dans la cuve par étapes de volume défini (typiquement 10  $\mu$ L sur un appareil VP-ITC), puis il mesure la puissance électrique qu'il doit fournir pour maintenir la température constante. Ceci revient à mesurer les échanges de chaleur entre le contenu de la cuve et l'enceinte : la courbe de la puissance fournie en fonction du temps s'appelle le thermogramme. L'intégration des pics du thermogramme fournit une mesure directe à chaque point d'injection de la variation d'enthalpie ( $\Delta H$ ) accompagnant la liaison des deux molécules. Les concentrations et les volumes de départ sont connus, et le volume final dans la cuve après chaque injection est facile à déterminer : il est donc possible de calculer les concentrations des deux molécules dans la cuve après chaque injection et le rapport molaire entre elles. Lorsque les concentrations de départ sont bien choisies, le tracé de la variation d'enthalpie en fonction du rapport molaire à chaque point d'injection donne une courbe sigmoïde qui permet d'obtenir par ajustement d'un modèle mathématique les valeurs de  $K_A$  (la constante d'équilibre d'association ;  $K_A = 1 / K_D$ ) et  $n$  (la stoechiométrie) associées à l'interaction. La valeur de  $K_A$  correspond au coefficient directeur de la tangente au point d'inflexion, tandis que la stoechiométrie correspond au rapport molaire auquel le point d'inflexion apparaît. La variation d'enthalpie libre associée à l'interaction ( $\Delta G$ ) peut être calculée avec l'équation  $\Delta G = - RT \ln(K_A)$  où  $R$  est la constante des gaz parfaits ( $R = 1,987 \text{ cal.K}^{-1}.\text{mol}^{-1}$ ) et  $T$  la température. Finalement, la variation d'entropie ( $\Delta S$ ) est calculée à partir de la relation  $\Delta G = \Delta H - T\Delta S$ . La titration calorimétrique est la seule technique à ce jour qui puisse fournir tous ces paramètres thermodynamiques en une expérience, et donc de caractériser en détails des interactions complexes.

Dans nos conditions expérimentales (cf. Matériel et méthodes), la titration calorimétrique de TRF2 par RAP1 produit de façon reproductible une courbe à deux transitions sigmoïdes. D'autre part, lorsque nous titrons TRF2 par le domaine C-terminal isolé de RAP1 (RAP1-RCT) nous n'observons que la transition majeure dans la courbe de titration (figure suivante), en accord avec les données déjà disponibles dans la littérature (Chen *et al*, 2011; Janoušková *et al*, 2015).

## RAP1 vs TRF2



## RAP1-RCT vs TRF2



**Figure 25 : Titration calorimétrique de TRF2 par RAP1 (panneau de gauche) et par RAP1-RCT (panneau de droite)**

La comparaison des différents paramètres thermodynamiques de ces interactions (tableau suivant) nous permet de confirmer que la transition majeure correspond bien à la liaison du domaine RCT de RAP1 à la région RBM de TRF2.

**Tableau 3 : Comparaison des paramètres thermodynamiques des interactions TRF2/RAP1-RCT et TRF2/RAP1 (a : transition majeure ; b : transition mineure)**

Paramètre	TRF2/RAP1-RCT	TRF2/RAP1
$K_A$ ( $M^{-1}$ )	$2.10^{10}$	$2.10^{10}$ (a) $2.10^6$ (b)
$\Delta G$ (kcal/mol)	-13,2	-13,2 (a) -8,1 (b)
$\Delta H$ (kcal/mol)	-7,4	-6,2 (a) 1 (b)
$T\Delta S$ (kcal/mol)	6,1	7,07 (a) 9,1 (b)

Nous obtenons les  $K_D$  suivants :

- TRF2/RAP1-RCT :  $K_D \simeq 0,5$  nM (la valeur est approximative car il y a peu de points dans la région d'inflexion de la sigmoïde),
- TRF2/RAP1 :  $K_D \simeq 0,5$  nM pour la transition majeure, et  $K_D \simeq 0,5$   $\mu$ M pour la transition mineure (la valeur est approximative car la transition mineure est en partie masquée par la transition majeure).

Ce phénomène montre que l'interaction entre TRF2 et RAP1 entières est plus complexe qu'un seul événement de liaison (nous l'appellerons l'interaction complète). Les observations par AFM des complexes TRF2/ADN, RAP1/TRF2/ADN et RAP1-RCT/TRF2/ADN, suggèrent quant à elles que la région N-terminale de RAP1 intervient dans les effets observés sur les propriétés de TRF2 (cf. introduction de la partie et article en annexe).

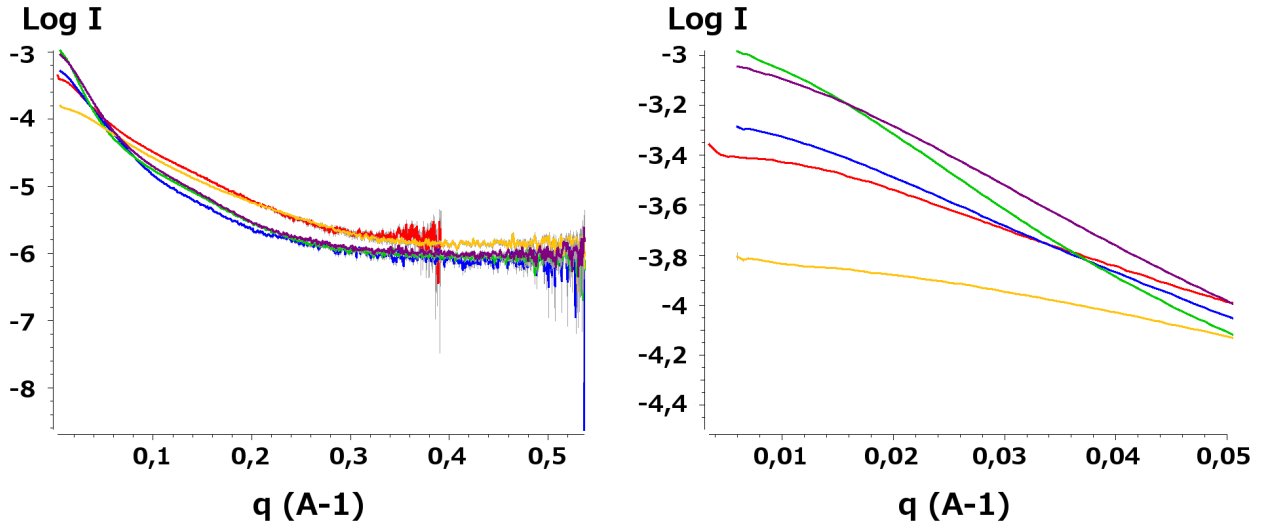
### 1.1.2 L'interaction est associée à des changements de conformation

Pour compléter cette information, nous avons voulu mieux comprendre l'interaction complète entre TRF2 et RAP1 d'un point de vue structural. La cristallographie n'est pas adaptée à l'étude de ces protéines entières car la flexibilité conformationnelle conférée par leurs longues régions non structurées les rend difficiles voire impossibles à cristalliser, et une structure cristallographique ne correspondrait de toute façon qu'à une seule conformation possible. D'autre part nous avons besoin d'une technique applicable sur des objets en solution, afin d'être dans des conditions expérimentales les plus proches possible de celles des mesures d'ITC dans lesquelles nous avons pu détecter l'interaction complète. Deux techniques de biologie structurale répondent bien à cette contrainte : la diffusion des rayons X aux petits angles (SAXS, pour *small-angle X-ray scattering*) et la résonance magnétique nucléaire (RMN). Nous avons utilisé le SAXS car cette technique est bien adaptée à la caractérisation de protéines et complexes flexibles (Rambo & Tainer, 2010). Elle est aussi plus facile et rapide à mettre en place car elle ne nécessite pas de marquage, et elle n'est pas limitée par la taille des protéines à étudier, contrairement à la RMN (la limite actuellement admise est autour de 30 à 50 kDa ; le complexe TRF2/RAP1 a une masse d'environ 200 kDa).

Un prérequis indispensable à une expérience de SAXS est la préparation d'une solution de la macromolécule d'intérêt concentrée (de l'ordre de 1 mg/mL) et bien homogène. La diffusion des rayons X est en effet sensible à la taille des objets présents dans la solution, et son intensité augmente non linéairement avec la taille. Ainsi une espèce de grande taille, comme par exemple un agrégat de la protéine étudiée, affecte fortement la diffusion même si cette espèce représente une population minoritaire dans la solution. La ligne de lumière SWING du synchrotron SOLEIL, où nous avons réalisé les acquisitions, dispose d'un système de SEC-SAXS (chromatographie d'exclusion stérique couplée à la mesure de diffusion des

rayons X aux petits angles, voir David & Pérez, 2009; Pérez & Koutsioubas, 2015) afin d'optimiser l'homogénéité de l'échantillon juste avant la mesure. Nous avons tiré parti de cette installation, et nous nous sommes aussi assurés au préalable de la haute pureté de nos préparations de TRF2 et RAP1 (cf. Matériel et méthodes, purification des protéines recombinantes).

Les courbes de SAXS des protéines TRF2, RAP1, RAP1-RCT et des complexes TRF2/RAP1 et TRF2/RAP1-RCT sont présentées dans la figure suivante :



**Figure 26 : Courbes de SAXS de TRF2 (bleu), RAP1 (rouge), RAP1-RCT (jaune), TRF2/RAP1 (vert) et TRF2/RAP1-RCT (violet). Le panneau de droite montre un agrandissement dans la région des petits angles.**

À partir de la courbe expérimentale ( $I = f(q)$ , souvent représentée comme  $\ln I = f(q)$ ), l'analyse de Guinier (1939) permet d'obtenir les valeurs du rayon de giration ( $R_g$ ) et de l'intensité de diffusion à l'angle nul ( $I(0)$ , aussi notée  $I_0$ ) à l'aide de la formule suivante (où  $q$  est l'angle de diffusion et  $I(q)$  l'intensité mesurée à cet angle) :

$$I(q) = I(0) \times \exp\left(-\frac{1}{3} R_g^2 q^2\right)$$

Cette formule se linéarise comme suit :

$$\ln I(q) = \ln I(0) - \frac{1}{3} R_g^2 q^2$$

Ainsi en représentant la courbe expérimentale comme  $\ln I(q) = f(q^2)$ , la portion de courbe aux très petits angles peut être modélisée par une droite d'ordonnée à l'origine  $\ln I(0)$  et de coefficient directeur  $-\frac{1}{3} R_g^2$ .

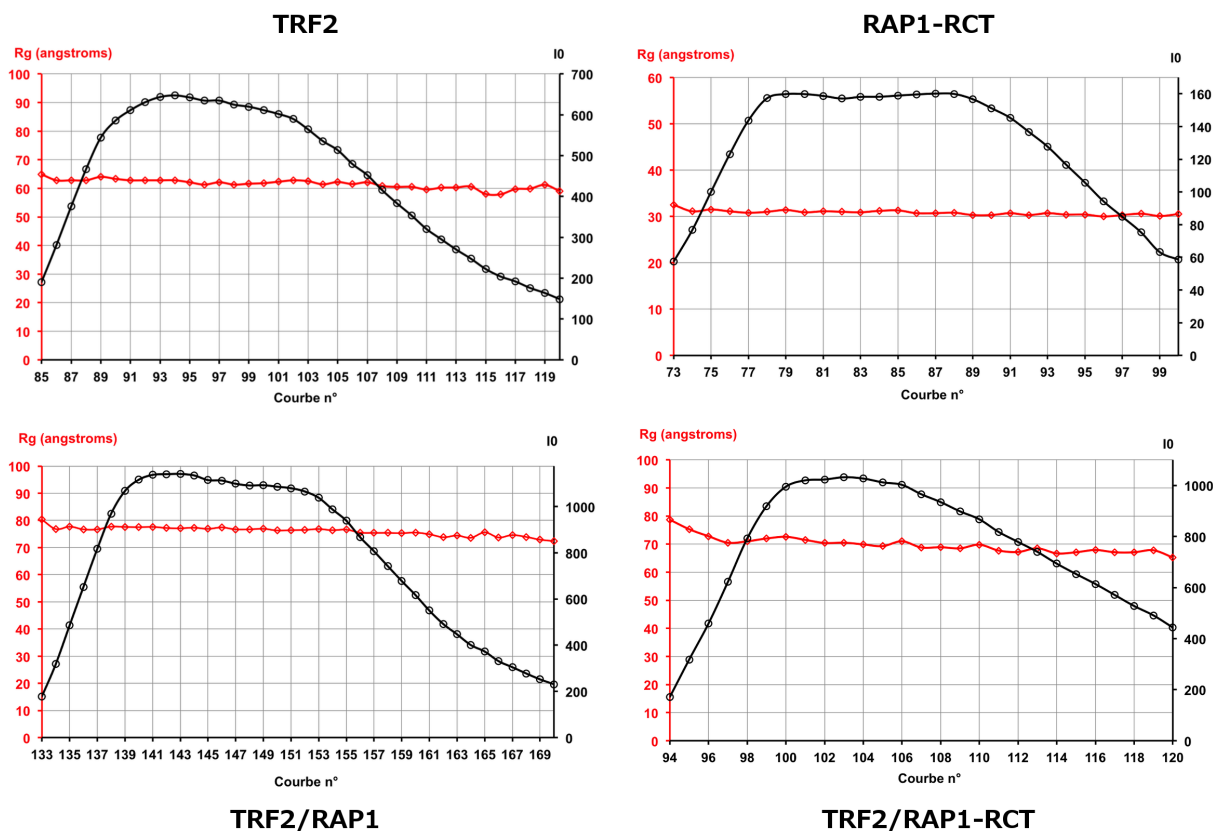
La valeur de  $I_0$  permet de calculer la masse moléculaire avec la formule suivante (Mylonas & Svergun, 2007) :

$$M_W = \frac{I_0 \times N_A}{c \times \Delta\rho_M^2}$$

où  $M_W$  est la masse moléculaire en Da,  $c$  est la concentration en protéine en mg/mL,  $I_0$  est l'intensité diffusée à l'angle nul en  $\text{cm}^{-1}$  (mise sur une échelle absolue en multipliant par le rapport  $(I_0)_{\text{eau, théorique}} / (I_0)_{\text{eau, expérimentale}}$  où  $(I_0)_{\text{eau, théorique}} = 1,632 \times 10^{-2}$  et  $(I_0)_{\text{eau, expérimentale}}$  sont les intensités diffusées à l'angle nul théorique et expérimentale de l'eau pure),  $N_A = 6,02 \times 10^{23} \text{ mol}^{-1}$  est le nombre d'Avogadro et  $\Delta\rho_M$  est le contraste de densité électronique entre la macromolécule et le tampon en  $\text{cm.g}^{-1}$  (pour les protéines,  $\Delta\rho_M \approx 2.10^{10} \text{ cm.g}^{-1}$ ).

La masse moléculaire peut aussi être calculée sans connaître la concentration en protéine (qui n'est pas mesurable lors des acquisitions couplées à la chromatographie d'exclusion stérique) par la formule suivante (Fischer *et al*, 2010) :  $M_W = 2\pi^2 \times \frac{I_0}{Q} \times \rho$  où  $I_0$  est sur une échelle relative (il n'est pas nécessaire de mesurer la diffusion de l'eau pour calibrer l'échelle d'intensité),  $Q = \int_0^\infty I(q)q^2 dq$  et  $\rho$  est la masse volumique de la protéine.

Lors des acquisitions couplées à la chromatographie d'exclusion stérique (qui a été utilisée pour toutes les protéines et complexes sauf RAP1, cf. Matériel et méthodes), plusieurs courbes de diffusion sont enregistrées tout le long du chromatogramme. L'analyse de Guinier est réalisée sur chacune de ces courbes pour déterminer  $R_g$  et  $I_0$ . Nous pouvons ensuite tracer les valeurs de ces deux paramètres en fonction du numéro de la courbe : le profil de  $I_0$  suit celui du chromatogramme puisque cette grandeur est proportionnelle à la concentration de l'échantillon. Les profils de  $R_g$  et  $I_0$  le long du chromatogramme sont représentés dans la figure suivante :

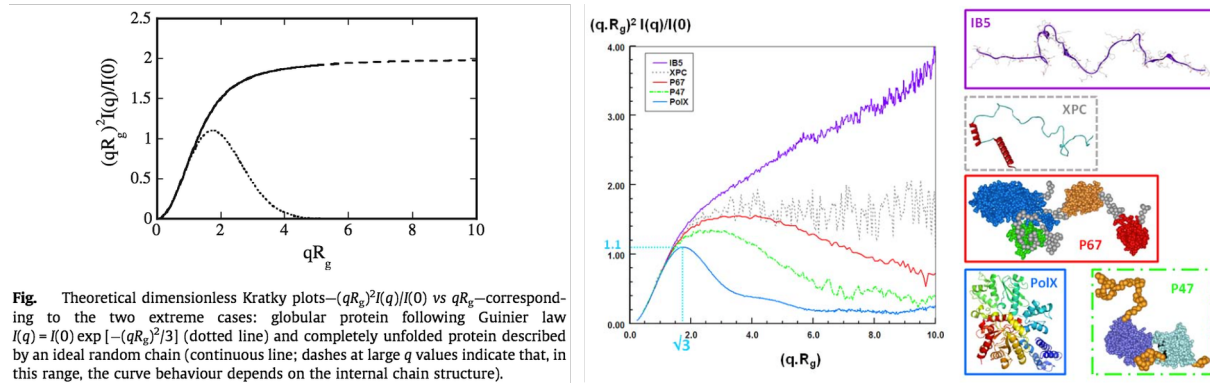


**Figure 27 : Profils de  $R_g$  (en rouge) et  $I_0$  (en noir) le long du chromatogramme de TRF2 (en haut à gauche), RAP1-RCT (en haut à droite), le complexe TRF2/RAP1 (en bas à gauche) et le complexe TRF2/RAP1-RCT (en bas à droite).**

La variation de  $R_g$  le long du chromatogramme nous renseigne sur l'homogénéité de l'échantillon, et permet de sélectionner uniquement les courbes donnant des valeurs de  $R_g$  sensiblement identiques. L'acquisition couplée à la chromatographie d'exclusion stérique apporte une grande redondance de données, ainsi nous pouvons sélectionner les courbes sur des critères très stricts. Par exemple pour TRF2 nous avons sélectionné les courbes 100 à 106, car la comparaison de toutes les courbes a révélé une très légère tendance à l'agrégation dans le début du pic chromatographique. Le  $R_g$  quasiment constant à travers tout le chromatogramme indique néanmoins une grande homogénéité des échantillons, et confirme la quasi absence d'agrégation et le haut degré de pureté que nous avons vérifiés lors des purifications (voir les chromatogrammes d'exclusion stérique et les analyses par SDS-PAGE dans la partie Matériel et méthodes). Les courbes de diffusion finales présentées précédemment ont été obtenues en calculant la moyenne des courbes sélectionnées par cette procédure.

La courbe expérimentale et les paramètres  $R_g$  et  $I_0$  permettent de représenter les données sous une forme modifiée comme  $(qR_g)^2 I(q) / I_0 = f(qR_g)$ , appelée la représentation de Kratky normalisée (Durand *et al*, 2010). Cette représentation accentue aux grands angles

les différences entre une chaîne repliée de façon compacte et une chaîne dépliée de conformation complètement aléatoire. La figure suivante montre les représentations de Kratky normalisées théoriques pour une chaîne parfaitement globulaire et compacte et pour une chaîne complètement dépliée, ainsi que quelques exemples de représentations de Kratky de protéines réelles.



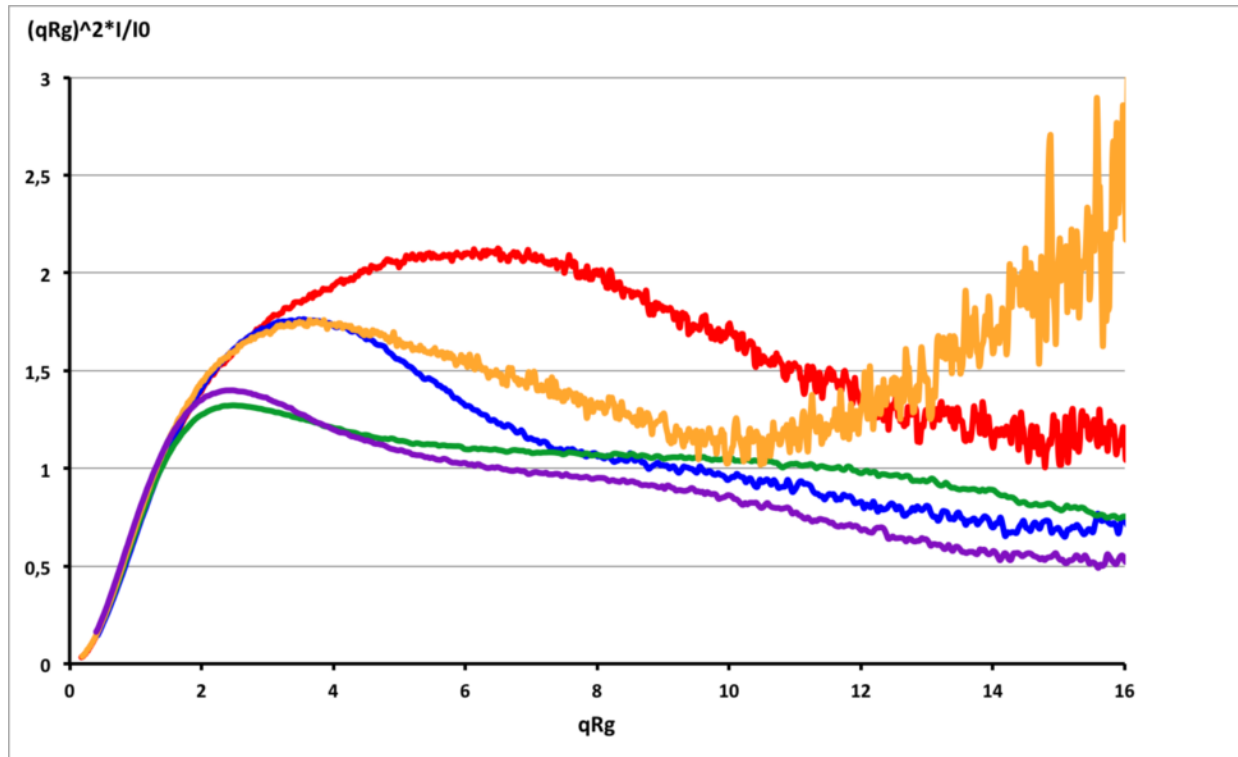
**Fig.** Theoretical dimensionless Kratky plots— $(qR_g)^2 I(q)/I(0)$  vs  $qR_g$ —corresponding to the two extreme cases: globular protein following Guinier law  $I(q) = I(0) \exp[-(qR_g)^2/3]$  (dotted line) and completely unfolded protein described by an ideal random chain (continuous line; dashes at large  $q$  values indicate that, in this range, the curve behaviour depends on the internal chain structure).

**Figure 28 : Représentations de Kratky normalisées théoriques d’une chaîne repliée compacte et d’une chaîne complètement dépliée (à gauche ; Durand *et al*, 2010). Représentations de Kratky normalisées de quelques protéines représentatives de différents degrés de structuration (à droite ; Receveur-Brechot & Durand, 2012) : globulaire compacte (bleu), multi-domaines compacte (vert), multi-domaines flexible (rouge) et dépliée (gris et violet).**

L’intérêt de cette représentation est qu’elle fournit directement à partir de la courbe expérimentale une information sur le degré de structuration de la protéine étudiée.

Les représentations de Kratky normalisées de nos protéines et complexes sont représentées dans la figure suivante.

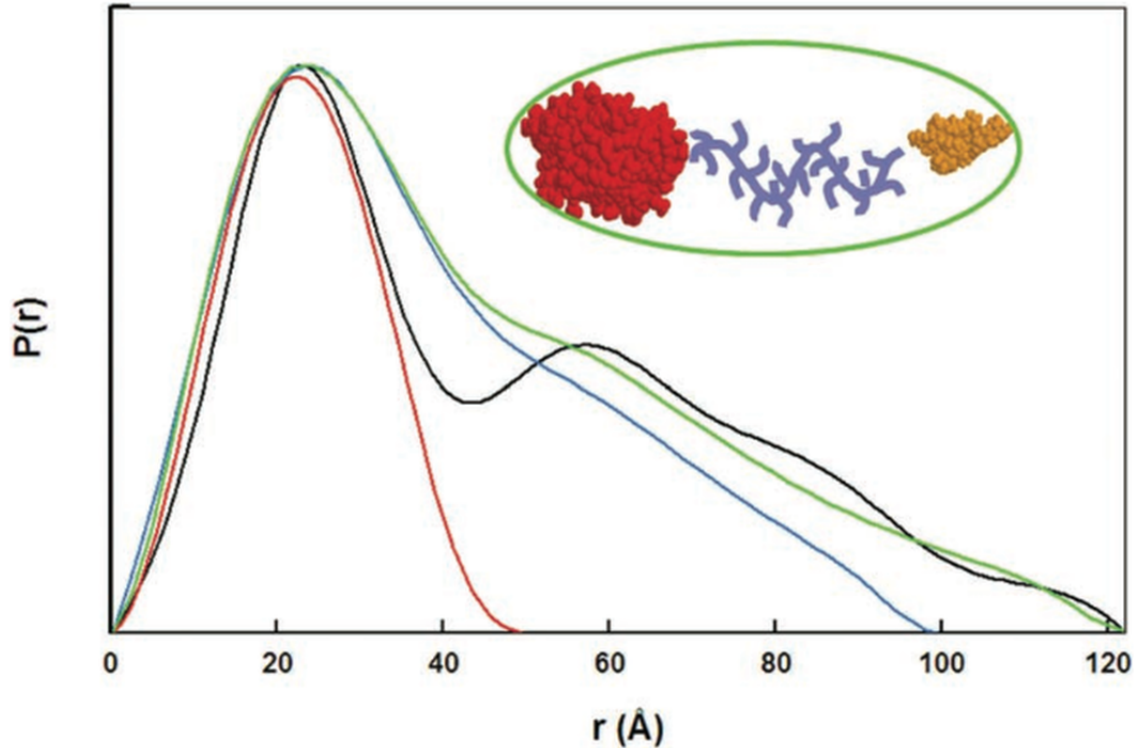




**Figure 29 : Représentations de Kratky normalisées de TRF2 (bleu), RAP1 (rouge), RAP1-RCT (jaune), TRF2/RAP1 (vert) et TRF2/RAP1-RCT (violet).**

Ces courbes nous indiquent que toutes les protéines et tous les complexes ont une conformation de type “collier de perles” correspondant à des domaines compacts connectés par des régions flexibles, ce qui confirme les prédictions concernant les régions non structurées des protéines.

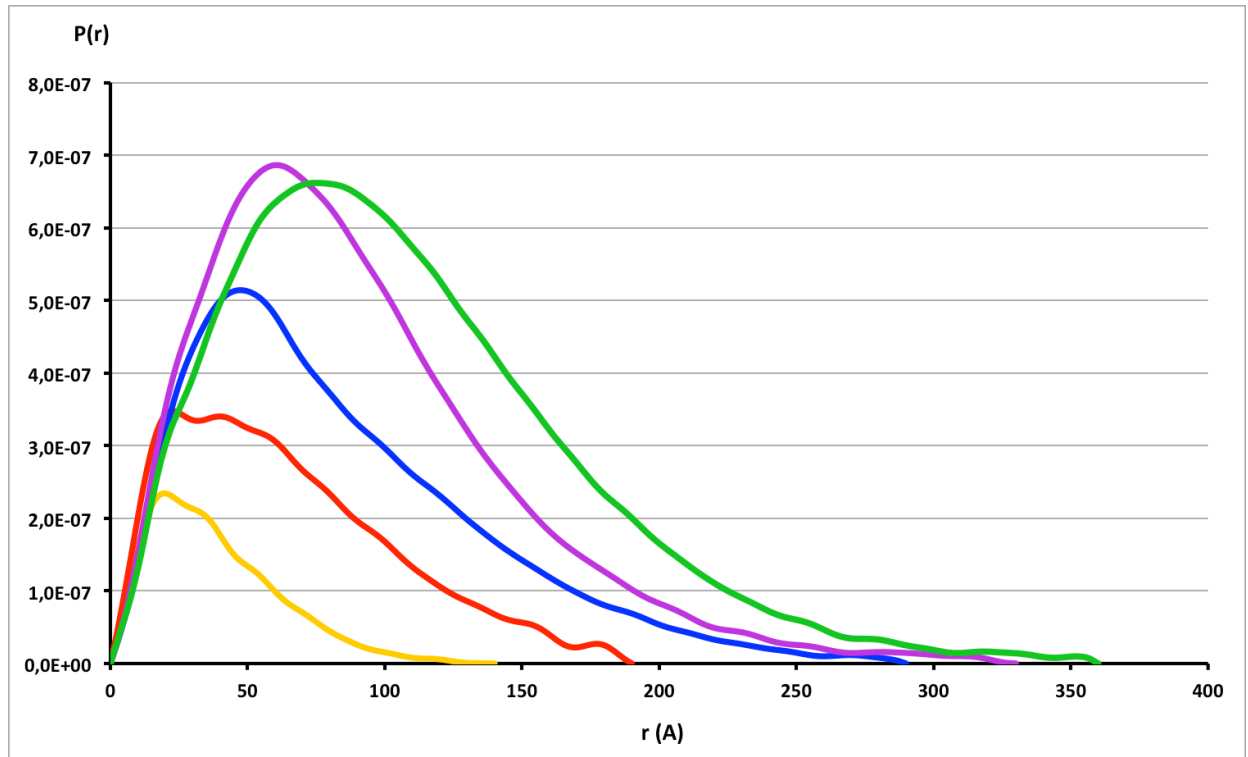
La courbe expérimentale de SAXS permet également de calculer la fonction de distribution des distances  $P(r)$ . Cette fonction s’obtient par la transformée de Fourier de la courbe de diffusion, et correspond à un histogramme des distances intramoléculaires. La fonction  $P(r)$  d’une protéine globulaire a la forme d’une courbe en cloche : son sommet fournit une autre estimation du  $R_g$  (obtenu à partir de l’ensemble des données, contrairement au  $R_g$  obtenu par l’analyse de Guinier qui n’utilise que les données aux très petits angles), et la valeur maximale de  $r$  pour laquelle  $P(r) > 0$  correspond au paramètre  $D_{max}$  : la plus grande distance présente dans la macromolécule. Comme le graphe de Kratky, la fonction  $P(r)$  contient des informations structurales intéressantes :



**Figure 30 : Formes caractéristiques de la fonction  $P(r)$  : en rouge pour un domaine globulaire, en bleu pour le domaine globulaire connecté à une extrémité dépliée, en vert pour la protéine à deux domaines connectés par une région flexible, en noir la même protéine à deux domaines avec un linker plus rigide (Receveur-Brechot & Durand, 2012).**

La fonction de distribution des distances nous renseigne moins sur la flexibilité des objets (contrairement au graphe de Kratky), mais son allure est caractéristique de la forme de l'objet. La fonction  $P(r)$  d'une protéine globulaire est une gaussienne quasi-parfaite. Pour une protéine de forme plus allongée, la  $P(r)$  est asymétrique et présente un épaulement du côté des grandes distances. Cet épaulement devient même un second pic dans le cas où la protéine comporte deux domaines distincts séparés par une région rigide.

Les fonctions de distribution des distances de nos protéines et complexes sont présentées dans la figure-ci dessous :



**Figure 31 : Fonctions de distribution des distances de TRF2 (bleu), RAP1 (rouge), RAP1-RCT (jaune), TRF2/RAP1 (vert) et TRF2/RAP1-RCT (violet).**

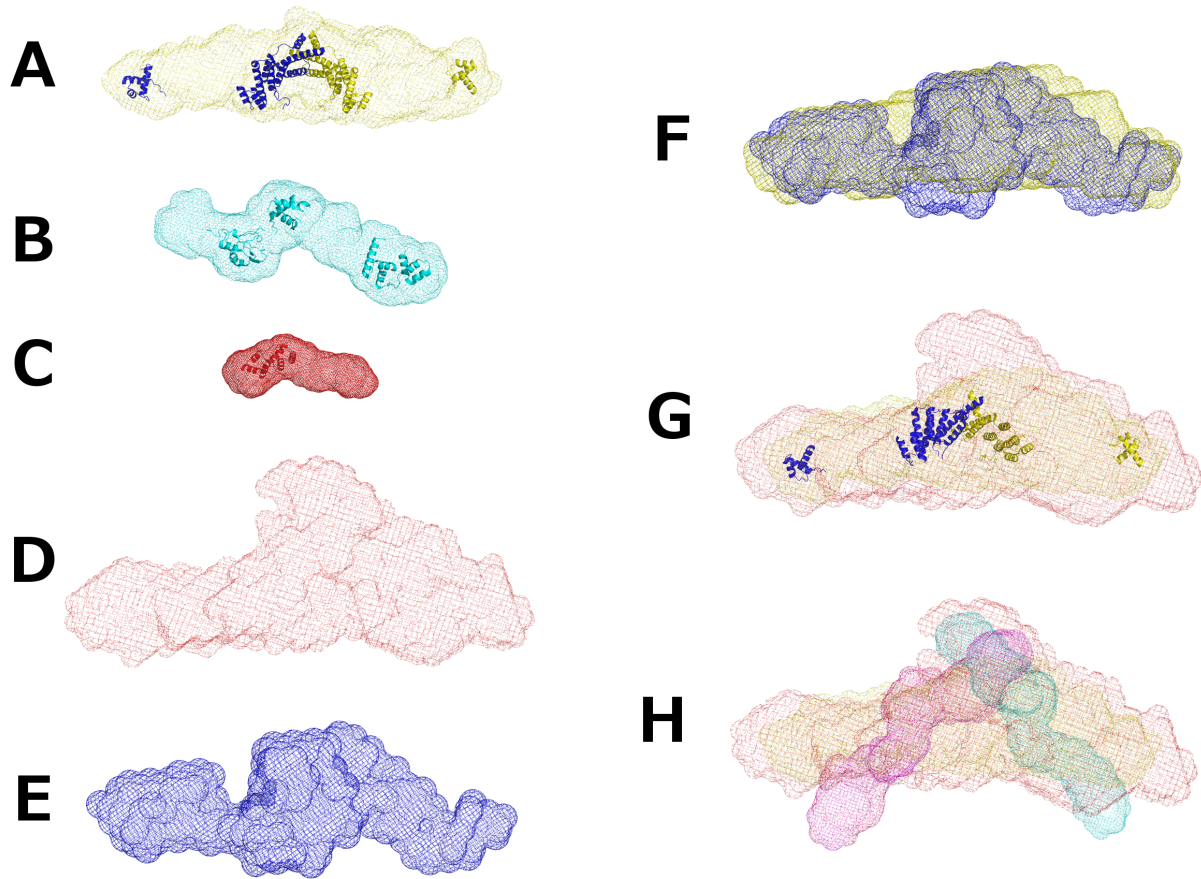
Ces courbes sont asymétriques, contrairement à la fonction  $P(r)$  attendue pour une protéine globulaire compacte, ce qui confirme la structuration en domaines connectés par des régions flexibles.

Les valeurs des masses moléculaires théoriques (calculées d'après la séquence des protéines) et expérimentales (calculées d'après la valeur de  $I_0$ ), et des paramètres  $R_g$  et  $D_{max}$  sont résumées dans le tableau suivant. Les valeurs de  $R_g$  présentées ici sont celles obtenues avec la fonction  $P(r)$  : elles sont plus grandes que celles obtenues par l'analyse de Guinier car cette dernière est basée sur l'approximation que les particules diffusant les rayons X sont globulaires (elle tend à sous-estimer le  $R_g$  pour des protéines flexibles, voir Receveur-Brechot & Durand, 2012).

**Tableau 4 :  $R_g$ ,  $D_{max}$  et masses moléculaires des protéines et des complexes (nd : non déterminé ; les valeurs de  $R_g$  et  $D_{max}$  sont obtenues par la fonction de distribution des distances)**

<b>Protéine ou complexe</b>	<b>Masse moléculaire théorique (kDa)</b>	<b>Masse moléculaire expérimentale (kDa)</b>	<b><math>R_g</math> (Å)</b>	<b><math>D_{max}</math> (Å)</b>
TRF2	56,6 x2	117	70,2	290
RAP1	44,5	42,5	53,1	190
RAP1-RCT	15,2	nd	32,0	130
TRF2/RAP1	56,6 x2 + 44,5 x2	200	85,4	345
TRF2/RAP1-RCT	56,6 x2 + 15,2 x2	nd	67,8	265

La fonction de distribution des distances sert ensuite de contrainte pour construire *ab initio* des enveloppes représentant la distribution statistique de la conformation tridimensionnelle en solution des protéines seules et des complexes. Les enveloppes de TRF2, RAP1, RAP1-RCT et des complexes TRF2/RAP1 et TRF2/RAP1-RCT sont présentées dans la figure suivante :



**Figure 32 : Enveloppes de TRF2 (A), de RAP1 (B), de RAP1-RCT (C), du complexe TRF2/RAP1 (D) et du complexe TRF2/RAP1-RCT (E). Superpositions des enveloppes du complexe TRF2/RAP1-RCT et de TRF2 (F), du complexe TRF2/RAP1 et de TRF2 (G), et du complexe TRF2/RAP1, de TRF2 et de deux RAP1 (H). Les domaines ont été positionnés manuellement dans les enveloppes et sont présentés comme simple repère de taille et proportions ; les superpositions d’enveloppes ont aussi été effectuées manuellement (avec le logiciel PyMOL ; Schrödinger LLC, 2010).**

En accord avec les graphes de Kratky, nous observons que TRF2 adopte une conformation étendue, avec une grande distance séparant ses domaines Myb (jusqu’à environ 230 Å ; figure ci-dessus, A), et que la conformation de RAP1 est également étendue (figure ci-dessus, B). Concernant RAP1-RCT, sa grande valeur de  $D_{\max}$  ainsi que son graphe de Kratky caractéristique d’une chaîne majoritairement dépliée peuvent s’expliquer par le fait que notre construction (réalisée avant la publication de la structure du domaine RCT par Chen *et al*, 2011) comporte en N-terminal 36 résidus de plus que le domaine minimal.

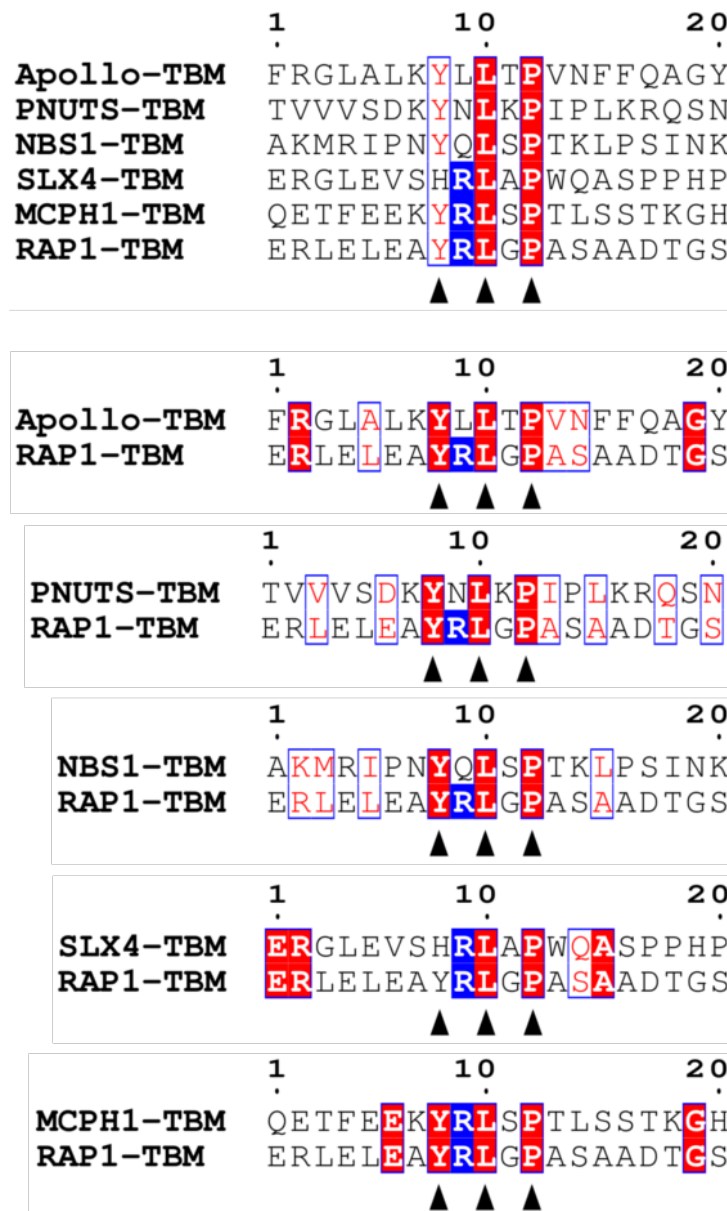
Puisque les enveloppes représentent une moyenne statistique de conformations, elles ont tendance à “écraser” les conformations minoritaires. En d’autres termes, la superposition de ces enveloppes deux à deux nous permet seulement de détecter des changements de

conformation globaux et qui se produisent pour l'ensemble des molécules présentes dans l'échantillon.

En comparant les enveloppes superposées de TRF2 et du complexe TRF2/RAP1 (figure ci-dessus, G), nous pouvons observer que le complexe TRF2/RAP1 est plus étendu que TRF2 seule, comme l'indique une plus grande valeur de  $D_{\max}$  pour le complexe TRF2/RAP1 comparé à TRF2 seule. Au contraire, la conformation du complexe TRF2/RAP1-RCT n'est pas significativement différente de celle de TRF2 seule, comme en témoignent les valeurs de  $R_g$  et  $D_{\max}$  relativement proches (voir aussi la figure ci-dessus, F). Enfin, de façon particulièrement intéressante, la comparaison de ces enveloppes révèle que celles de TRF2 et RAP1 ne peuvent pas se superposer à celle du complexe TRF2/RAP1 (figure ci-dessus, H : un exemple parmi tous nos essais), ce qui indique que la formation du complexe fait intervenir un changement de conformation majeur des deux protéines.

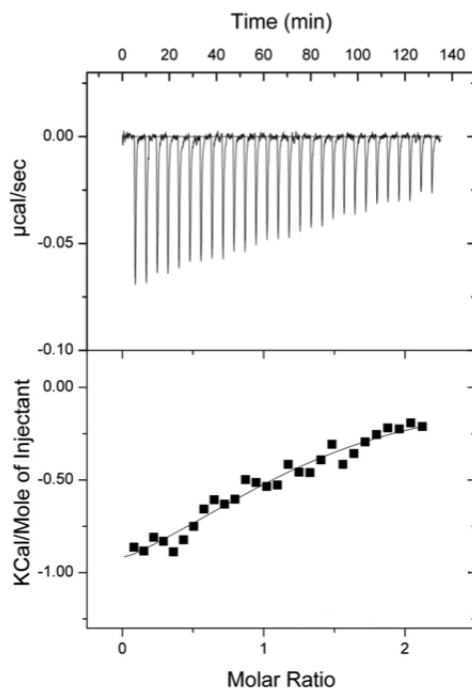
### 1.1.3 L'interaction entre TRF2 et RAP1 implique un motif TBM

À ce stade nous savons donc d'après les titrations calorimétriques que l'interaction entre TRF2 et RAP1 est plus complexe qu'un simple événement de liaison, et que cette interaction complexe dépend de la région N-terminale de RAP1 puisque nous n'observons qu'une simple liaison avec le domaine RCT isolé de RAP1. D'autre part nous savons d'après les enveloppes calculées avec les données de SAXS que cette interaction complexe est associée à un changement de conformation significatif des deux protéines. Nous avons donc cherché un autre site d'interaction potentiel entre TRF2 et RAP1. Puisque le domaine TRFH est un domaine d'interaction protéine-protéine par lequel TRF2 recrute des partenaires en liant un motif TBM (cf. Introduction), il est possible que ce domaine intervienne dans l'interaction avec RAP1 : nous avons donc cherché le motif consensus [YH]RLxP dans la séquence de RAP1 et l'avons trouvé dans la région N-terminale, aux résidus 99 à 103. La figure suivante montre l'alignement du motif YRLGP (TBM candidat) de RAP1 aux motifs TBM connus d'Apollo et NBS1 (Chen *et al*, 2008), de PNUIS et MCPH1 (Kim *et al*, 2009), et de SLX4 (Wan *et al*, 2013) :



**Figure 33 : Aligement du motif YRLGP de RAP1 avec les motifs TBM connus d’Apollo, NBS1, PNUTS, MCPH1 et SLX4. Les résidus du motif sont indiqués par les triangles noirs, le résidu R du motif consensus défini par Kim *et al* (2009) est surligné en bleu, les résidus identiques sont surlignés en rouge et les résidus similaires sont encadrés en bleu. Les alignements deux à deux sont aussi présentés.**

Ce motif de RAP1 est non seulement très similaire aux cinq motifs TBM validés expérimentalement déjà rapportés dans la littérature, mais sa séquence est également la plus proche de la séquence consensus. Il est situé en C-terminal du domaine BRCT au début d’une région prédite comme non structurée, ce qui suggère qu’il est suffisamment accessible pour être engagé dans une interaction. Nous avons voulu vérifier si ce motif était

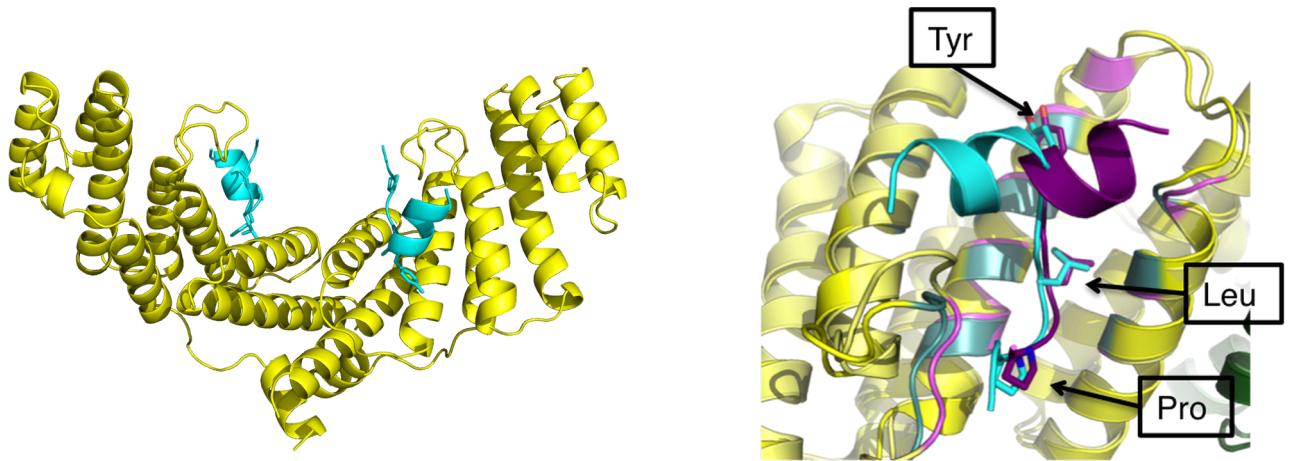


effectivement capable de lier le domaine TRFH. Pour cela nous avons testé par ITC l'interaction d'un peptide de RAP1 contenant ce motif YRLGP (résidus 89 à 106 de RAP1, appelés par la suite RAP1-TBM) avec le domaine TRFH de TRF2 (figure suivante) : dans nos conditions expérimentales, nous obtenons un  $K_D \approx 13 \mu\text{M}$ .

**Figure 34 : Titration calorimétrique du domaine TRFH de TRF2 par le peptide RAP1-TBM**

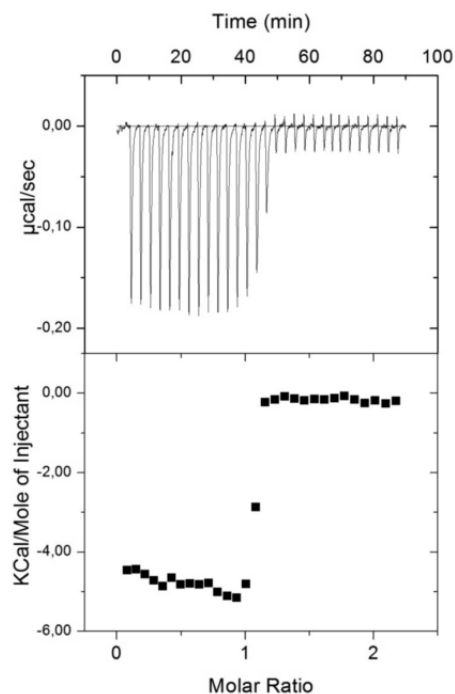
La structure cristallographique à 2,44 Å du complexe TRFH/RAP1-TBM a par ailleurs été résolue au laboratoire, par remplacement moléculaire avec comme modèle un monomère du domaine TRFH (entrée PDB 3BUA). Les statistiques de collecte des données, de résolution de la structure et d'affinement du modèle sont disponibles dans l'article présenté en annexe. Le modèle final est représenté dans la figure suivante :





**Figure 35 : Structure cristallographique du complexe TRFH/RAP1-TBM. Vue d'ensemble du complexe (à gauche). Vue détaillée du site de liaison du motif TBM (à droite), en superposition avec la structure du complexe contenant le TBM d'Apollo. Le peptide RAP1-TBM est coloré en cyan, le peptide Apollo-TBM est coloré en violet, le domaine TRFH est coloré en jaune (ses résidus engagés dans l'interaction avec les motifs TBM sont colorés de la même couleur que les motifs).**

Afin de confirmer que ce motif TBM est bien engagé dans l'interaction entre RAP1 et TRF2, nous avons muté en alanine les trois résidus Y, L et P dans la séquence YRLGP de RAP1 entière, donnant le motif muté ARAGA. Ce mutant est appelé RAP1-ARAGA. Nous avons mesuré l'interaction de ce mutant avec TRF2 par ITC (figure suivante) :

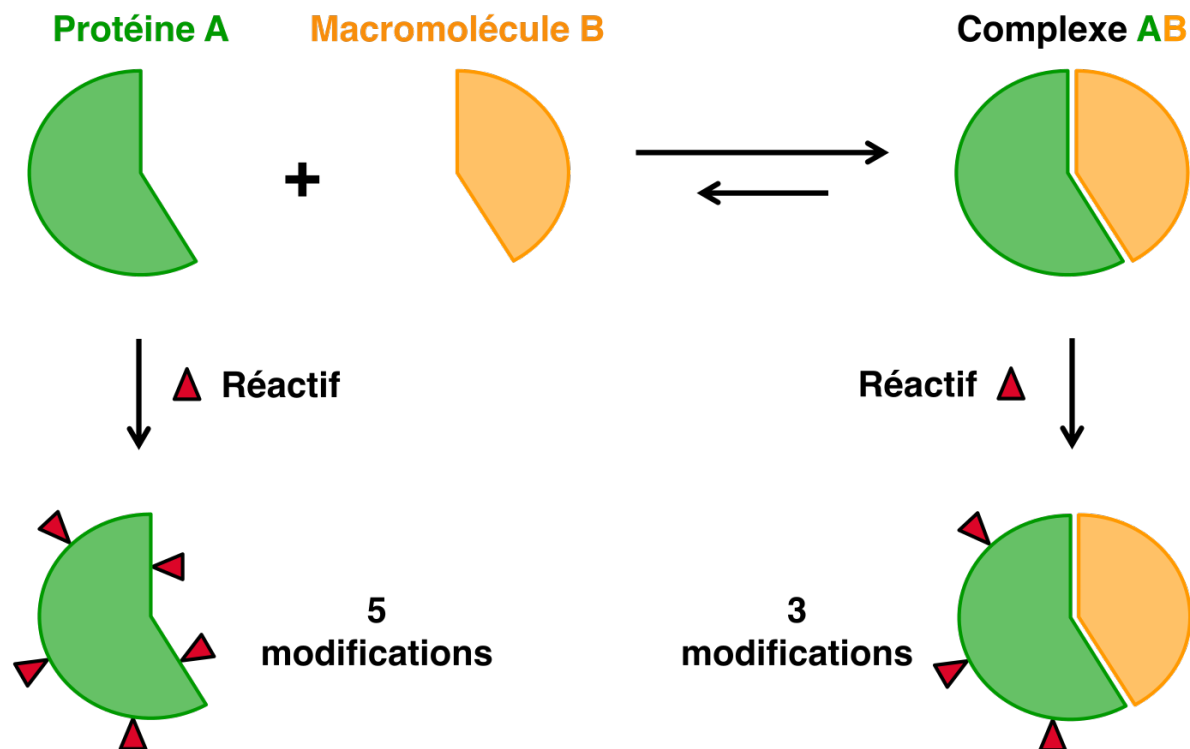


**Figure 36 : Titration calorimétrique de TRF2 par RAP1-ARAGA**

Comme attendu, nous retrouvons la transition majeure correspondant à l'interaction de haute affinité ( $K_D \approx 0,5$  nM) entre le domaine RCT de RAP1 et TRF2. Sans être totalement abolie, la transition mineure est nettement réduite, et il n'est plus possible de calculer le  $K_D$  correspondant. Ceci démontre que le motif TBM de RAP1 participe bien à l'interaction complète avec TRF2, même si le signal résiduel suggère que la surface d'interaction est plus étendue.

#### **1.1.4 Les surfaces d'interaction entre TRF2 et RAP1 ne sont pas limitées au motif TBM**

Nous connaissons maintenant la région minimale des deux protéines impliquée dans l'interaction complète : le motif TBM de RAP1 et la poche de liaison des motifs TBM dans le domaine TRFH de TRF2. Cependant, cette interaction complète affecte probablement des régions plus grandes dans les protéines entières. Afin de déterminer quelles sont ces régions, nous avons réalisé des expériences d'empreinte protéique par acétylation des lysines. Le principe de cette méthode est schématisé dans la figure suivante.



**Figure 37 : Principe de l'expérience d'empreinte protéique**

La méthode d'empreinte protéique repose sur la possibilité d'identifier par spectrométrie de masse des lysines modifiées. En comparant les spectres de masse de la protéine traitée et non traitée par le réactif, nous pouvons identifier les lysines acétylées, c'est-à-dire celles qui sont accessibles au réactif, donc exposées en surface de la protéine. Le même traitement est ensuite réalisé sur la protéine en complexe avec son partenaire, et permet d'identifier les lysines non acétylées (ou moins acétylées) en présence du partenaire, c'est-à-dire celles qui sont protégées par la présence du partenaire, donc engagées dans l'interaction (figure ci-dessus).

Cette méthode a été utilisée avec succès sur des complexes protéine/ADN (voir par exemple Kvaratskhelia *et al*, 2002) et est complémentaire des techniques d'empreinte sur l'ADN (par exemple l'empreinte à la DNase I) puisqu'elle indique l'empreinte de l'ADN sur la protéine. Dans le cas d'un complexe protéine/protéine comme le complexe TRF2/RAP1, nous avons dû réaliser deux expériences : une expérience en présence d'un excès de RAP1 pour analyser TRF2 en étant certain qu'elle est engagée dans le complexe à 100 %, et l'expérience réciproque en présence d'un excès de TRF2 pour analyser RAP1 en étant certain qu'elle est engagée dans le complexe à 100 %. Du fait du nombre de lysines différent dans les deux protéines, ces excès de l'une ou l'autre protéine conduisent à des nombres de lysines différents dans les différentes expériences. Pour en tenir compte, nous avons donc dû ajuster la concentration en réactif d'une expérience à l'autre (le réactif doit toujours être

en excès par rapport au nombre total de lysines). Les détails de ces ajustements sont indiqués dans la partie Matériel et méthodes.

Les lysines présentes dans TRF2 (44 par monomère) et RAP1 (25) sont indiquées dans la figure suivante, avec leur localisation spatiale pour celles situées dans les domaines dont une structure tridimensionnelle est disponible.



**Figure 38 : Positions des lysines dans les séquences et structures de TRF2 et RAP1 (les lysines sont indiquées en rouge)**

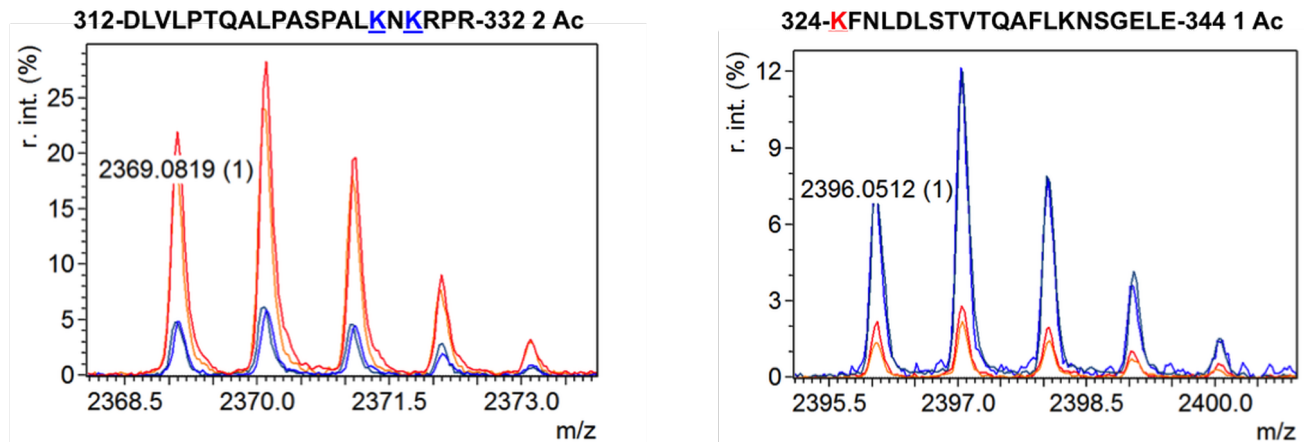
Pour chacune des deux protéines, les lysines sont réparties dans toute la séquence (sauf dans le domaine B de TRF2) et sont globalement exposées en surface des domaines repliés, ce qui les rend bien adaptées à notre expérience.

Nous avons comparé les profils d'acétylation des lysines des complexes TRF2/RAP1 et TRF2/RAP1-ARAGA avec ceux des protéines libres TRF2, RAP1 et RAP1-ARAGA. Nous détectons plusieurs types de lysines :

1. des lysines acétylées dans toutes les conditions, ce qui indique une région constamment accessible et non engagée dans l'interaction ;
2. des lysines acétylées dans les protéines libres mais non acétylées dans le complexe (indifféremment entre RAP1 et RAP1-ARAGA), indiquant des régions impliquées dans l'interaction ;
3. des lysines acétylées dans les protéines libres et dans le complexe TRF2/RAP1-ARAGA, mais non acétylées dans le complexe TRF2/RAP1, qui indiquent une région d'interaction affectée par la mutation ;

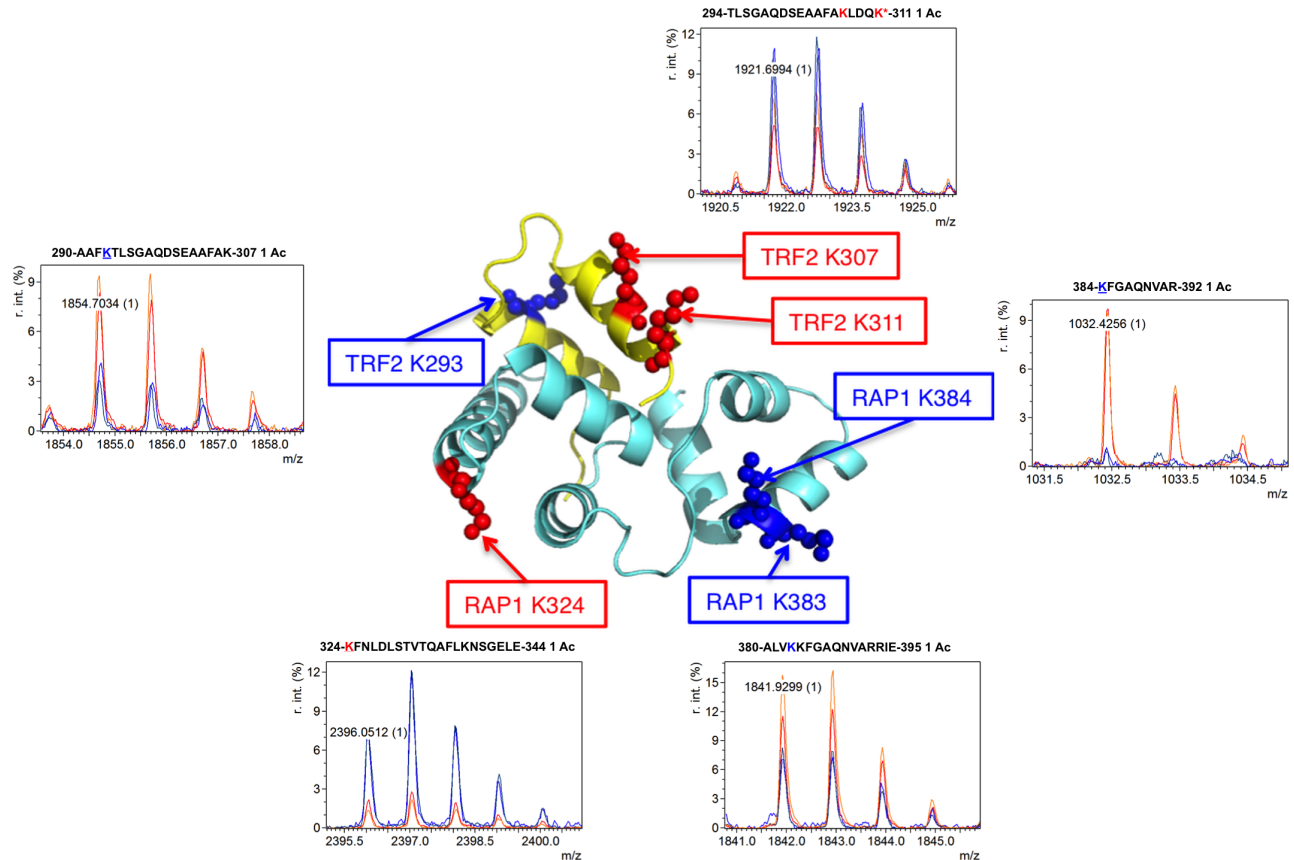
- des lysines non acétylées dans les protéines libres mais acétylées dans les complexes, pouvant correspondre à des régions indirectement affectées par l'interaction qui se trouvent exposées suite au changement de conformation.

La figure suivante indique comment nous avons interprété les spectres de masse :



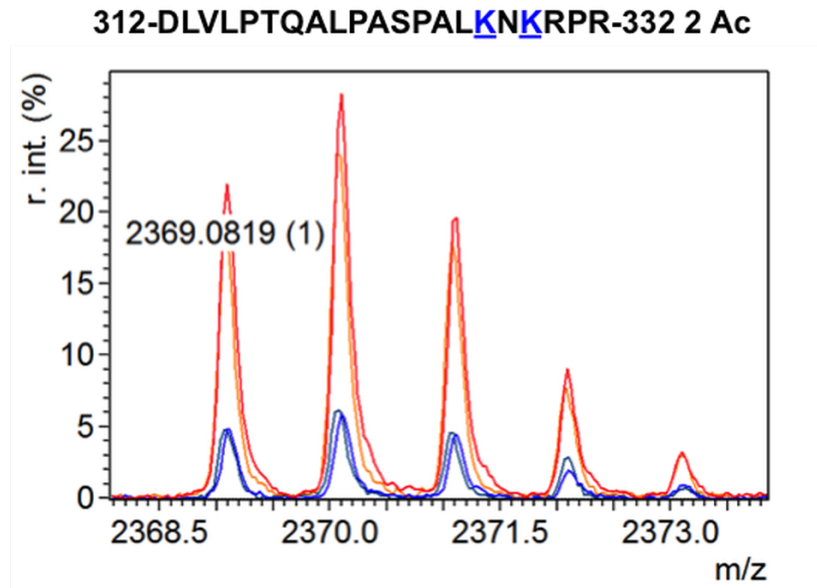
**Figure 39 : Interprétation des spectres de masse. Les spectres orange et rouge (deux réplicats) correspondent à la protéine libre ; les spectres bleu clair et foncé (deux réplicats) correspondent à la protéine dans le complexe. Les fragments des protéines à l'origine des pics présentés sont indiqués au-dessus des spectres, avec le nombre d'acétylations. À gauche, exemple de lysines protégées dans le complexe : le fragment acétylé est environ 4 fois moins abondant dans l'échantillon obtenu à partir de la protéine en complexe que dans celui obtenu à partir de la protéine libre. À droite, exemple d'une lysine plus accessible dans le complexe : le fragment acétylé est environ 3 fois plus abondant dans l'échantillon obtenu à partir de la protéine en complexe que dans celui obtenu à partir de la protéine libre.**

Comme attendu, nous pouvons détecter l'interaction de haute affinité déjà décrite entre le domaine RAP1-RCT et la région TRF2-RBM, ce qui valide notre méthode. La lysine K293 de TRF2 (dans la région RBM) est protégée dans le complexe, tandis que les lysines K307 et K311 sont plus accessibles dans le complexe. Du côté de RAP1, les lysines K383 et K384 sont protégées dans le complexe, tandis que la lysine K324 est plus accessible dans le complexe. Ces résultats sont résumés dans la figure suivante :



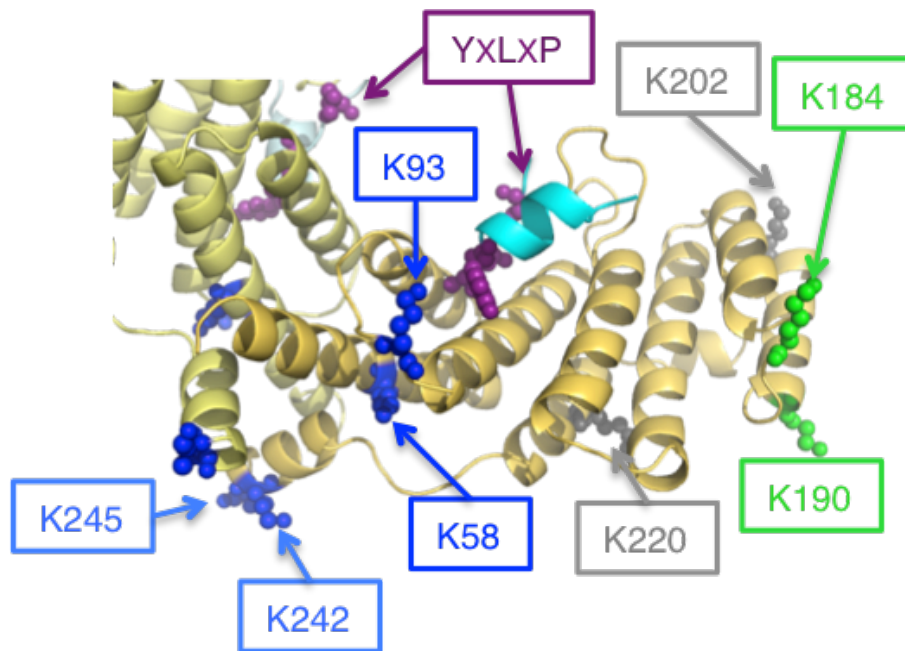
**Figure 40 : Empreintes de TRF2 sur RAP1 et de RAP1 sur TRF2 dans la région de l'interaction de haute affinité (TRF2-RBM et RAP1-RCT). TRF2 est représentée en jaune, RAP1 en cyan. Les lysines sont représentées en sphères, celles qui sont protégées dans le complexe sont colorées en bleu, celles qui sont plus accessibles dans le complexe sont colorées en rouge. La figure a été réalisée avec l'entrée PDB 3K6G. Les pics issus des spectres de masse (deux répliqués) sont indiqués pour chaque lysine (spectres orange et rouge : protéine libre ; spectres bleus : protéine dans le complexe). Le fragment de la protéine correspondant à chaque pic est indiqué au-dessus avec le nombre d'acétylations et les lysines représentées avec le même code couleur que dans la structure.**

Les lysines K327 et K329 de TRF2 sont aussi protégées dans le complexe (cf. figure suivante), ce qui suggère que le motif RBM de TRF2 dans le contexte des protéines entières est en réalité plus long d'une vingtaine de résidus que le motif minimal présent dans l'entrée PDB 3K6G (Chen *et al*, 2011).



**Figure 41 : Protection des lysines K327 et K329 de TRF2 dans le complexe TRF2/RAP1. Les spectres orange et rouge (deux réplicats) correspondent à la protéine libre ; les spectres bleu clair et foncé (deux réplicats) correspondent à la protéine dans le complexe. Le fragment de la protéine à l'origine du pic présenté est indiqué au-dessus des spectres, avec le nombre d'acétylations.**

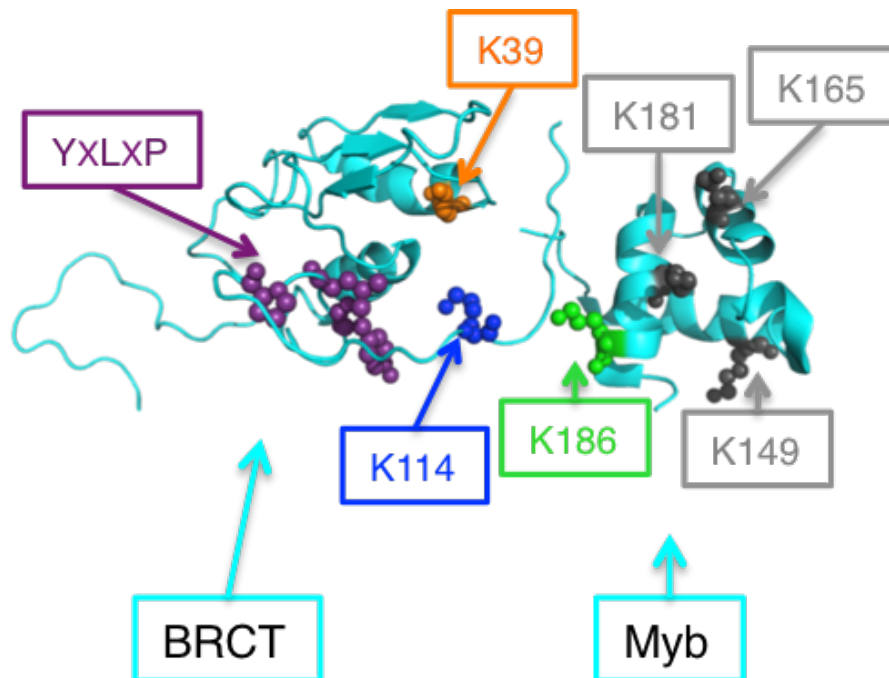
Le domaine TRFH contient 18 lysines par monomère (cf. Figure 38), dont 8 pour lesquelles nous avons pu détecter l'acétylation. Pour six d'entre elles l'acétylation est diminuée en présence de RAP1 : K242 et/ou K245 situées à la base du domaine TRFH, K58 et K93 situées proches de la poche de liaison du motif TBM, et K184 et K190 situées plus loin de cette poche. Il est intéressant de noter que K184 et K190 sont protégées de l'acétylation uniquement par RAP1, et ne sont pas protégées par RAP1-ARAGA, ce qui suggère que l'engagement de cette région du domaine TRFH dans l'interaction dépend de la liaison du motif TBM de RAP1. Le fait que les autres lysines (K242, K245, K58 et K93) soient protégées de la même façon par RAP1 et RAP1-ARAGA est en accord avec les mesures d'ITC qui montrent que la mutation n'abolit pas totalement l'interaction secondaire. Les lysines K202 et K220, situées sur l'autre face du domaine par rapport aux précédentes, sont acétylées indifféremment en présence ou absence de RAP1, ce qui indique que cette face du domaine n'est pas engagée dans l'interaction avec RAP1. L'analyse détaillée des spectres de masse est présentée dans les figures supplémentaires de l'article en annexe. L'empreinte de RAP1 sur le domaine TRFH est représentée dans la figure suivante :



**Figure 42 : Empreinte de RAP1 sur le domaine TRFH de TRF2. Le motif TBM de RAP1 est représenté à sa position dans le complexe, coloré en cyan avec les résidus Y, L et P en sphères violettes. Les lysines sont représentées en sphères, celles qui sont protégées sont colorées en bleu (bleu clair lorsqu'il existe une ambiguïté sur la position de l'acétylation due à la présence de plusieurs lysines dans le peptide analysé), celles qui sont non affectées sont colorées en gris. Les lysines colorées en vert sont protégées en présence de RAP1 sauvage, mais pas en présence de RAP1-ARAGA.**

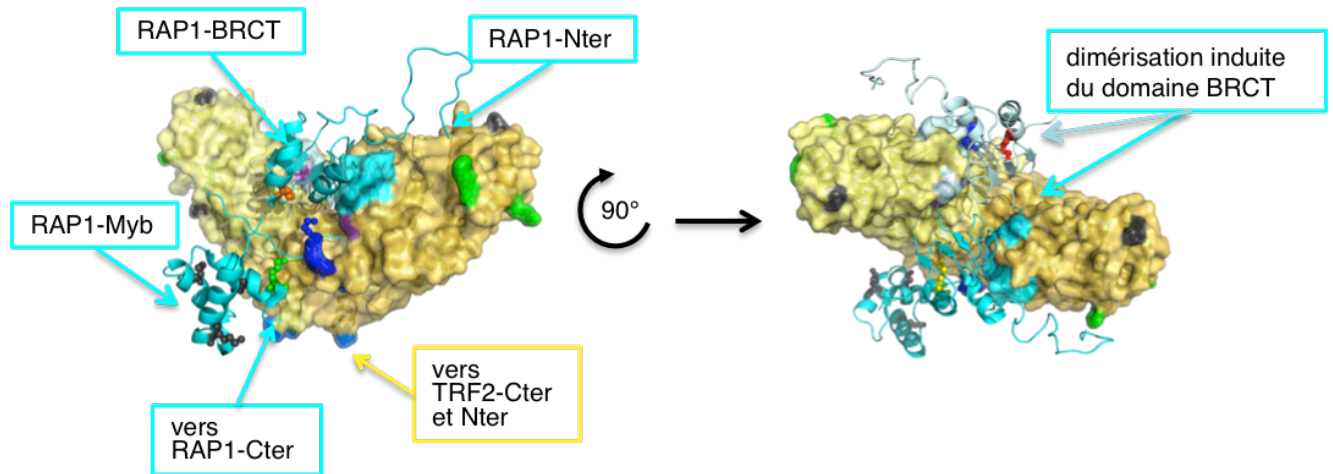
La région N-terminale de RAP1 (résidus 1 à 230) contient 13 lysines dont 6 pour lesquelles nous pouvons détecter une acétylation. Pour trois d'entre elles l'acétylation est affectée par la formation du complexe. L'analyse détaillée des spectres de masse est présentée dans les figures supplémentaires de l'article en annexe. L'empreinte de TRF2 sur la région N-terminale de RAP1 est représentée dans la figure suivante :





**Figure 43 : Empreinte de TRF2 sur la région N-terminale de RAP1. Le domaine BRCT est un modèle par homologie construit d'après la structure du domaine équivalent chez *S.cerevisiae* (entrée PDB 2L42). Les résidus du motif TBM sont représentés en sphères violettes. Les lysines sont représentées en sphères, K114 protégée dans le complexe est colorée en bleu, K186 protégée dans le complexe uniquement avec RAP1 sauvage (et pas avec RAP1-ARAGA) est colorée en vert, K39 rendue plus accessible dans le complexe uniquement avec RAP1 sauvage (et pas avec RAP1-ARAGA) est colorée en orange. Les lysines dont l'acétylation n'est pas modifiée en présence de TRF2 sont colorées en gris.**

Sur la base des résultats des expériences d'empreinte et de l'interaction du motif TBM de RAP1 avec le domaine TRFH de TRF2, nous avons construit manuellement (avec le logiciel PyMOL ; Schrödinger LLC, 2010) un modèle du positionnement relatif du domaine TRFH de TRF2 et de la région N-terminale de RAP1 dans le complexe formé par les protéines entières. Ce modèle est présenté dans la figure suivante :



**Figure 44 : Modèle du complexe TRF2/RAP1 au niveau du domaine TRFH de TRF2 et de la région N-terminale de RAP1. TRF2 est représentée en jaune et RAP1 en cyan. Les lysines et les résidus du motif TBM de RAP1 sont colorés selon le même code que dans les figures précédentes.**

## 1.2 Discussion

Notre étude de l'interaction entre TRF2 et RAP1 entières a révélé un mode d'interaction complexe qui n'était pas anticipé d'après les données disponibles sur les domaines isolés (Chen *et al*, 2011). Au cours de cette étude nous avons mis en œuvre des approches complémentaires sans lesquelles la compréhension de l'interaction entre les protéines entières aurait été très difficile. La grande convergence des résultats obtenus leur confère une robustesse qui nous a permis de construire un modèle complet de l'interaction entre ces deux protéines. Ces approches sont discutées dans les paragraphes suivants, ainsi que quelques voies de recherche potentiellement ouvertes par nos résultats.

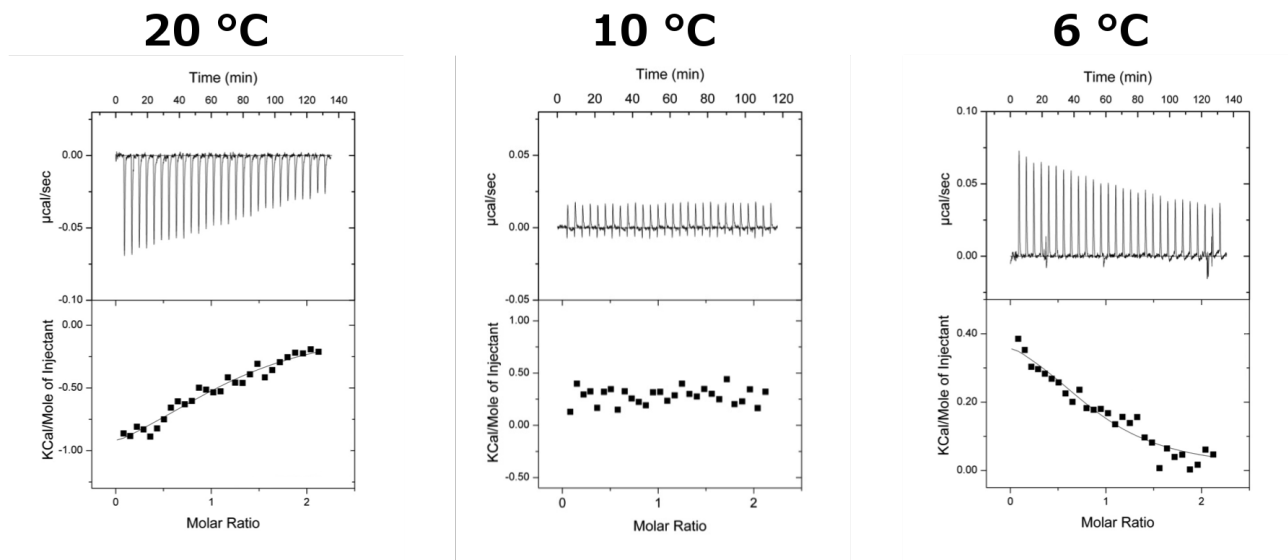
### 1.2.1 Mise en évidence de la transition mineure dans la courbe de titration

Un article récemment publié (Janoušková *et al*, 2015) rapporte l'étude de l'interaction entre TRF2 et RAP1 entières et montre une courbe de titration calorimétrique à une seule transition, typique d'une interaction à un seul site de liaison. Il est intéressant de remarquer les différences entre les conditions expérimentales de cette titration et de notre titration, résumées dans le tableau suivant :

**Tableau 5 : Conditions expérimentales de la titration calorimétrique de TRF2 par RAP1**

Condition	Janoušková <i>et al</i> (2015)	Notre étude
[TRF2] (monomères)	5 $\mu$ M (cuve)	13 $\mu$ M (cuve)
[RAP1]	44 $\mu$ M (seringue)	130 $\mu$ M (seringue)
Injections	10 $\mu$ L	10 $\mu$ L
Température	25 °C	10 °C
Tampon	phosphate de sodium 20 mM pH = 7,0 NaCl 50 mM	Tris-HCl 50 mM pH = 7,5 NaCl 150 mM 2-mercaptoethanol 10 mM

Les différences notables de concentrations en protéines, de force ionique et de température sont très probablement responsables du fait que la transition mineure de la courbe n'était pas détectable dans les conditions expérimentales de cette étude. Parmi ces paramètres, la température a notamment un effet très important sur les thermogrammes, comme nous avons pu l'observer en mesurant l'interaction entre RAP1-TBM et TRFH à plusieurs températures (figure suivante) :



**Figure 45 : Titration calorimétrique de TRFH par RAP1-TBM à 20 °C (panneau de gauche), 10 °C (panneau central) et 6 °C (panneau de droite).**

Pour cette raison nous avons réalisé toutes nos expériences dans des conditions de température et de force ionique les plus proches possible de celles de notre titration calorimétrique. Les concentrations en protéines plus élevées dans nos conditions sont aussi plus favorables à la détection d'une interaction secondaire de basse affinité.

Ces différences montrent clairement que l'étude d'assemblages protéiques complexes par calorimétrie nécessite d'évaluer systématiquement les effets de la température et des concentrations en protéines pour pouvoir détecter toutes les transitions associées au mode d'assemblage. D'autre part, la température physiologique est difficile à atteindre dans de telles études *in vitro* du fait de la stabilité thermique limitée des protéines purifiées.

### 1.2.2 Convergence des résultats et corrélation avec la conservation des protéines

Les résultats obtenus par nos différentes approches dans des conditions expérimentales proches convergent, ce qui augmente leur robustesse. Nos résultats sont également appuyés par la conservation du motif TBM de RAP1 : le consensus YRL défini par Kim *et al* (2009) est présent chez 12 des 14 espèces de vertébrés dont nous avons aligné les séquences (conservé chez les mammifères et les amphibiens, absent chez les poissons et les oiseaux). La conservation du motif TBM de RAP1 est présentée dans la figure suivante :

	1	10	20
Apollo-TBM	FRGLALK	YLT	PVNFFQAGY
PNUTS-TBM	TVVVS DK	YNL	KPIPLKRQSN
NBS1-TBM	AKMRIPN	YQL	SPTKLPSINK
SLX4-TBM	ERGLEVSH	RRL	APWQASPPHP
MCPH1-TBM	QETFEEK	YRL	SPTLSSTKGH
RAP1-TBM	ERLELEA	YRL	GPASAADTGS

▲ ▲ ▲

Homo_sapiens	92	ERLELEA	YRL	GPASAADTGS
Gallus_gallus	77	RRLPLDL	YRH	SGP.....
Xenopus_laevis	82	ELLHVD	YRL	DGSHGRKRQS
Xenopus_tropicalis	82	KLLDVD	YRL	DGSHGRPRKS
Callithrix_jacchus	92	ERLELET	YRL	GPASAADTGS
Oryctolagus_cuniculus	92	ERLELEA	YRL	GLASGAGPAP
Bos_taurus	92	EKLELEA	YRL	GPAPAAAYQAP
Rattus_norvegicus	92	EKLELEA	YRL	GLTEQ...AS
Mus_musculus	92	EKLDLEA	YRL	GLTEQ...AS
Macaca_mulatta	92	ERLELEA	YRL	GSASAADTGS
Monodelphis_domestica	92	ERLELED	YRL	GRRDSES...
Canis_familiaris	92	ERLDLEA	YRL	GSAAGQ..AP
Taeniopygia_guttata	76	GH.....		
Danio_rerio	89	QQLDPN	YAI	IIGPSVQ...

▲ ★ ▲ ▲

**Figure 46 : Région du motif TBM dans l’alignement des séquences de RAP1 de 14 espèces de vertébrés. Les motifs TBM connus sont rappelés dans le panneau du haut. Les résidus du motif sont indiqués par les triangles noirs, le résidu R du motif consensus défini par Kim *et al* (2009) est surligné en bleu et indiqué par une étoile. Les résidus identiques sont surlignés en rouge, les résidus similaires sont encadrés en bleu.**

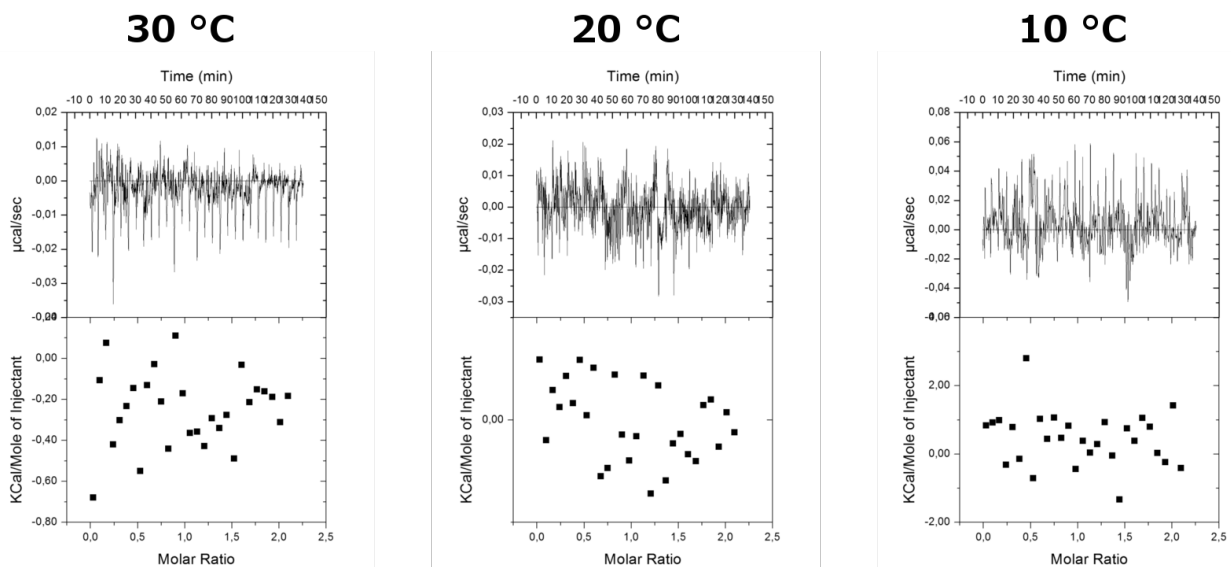
Certaines des lysines protégées dans le complexe TRF2/RAP1, mises en évidence par notre expérience d’empreinte, sont également bien conservées (voir les alignements complets en annexe). Chez RAP1, les lysines K114 (protégée chez RAP1 sauvage et le mutant ARAGA) et K186 (protégée uniquement chez RAP1 sauvage) dans la région N-terminale sont conservées respectivement chez 9 et 11 espèces parmi les 14 considérées. Chez TRF2, les lysines K58 et K93 dans le domaine TRFH sont toutes les deux conservées chez 9 de ces 14 espèces. De façon encore plus intéressante, les lysines K184 et K190 du domaine TRFH, protégées uniquement par RAP1 sauvage, sont conservées respectivement chez 11 et 13 de ces 14 espèces. Les lysines K353, K327, K329 et K333 (situées dans le *linker*) sont elles aussi bien conservées (respectivement chez 12, 11, 11 et 7 espèces parmi les 14).

Cette conservation du motif TBM de RAP1 et des lysines protégées dans le complexe TRF2/RAP1 suggère clairement que l’interaction complète entre TRF2 et RAP1 est un

phénomène conservé au moins chez les mammifères. Ceci nous conduit à émettre l'hypothèse que cette interaction complète a un rôle fonctionnel important, et pose bien entendu la question d'identifier ce rôle.

### 1.2.3 Importance de l'étude des protéines entières

L'interaction complexe que nous avons mise en évidence a été détectée pour la première fois avec les deux protéines entières, puis avec un peptide de RAP1 et un domaine isolé de TRF2. Mais curieusement nous n'avons jamais pu la détecter avec des domaines isolés des deux protéines. Nous avons notamment cherché à former un complexe entre le domaine TRFH de TRF2 et la région N-terminale de RAP1 dans l'espoir de le cristalliser pour résoudre sa structure, mais sans succès. Pour comprendre ces échecs nous avons voulu vérifier à nouveau l'interaction en titrant le domaine TRFH de TRF2 par la région N-terminale de RAP1 (RAP1\_1-208), et nous avons également voulu vérifier si l'extrémité N-terminale basique de TRF2 (domaine B) pouvait être impliquée dans l'interaction. Nous n'avons pas pu détecter l'interaction avec ces constructions, aux trois températures testées (figure suivante pour la construction B-TRFH ; les résultats sont identiques avec la construction TRFH).



**Figure 47 : Titration calorimétrique de B-TRFH par RAP1\_1-208 à 30 °C (panneau de gauche), 20 °C (panneau central) et 10 °C (panneau de droite).**

Une explication possible est que l'interaction de basse affinité entre la région N-terminale de RAP1 et le domaine TRFH de TRF2 est dépendante du rapprochement spatial permis par l'interaction de haute affinité entre le domaine RCT de RAP1 et la région RBM de TRF2. D'autre part, la liaison de TRF2 à l'ADN semble modifier la capacité de la région N-terminale de RAP1 à interagir avec TRF2. En effet nos collaborateurs au laboratoire d'Eric Gilson ont

vérifié que la construction RAP1\_1-208 ajoutée à TRF2 a un effet équivalent sur la protection de la jonction de Holliday comparé à RAP1 entière ajoutée au mutant de TRF2 délété de la région RBM (voir article en annexe).

Plus généralement, l'expérience acquise par l'étude du complexe TRF2/RAP1 souligne bien l'intérêt de combiner des approches expérimentales pour l'étude de telles protéines multi-domaines, et l'importance d'analyser les protéines entières plutôt que leurs domaines isolés. Il est en effet probable que de telles interactions faibles favorisées par un rapprochement spatial au sein d'un complexe (assemblé par d'autres interactions plus fortes) soient fréquentes dans les divers complexes protéiques présents dans la cellule. Ces interactions sont nécessairement indétectables par l'approche structurale classique s'intéressant en priorité aux domaines isolés.

#### **1.2.4 Effets de RAP1 sur les propriétés d'interaction de TRF2 avec l'ADN**

Nous avons montré que la région N-terminale de RAP1 participe à l'interaction avec TRF2, ce qui constitue une première étape dans la compréhension des effets de RAP1 sur TRF2 : nous savons maintenant que l'effet observé sur la condensation de l'ADN par TRF2 est très probablement la conséquence directe de la présence de la région N-terminale de RAP1 et de son interaction physique avec le domaine TRFH de TRF2. Cependant notre étude a été faite en absence d'ADN car il était très difficile d'obtenir des échantillons homogènes en formant le complexe avec l'ADN, or l'homogénéité est un prérequis indispensable pour le SAXS (qui analyse une population de molécules en solution et fournit un résultat correspondant à une moyenne sur toute la population), et l'empreinte protéique devait être réalisée dans les mêmes conditions expérimentales pour nous assurer que nous observions bien le même phénomène.

L'étude de Janoušková *et al* (2015) apporte des informations intéressantes sur l'effet de RAP1 sur la liaison de TRF2 à l'ADN. RAP1 réduit la contribution électrostatique de l'interaction (probablement du fait d'une région riche en résidus acides dans sa séquence, qui masquerait les charges positives du domaine B de TRF2), ce qui a pour conséquence de réduire globalement l'affinité de TRF2 pour l'ADN (en réduisant la liaison non spécifique par le domaine B) tout en augmentant sa spécificité pour les répétitions TTAGGG. Cependant ces expériences ont été réalisées avec des ADN courts (deux et cinq répétitions), et ne nous renseignent donc pas sur les propriétés de condensation de l'ADN.

Notre expérience d'empreinte peut s'envisager en présence d'un ADN long comme celui des expériences d'AFM (650 pb) mais nécessiterait une mise au point importante et de nombreux contrôles, le premier étant de s'assurer que le complexe condensé d'intérêt est bien formé. Le contrôle idéal serait l'observation directe par AFM de l'échantillon juste avant la réaction d'empreinte.

L'étude des complexes TRF2/ADN et RAP1/TRF2/ADN par microscopie électronique pourrait aussi permettre d'obtenir des informations complémentaires de celles obtenues par les expériences d'AFM de nos collaborateurs, et aiderait certainement à mieux comprendre comment le mode d'interaction que nous avons mis en évidence entre TRF2 et RAP1 module la capacité de TRF2 à condenser l'ADN.

### 1.2.5 Rôle du domaine BRCT de RAP1

Notre modèle de l'interaction complète entre TRF2 et RAP1 suggère un rapprochement spatial important des domaines BRCT des deux molécules de RAP1 liées à un dimère de TRF2. Il a été montré que des domaines BRCT répétés en tandem sont capables de lier des motifs phosphorylés (Manke *et al*, 2003; Yu *et al*, 2003; Rodriguez *et al*, 2003), en revanche la capacité d'un domaine BRCT isolé à lier ces motifs est encore mal caractérisée. L'hypothèse que des dimères de domaines BRCT (et non des domaines répétés en tandem dans la même chaîne polypeptidique) pourraient également lier des motifs phosphorylés a été émise (Rodriguez *et al*, 2003) et nous semble pertinente dans le cas de la dimérisation induite du domaine BRCT de RAP1 dans le complexe TRF2/RAP1.

Il est aussi intéressant de remarquer que le domaine BRCT de RAP1 est le plus conservé dans toute sa séquence, même plus conservé que son domaine RCT nécessaire à l'interaction avec TRF2 (voir en annexes l'alignement des séquences de RAP1 de 14 espèces de vertébrés possédant un complexe TRF2/RAP1). D'autre part, au cours de l'évolution jusqu'aux levures, le domaine Myb de RAP1 a subi une duplication et un changement radical de propriétés d'interaction, en acquérant la capacité d'interagir avec l'ADN (chez *S. cerevisiae*). Le domaine RCT a aussi complètement changé de propriétés d'interaction chez *S. cerevisiae*, en perdant la capacité à lier Tbf1 (l'orthologue de TRF2) et en gagnant la capacité à recruter d'autres protéines (les protéines Rif et Sir). Le domaine RCT a aussi été complètement perdu, par exemple chez *Candida albicans* (Yu *et al*, 2010). Au contraire, le domaine BRCT est conservé chez des RAP1 appartenant à des espèces évolutivement distantes comme les levures *S. cerevisiae* et *S. pombe*. Cette conservation extrême, alors même que les domaines BRCT entre eux sont relativement variables (Callebaut & Mornon, 1997), suggère des fonctions importantes possiblement associées à la phosphorylation des protéines partenaires de RAP1.

La résolution de la structure tridimensionnelle du domaine BRCT de RAP1 nous semble une étape importante dans la compréhension de ses fonctions (en permettant de le comparer aux autres domaines BRCT dont une structure est disponible), et des fonctions de l'interaction complète entre TRF2 et RAP1. Nous avons tenté sans succès de cristalliser la construction 1-208 de RAP1, qui s'avère extrêmement soluble (les cribles de conditions de cristallisation contiennent majoritairement des gouttes claires, même à une concentration en protéine d'environ 100 mg/mL). En revanche cette grande solubilité fait de cette



construction un outil bien adapté à une étude par RMN, d'autant plus que la structure du domaine Myb (résidus 132 à 191) a déjà été résolue par RMN (Hanaoka *et al*, 2001) et son attribution est disponible.

Ces considérations sur le domaine BRCT de RAP1 et sur son rôle à travers l'interaction complète entre TRF2 et RAP1 en font selon nous un axe de recherche particulièrement intéressant à développer pour le futur.

### **1.2.6 Étude en présence de RAP1 de l'interaction entre TRF2 et ses partenaires connus**

Finalement, l'interaction complète entre TRF2 et RAP1 suggère que RAP1 peut aussi moduler l'affinité de TRF2 pour ses autres partenaires protéiques, notamment ceux qui interagissent avec TRF2 par un motif TBM comme Apollo, SLX4, PNUITS, MCPH1 ou NBS1, mais potentiellement tous ses partenaires. Puisque RAP1 est un partenaire constitutif de TRF2 *in vivo*, il nous a paru pertinent d'étudier *in vitro* l'interaction entre TRF2 et quelques-uns de ses partenaires en présence de RAP1. Ces résultats sont abordés dans la partie suivante.

## 2 Partie 2 : Partenaires de TRF2 et du complexe TRF2/RAP1

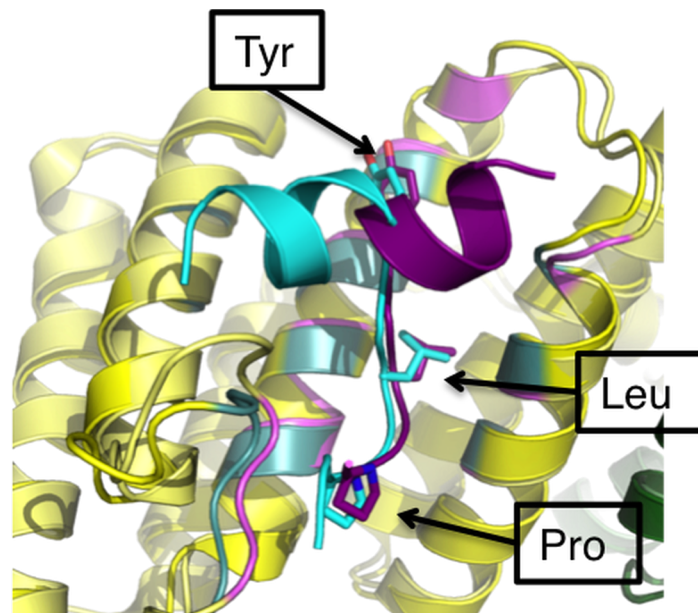
Afin de compléter l'étude de l'interaction entre TRF2 et RAP1 présentée dans la première partie et pour évaluer l'effet de RAP1 sur les propriétés d'interaction de TRF2, nous avons cherché à caractériser les interactions de TRF2 et du complexe TRF2/RAP1 avec des partenaires connus ou supposés.

Nous avons commencé par déterminer si la présence de RAP1 pouvait empêcher la liaison du motif TBM de la protéine Apollo, un partenaire connu de TRF2 dont le mode d'interaction a été bien décrit (Chen *et al*, 2008). Nous avons ensuite testé l'hypothèse de la présence d'un motif TBM dans la protéine WRN, interagissant directement avec TRF2 *in vitro* (Opresko *et al*, 2002). Nous avons aussi commencé à étudier l'interaction de TRF2 avec la protéine Ku. Finalement, une expérience de *pull down* nous suggère que TRF2 et le complexe TRF2/RAP1 ont globalement des spécificités de liaison différentes.

## 2.1 Résultats

### 2.1.1 Apollo

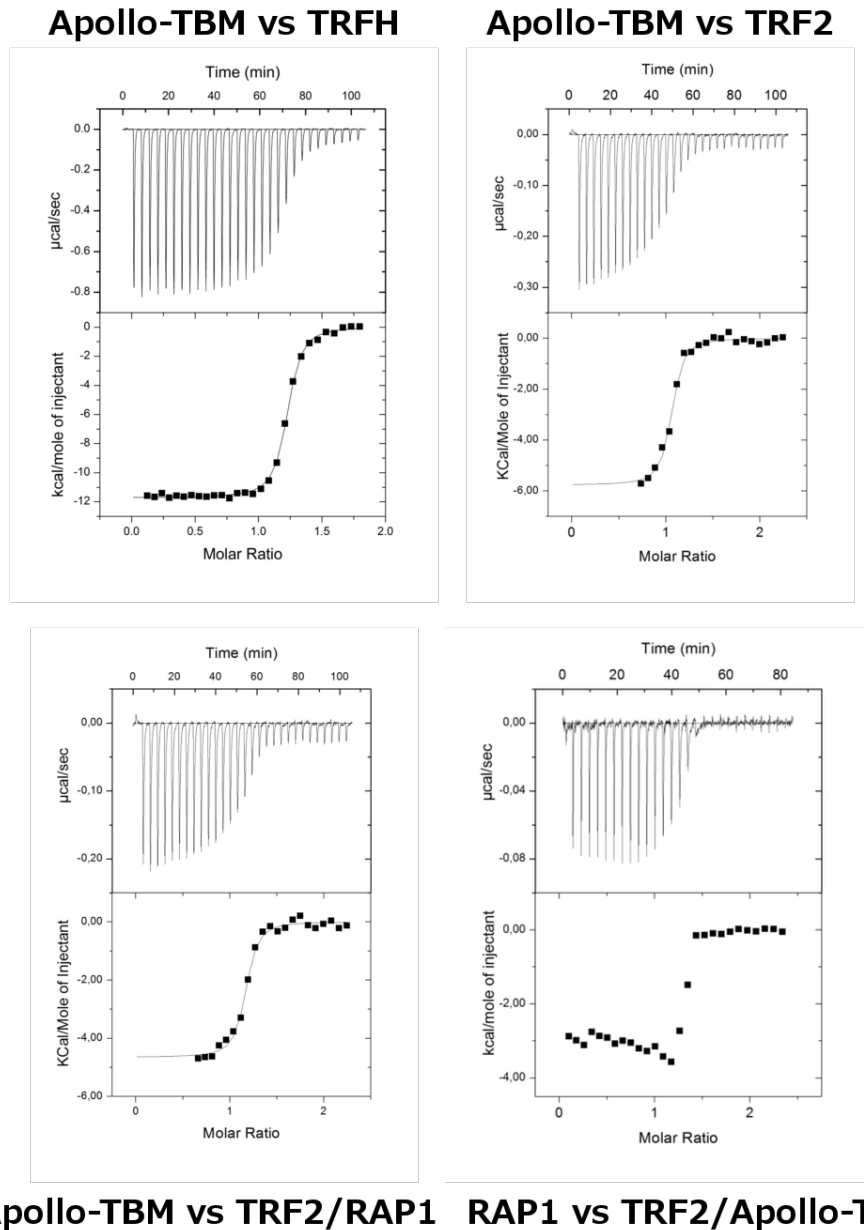
La superposition de notre structure cristallographique du complexe TRFH/RAP1-TBM avec celle du complexe TRFH/Apollo-TBM (entrée PDB 3BUA, Chen *et al*, 2008) montre que les motifs TBM de ces deux protéines occupent le même site de liaison dans le domaine TRFH, comme attendu. La superposition est présentée dans la figure suivante :



**Figure 48 : Superposition des structures TRFH/RAP1-TBM et TRFH/Apollo-TBM. Le peptide RAP1-TBM est coloré en cyan, le peptide Apollo-TBM est coloré en violet, le domaine TRFH est coloré en jaune (ses résidus engagés dans l'interaction avec les motifs TBM sont colorés de la même couleur que les motifs).**

Ceci suggère une compétition possible entre RAP1 et Apollo, et plus généralement entre RAP1 et les autres partenaires de TRF2 qui interagissent *via* un motif TBM. L'affinité du motif TBM d'Apollo pour le domaine TRFH est bien meilleure que celle du motif TBM de RAP1 : leurs constantes d'équilibre de dissociation sont respectivement de 120 nM (Chen *et al*, 2008) pour le motif d'Apollo, ou 59 nM lorsque nous avons reproduit cette mesure, contre 13  $\mu$ M pour le motif de RAP1 (cf. première partie). Néanmoins, la liaison de RAP1 par son domaine RCT à TRF2 a une affinité très forte (son  $K_D$  est de l'ordre de 1 nM, cf. première partie) et favorise probablement la liaison du motif TBM du fait de son rapprochement spatial de son site de liaison, comme le suggère le  $K_D$  d'environ 0,5  $\mu$ M obtenu pour cette interaction secondaire avec les protéines entières. Pour déterminer si la proximité du motif TBM de RAP1 pourrait compromettre la liaison d'Apollo, nous avons

mesuré l'affinité du motif TBM d'Apollo pour le complexe TRF2/RAP1 préformé, et réciproquement l'affinité de RAP1 pour le complexe TRF2/Apollo-TBM préformé (figure suivante).



**Figure 49 : Titrations calorimétriques de TRFH par Apollo-TBM (en haut à gauche), de TRF2 par Apollo-TBM (en haut à droite), du complexe TRF2/RAP1 préformé par Apollo-TBM (en bas à gauche) et du complexe TRF2/Apollo-TBM préformé par RAP1 (en bas à droite).**

Nous obtenons les  $K_D$  suivants :

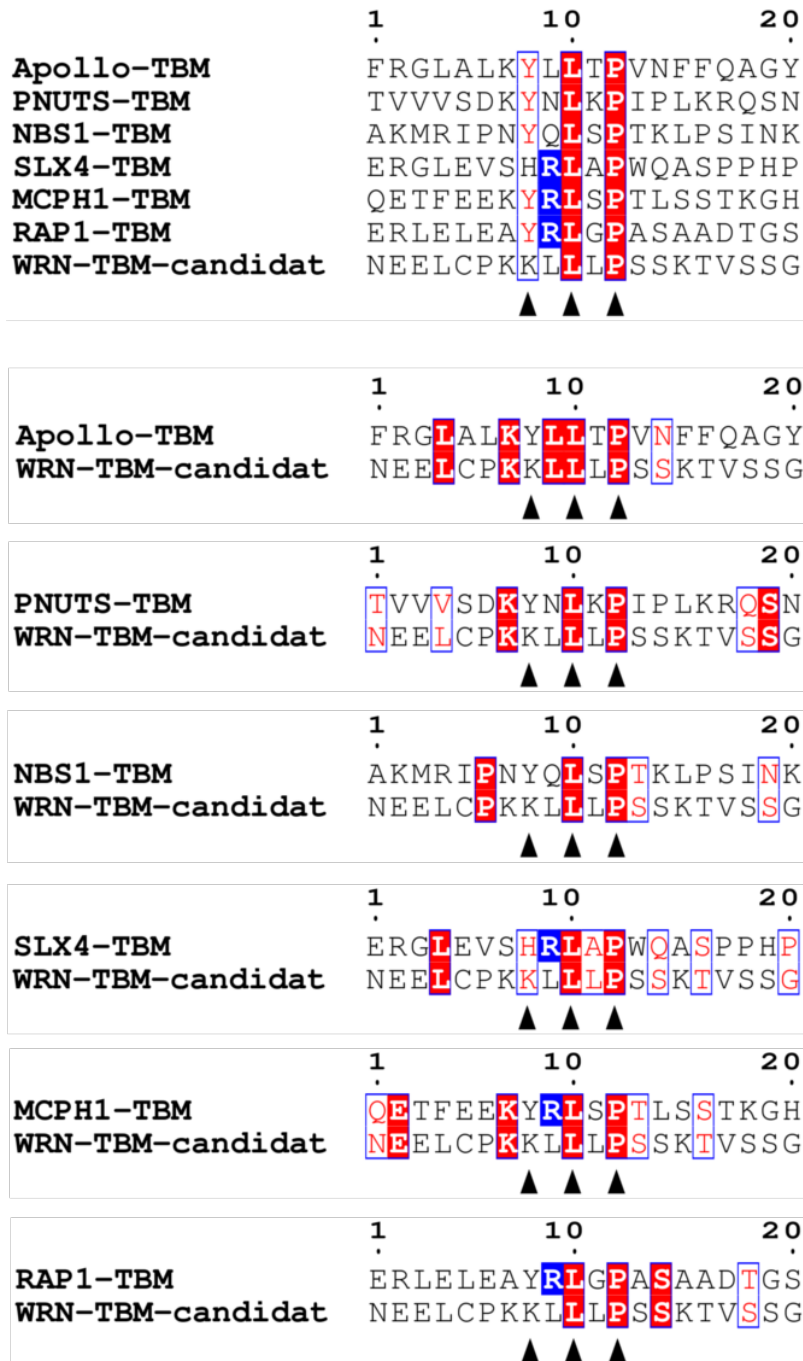
- Apollo-TBM contre TRFH :  $K_D \simeq 59$  nM,
- Apollo-TBM contre TRF2 :  $K_D \simeq 46$  nM,
- Apollo-TBM contre TRF2/RAP1 préformé :  $K_D \simeq 47$  nM,
- RAP1 contre TRF2/Apollo-TBM préformé :  $K_D \simeq 1$  nM pour la transition majeure, la transition mineure est plus faible et ne permet plus de calculer un  $K_D$ .

Dans le cas de la titration du complexe TRF2/Apollo-TBM par RAP1, nous observons toujours une courbe à deux transitions, mais dont la transition mineure n'est plus suffisamment marquée pour calculer un  $K_D$ . De façon remarquable, cette courbe de titration est très similaire à celle que nous obtenions avec la titration de TRF2 par RAP1-ARAGA (partie 1, Figure 36), ce qui identifie clairement le motif TBM de RAP1 comme le déterminant principal de l'interaction secondaire à l'origine de la transition mineure dans la courbe.

Ces données montrent qu'Apollo est capable d'interagir avec TRF2 même en présence de RAP1, en accord avec son recrutement aux télomères par TRF2 *in vivo* (van Overbeek & de Lange, 2006; Lenain *et al*, 2006).

### 2.1.2 WRN

L'interaction directe *in vitro* entre TRF2 et WRN est décrite dans la littérature (Opresko *et al*, 2002). D'après cette étude, le fragment 949-1092 de WRN est suffisant pour lier TRF2 entière. Ce fragment contient un motif KLLLP aux positions 1073 à 1077, qui est très semblable aux motifs TBM de liaison au domaine TRFH déjà mis en évidence (partie 1 et Chen *et al*, 2008; Kim *et al*, 2009; Wan *et al*, 2013). L'alignement de ce motif de WRN aux motifs TBM connus d'Apollo et NBS1 (Chen *et al*, 2008), de PNUTS et MCPH1 (Kim *et al*, 2009), de SLX4 (Wan *et al*, 2013), et de RAP1 (Partie 1), est représenté dans la figure suivante :

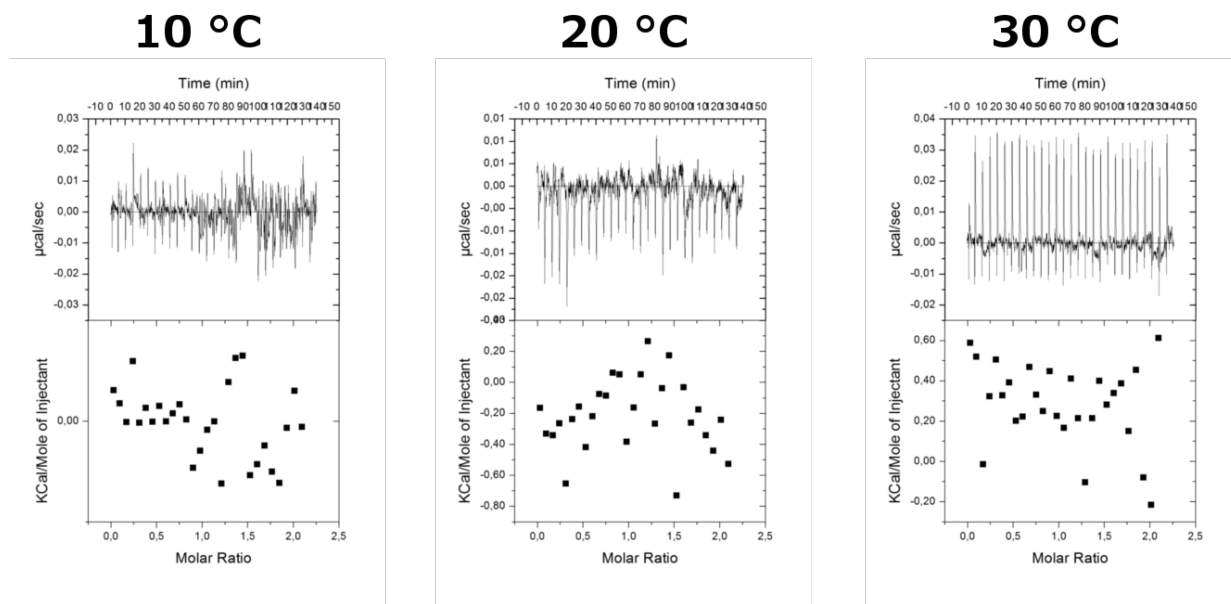


**Figure 50 : Alignement du motif KLLLP de WRN avec les motifs TBM connus d’Apollo, NBS1, PNUTS, MCPH1, SLX4 et RAP1. Les résidus du motif sont indiqués par les triangles noirs, le résidu R du motif consensus défini par Kim *et al* (2009) est surligné en bleu, les résidus identiques sont surlignés en rouge et les résidus similaires sont encadrés en bleu. Les alignements deux à deux sont aussi présentés.**

Une structure RMN du fragment 949-1092 montre que ce motif est situé dans une région non structurée de la protéine (Hu *et al*, 2005), ce qui en fait un candidat pertinent pour être

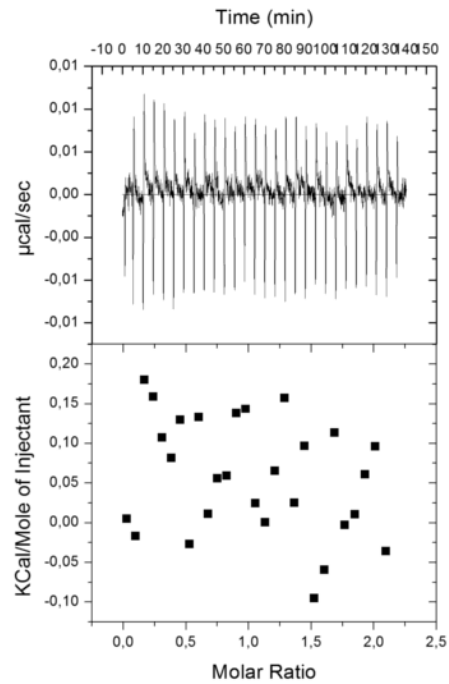
un motif TBM. Les auteurs de cette structure RMN indiquent que cette construction a été difficile à produire de façon soluble, en revanche une étude plus récente a montré que le fragment 949-1079 est stable dans une expérience de protéolyse ménagée, peut être produit et purifié aisément, et est bien soluble (Kitano *et al*, 2010). Ce dernier fragment contient toujours le motif TBM putatif. Nous avons donc exprimé et purifié cette construction appelée WRN-RQC (cf. Matériel et méthodes) afin de tester son interaction avec le domaine TRFH de TRF2. Nous avons privilégié cette approche plutôt que l'étude des protéines entières car en cas d'interaction nous obtiendrions un complexe compact, compatible avec une étude cristallographique, ce qui ne serait pas le cas avec les protéines entières.

La titration calorimétrique du domaine TRFH de TRF2 par WRN-RQC ne montre aucune interaction détectable entre ces deux protéines dans nos conditions expérimentales (figure suivante), ce qui remet en question notre hypothèse que le motif KLLLP dans la séquence de WRN soit un motif TBM.



**Figure 51 : Titrations calorimétriques du domaine TRFH de TRF2 par WRN-RQC. À 10 °C (panneau de gauche), 20 °C (panneau central) et 30 °C (panneau de droite).**

Puisque la délimitation des régions d'interaction n'a été faite que du côté de WRN (Opresko *et al*, 2002), il est possible que le domaine WRN-RQC interagisse avec TRF2 dans une région située hors du domaine TRFH (comme c'est le cas pour RAP1 dont le site principal d'interaction est situé dans le *linker* de TRF2). Nous avons donc testé l'interaction de WRN-RQC avec TRF2 entière, et avons également obtenu un résultat négatif (figure suivante).

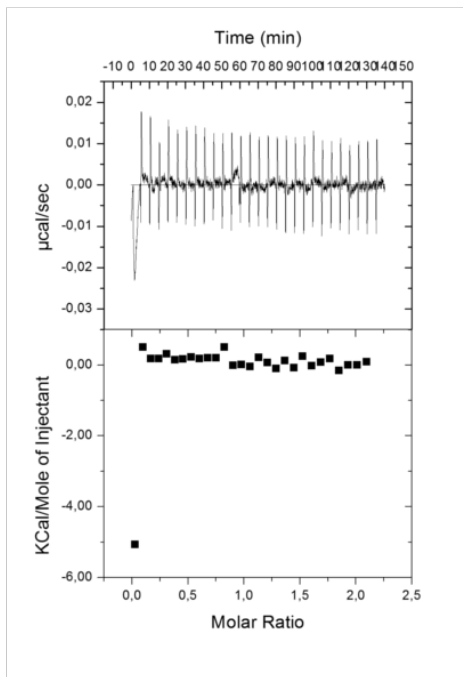


**Figure 52 : Titration calorimétrique de TRF2 par WRN-RQC**

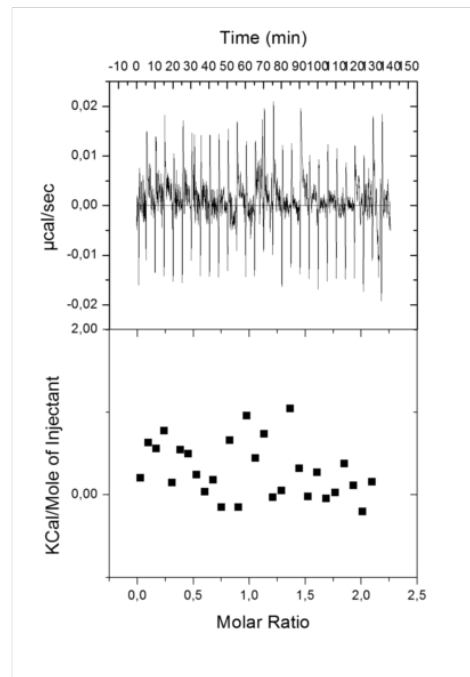
Finale­ment, il est possible que l'interaction dépende de la présence de RAP1 puisque nous avons montré dans la partie précédente que celle-ci est susceptible de modifier les propriétés d'interaction de TRF2. Nous avons donc testé l'interaction de WRN-RQC avec le complexe TRF2/RAP1, et avec RAP1 seule comme contrôle. Là encore nous n'avons obtenu que des résultats négatifs (figure suivante).



## WRN vs TRF2/RAP1

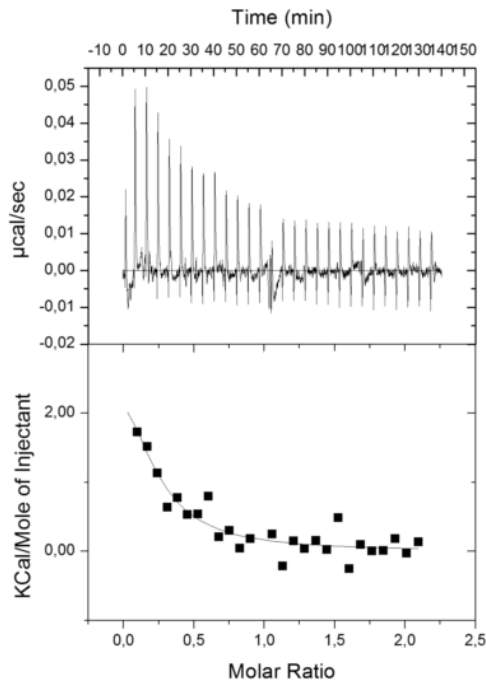


## WRN vs RAP1



**Figure 53 : Titration calorimétrique du complexe TRF2/RAP1 (gauche) et de RAP1 (droite) par WRN-RQC**

Afin de vérifier le bon repliement du domaine WRN-RQC, nous avons finalement testé son interaction avec l'ADN (cf. Matériel et méthodes) et avons retrouvé un  $K_D$  d'environ 2,15  $\mu\text{M}$ , de l'ordre de grandeur de celle rapportée dans la littérature (Kitano *et al*, 2010). Ce dernier thermogramme est présenté dans la figure suivante :



**Figure 54 : Titration calorimétrique de WRN-RQC par l'oligonucléotide T2-S5S6**

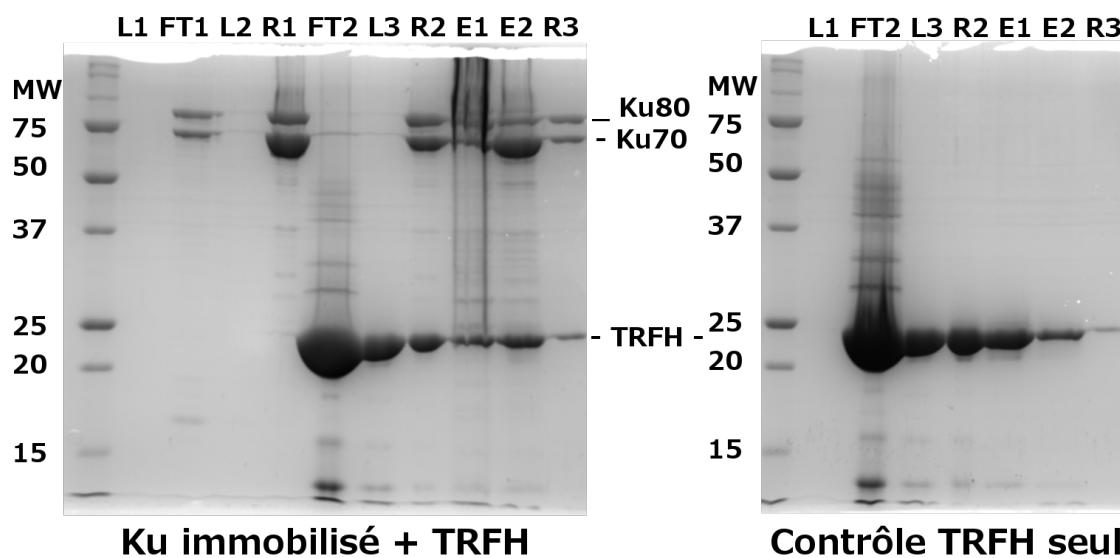
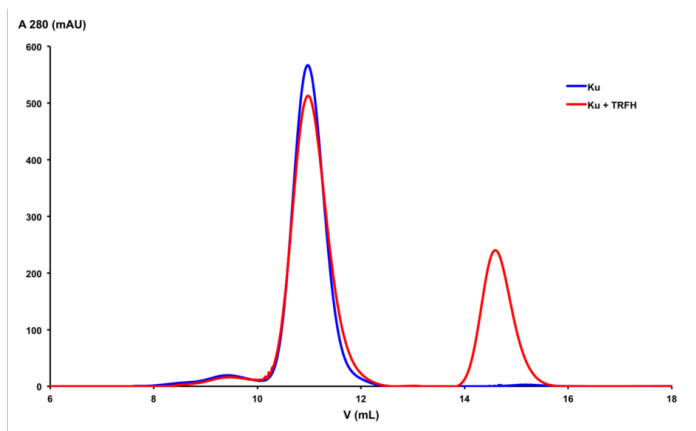
Ceci démontre que la protéine est correctement repliée, et écarte la possibilité que nous ne pouvions pas détecter l'interaction avec TRF2 à cause d'un problème de préparation de l'échantillon.

### 2.1.3 Ku

Il a été montré par des expériences de double-hybride puis confirmé *in vitro* avec des protéines recombinantes purifiées que TRF2 interagit avec Ku70 par son domaine TRFH (Song *et al*, 2000). Cette interaction a été confirmée *in vivo* plus récemment (Ribes-Zamora *et al*, 2013) par la technique de PCA (*protein complementation assay*), qui consiste à exprimer des protéines de fusion portant chacune un fragment de la YFP (*yellow fluorescent protein*) : lorsque ces protéines interagissent, les deux fragments sont suffisamment rapprochés pour reconstituer le fluorophore et on peut donc détecter par microscopie l'émission de fluorescence résultante, tandis que ces fragments exprimés seuls (ou en fusion avec des protéines qui n'interagissent pas entre elles) ne sont pas capables d'interagir l'un avec l'autre et n'émettent donc aucune fluorescence.

Puisque l'hétérodimère Ku70/80 (appelé simplement Ku par la suite, pour parler du dimère entier) est produit par une autre équipe du laboratoire (équipe de Jean-Baptiste Charbonnier), nous avons cherché à vérifier cette interaction avec l'objectif de produire un complexe minimal compatible avec une étude cristallographique. Nous avons donc dans un

premier temps testé l'interaction de Ku avec le domaine TRFH de TRF2, par chromatographie d'exclusion stérique et par rétention sur résine de nickel. Nous n'avons pas pu confirmer l'interaction dans ces conditions (figure suivante) :

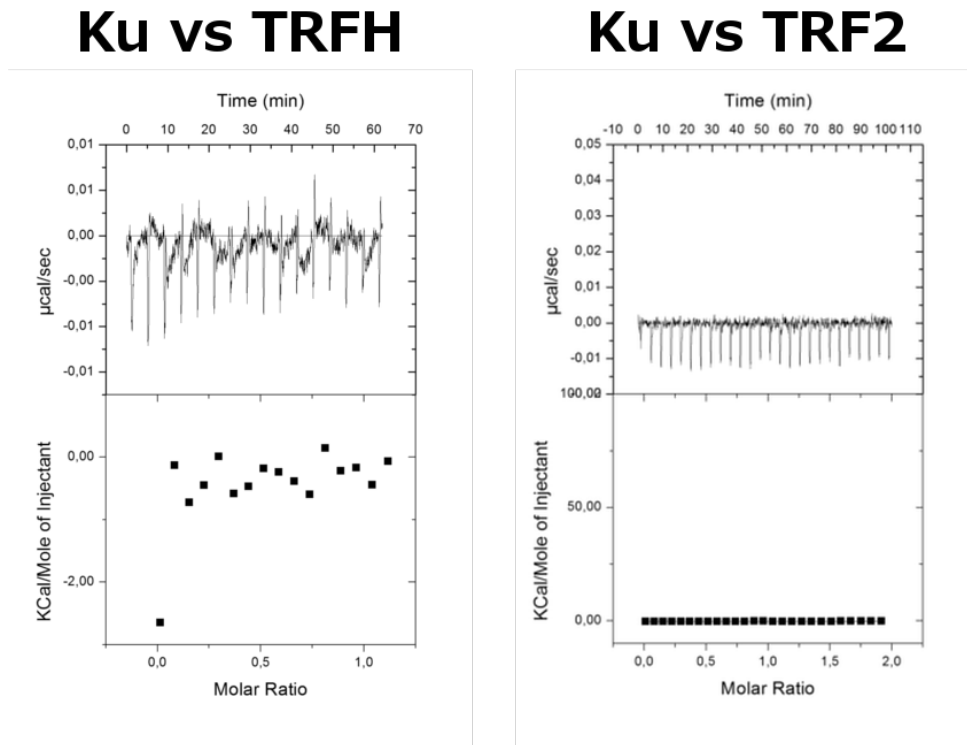


**Figure 55 : Tests d'interaction entre Ku et TRFH. En haut, chromatogrammes d'exclusion stérique (Ku seul en bleu, mélange Ku + TRFH en rouge). En bas, analyse SDS-PAGE de la rétention sur résine de nickel.**

En présence du domaine TRFH de TRF2 (chromatogramme rouge dans la figure ci-dessus), nous n'observons aucun pic dont le volume d'élué est inférieur à celui de Ku (11 mL, chromatogramme bleu dans la figure ci-dessus) ce qui signifie que dans ces conditions aucun complexe ne se forme. L'expérience de rétention sur résine de nickel ne montre pas non plus d'interaction puisque la quantité de TRFH élué est identique en absence ou en présence de Ku immobilisé sur la résine (comparer les pistes E1 et E2 correspondant aux fractions d'élué, voir la partie Matériel et méthodes pour les détails de l'expérience) et correspond donc à la fraction de TRFH retenue non spécifiquement par la résine. Si TRFH

interagissait avec Ku, il serait enrichi dans la fraction d'élution dans l'expérience où Ku est présent.

Nous avons également réalisé la titration calorimétrique de Ku par TRFH et par TRF2, et nous confirmons que dans nos conditions expérimentales *in vitro* aucune interaction n'est détectée entre Ku et TRF2 (figure suivante) :

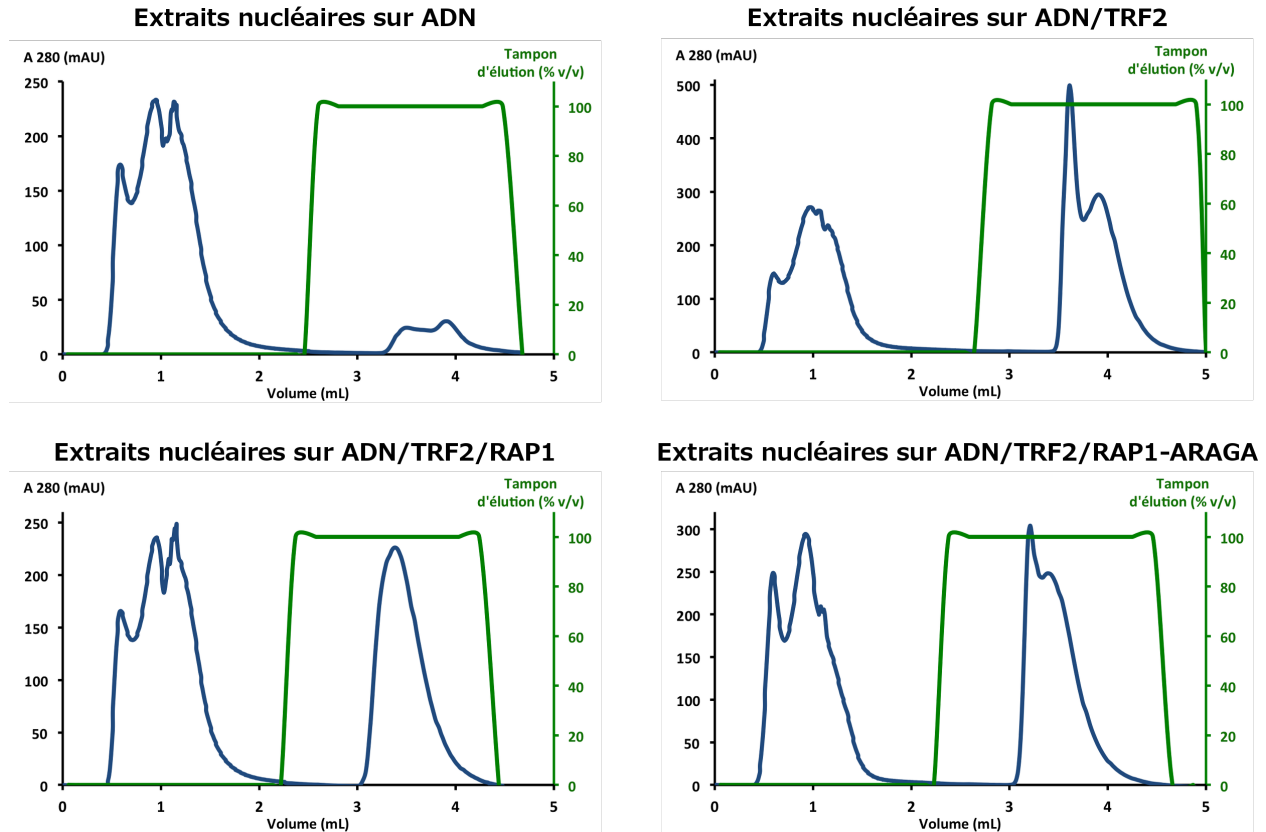


**Figure 56 : Titrations calorimétriques de Ku par TRFH (à gauche) et par TRF2 (à droite)**

#### 2.1.4 Pull down streptavidine-ADN

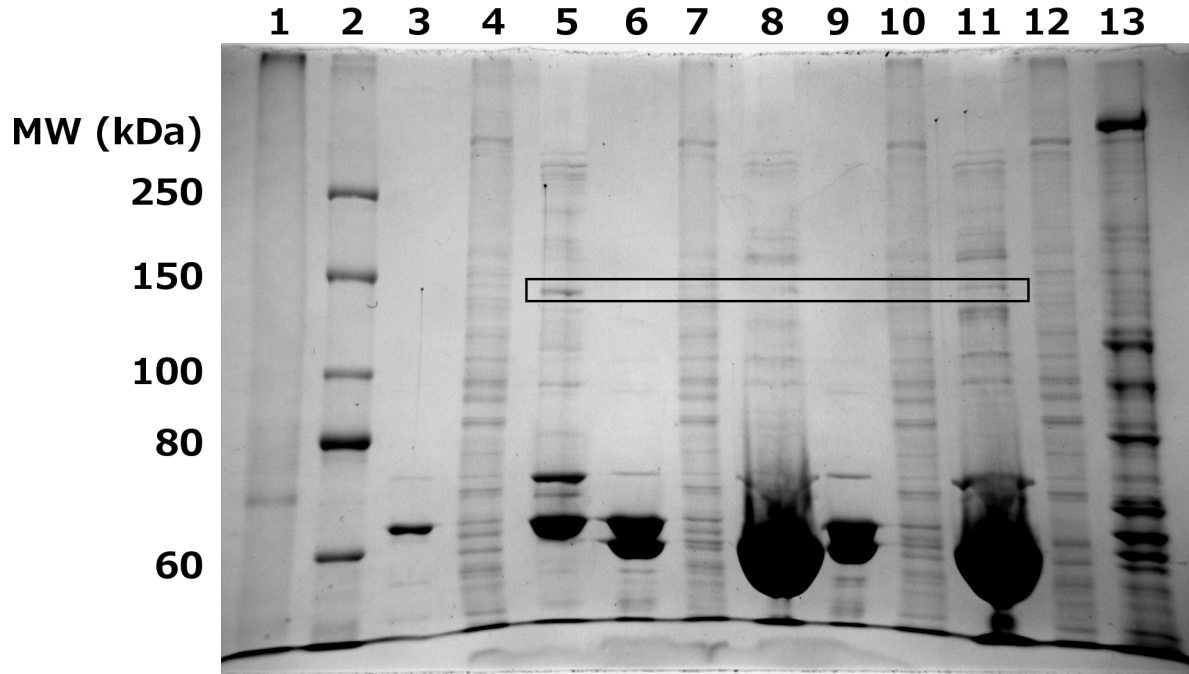
Pour mettre en évidence de façon plus générale l'effet de RAP1 sur les propriétés d'interaction de TRF2 avec ses autres partenaires protéiques, nous avons réalisé des expériences de *pull down* (cf. Matériel et méthodes). Dans ces expériences un oligonucléotide contenant un site de liaison de TRF2 est immobilisé sur un support chromatographique par un système streptavidine/biotine. Une quantité saturante de TRF2 ou de complexe TRF2/RAP1 ou TRF2/RAP1-ARAGA est alors injectée à travers la colonne. Puis nous injectons des extraits nucléaires de cellules HeLa dans la colonne contenant le complexe immobilisé ADN/TRF2 ou ADN/TRF2/RAP1 ou ADN/TRF2/RAP1-ARAGA. Enfin, les protéines sont éluées par augmentation de la force ionique (pour dissocier TRF2 de l'ADN), et nous analysons le contenu des éluats par SDS-PAGE pour déterminer quelles protéines des extraits nucléaires sont retenues par les différents complexes.

Nous observons de façon reproductible des chromatogrammes d'éluion différents selon que les extraits nucléaires ont été incubés avec TRF2 seul, le complexe TRF2/RAP1, le complexe TRF2/RAP1-ARAGA ou l'ADN seul (figure-ci dessous), alors que les quantités de protéines et d'extraits nucléaires sont identiques dans tous les essais.



**Figure 57 : Chromatogrammes d'éluion des pull down (courbes bleues). ADN seul (en haut à gauche), ADN saturé en TRF2 (en haut à droite), ADN saturé en TRF2/RAP1 (en bas à gauche), ADN saturé en TRF2/RAP1-ARAGA (en bas à droite). Le premier pic autour de 1 mL correspond à la fraction non retenue des extraits nucléaires, le pic d'éluion se trouve dans la région où le tampon d'éluion est à 100 % (courbe verte).**

L'analyse par SDS-PAGE est présentée dans la figure suivante :



**Figure 58 : Analyse SDS-PAGE des éluats de la figure précédente. 1 : extrait nucléaire (en entrée), 2 : marqueur de masse moléculaire, 3 : excès de TRF2 non fixé sur la colonne, 4 : excès d'extraits nucléaires non retenus par TRF2, 5 : éluat TRF2, 6 : excès de TRF2/RAP1 non fixé sur la colonne, 7 : excès d'extraits nucléaires non retenus par TRF2/RAP1, 8 : éluat TRF2/RAP1, 9 : excès de TRF2/RAP1-ARAGA non fixé sur la colonne, 10 : excès d'extraits nucléaires non retenus par TRF2/RAP1-ARAGA, 11 : éluat TRF2/RAP1-ARAGA, 12 : extrait nucléaires non retenus sur l'ADN, 13 : extraits nucléaires retenus sur l'ADN.**

Les éluats contiennent des protéines différentes (comparer les pistes 5, 8 et 11), par exemple dans la région encadrée : une protéine retenue par TRF2 n'est pas retenue par le complexe TRF2/RAP1, et seulement peu retenue par le complexe TRF2/RAP1-ARAGA.

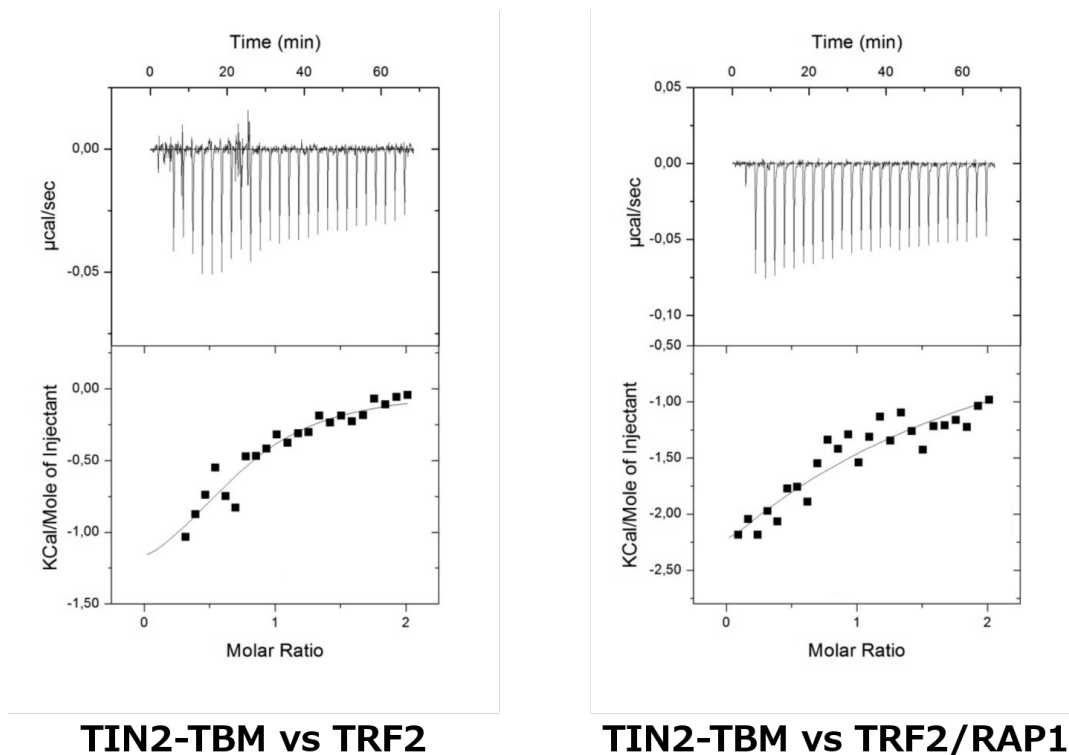
Ce résultat montre que les spécificités d'interaction protéine/protéine du complexe TRF2/RAP1 et de TRF2 sont effectivement différentes, et de façon potentiellement dépendante de la région N-terminale de RAP1.

## 2.2 Discussion

### 2.2.1 Apollo

Nous avons montré que la présence de RAP1 n'empêche pas la liaison du motif TBM d'Apollo au domaine TRFH de TRF2, en accord avec le fait que TRF2 et Apollo interagissent ensemble *in vivo* (van Overbeek & de Lange, 2006; Lenain *et al*, 2006). Ceci est également en accord avec la meilleure affinité du domaine TRFH pour le motif TBM d'Apollo ( $K_D \approx 0,12 \mu\text{M}$ , Chen *et al* (2008) ; ou  $0,059 \mu\text{M}$  dans nos mains) que pour celui de RAP1 ( $K_D \approx 0,5 \mu\text{M}$  pour la transition mineure, cf. partie 1, Figure 25). Nous pouvons donc raisonnablement supposer que la protéine Apollo entière est également capable de se lier au complexe TRF2/RAP1 *in vitro*, bien que cette expérience reste à réaliser. Les mêmes considérations s'appliquent à la protéine SLX4, qui interagit elle aussi avec TRF2 par un motif TBM (Wan *et al*, 2013). L'affinité du domaine TRFH pour le motif TBM de SLX4 ( $K_D \approx 0,75 \mu\text{M}$ , Wan *et al*, 2013) est du même ordre de grandeur que son affinité pour le motif TBM de RAP1, ainsi l'influence de RAP1 sur l'interaction entre TRF2 et SLX4 reste encore à déterminer.

La protéine TRF1 recrute elle aussi des partenaires possédant un motif TBM légèrement différent, de séquence consensus FxLxP. Il a été montré qu'un motif TBM de ce type est présent dans la séquence de la protéine TIN2 et est spécifique du domaine TRFH de TRF1 ( $K_D \approx 0,3 \mu\text{M}$  ; Chen *et al*, 2008), mais peut également se lier non spécifiquement au domaine TRFH de TRF2 ( $K_D \approx 6 \mu\text{M}$  selon Chen *et al*, 2008; ou  $3,5 \mu\text{M}$  selon Kim *et al*, 2009). Un des rôles fonctionnels du motif TBM de RAP1 pourrait ainsi être d'empêcher la liaison non spécifique des motifs FxLxP au domaine TRFH de TRF2. Pour tester cette hypothèse, nous avons mesuré l'interaction entre le motif TBM de TIN2 (peptide TIN2-TBM) et TRF2 d'une part, et entre TIN2-TBM et le complexe TRF2/RAP1 préformé d'autre part. Les thermogrammes et courbes de titration sont présentés dans la figure suivante :



**Figure 59 : Titration calorimétrique de TRF2 (à gauche) et du complexe TRF2/RAP1 préformé (à droite) par TIN2-TBM**

Nous obtenons les  $K_D$  suivants :

- TIN2-TBM contre TRF2 :  $K_D \simeq 6,5 \mu\text{M}$  (en accord avec le résultat de Chen *et al*, 2008),
- TIN2-TBM contre TRF2/RAP1 préformé :  $K_D \simeq 123,9 \mu\text{M}$ .

L'affinité de TRF2 pour le motif TBM de TIN2 est donc réduite d'un facteur environ 19 en présence de RAP1. Même si la mesure est moins précise dans cette gamme d'affinités, ceci nous conforte dans notre hypothèse selon laquelle une des fonctions de RAP1 serait d'empêcher la liaison non spécifique à TRF2 de protéines contenant un motif TBM spécifique de TRF1, ou contenant simplement un motif ressemblant au TBM mais sans rapport avec la fonction de recrutement aux télomères (de tels motifs courts sont fréquents : par exemple le motif FxLxP est présent dans plus de 7700 séquences de protéines humaines de la base de données Uniprot). Cette expérience serait intéressante à réaliser avec RAP1-ARAGA, afin de déterminer si le motif TBM de RAP1 est critique pour cette fonction.

### 2.2.2 WRN

Nous avons démontré que l'interaction entre TRF2 et WRN, précédemment mise en évidence avec les protéines entières (Opresko *et al*, 2002), ne peut pas être détectée par



calorimétrie dans nos conditions expérimentales en utilisant la région minimale de WRN supposée suffisante pour l'interaction. Ceci peut signifier que le motif KLLLP présent dans la séquence de WRN aux positions 1073-1077 n'est pas un motif TBM et que l'interaction est due à d'autres régions des protéines, ou bien que le motif n'est pas suffisant pour conférer une affinité détectable comme c'est le cas pour le motif TBM de RAP1 (voir la première partie : nous ne pouvons pas détecter l'interaction entre les domaines isolés RAP1\_1-208 et TRFH, alors que nous pouvons la détecter avec un peptide du motif TBM et avec les protéines entières). À cet égard, l'étude biophysique détaillée de l'interaction entre TRF2 et WRN entières serait intéressante et complémentaire de notre étude sur RAP1 : l'étude des facteurs télomériques accessoires est en effet une suite logique de l'étude des protéines du *shelterin*.

Les protéines entières de l'étude d'Opresko *et al* (2002) ont été exprimées en cellules d'insectes, elles portent donc potentiellement des modifications post-traductionnelles. Au contraire, toutes nos constructions sont exprimées dans *E. coli* et ne portent *a priori* aucune modification. Ceci peut expliquer les résultats divergents que nous obtenons. Il n'existe pas de site de modification dans la région 949-1079 de WRN (selon Uniprot), mais la modification nécessaire se trouve peut-être du côté de TRF2 puisque cette dernière possède de nombreux sites de modification (Walker & Zhu, 2012).

Dans l'objectif d'identifier les déterminants structuraux de l'interaction entre TRF2 et WRN, il serait intéressant de réaliser les expériences suivantes :

Afin de déterminer si le motif KLLLP de WRN (TBM putatif) est engagé dans l'interaction, il faudrait tester sa liaison au domaine TRFH, et la liaison à TRF2 entière de WRN entière mutée au niveau de ce motif. Comme dans le cas de RAP1 il se peut que l'interaction soit détectable uniquement avec le peptide ou les protéines entières, mais pas avec des domaines isolés. En cas de résultat positif pour l'interaction TRFH/TBM-putatif, il serait de plus possible de cristalliser le complexe pour résoudre sa structure tridimensionnelle, ce qui permettrait d'affiner notre compréhension de la reconnaissance des motifs TBM par le domaine TRFH.

Dans le cas où l'implication de ce motif TBM putatif serait réfutée, des expériences de protéolyse ménagée sur le complexe des protéines entières pourraient permettre d'isoler la région minimale d'interaction et potentiellement générer un complexe tronqué cristallisable.

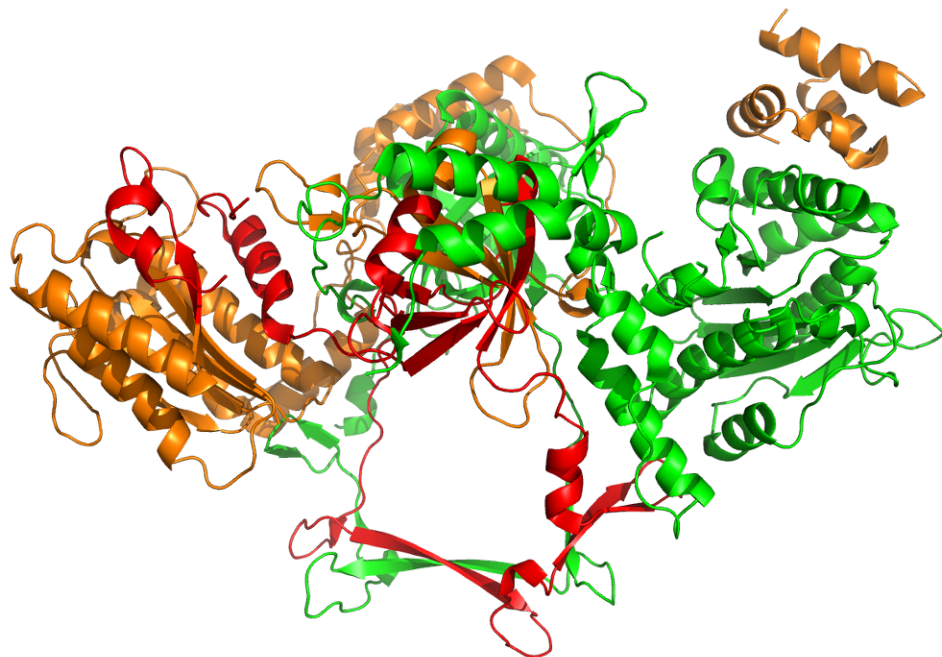
Que l'interaction soit médiée ou non par ce motif TBM, il est important de l'aborder avec les protéines entières. L'étude de l'interaction avec les protéines entières exprimées dans *E. coli* permettrait ainsi de déterminer si elle dépend ou non de modifications post-traductionnelles. Il serait aussi intéressant de phosphoryler ces protéines par leurs kinases spécifiques (dont certaines sont déjà connues, voir par exemple Walker & Zhu, 2012) pour

tester l'interaction avec des phosphorylations bien déterminées. Enfin, il nous semble aussi important de tester l'interaction de WRN entière avec le complexe TRF2/RAP1 ou TRF2/RAP1-ARAGA préformé, afin de déterminer si RAP1 module l'interaction de TRF2 avec WRN.

### 2.2.3 Ku

Dans nos conditions expérimentales aucune interaction n'a pu être détectée entre Ku et TRF2. Ce résultat est en désaccord avec des résultats précédents (Song *et al*, 2000; Ribes-Zamora *et al*, 2013), et nous allons tenter d'y apporter une explication.

Concernant l'interaction de Ku70 avec le domaine TRFH (Song *et al*, 2000), il convient d'aborder ce résultat avec un regard critique. Les auteurs précisent qu'ils utilisent pour les expériences de double-hybride une construction tronquée de Ku70 contenant les résidus 200-385, pour éviter de détecter uniquement l'interaction avec Ku80. Il est important de remarquer que cette étude est antérieure à la publication de la structure tridimensionnelle de l'hétérodimère Ku par Walker *et al* (2001). Cette structure nous apprend que le repliement de chacun des deux monomères est très dépendant de la présence de l'autre. La figure suivante montre dans la structure de Ku (entrée PDB 1JEQ) la région correspondant à la construction Ku70\_200-385 :



**Figure 60 : Structure tridimensionnelle de l'hétérodimère Ku. Ku80 en vert, Ku70 en orange, Ku70\_200-385 en rouge.**

Les délimitations de cette construction ne correspondent clairement pas aux délimitations naturelles entre les domaines de la protéine. Le repliement correct de cette construction n'a pas été vérifié par les auteurs de l'étude, et nous semble peu probable. Cette construction ne correspond donc à aucune réalité biologique ni d'un point de vue structural (cf. figure ci-dessus), ni d'un point de vue fonctionnel puisque cette construction de Ku70 est incapable de s'associer avec Ku80, or *in vivo* les deux sous-unités dépendent l'une de l'autre pour leur stabilité (Fell & Schild-Poulter, 2015, et références associées). Sur la base de ces considérations, l'interaction rapportée par Song *et al* (2000) pourrait être un faux positif.

L'interaction mise en évidence *in vivo* (Ribes-Zamora *et al*, 2013) par des expériences de PCA (*protein complementation assay*) semble plus plausible. Ce type d'expérience n'exclut pas la possibilité d'une interaction indirecte, et puisque TRF2 et Ku sont toutes les deux des protéines de liaison à l'ADN il est possible que leur interaction dépende de la présence d'ADN. Il serait donc intéressant de tester l'interaction de Ku avec TRF2 en présence d'ADN, par exemple en préformant un complexe Ku/ADN puis en le titrant par TRF2. L'équipe de Jean-Baptiste Charbonnier a déjà mis au point des expériences de SPR (résonance plasmonique de surface) pour mesurer l'interaction de Ku avec des ADN immobilisés sur une puce, ce qui constitue un bon outil pour mesurer l'interaction de Ku fixé sur l'ADN avec ses partenaires protéiques. Il est également possible que Ku soit recruté par RAP1, comme le suggèrent des expériences de co-immunoprécipitation dans lesquelles Ku co-précipite avec RAP1 délété de son domaine RCT, donc indépendamment de TRF2 (O'Connor *et al*, 2004). Une autre expérience directement réalisable, puisque nous disposons déjà des constructions protéiques, est la titration de Ku par RAP1\_1-208 et par RAP1 entière.

D'autre part, l'interaction directe *in vitro* entre Ku et TRF1 a été mise en évidence par des expériences de SPR (Hsu *et al*, 2000), ce qui peut signifier deux choses : Ku pourrait avoir différentes fonctions aux télomères, selon la protéine télomérique par laquelle elle est recrutée, ou bien la présence de Ku aux télomères est critique et son recrutement est donc assuré de façon redondante par plusieurs des protéines du *shelterin*.

#### 2.2.4 Pull down streptavidine-ADN

Notre expérience de *pull down* confirme notre hypothèse selon laquelle TRF2 et le complexe TRF2/RAP1 recrutent des partenaires protéiques différents. Nous pouvons aussi observer que le complexe TRF2/RAP1-ARAGA a une spécificité de liaison intermédiaire entre celle de TRF2 et celle du complexe TRF2/RAP1, ce qui est en accord avec nos autres données relatives à ce mutant de RAP1 (voir la première partie).

Nous avons tenté d'identifier par spectrométrie de masse des bandes intéressantes présentes dans les éluats TRF2 et TRF2/RAP1-ARAGA et absentes dans l'éluat TRF2/RAP1 (correspondant à des partenaires de TRF2 dont le recrutement est inhibé par RAP1), ou des bandes absentes dans les éluats TRF2 et TRF2/RAP1-ARAGA et présentes dans l'éluat

TRF2/RAP1 (correspondant à des partenaires du complexe TRF2/RAP1 et dont le recrutement dépend de la présence de RAP1). Malheureusement le faible enrichissement de ces protéines et la complexité des mélanges nous ont empêché d'identifier ces protéines de façon fiable.

Il serait intéressant de reproduire cette expérience en éluant avec un gradient de force ionique plutôt qu'avec une marche directement à 1 M en NaCl, ce qui pourrait permettre d'éliminer les protéines retenues non spécifiquement avant d'éluer les protéines d'intérêt (ces dernières se trouveraient alors plus enrichies). Une autre possibilité consisterait à réaliser l'expérience sur le même support chromatographique utilisé directement dans des tubes plutôt que dans une colonne : ceci permet d'éluer avec un volume très faible et donc d'obtenir un éluat dans lequel les protéines d'intérêt sont plus concentrées (en revanche ce système ne permet pas d'éluer par un gradient de concentration en sel, seuls des lavages à des concentrations définies seraient possibles).

### 2.2.5 Perspectives sur l'étude des partenaires de TRF2 et RAP1

L'ensemble de ces études nous montre que la plasticité du complexe TRF2/RAP1 rend délicate la caractérisation *in vitro* des interactions mises en évidence *in vivo*. Le cas de l'interaction avec Ku en est un bon exemple : elle a été démontrée *in vivo* avec TRF2 (Ribes-Zamora *et al*, 2013) et RAP1 (O'Connor *et al*, 2004). Nos résultats montrent cependant que TRF2 seule n'est pas suffisante pour recruter Ku directement, et suggèrent que l'interaction dépend de la fixation de TRF2 et/ou de Ku sur l'ADN, et de la présence de RAP1 en complexe avec TRF2.

L'ensemble des expériences que nous avons réalisées dans cette partie révèle surtout que la caractérisation des interactions de TRF2 avec ses partenaires doit s'envisager dans un contexte plus complet, c'est-à-dire non seulement en présence de RAP1 (son partenaire constitutif), mais aussi dans la situation où TRF2 se trouve liée à l'ADN. La grande flexibilité de TRF2 que nous avons mise en évidence dans la première partie suggère en effet que sa liaison à l'ADN modifie ses propriétés d'interaction. Il semble donc important de comprendre finement le mode d'interaction de TRF2 avec l'ADN. La troisième partie présente les premiers résultats d'une étude en cours visant à compléter la description de ce mode d'interaction avec l'ADN au niveau moléculaire.



### 3 Partie 3 : Mode d'interaction de TRF2 avec l'ADN

Comme nous l'avons vu dans l'introduction, le mode d'interaction de TRF2 avec l'ADN est complexe et ses propriétés ne s'expliquent pas uniquement par la liaison du domaine Myb à l'ADN. L'interaction engage en effet plusieurs régions de la protéine : au moins son domaine Myb et son extrémité N-terminale basique. L'observation par AFM du complexe entre TRF2 et un ADN télomérique long (650 pb) a mis en évidence que TRF2 condense l'ADN (Poulet *et al*, 2012). De façon intéressante, l'AFM a aussi permis d'observer qu'en augmentant le rapport molaire de TRF2 face à l'ADN, la longueur apparente de l'ADN diminue par pas réguliers d'une trentaine de nanomètres et en corrélation avec l'augmentation de la circonférence apparente des complexes (communication personnelle de Marie-Josèphe Giraud-Panis). Ceci suggère que TRF2 est capable d'enrouler l'ADN autour d'elle. Une expérience d'empreinte protéique par acétylation des lysines montre par ailleurs que de nombreuses lysines de TRF2 sont protégées par l'ADN dans le complexe condensé : les lysines du domaine Myb comme attendu, mais également des lysines dans le domaine TRFH et dans le *linker* (communication personnelle de Marie-Josèphe Giraud-Panis). Les lysines du *linker* situées directement en N-terminal du domaine Myb et significativement protégées par l'ADN sont les suivantes : K402, K405, K408, K410 et K420 (en rouge dans les séquences présentées ci-dessous). De façon remarquable, ces cinq lysines sont bien conservées chez les espèces dont nous avons aligné les séquences (voir l'alignement en annexes) : chez 10 à 12 espèces parmi les 14 considérées.

Dans le but d'affiner nos connaissances sur les déterminants structuraux de l'interaction entre la région C-terminale de TRF2 et l'ADN, nous avons donc produit une construction du domaine Myb allongée de 46 résidus en N-terminal (appelée Myb-long) incluant les lysines protégées dans l'expérience d'empreinte de nos collaborateurs, et la construction classique correspondant au domaine présent dans l'entrée PDB 1W0U (appelée Myb-court) afin de comparer leurs modes d'interaction avec l'ADN. Les séquences alignées de ces deux constructions sont indiquées ci-dessous :

```
Myb-court 445 -----GKKQKWTVEESEWVK 459
Myb-long 400 GEKNPKVPKGKWNSNGVEEKETWVEEDELFQVQAAPDEDSTTNITKKQKWTVEESEWVK 459
                                             *****
```

```
Myb-court 460 AGVQKYGEGNWA AISKNYPFVNRTAVMIKDRWRMTMKRLGMN 500
Myb-long 460 AGVQKYGEGNWA AISKNYPFVNRTAVMIKDRWRMTMKRLGMN 500
               *****
```

Notre objectif est donc de comprendre si la région supplémentaire incluant les lysines protégées modifie les propriétés d'interaction du domaine Myb de TRF2 avec l'ADN, et de caractériser cette interaction le cas échéant. Nous avons décidé de combiner l'ITC, qui nous fournira les paramètres thermodynamiques de l'interaction avec l'ADN, le SAXS qui nous

permettra de comparer les degrés de compaction respectifs des complexes Myb-long/ADN et Myb-court/ADN, et enfin la cristallographie pour tenter d'obtenir des informations à l'échelle moléculaire sur les déterminants structuraux de l'interaction dans le complexe Myb-long/ADN.

## 3.1 Résultats

### 3.1.1 Calorimétrie

Les titrations calorimétriques des deux constructions Myb-court et Myb-long par l'oligonucléotide double-brin T2-S5S7 (voir Matériel et méthodes) contenant deux sites TAGGG sont présentées dans la figure suivante :

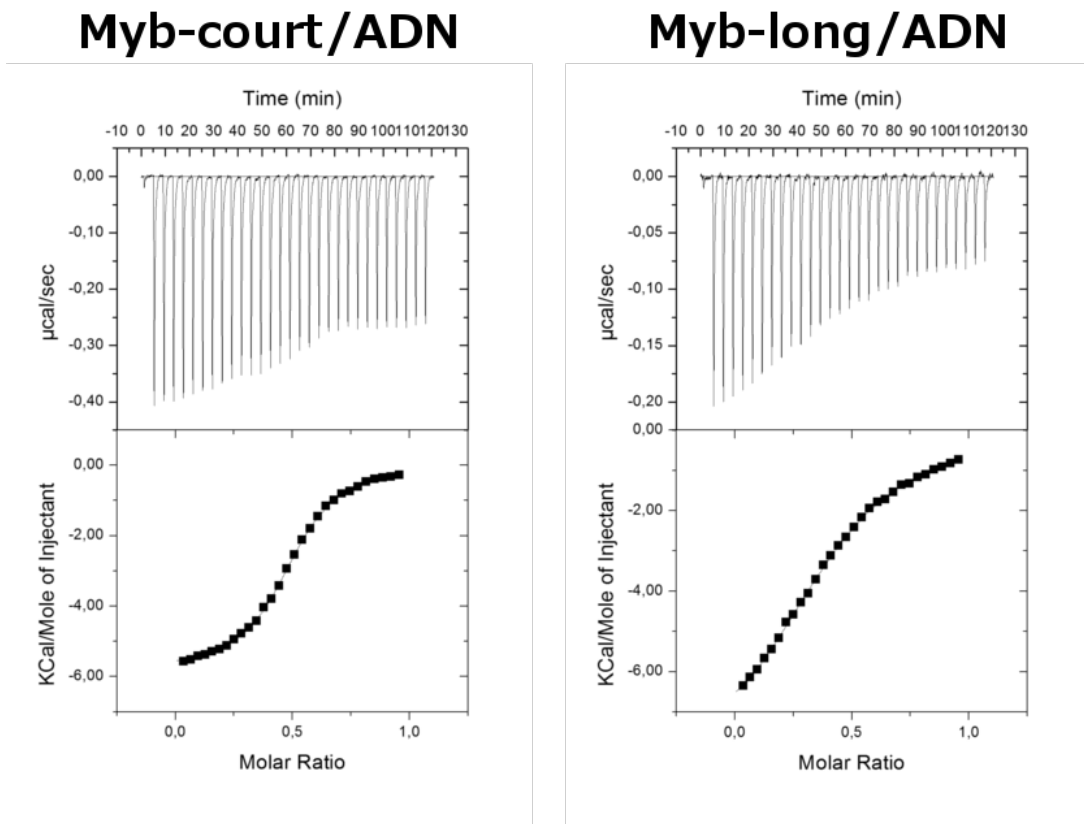


Figure 61 : Titrations calorimétriques des constructions Myb-court (à gauche) et Myb-long (à droite) par l'oligonucléotide T2-S5S7



Les paramètres thermodynamiques obtenus sont résumés dans le tableau suivant :

**Tableau 6 : Comparaison des paramètres thermodynamiques des interactions Myb-court/ADN et Myb-long/ADN**

Paramètre	Myb-court/ADN	Myb-long/ADN
N	0,496	0,420
$K_A$ ( $M^{-1}$ )	$2,21.10^6 \pm 4,53.10^4$	$3,89.10^5 \pm 9,89.10^3$
$K_D$ ( $\mu M$ )	$0,45 \pm 0,01$	$2,47 \pm 0,07$
$\Delta G$ (cal/mol)	$-8221,50 \pm 11,53$	$-7248,92 \pm 28,60$
$\Delta H$ (cal/mol)	$-5816 \pm 17,77$	$-8596 \pm 89,37$
$-T\Delta S$ (cal/mol)	$-2405,5 \pm 29,30$	$1347,08 \pm 117,97$

Puisque l'oligonucléotide à deux sites se trouve dans la seringue et qu'une molécule d'oligonucléotide se lie à deux molécules de protéine présentes dans la cuve, la stoechiométrie attendue pour l'interaction est de 0,5 (1 ligand pour 2 récepteurs, soit 0,5 ligand par récepteur). Le paramètre N déterminé par l'analyse du thermogramme représente la stoechiométrie à condition que les concentrations des deux molécules soient précisément déterminées, et à condition que chacune des deux molécules soit active à 100 %. Le paramètre N peut être calculé par la formule suivante (Milev, 2015) :

$$N = S \times \frac{FA_{cuve}}{FA_{seringue}}$$

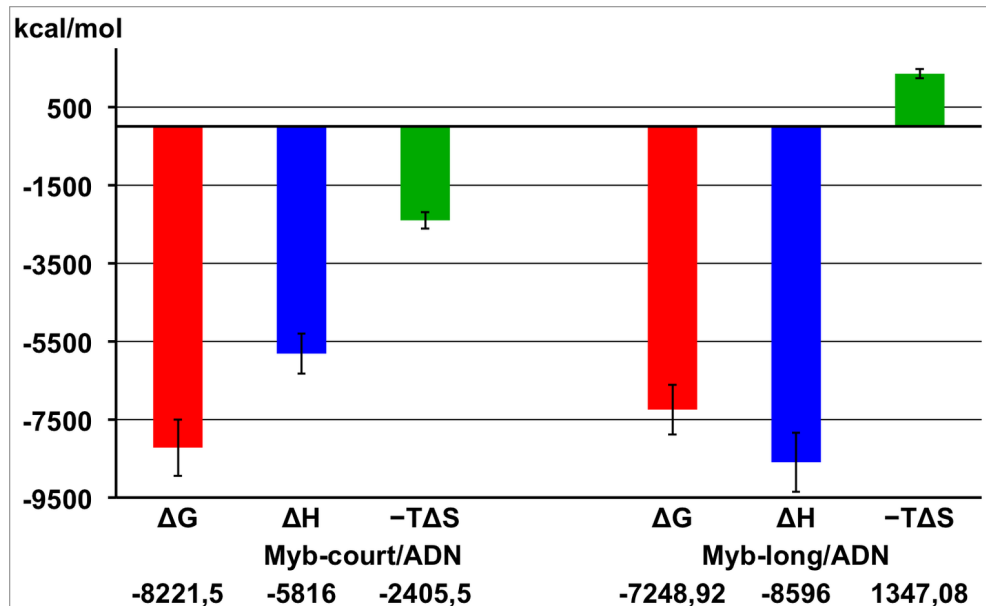
où S est la stoechiométrie et FA la fraction active telle que

$$FA = \frac{[molécule]_{active}}{[molécule]_{totale}}$$

Nous pouvons raisonnablement faire l'hypothèse que l'oligonucléotide est actif à 100 % car nous avons vérifié son hybridation correcte. La valeur de N inférieure à 0,5 est par conséquent probablement attribuable à une fraction active de protéine inférieure à 100 %, ce qui est vraisemblable étant donné qu'il est en pratique impossible de purifier parfaitement une protéine. La formule précédente donne des fractions actives de 99,2 % et 84 % pour Myb-court et Myb-long, respectivement.

L'affinité de Myb-long pour l'ADN est environ 5 fois moins bonne que celle de Myb-court, mais est tout de même forte car dans la gamme micromolaire basse. La comparaison des paramètres thermodynamiques  $\Delta G$  (variation d'enthalpie libre),  $\Delta H$  (variation d'enthalpie) et  $\Delta S$  (variation d'entropie) renseigne sur le mode de liaison.  $\Delta G$  est négatif pour une interaction favorable, il est directement lié à la constante d'équilibre de dissociation par la relation  $\Delta G = RT \ln K_D$  ( $R = 1,987 \text{ cal.K}^{-1}.\text{mol}^{-1}$  est la constante des gaz parfaits, T est la

température de l'expérience en K). D'autre part,  $\Delta H$  et  $\Delta S$  contribuent à  $\Delta G$  selon la relation  $\Delta G = \Delta H - T\Delta S$ . Ainsi, une variation d'enthalpie libre favorable à l'interaction (donc négative) nécessite une contribution enthalpique ( $\Delta H$ ) négative et une contribution entropique ( $\Delta S$ ) positive. Les valeurs des paramètres  $\Delta G$ ,  $\Delta H$  et  $-T\Delta S$  des deux interactions étudiées sont représentées graphiquement dans la figure suivante :



**Figure 62 : Comparaison des contributions enthalpique et entropique dans les interactions Myb-court/ADN et Myb-long/ADN. Variation d'enthalpie libre en rouge, variation d'enthalpie en bleu et variation d'entropie en vert.**

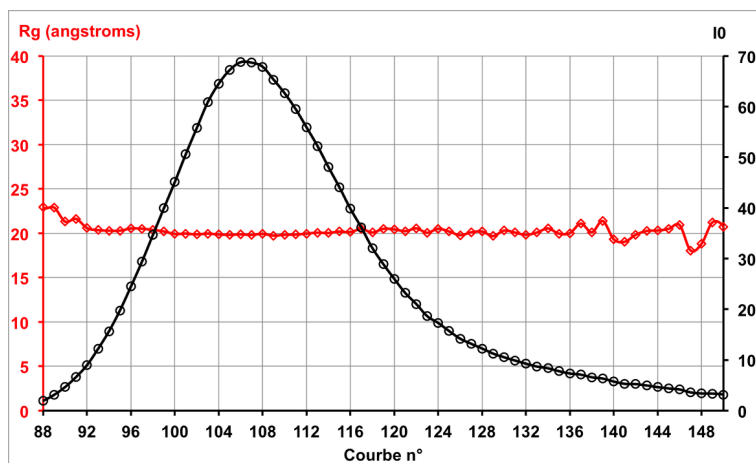
La contribution enthalpique ( $\Delta H$ ) est principalement due aux liaisons hydrogène établies et rompues au cours de l'interaction (Leavitt & Freire, 2001) : une valeur négative de  $\Delta H$  indique que des liaisons ont été établies. Nous observons que la contribution enthalpique est plus favorable pour Myb-long que pour Myb-court (d'un facteur environ 1,5), ce qui signifie que Myb-long établit plus de liaisons hydrogène avec l'ADN. Cette observation est en accord avec les résultats de l'expérience d'empreinte de nos collaborateurs.

La contribution entropique ( $\Delta S$ ) représente les changements de solvatation et de conformation (Leavitt & Freire, 2001). Notamment,  $\Delta S$  est négatif (la contribution à  $\Delta G$  est donc défavorable) lorsqu'une partie de la chaîne protéique perd des degrés de liberté conformationnelle au cours de l'interaction. La contribution entropique est défavorable pour Myb-long, comparé à Myb-court. Nous pouvons en déduire que la portion de *linker* (résidus 400 à 445) adopte probablement une conformation plus contrainte dans le complexe avec l'ADN, comparée à celle qu'elle adopte dans la forme libre de la protéine. Ceci est également en accord avec l'hypothèse de nos collaborateurs sur l'implication de cette région de TRF2 dans l'interaction avec l'ADN.

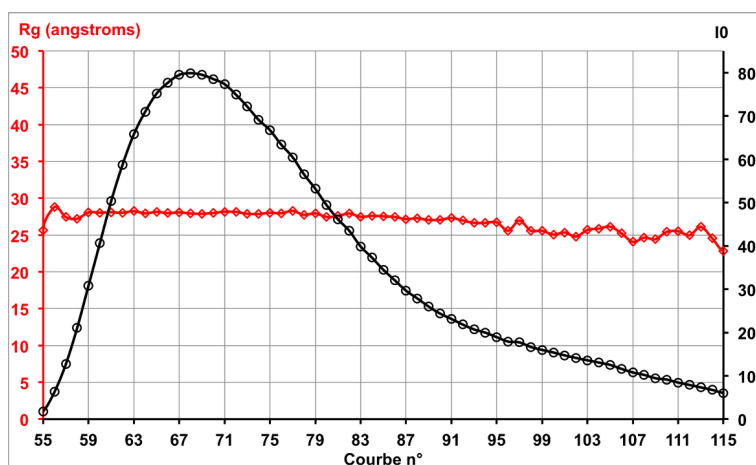
Il résulte de ces différentes contributions enthalpique et entropique que la variation d'enthalpie libre de l'interaction Myb-long/ADN est presque égale à celle de l'interaction Myb-court/ADN (cette dernière est plus favorable d'un facteur environ 1,1), ce qui explique le même ordre de grandeur des constantes d'équilibre de dissociation correspondantes.

### 3.1.2 SAXS

Nous avons réalisé des mesures de SEC-SAXS (SAXS couplé à la chromatographie d'exclusion stérique, voir David & Pérez, 2009; Pérez & Koutsioubas, 2015) sur les deux complexes Myb-court/ADN et Myb-long/ADN (cette fois avec l'oligonucléotide double-brin T2-S5S8, voir Matériel et méthodes). Deux réplicats ont été mesurés pour chaque complexe. Pour toutes les acquisitions le  $R_g$  est constant tout le long du chromatogramme, ce qui montre la bonne homogénéité des échantillons (un des deux réplicats de chaque complexe est présenté dans la figure suivante) :



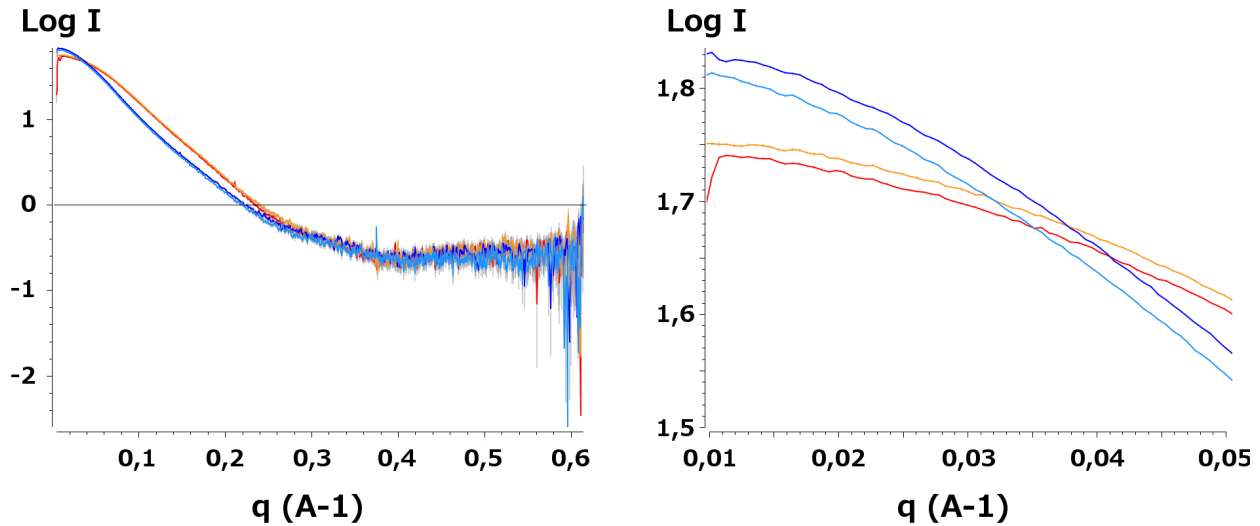
**Myb-court/ADN**



**Myb-long/ADN**

**Figure 63 : Variation des paramètres  $R_g$  (en rouge) et  $I_0$  (en noir) au cours de l'éluion des complexes Myb-court/ADN (en haut) et Myb-long/ADN (en bas)**

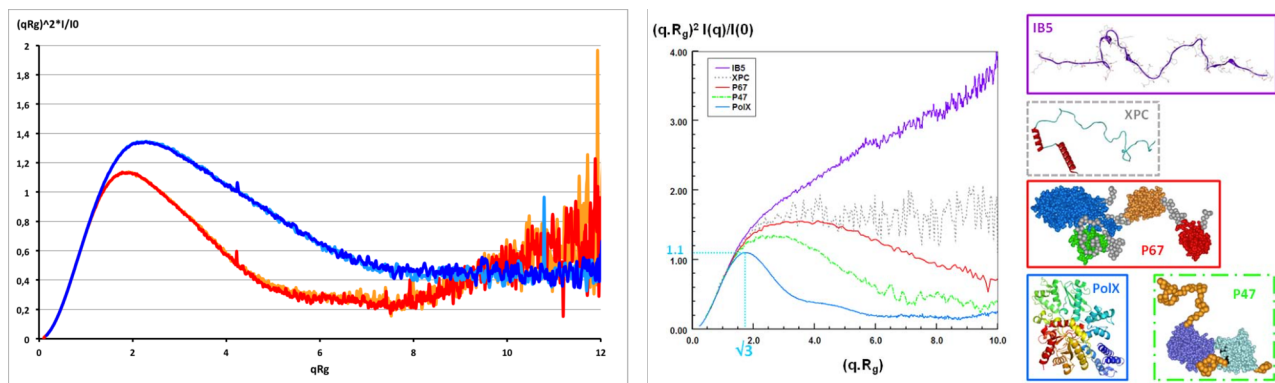
Les courbes finales correspondant à ces acquisitions ont été obtenues en calculant la moyenne des courbes 100 à 116 pour le complexe Myb-court/ADN (courbe orange dans la figure suivante), et la moyenne des courbes 60 à 76 pour le complexe Myb-long/ADN (courbe bleue foncée dans la figure suivante). Les courbes correspondant à l'autre réplicat (courbes rouge et bleu clair dans la figure suivante) ont été obtenues en calculant la moyenne des courbes dans la même région du chromatogramme.



**Figure 64 : Courbes de SAXS des complexes Myb-court/ADN (deux réplicats en rouge et orange) et Myb-long/ADN (deux réplicats en bleu foncé et clair). Le panneau de droite montre un agrandissement dans la région des petits angles.**

L'analyse de Guinier montre que les complexes Myb-court/ADN et Myb-long/ADN ont respectivement des  $R_g$  de 20 et 28 Å. La différence de  $R_g$  de 8 Å est clairement trop faible pour correspondre aux 46 résidus supplémentaires dans un état désordonné (qui représenteraient environ 80 Å de long dans une conformation complètement étendue). Nous supposons donc à ce stade que le *linker* est au moins partiellement stabilisé, soit par un repliement plus important que les prédictions, soit par une interaction avec l'ADN.

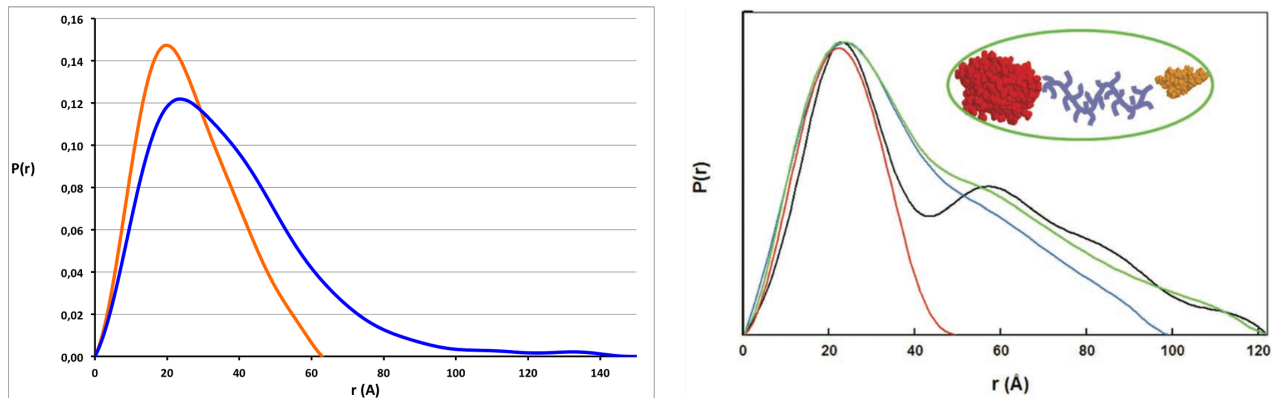
La représentation de Kratky normalisée (Durand *et al*, 2010) des courbes de diffusion des complexes Myb-court/ADN et Myb-long/ADN est montrée dans la figure suivante :



**Figure 65 : Représentation de Kratky normalisée des courbes de diffusion des complexes Myb-court/ADN (rouge et orange) et Myb-long/ADN (bleu foncé et bleu clair). Le panneau de droite indique les courbes typiques de différents niveaux de structuration (Receveur-Brechot & Durand, 2012) : globulaire compacte (bleu), multi-domaines compacte (vert), multi-domaines flexible (rouge) et dépliée (gris et violet).**

Ces représentations indiquent que les deux complexes sont relativement compacts. En particulier, si nous comparons les graphes de Kratky des complexes Myb-long/ADN et Myb-court/ADN à celui que nous obtenions avec le domaine RCT de RAP1 (présentant une grande flexibilité, voir partie 1), nous n’observons pas une flexibilité notablement plus importante pour le complexe Myb-long/ADN. Ceci conforte l’interprétation selon laquelle le *linker* adopte une conformation non aléatoire, en accord avec son interaction avec l’ADN montrée par l’expérience d’empreinte de nos collaborateurs.

Les fonctions de distribution des distances des deux complexes (deuxième réplicat) sont présentées dans la figure suivante :



**Figure 66 : Fonctions de distribution des distances des complexes Myb-court/ADN (orange) et Myb-long/ADN (bleu). Le panneau de droite rappelle les formes caractéristiques que peut prendre cette fonction : en rouge pour un domaine globulaire, en bleu pour le domaine globulaire connecté à une extrémité dépliée, en vert pour la protéine à deux domaines connectés par une région flexible, en noir la même protéine à deux domaines avec un linker plus rigide (Receveur-Brechot & Durand, 2012).**

Ces courbes ne donnent pas d’indication sur la flexibilité des objets. Les paramètres  $R_g$  et  $D_{max}$  calculés à partir des fonctions  $P(r)$  sont présentés dans le tableau suivant :

**Tableau 7 :  $R_g$  et  $D_{max}$  des complexes Myb-court/ADN et Myb-long/ADN**

Complexe	$R_g$ (Å)	$D_{max}$ (Å)
Myb-court/ADN	20,47	63
Myb-long/ADN	30,33	150

L’allure de la fonction  $P(r)$  du complexe Myb-court/ADN est caractéristique d’un objet globulaire légèrement allongé, avec un rapport d’ellipticité  $R_g/D_{max}$  proche de 1/3 (ce rapport vaut 1/2 pour une sphère parfaite). Dans le cas du complexe Myb-long/ADN, le rapport  $R_g/D_{max}$  est compris entre 1/5 et 1/4, mais sa fonction  $P(r)$  a globalement elle aussi

une allure caractéristique d'un objet globulaire allongé car les grandes distances sont très faiblement représentées.

### 3.1.3 Cristallographie

Afin de caractériser à l'échelle moléculaire le mode d'interaction du Myb-long avec l'ADN, nous avons étudié par cristallographie la structure tridimensionnelle de ce complexe.

#### 3.1.3.1 Cristallisation

Un premier crible de 7x 96 conditions de cristallisation (certaines redondantes) a été réalisé à la plateforme de cristallogénèse de l'Institut Pasteur (Paris) avec trois complexes différents basés sur les oligonucléotides de 16 pb T2-S1S2, T2-S1S3 et T2-S1S4 (cf. Matériel et méthodes). L'optimisation au laboratoire des meilleures pistes obtenues avec ces cribles a conduit à la condition de cristallisation suivante :

- extrémités 3'-sortantes (oligonucléotide T2-S1S3),
- 17 °C,
- ~23 mg/mL en complexe,
- solution de cristallisation :
  - HEPES 100 mM pH = 7,5
  - PEG 8000 3-10 % m/v
  - Ca acétate 50 mM

Cette condition n'a malheureusement produit que des cristaux irréguliers qui diffractaient seulement à très basse résolution (environ 20 Å).

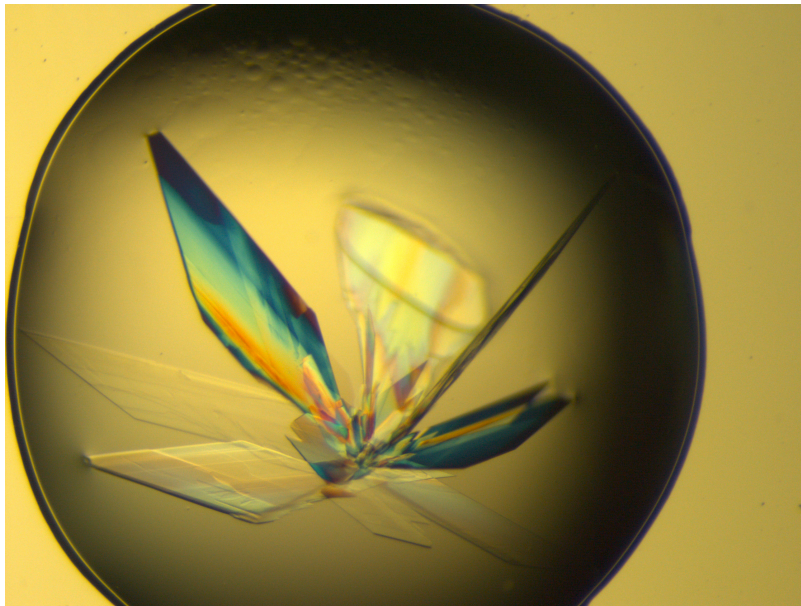
La cristallisation des complexes protéine/ADN est affectée par la longueur et le type d'extrémités (donc par la séquence) de l'oligonucléotide utilisé, et il peut être plus efficace de cribler peu de conditions avec plusieurs types d'oligonucléotides plutôt que de très nombreuses conditions avec un seul oligonucléotide défavorable (Hollis, 2007). Nous avons donc cherché à optimiser les oligonucléotides pour faire correspondre leur longueur et leurs extrémités (nombre et nature des bases sortantes) à celles qui sont le plus souvent observées dans les complexes protéine/ADN disponibles dans la PDB (Hol, 2011; Brown & Freemont, 1996). Cette démarche nous a conduit à concevoir des oligonucléotides de 20 pb avec une seule base sortante (A ou T) : T2-S5S6, T2-S5S7 et T2-S5S8 (cf. Matériel et méthodes).

Afin de tester uniquement l'effet de l'oligonucléotide dans un premier temps, nous avons commencé par tester la meilleure condition dont nous disposions alors (détaillée ci-dessus) avec ces nouveaux oligonucléotides. Ceci nous a confirmé que les cristaux les plus réguliers se forment avec l'oligonucléotide comportant des extrémités sortantes en 3' (T2-S5S7), et à 17 °C, mais le pouvoir de diffraction de ces cristaux n'était pas amélioré. Nous avons donc

réalisé un nouveau crible plus restreint, uniquement avec l'oligonucléotide T2-S5S7 et à 17 °C, en utilisant les 2x 48 conditions décrites par Pryor *et al* (2012) et Le Bihan *et al* (2013) et optimisées pour la cristallisation de complexes protéine/ADN et domaines Myb/ADN, respectivement. Parmi ces conditions, seule la n° 48 du crible décrit par Pryor *et al* (2012) nous a permis d'obtenir des cristaux. Ces cristaux ont une apparence meilleure que les précédents et apparaissent de façon bien reproductible :

- extrémités 3'-sortantes (oligonucléotide T2-S5S7),
- 17 °C,
- ~23 mg/mL en complexe,
- solution de cristallisation :
  - HEPES 100 mM pH = 7,5
  - PEG 4000 25-29 % m/v
  - CaCl<sub>2</sub> 170-180 mM

L'apparence typique de ces cristaux est montrée dans la photo suivante :



**Figure 67 : Cristaux du complexe Myb-long/T2-S5S7**

### 3.1.3.2 Collecte et traitement des données

Ces derniers cristaux nous ont permis d'obtenir plusieurs jeux de données à des résolutions allant de 1,8 à 2,2 Å, tous du groupe d'espace P 1 2<sub>1</sub> 1. Ces jeux de données sont présentés dans le tableau suivant, avec les statistiques de traitement des données :



**Tableau 8 : Jeux de données cristallographiques et statistiques de traitement des données (les informations entre parenthèses concernent la tranche de plus haute résolution).**

<b>Cristal</b>	<b>Myb2</b>	<b>Myb9</b>	<b>Myb10</b>	<b>Myb31</b>	<b>Myb32</b>	<b>Myb34</b>
<b>Maille</b> a, b, c	87,87	89,23	89,12	89,07	89,0 28,37	89,16
(Å) α, β, γ (°)	27,21	28,43	28,47	28,22	110,62 90	28,45
	108,74 90	110,6 90	110,44 90	110,71 90	99,479 90	110,42 90
	98,29 90	99,5 90	99,42 90	99,49 90		99,36 90
<b>Résolution (Å)</b>	49,024 –	44,003 –	43,959 –	43,926 –	43,89 –	43,986 –
	2,146	1,991	2,093	2,198	1,789	1,788
	(2,28 –	(2,12 –	(2,22 –	(2,33 –	(1,91 –	(1,90 –
	2,15)	2,00)	2,09)	2,20)	1,80)	1,79)
<b>Nombre de réflexions uniques</b>	28718	38419	32274	28222	52467	51069
<b>Complétude (%)</b>	99,49	99,35	96,79	98,60	99,44	96,42
<b>I/σ(I)</b>	13,79	15,76	12,35	8,00	12,78	11,02
	(1,77)	(1,60)	(1,72)	(1,12)	(0,95)	(0,93)
<b>R<sub>meas</sub> (%)</b>	8,5 (96,9)	6,5	9,9	12,8	5,9	5,3
		(119,1)	(108,8)	(154,7)	(176,7)	(116,4)
<b>CC1/2 (%)</b>	99,9	99,9	99,9	99,7	99,9	99,8
	(67,7)	(81,6)	(70,0)	(53,2)	(63,0)	(54,3)

Les paramètres de maille proches (écart de moins de 2 % environ) indiquent que tous ces cristaux sont isomorphes.

Le facteur  $I/\sigma(I)$  est le rapport signal sur bruit moyen de l'ensemble des réflexions.

Le facteur  $R_{meas}$  mesure pour chaque réflexion l'écart entre son intensité et la moyenne des intensités obtenue avec toutes les mesures indépendantes de cette même réflexion, et ce de façon normalisée par rapport à la redondance des données (Diederichs & Karplus, 1997). Il se calcule par la formule suivante :

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)}$$

où  $I_i(hkl)$  est l'intensité mesurée pour la  $i$ -ème fois pour la réflexion d'indices (hkl),  $\bar{I}(hkl)$  est la moyenne des intensités mesurées pour la réflexion d'indices (hkl) et  $n$  est la

multiplicité (ou redondance) du jeu de données. Ce facteur est donc une mesure de la cohérence interne d'un jeu de données.

Le facteur  $CC1/2$  est le coefficient de corrélation entre deux demi-jeux de données sélectionnés au hasard depuis le jeu de données analysé (Karplus & Diederichs, 2012). À ce titre il est également une mesure de la cohérence interne d'un jeu de données. Il vaut 1 (100 %) à basse résolution où le rapport signal sur bruit est grand, et décroît vers la haute résolution.

Puisque l'intensité diffractée diminue avec la résolution, la limite de résolution exploitable est habituellement déterminée à partir des seuils suivants (Evans, 2011) :

- la résolution à partir de laquelle  $R_{meas}$  dépasse 60-80 %,
- la résolution à partir de laquelle le rapport signal sur bruit  $I/\sigma(I)$  tombe en dessous de 2 (pour des données non fusionnées) ou 1 (pour des données fusionnées),
- la résolution à partir de laquelle le coefficient  $CC1/2$  tombe en dessous de 50 %.

Selon ces critères, les limites de résolution de nos jeux de données semblent raisonnables. Le coefficient  $CC1/2$  dans la tranche de plus haute résolution suggère même que nous avons été conservateurs pour le choix de la limite de résolution exploitable des jeux de données Myb2, Myb9, Myb10 et Myb32.

### 3.1.3.3 Résolution de la structure et affinement

La structure du complexe Myb-court/ADN est disponible dans la PDB (entrée 1W0U), à 100 % d'identité avec notre complexe, mais cristallise dans un groupe d'espace différent ( $P 2_1 2_1 2_1$ ). Nous avons donc utilisé la méthode du remplacement moléculaire pour déterminer les phases. La préparation du modèle à partir de l'entrée PDB 1W0U est décrite dans la partie Matériel et méthodes.

Les coefficients de Matthews (Matthews, 1968) calculés pour le cristal Myb10 en supposant 1, 2 et 3 copies du complexe dans l'unité asymétrique sont présentés ci-dessous (les résultats pour les autres cristaux sont similaires) :

Cell volume:	276434.844			
For estimated molecular weight	35615.3			
Nmol/asym	Matthews	Coeff	%solvent	P(tot)
1	3.88	71.72	0.60	
2	1.94	43.44	0.39	
3	1.29	15.16	0.01	

Ces valeurs suggèrent que l'unité asymétrique du cristal contient deux copies du complexe. Sachant que le complexe est constitué d'un oligonucléotide à deux sites lié par deux

molécules de Myb-long (une par site), ceci nous indique le nombre de copies du modèle à chercher dans l'unité asymétrique par la procédure de remplacement moléculaire (voir la partie Matériel et méthodes).

Les résultats du remplacement moléculaire (réalisé avec le logiciel Phaser ; McCoy *et al*, 2007) et du premier affinement en corps rigide sont présentés dans le tableau suivant.

**Tableau 9 : Résultats du remplacement moléculaire et du premier affinement**

<b>Cristal</b>	<b>Myb2</b>	<b>Myb9</b>	<b>Myb10</b>	<b>Myb31</b>	<b>Myb32</b>	<b>Myb34</b>
<b>Nombre de solutions</b>	1	1	1	1	1	1
<b>Groupe d'espace</b>	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1
<b>Top LLG</b>	2755,010	3841,366	3565,377	2989,351	4686,503	4706,498
<b>Top TFZ (&gt;8)</b>	8,1	21,1	20,1	20,0	35,9	20,5
<b>R<sub>work</sub> (%)</b>	34,58	36,26	34,22	33,65	36,12	34,48
<b>R<sub>free</sub> (%)</b>	40,40	40,02	38,38	37,23	38,94	38,17

Le score LLG (*Log Likelihood Gain*) est une mesure de l'accord entre le modèle et les données expérimentales relativement à un modèle de même composition en atomes mais aux coordonnées aléatoires. Sa valeur augmente au cours de la procédure de remplacement moléculaire dans le cas d'une solution correcte (la valeur minimale visée par Phaser est 120).

Le score TFZ (*translation function Z-score*) est un indicateur du résultat de la recherche par translation. Ce score compare la valeur de LLG obtenue après une translation donnée avec la valeur moyenne des LLG obtenus après des translations aléatoires. La documentation de Phaser indique qu'un score TFZ supérieur ou égal à 8 en fin de remplacement moléculaire désigne une solution certaine.

Le facteur R est une mesure de l'accord entre les amplitudes des facteurs de structure calculés à partir du modèle (appelés  $F_{calc}$ ) et celles des facteurs de structure expérimentaux (appelés  $F_{obs}$ ). Il est calculé selon la formule suivante :

$$R = \frac{\sum_{000}^{hkl} ||F_{obs}|_{hkl} - |F_{calc}|_{hkl}|}{\sum_{000}^{hkl} |F_{obs}|_{hkl}}$$

R<sub>free</sub> est calculé avec une petite fraction (≤ 5 %) des réflexions prises aléatoirement dans le jeu de données avant l'affinement, et qui par la suite ne sont pas utilisées pour l'affinement. R<sub>free</sub> est ainsi une mesure non biaisée de l'accord entre le modèle et les données expérimentales. R<sub>work</sub> est calculé selon la même formule, avec le reste du jeu de données

(c'est-à-dire les réflexions utilisées pour l'affinement). La valeur des facteurs R est d'autant plus petite que l'accord entre le modèle et les données expérimentales est bon.

À ce stade, les quatre copies du modèle utilisé pour le remplacement moléculaire semblent correctement positionnées les unes par rapport aux autres. Les cartes de densité électronique sont suffisamment interprétables pour permettre la construction complète de l'oligonucléotide, ce qui confirme que les positions relatives des modèles de départ sont correctes.

Malheureusement, malgré tous nos efforts d'affinement, nous n'observons que des pics de densité électronique discontinus dans la carte de différence mFo-DFc. Une explication possible vient du fait que les cristaux utilisés pour collecter ces données ont été obtenus plusieurs jours après la préparation du complexe (le temps nécessaire pour obtenir les premiers cristaux et affiner les conditions de cristallisation), ainsi il se peut que le *linker* se soit partiellement dégradé pendant la conservation à 4 °C du complexe en solution, et que nos cristaux ne contiennent qu'une forme dégradée de la construction Myb-long. Maintenant que la condition de cristallisation est bien définie, il serait possible d'obtenir de nouveaux cristaux en préparant les gouttes de cristallisation immédiatement après la préparation du complexe, afin de minimiser les risques de dégradation.

## 3.2 Discussion

Les résultats de calorimétrie et de SAXS suggèrent que les 46 résidus supplémentaires en N-terminal dans la construction Myb-long sont engagés au moins en partie dans l'interaction avec l'ADN. Ces résultats sont en accord avec ceux de l'expérience d'empreinte de nos collaborateurs. L'analyse cristallographique du complexe Myb-long/ADN est en cours et révèle que la région N-terminale est au moins partiellement désordonnée dans le cristal. Nous observons des pics positifs dans la carte de densité électronique, mais ils sont essentiellement en discontinuité du domaine Myb et leur modélisation est difficile.

## 4 Conclusion

Les objectifs que nous nous étions fixés étaient d'apporter des éléments de réponse aux questions suivantes :

- Comment l'architecture de TRF2 peut-elle expliquer ses fonctions si nombreuses et variées ?
- Quels sont les effets de RAP1 sur TRF2 au niveau conformationnel et fonctionnel ?
- Comment ces deux protéines s'influencent-elles mutuellement concernant le recrutement de leurs nombreux partenaires ?
- Quels déterminants structuraux de TRF2 permettent ses modes d'interaction avec l'ADN ?

La combinaison de méthodes biophysiques, biochimiques et de biologie structurale que nous avons mise en place nous a permis d'analyser les protéines entières, sans découpage en domaines isolés. Les résultats présentés dans la première partie documentent à l'échelle moléculaire l'interaction entre TRF2 et RAP1. Nous avons montré que TRF2 est hautement flexible, au même titre que le complexe TRF2/RAP1. Nous avons également montré que l'assemblage du complexe s'accompagne d'importants ajustements de conformation des deux protéines, et implique une interaction de basse affinité qui engage de grandes régions des deux protéines, et affecte leurs propriétés d'interactions. Ces résultats sont actuellement soumis à publication. Enfin, nous avons commencé l'étude de la région C-terminale de TRF2. Les premiers résultats obtenus sont encourageants, et cette étude est en bonne voie d'aboutissement et donnera lieu à une deuxième publication dans les prochains mois.

Nos résultats alimentent le mystère qui entoure RAP1. Nous savons maintenant que la conservation de sa région N-terminale est probablement associée, au moins chez les vertébrés, à son mode d'interaction avec TRF2. Cependant les implications biologiques de ce mode d'interaction restent hypothétiques : prévention des interactions non spécifiques du domaine TRFH de TRF2, dimérisation induite du domaine BRCT de RAP1 autorisant des liaisons avec des partenaires phosphorylés. Il serait intéressant de rechercher si ce mode d'interaction est conservé chez la levure *S. pombe*, entre les protéines Taz1 (orthologue des TRF) et Rap1. Si c'est le cas, ceci apporterait un argument supplémentaire pour avancer que la conservation de RAP1 est due à ses fonctions télomériques. Le même type d'étude ciblée pourrait être appliqué au complexe de *S. pombe*, à la différence près que nous ne savons pas encore si Taz1 possède le même mécanisme de recrutement par des motifs TBM (et si c'est le cas, la séquence du TBM consensus chez *S. pombe* n'est pas encore connue).

Du point de vue méthodologique, nos résultats sur l'assemblage du complexe TRF2/RAP1 rappellent l'importance d'étudier des protéines entières, et par conséquent la nécessité de combiner des approches expérimentales complémentaires. Les protéines multi-domaines représentent en effet plus des deux tiers du protéome chez les eucaryotes (Vogel *et al*, 2004), et il est clair que la compréhension de leurs fonctions à l'échelle moléculaire fait nécessairement appel à des combinaisons de techniques expérimentales telle que celle présentée ici. Il est important de remarquer que notre stratégie a été constamment adaptée aux nouvelles questions qui émergeaient au cours de l'étude, et qu'il nous aurait été difficile de prévoir à l'avance quelles techniques nous allions employer. Pour répondre aux problématiques actuelles en biologie structurale, une veille scientifique et technique assidue est essentielle afin de cultiver une connaissance large de méthodes expérimentales, d'entretenir une expertise dans l'une de ces méthodes (au moins) et d'entretenir un niveau d'utilisateur averti dans les autres pour connaître leurs avantages et leurs limites et être capable de les mettre en œuvre.

*« Je ne sais pas le reste. »*

## 5 Matériel et méthodes

### 5.1 Constructions utilisées

#### 5.1.1 Protéines

Les constructions protéiques utilisées dans nos études sont listées ci-dessous. Les masses molaires, points isoélectriques et coefficients d'extinction molaire à 280 nm ont été déterminés à l'aide de l'outil ProtParam <http://web.expasy.org/protparam/>. Les étiquettes poly-histidine sont indiquées en gras jusqu'au site de coupure par la TEV ou la thrombine, la barre verticale délimite les résidus N-terminaux laissés par le vecteur et qui ne font pas partie de la séquence naturelle de la protéine.

##### 5.1.1.1 TRF2

**TRF2** (1-500)

**GHHHHHHHHHTGQQMGRDLENLYFQG**|AGGGGSSDGSRAAGRASRSSGRARRGRHEPGLGGPAERGAGEARLEEAVNRW  
VLKIFYHEALRAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEENLDCSFDMEALTPLESAINVLEMIKTEFTL  
TEAVVSSRKLKVEAAVIICIKNKEFEKASKILKKHMSKDPTTQKLRNDLLNIIEKNLAHPVIQNFYSYETFQQKMLRFLESHLDDAEPYL  
LTMAKKALKSESAASSTGKEDKQPAPGPVEKPPREPARQLRNPPTTIGMMTLKAAFKTLGSAQDSEAAFAKLDQKDLVLPQTALPAS  
PALKNKRPRKDENESSAPADGEGGSELQPKNKRMTISRVLLEEDSQSTEPSAGLNSSQEASAPPSKPTVLNQPLPGEKNPKVPGK  
WNSSNGVEEKETWVEEDELQVQAAPDEDSTTNITKKQKWTVEESEWVKAGVQKYGEGNWAASKNYPFVNRTAVMIKDRWRT  
MKRLGMN

525 résidus ; pI = 9,14 ; MW = 58470,1 Da ;  $\epsilon_{280\text{ nm}}$  = 48930 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après TEV : 500 résidus ; pI = 9,22 ; MW = 55476,9 Da ;  $\epsilon_{280\text{ nm}}$  = 47440 M<sup>-1</sup>.cm<sup>-1</sup>.

**B-TRFH** (2-245)

**AGAASHHHHHHHHTGENLYFQG**AGGGGSSDGSRAAGRASRSSGRARRGRHEPGLGGPAERGAGEARLEEAVNRWVLKIFY  
HEALRAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEENLDCSFDMEALTPLESAINVLEMIKTEFTL  
TEAVV  
ESSRKLKVEAAVIICIKNKEFEKASKILKKHMSKDPTTQKLRNDLLNIIEKNLAHPVIQNFYSYETFQQKMLRFLESHLDDAEPYLLTMAK  
KALK

267 résidus ; pI = 9,42 ; MW = 30065,3 Da ;  $\epsilon_{280\text{ nm}}$  = 12950 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après TEV : 245 résidus ; pI = 9,53 ; MW = 27600,7 Da ;  $\epsilon_{280\text{ nm}}$  = 11460 M<sup>-1</sup>.cm<sup>-1</sup>.

**TRFH** (43-245)

**AGAASHHHHHHHHTGENLYFQG**AGEARLEEAVNRWVLKIFYHEALRAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVM  
QCLSRIEENLDCSFDMEALTPLESAINVLEMIKTEFTL  
TEAVV  
ESSRKLKVEAAVIICIKNKEFEKASKILKKHMSKDPTTQKLRNDLL  
NIIEKNLAHPVIQNFYSYETFQQKMLRFLESHLDDAEPYLLTMAK  
KALK

226 résidus ; pI = 8,41 ; MW = 26151,1 Da ;  $\epsilon_{280\text{ nm}}$  = 12950 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après TEV : 204 résidus ; pI = 8,7 ; MW = 23686,5 Da ;  $\epsilon_{280\text{ nm}}$  = 11460 M<sup>-1</sup>.cm<sup>-1</sup>.

**Myb-court** (447-500)

**MAGAASHHHHHHHHTGENLYFQG**|KQKWTVEESEWVKAGVQKYGEGNWAASKNYPFVNRTAVMIKDRWRTMKRLGMN

77 résidus ; pI = 9,82 ; MW = 8940 Da ;  $\epsilon_{280\text{ nm}}$  = 26470 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après TEV : 55 résidus ; pI = 10,09 ; MW = 6475,4 Da ;  $\epsilon_{280\text{ nm}}$  = 24980 M<sup>-1</sup>.cm<sup>-1</sup>.



**Myb-long**

(400-500)

**MGHHHHHHHHGTGQMGRDLENLYFQ**GEKNPKVPGKWNSSNGVEEKETWVEEDELQVQAAPDEDSTTNITKKQKWTVEE  
SEWVKAGVQKYGEGNWAASKNYPFVNRTAVMIKDRWRMTMKRLGMN

127 résidus ; pI = 7,11 ; MW = 14815,4 Da ;  $\epsilon_{280\text{ nm}}$  = 37470 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après TEV : 101 résidus ; pI = 8,12 ; MW = 11691,1 Da ;  $\epsilon_{280\text{ nm}}$  = 35980 M<sup>-1</sup>.cm<sup>-1</sup>.

## Comparaison des constructions Myb-court et Myb-long :

Myb-court 445 -----GKKQKWTVEESEWVK 459  
Myb-long 400 GEKNPKVPGKWNSSNGVEEKETWVEEDELQVQAAPDEDSTTNITKKQKWTVEESEWVK 459  
\*\*\*\*\*

Myb-court 460 AGVQKYGEGNWAASKNYPFVNRTAVMIKDRWRMTMKRLGMN 500  
Myb-long 460 AGVQKYGEGNWAASKNYPFVNRTAVMIKDRWRMTMKRLGMN 500  
\*\*\*\*\*

**5.1.1.2 RAP1****RAP1**

(1-399)

**MGHHHHHHHHHPTTENLYFQ**GHT|MAEAMDGLGKDPNGPHTSSTLFVRDDGSSMSFYVRPSPAKRRLSTLILHGGGTVCRVQEPG  
AVLLAQPGALAEASGDFISTQYILDCVERNERLELEAYRLGPASAADTGSEAKPGALAEGAAEPEPQRHAGRIAFTDADDVAILTYVK  
ENARSPSSVTGNALWKAMEKSSLTQHSWQSLKDRYLKHLRGQEHKYLKLLGDAPVSPSSQKLKRKAEDPEAADSSEGPQNKRTPDLP  
EYVKEEIQENEEAVKMLVEATREFEVEVVDESPPDFEIHITMCDDDPPTPEEDSETQPDEEEEEEEKVSQPEVGAAIKIIRQLMEKF  
NLDLSTVTQAFKNSGELEATSFLASGQRADGYPIWSRQDDIDLQKDEDEDREALVKKFQAQNVARRIEFRKK

422 résidus ; pI = 4,87 ; MW = 47071,9 Da ;  $\epsilon_{280\text{ nm}}$  = 29910 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après TEV : 402 résidus ; pI = 4,66 ; MW = 44555,2 Da ;  $\epsilon_{280\text{ nm}}$  = 28420 M<sup>-1</sup>.cm<sup>-1</sup>.

**RAP1-ARAGA** (1-399 et mutation indiquée en gras)

**MGHHHHHHHHHPTTENLYFQ**GHT|MAEAMDGLGKDPNGPHTSSTLFVRDDGSSMSFYVRPSPAKRRLSTLILHGGGTVCRVQEPG  
AVLLAQPGALAEASGDFISTQYILDCVERNERLEEA**ARAGA**ASAADTGSEAKPGALAEGAAEPEPQRHAGRIAFTDADDVAILTYV  
KENARSPSSVTGNALWKAMEKSSLTQHSWQSLKDRYLKHLRGQEHKYLKLLGDAPVSPSSQKLKRKAEDPEAADSSEGPQNKRTPDLP  
EEEVYKKEIQENEEAVKMLVEATREFEVEVVDESPPDFEIHITMCDDDPPTPEEDSETQPDEEEEEEEKVSQPEVGAAIKIIRQLMEKF  
FNLDLSTVTQAFKNSGELEATSFLASGQRADGYPIWSRQDDIDLQKDEDEDREALVKKFQAQNVARRIEFRKK

422 résidus ; pI = 4,87 ; MW = 46911,7 Da ;  $\epsilon_{280\text{ nm}}$  = 28420 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après TEV : 402 résidus ; pI = 4,66 ; MW = 44395 Da ;  $\epsilon_{280\text{ nm}}$  = 26930 M<sup>-1</sup>.cm<sup>-1</sup>.

**RAP1\_1-208**

(1-208)

**MGHHHHHHHHHPTTENLYFQ**GHT|MAEAMDGLGKDPNGPHTSSTLFVRDDGSSMSFYVRPSPAKRRLSTLILHGGGTVCRVQEPG  
AVLLAQPGALAEASGDFISTQYILDCVERNERLELEAYRLGPASAADTGSEAKPGALAEGAAEPEPQRHAGRIAFTDADDVAILTYVK  
ENARSPSSVTGNALWKAMEKSSLTQHSWQSLKDRYLKHLRGQEHKYLKLLGDAPVSPSSQK

231 résidus ; pI = 6,33 ; MW = 25218 Da ;  $\epsilon_{280\text{ nm}}$  = 21430 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après TEV : 211 résidus ; pI = 5,99 ; MW = 22701,3 Da ;  $\epsilon_{280\text{ nm}}$  = 19940 M<sup>-1</sup>.cm<sup>-1</sup>.

**RAP1-RCT**

(270-399)

**GSSHHHHHHSSGLVPRGSHM**|PPDFEIHITMCDDDPPTPEEDSETQPDEEEEEEEKVSQPEVGAAIKIIRQLMEKFNLDSLSTVTQAF  
LKNSGELEATSFLASGQRADGYPIWSRQDDIDLQKDEDEDREALVKKFQAQNVARRIEFRKK

150 résidus ; pI = 4,77 ; MW = 16991,6 Da ;  $\epsilon_{280\text{ nm}}$  = 6990 M<sup>-1</sup>.cm<sup>-1</sup>.  
Après thrombine : 134 résidus ; pI = 4,37 ; MW = 15240,7 Da ;  $\epsilon_{280\text{ nm}}$  = 6990 M<sup>-1</sup>.cm<sup>-1</sup>.

### 5.1.1.3 WRN

#### WRN-RQC

(949-1079)

**MGHHHHHHHGTGQMQMRDLENLYFQGDDSEDTSWDFGQAFKLLSAVDILGEKFGIGLPILFLRGSNSQRLADQYRRHSFLGT  
GKDQTESWWKAFSRQLITEGFLVEVSRYNKFMKICALTKKGRNWLHKANTESQSLILQANEELCPKLLLPSS**

157 résidus ; pI = 8,57 ; MW = 18020,3 Da ;  $\epsilon_{280\text{ nm}} = 26470 \text{ M}^{-1}.\text{cm}^{-1}$ .

Après TEV : 131 résidus ; pI = 8,96 ; MW = 14896 Da ;  $\epsilon_{280\text{ nm}} = 24980 \text{ M}^{-1}.\text{cm}^{-1}$ .

### 5.1.1.4 Ku

#### Ku70

MSGWESYYKTEGDEEAEQEEENLEASGDYKYSGRDSLIFLVDASKAMFESQSEDELTPFDMSIQCIQSVYISKIISDRDLLAVFYGT  
EKDKNSVNFKNIVYLQELDNPGAKRILELDQFKGQQGQKRFQDMMGHGSDYSLSEVLWVCANLFSQVQKMSHKRIMLFTNEDN  
PHGNDSAKASRARTKAGDLRDTGIFLDMHLKPKGGFDISLFYRDIISIAEDEDLRVHFEESKLEDLLRKVRKETRKRALSRLKLNK  
DIVISVGIYNLVQKALKPPPIKLYRETNEPVKTKTRTFNTSTGGLLPSTKRSQIYGSRQIILEKEETEELKRFDDPGLMLMGFKPLVLLK  
KHHYLRPSLFVYPEESLVIGSSTLFSALLIKCLEKEVAALCRYTPRRNIPPYFVALVPEEELDDQKIQTTPPGFQLVFLPFADDKRKMPFT  
EKIMATPEQVGMKAIVEKLRFTYRSDSFENPVLQQHFRNLEALALDLMEPEQAVDLTLPKVEAMNKRLGSLVDEFKELVYPPDYNP  
EGKVTKRKHDNEGSGSKRPKVEYSEEELKTHISKGTGKFTVPMLEACRAYGLKSGLKKQELLEALTKHFQD

#### 10-His-Ku80

MHHHHHHHHHHENLYFQGVRSNKAADVLCMDVGFMTSNSIPGIESPFQAKKVITMFVQRQVFAENKDEIALVLFGTGTDNP  
LSGGDQYQNTVHRHMLPDLLELIEDIESKIQPGSQADFLDALIVSMDVIQHETIGKKFEKRHIEIFDLSSRFSKSLDIIHSLKKCDI  
SLQFFLPFSLGKEDGSGDRGDGPFRLGGHGPSFPLKGITEQQKEGLEIVKMVMISLEGEDGLDEIYSFSESLRKLKCVFKKIERHSIHWPC  
RLTIGSNLSIRIAAYKSILQERVKKTWTVVDAKTLKKEDIQKQETVYCLNDDDETEVLKEDIQGFYRYSQVDFVSKVDEEQMKYKSEKCF  
SVLGFCKSSQVQRFFMGNQVLKVFVAARDDEAAVALSSLIHALDDLDMVAIVRYAYDKRANPQVGVAFPHIKHNYECLVYVQLPF  
MEDLRQYMFSSLKNKSKYAPTEAQLNAVDALIDSMSLAKKDEKTDLEDLFPPTKIPNPRFQRLQCLLHRALHPREPLPIQHIWN  
MLNPPAEVTTKSQIPLSKIKTLFPLIEAKKKDQVTAQEIQDNHEDGPTAKKLTQEGGAHFSVSSLAEGSVTSVGSVNPAENFRVLVK  
QKKASFEEASNQLINHIEQFLDTNETPYFMKSIDCIRAFREEAIKFSEEQRFNNFLKALQEKVEIKQLNHFWEIFVQDQGITLITKEEASGS  
SVTAAEAKKFLAPKDKPSGDAAVFEEGGDVDDLDMI

Pour le dimère : pI  $\approx$  5,94 ; MW = 69843 + 84927,8 = 154770,8 Da ;  $\epsilon_{280\text{ nm}} = 42290 + 42860 = 85150 \text{ M}^{-1}.\text{cm}^{-1}$ .

## 5.1.2 Oligonucléotides

Les oligonucléotides proviennent tous du fournisseur Eurogentec.

### 5.1.2.1 Première génération

Ces oligonucléotides ont été conçus pour pouvoir les assembler en duplexes de 16 pb avec deux sites TTAGGG, avec des extrémités franches ou de deux bases sortantes en 5' ou en 3' de façon à produire des extrémités cohésives terminées par un appariement GC. Ces extrémités cohésives peuvent établir des interactions que l'on espère favorables pour la cristallisation du complexe Myb/ADN.

#### T2-S1 (brin sens)

5' - CTAGGGTTAGGGTTAG -3'

#### T2-S2

5' - CTAACCCTAACCTAG -3'

### **T2-S3**

5' - AACCTAACCTAGCT -3'

### **T2-S4**

5' - AGCTAACCTAACCT -3'

L'hybridation de ces oligonucléotides produit les duplexes suivants :

### **T2-S1S1 (bouts francs)**

5' - CTAGGGTTAGGGTTAG -3'

3' - GATCCCAATCCCAATC -5'

### **T2-S1S3 (3'-sortant)**

5' - CTAGGGTTAGGGTTAG -3'

3' - TCGATCCCAATCCCAA -5'

### **T2-S1S4 (5'-sortant)**

5' - CTAGGGTTAGGGTTAG -3'

3' - TCCCAATCCCAATCGA -5'

### **5.1.2.2 Seconde génération**

Ces oligonucléotides ont été conçus sur la base des précédents, en tenant compte du fait que les complexes protéine/ADN disponibles dans la PDB ont le plus souvent un ADN de 20 pb et des extrémités sortantes d'une seule base formant une paire AT, avec un la base T en 5' (Hol, 2011; Brown & Freemont, 1996). D'autre part, l'ADN en conformation B possède en moyenne 10 pb par tour d'hélice, donc un oligonucléotide d'une longueur multiple de 10 donne un nombre entier de tours d'hélice, ce qui peut favoriser l'empilement des complexes en une hélice semi-continue.

### **T2-S5 (brin sens)**

5' - TACTAGGGTTAGGGTTAGCA -3'

### **T2-S6**

5' - TGCTAACCTAACCTAGTA -3'

### **T2-S7**

5' - GCTAACCTAACCTAGTAT -3'

### **T2-S8**

5' - ATGCTAACCTAACCTAGT -3'

L'hybridation de ces oligonucléotides produit les duplexes suivants :

**T2-S5S6** (bouts francs)

5' - TACTAGGGTTAGGGTTAGCA -3'  
3' - ATGATCCCAATCCCAATCGT -5'

**T2-S5S7** (3'-sortant)

5' - TACTAGGGTTAGGGTTAGCA -3'  
3' - TATGATCCCAATCCCAATCG -5'

**T2-S5S8** (5'-sortant)

5' - TACTAGGGTTAGGGTTAGCA -3'  
3' - TGATCCCAATCCCAATCGTA -5'

### 5.1.3 Peptides

Les peptides proviennent tous du fournisseur Eurogentec.

**Apollo-TBM** (494-511) ATEFRGLALKYLLTPVNF

**RAP1-TBM** (89-106) ENRERLELEAYRLGPASA

**TIN2-TBM** (248-265) THPEPLAGRHFNLAPLGR

## 5.2 Expression des protéines

Les protéines étudiées sont surexprimées de façon recombinante dans *E. coli* selon le protocole suivant (les détails spécifiques à chaque protéine sont détaillés par la suite). **Toutes les procédures de microbiologie sont réalisées en conditions stériles.**

### 5.2.1 Transformation

La transformation des bactéries chimiocompétentes est réalisée selon le protocole suivant :

- décongélation à 4 °C d'un aliquot de 100 µL de bactéries compétentes
- ajout de 2-mercaptoethanol qsp 25 mM (améliore l'efficacité de transformation)
- incubation 10 min à 4 °C
- ajout de 1 µL de solution de plasmide (typiquement à une concentration de l'ordre de 100 ng/µL)
- incubation au moins 30 min à 4 °C
- choc thermique : 45 s à 42 °C
- incubation au moins 2 min à 4 °C
- ajout de 250 µL de milieu SOC
- incubation au moins 1h à 37 °C sous agitation (900 rpm)
- étalement sur un milieu solide LB agar + antibiotique approprié (20 et 200 µL)
- incubation une nuit à 37 °C

Les transformations faites selon ce protocole produisent typiquement quelques colonies à quelques dizaines de colonies pour un étalement de 20 µL.

### 5.2.2 Culture des bactéries et induction de l'expression

Une colonie obtenue à l'étape précédente est mise en culture dans 100 mL de milieu LB additionné de l'antibiotique approprié, cette pré-culture est incubée une nuit à 37 °C sous agitation.

Le lendemain, des erlens de 800 mL de milieu LB additionné de l'antibiotique approprié sont ensemencés chacun avec 10 mL de la pré-culture saturée. Ces cultures sont incubées à 37 °C sous agitation (180 rpm) jusqu'à atteindre une densité de bactéries de  $DO_{600\text{ nm}} = 0,8 - 1,0$ . À cette densité, la culture est refroidie à 20 °C et l'expression de la protéine est induite par ajout d'IPTG qsp 0,5 mM final. La culture est alors poursuivie à 20 °C jusqu'au lendemain matin. Les cultures atteignent typiquement une densité finale de l'ordre de  $DO_{600\text{ nm}} = 1,7 - 3$ .

Les cultures sont finalement centrifugées à 3000 g pendant 35 min. Chaque culot est rincé par resuspension dans 40 mL de PBS 1X, et ces suspensions sont finalement centrifugées

une dernière fois à 3000 g pendant 20 min pour obtenir les culots finaux. Ces culots sont alors congelés rapidement dans l'azote liquide et stockés à -80 °C.

### 5.2.3 Détails spécifiques aux protéines

#### 5.2.3.1 TRF2, TRFH, B-TRFH, WRN\_949-1079, Myb-court, Myb-long

Ces protéines sont exprimées dans la souche *E. coli* BL21 DE3 star, à partir d'un plasmide pETM-13 contenant une séquence optimisée pour l'expression dans *E. coli* (achetés au fournisseur GenScript <http://www.genscript.com/> pour TRF2 et WRN, et sous-clonés à partir du plasmide TRF2 pour les domaines isolés de TRF2) et un gène de résistance à la kanamycine. Dans toutes les étapes décrites ci-dessus qui mentionnent un antibiotique, c'est la kanamycine à 30 µg/mL qui a été utilisée pour sélectionner les bactéries transformées par ces plasmides.

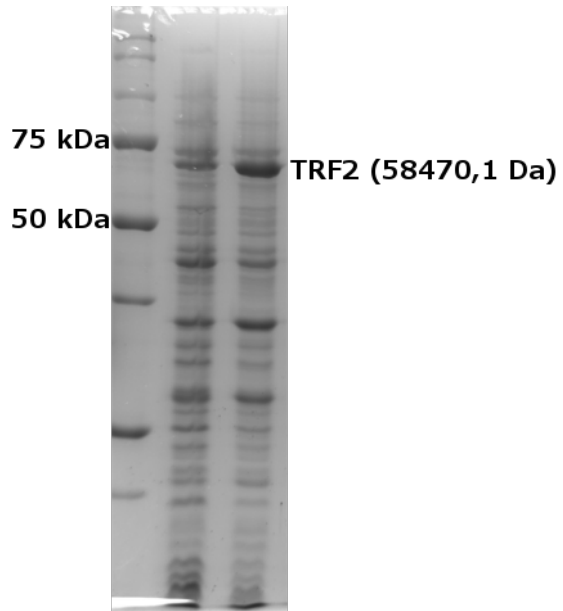
#### 5.2.3.2 RAP1, RAP1-ARAGA, RAP1\_1-208

Ces protéines sont exprimées dans la souche *E. coli* Rosetta, à partir d'un plasmide pET-16b contenant une séquence sauvage ou mutée (la mutation ARAGA et le codon stop permettant d'arrêter la synthèse au résidu 208 ont été introduits par le protocole QuickChange) et un gène de résistance à l'ampicilline. Dans toutes les étapes décrites ci-dessus qui mentionnent un antibiotique, c'est l'ampicilline à 100 µg/mL qui a été utilisée pour sélectionner les bactéries transformées par ces plasmides. La souche Rosetta contient par ailleurs un plasmide appelé pRARE, qui code pour les ARNt spécifiques de certains codons rares chez *E. coli* et permet ainsi un meilleur taux d'expression de protéines eucaryotes exprimées dans ce système. Ce plasmide contient un gène de résistance au chloramphénicol, par conséquent cet antibiotique a également été utilisé pour garantir que les bactéries transformées par le plasmide pET-16b ne perdent pas le plasmide pRARE. Le chloramphénicol a été utilisé à une concentration de 35 µg/mL.

### 5.2.4 Résultats des expressions

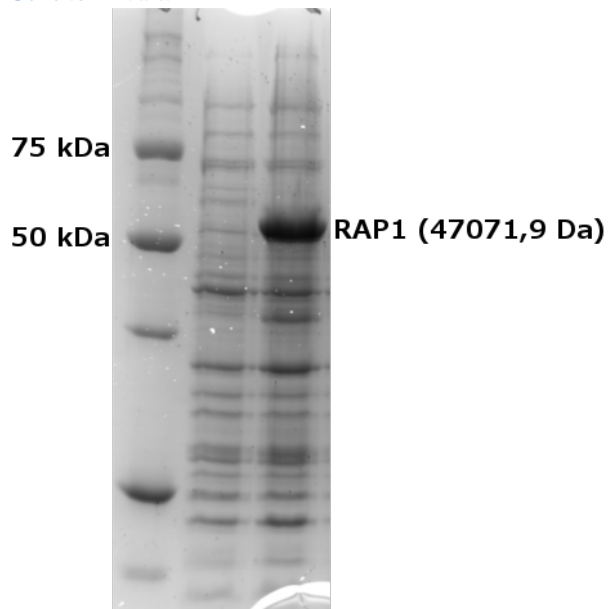
Les résultats typiques obtenus pour les expressions des protéines sont indiqués ci-dessous. Les pistes des gels sont, de gauche à droite : un marqueur de masse moléculaire, l'extrait bactérien total avant induction et l'extrait bactérien total après induction (les quantités de matériel déposées sont identiques entre ces deux pistes).

#### 5.2.4.1 TRF2



**Figure 68 : Expression de TRF2**

#### 5.2.4.2 RAP1



**Figure 69 : Expression de RAP1**

#### 5.2.4.3 RAP1-ARAGA

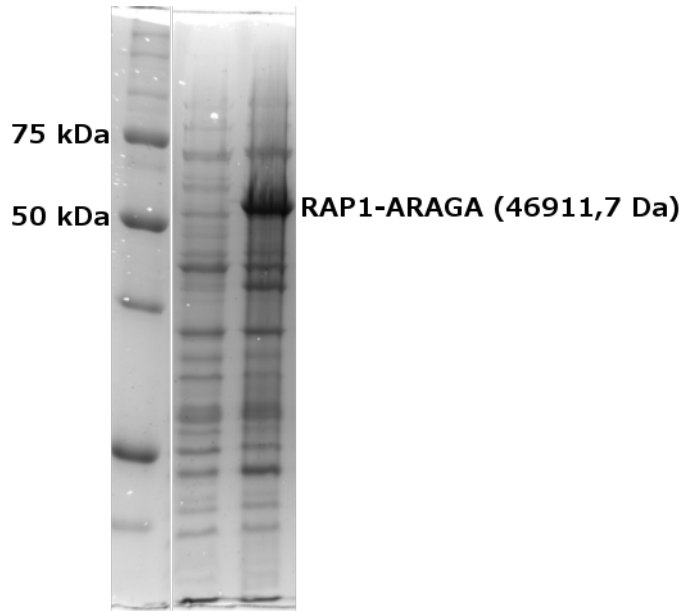


Figure 70 : Expression de RAP1-ARAGA

#### 5.2.4.4 RAP1\_1-208

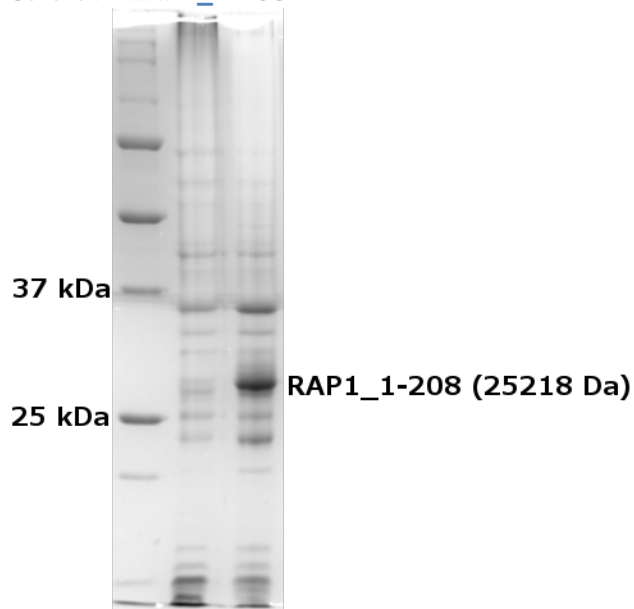
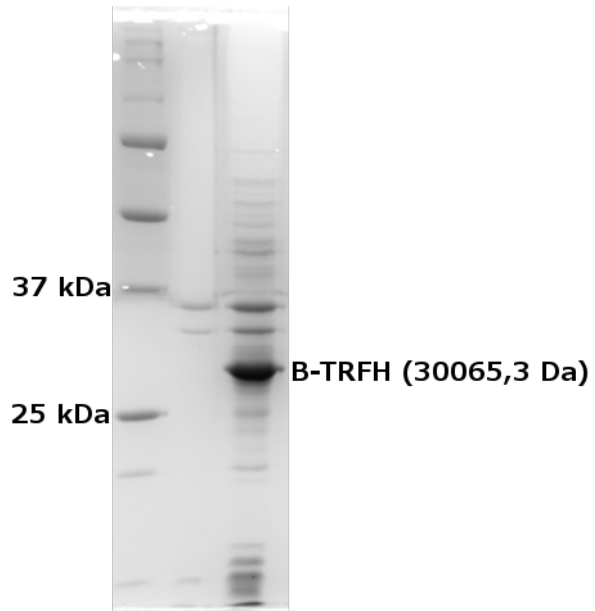


Figure 71 : Expression de RAP1\_1-208



#### 5.2.4.5 B-TRFH



**Figure 72 : Expression de B-TRFH**

5.2.4.6 WRN\_949-1079

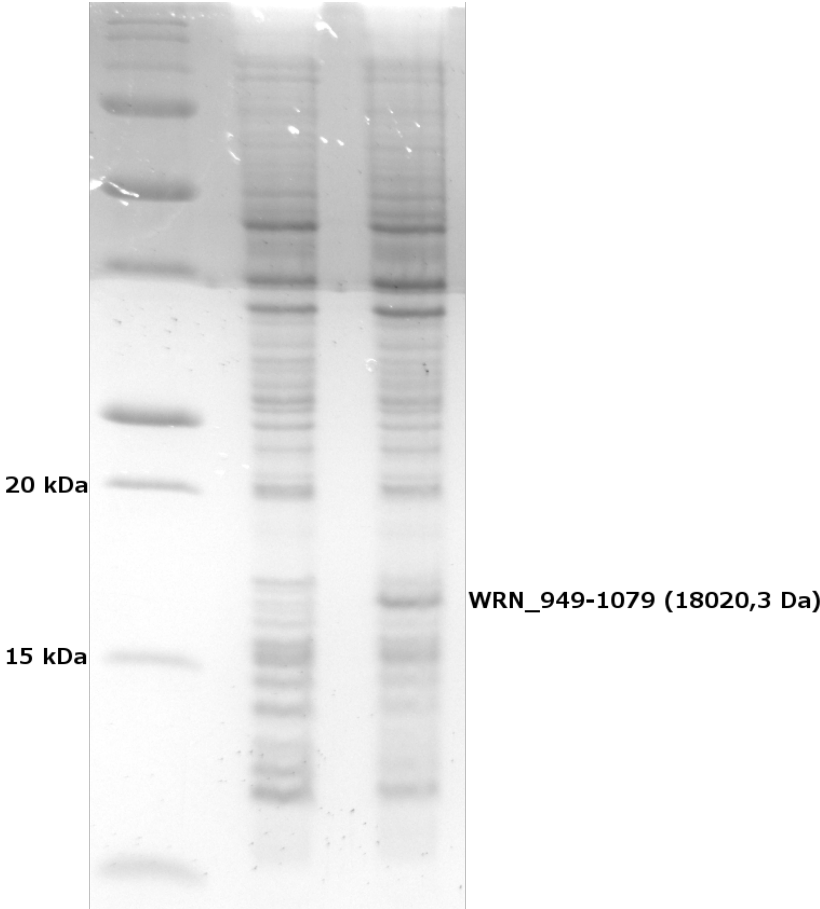
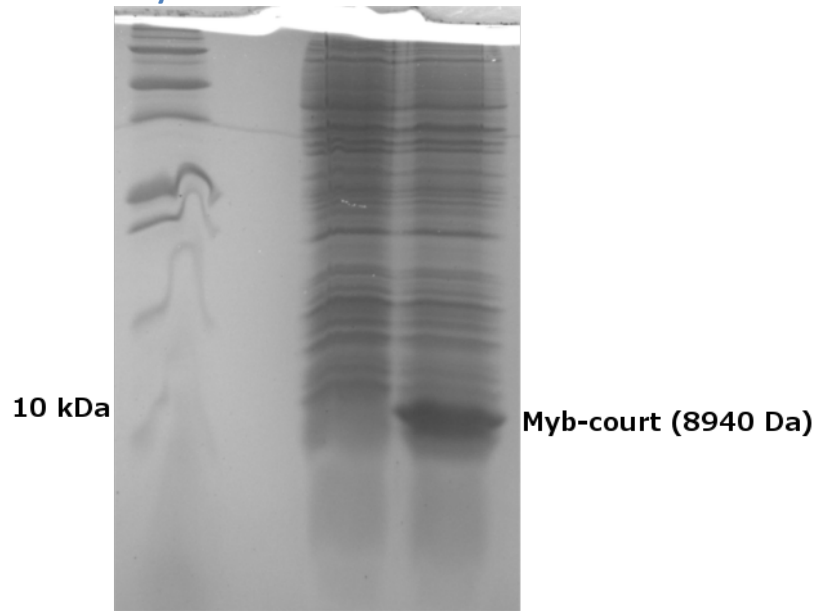


Figure 73 : Expression de WRN\_949-1079

#### 5.2.4.7 Myb-court



**Figure 74 : Expression du Myb-court**

## 5.3 Purification des protéines recombinantes

Le protocole suivant est suivi pour toutes les purifications de protéines recombinantes. Les détails spécifiques de chaque protéine sont indiqués par la suite.

### 5.3.1 Protocole

#### 5.3.1.1 Tampons

- **tampon NiA**
  - $\text{NaH}_2\text{PO}_4$  /  $\text{Na}_2\text{HPO}_4$  20 mM pH = 8
  - NaCl 500 mM
  - imidazole 60 mM
  - 2-mercaptoethanol 10 mM
- **tampon NiB** : idem NiA sauf imidazole 1 M
- **tampon Desalting**
  - Tris-HCl 20 mM pH = 7,5
  - NaCl 150 mM
  - 2-mercaptoethanol 5 mM
- **tampon Ni ap TEV**
  - Tris-HCl 20 mM pH = 7,5
  - NaCl 200 mM
  - imidazole 20 mM
  - 2-mercaptoethanol 5 mM

#### 5.3.1.2 Lyse des bactéries

Chaque culot bactérien est resuspendu dans 40 mL de tampon NiA additionné de PMSF qsp 1 mM final et de Triton-X-100 qsp 0,5 % v/v final. La suspension bactérienne résultante est additionnée de lysozyme à raison de 20 mg par culot et incubée 15 min à température ambiante sous agitation. Les bactéries sont alors lysées par sonication (sans dépasser 10 °C) : 60 pulses de 1 s espacées par 1 s de pause, à 35, 45, 55, 65 et enfin 75 % de la puissance du sonicateur. Le lysat est additionné de  $\text{MgCl}_2$  qsp 1,5 mM final et de benzonase à raison de 1  $\mu\text{L}$  par culot puis incubé 15 min à température ambiante sous agitation. Le lysat est finalement clarifié par centrifugation à 48 000 g et filtration 0,45  $\mu\text{m}$ .

#### 5.3.1.3 Purification des protéines

Les protéines sont purifiées par chromatographie d'affinité pour le nickel sur une colonne HisTrap FF 5 mL (GE Healthcare). L'extrait brut est chargé à 2 mL/min sur cette colonne préalablement équilibrée en tampon NiA, puis la colonne est lavée au même débit avec au

moins 20 CV de tampon NiA. La protéine est ensuite éluée par un gradient de 0 à 100 % v/v de tampon NiB en 15 CV (toujours à 2 mL/min), et l'éluat est collecté par fractions de 1 mL.

Après analyse par SDS-PAGE, les fractions d'intérêt sont réunies et dessalées par passage sur une colonne HiPrep 26/10 Desalting (GE Healthcare) préalablement équilibrée en tampon Desalting (tout à un débit de 1 mL/min). Après estimation de la concentration en protéine (par l'absorbance à 280 nm), la solution protéique est alors additionnée de TEV (produite et purifiée au laboratoire) à 4 % m/m et incubée une nuit à 4 °C afin de cliver le tag poly-His.

Le lendemain matin, la solution protéique est ajustée à 200 mM en NaCl (ou 400 mM pour RAP1 et ARAGA) et 20 mM en imidazole puis chargée à 2 mL/min sur la colonne HisTrap préalablement régénérée et équilibrée en tampon Ni ap TEV. La récupération de charge est collectée car elle contient la protéine d'intérêt. Pour contrôle, on collecte également l'éluat chassé par 100 % de tampon NiB pendant 1 CV.

À cette étape, le protocole diffère selon la protéine :

- Pour les domaines TRFH, B-TRFH et Myb-long de TRF2, et le domaine 1-208 de RAP1, la purification est terminée à cette étape.
- Pour RAP1 et RAP1-ARAGA, la purification est complétée par une étape de chromatographie d'échange d'anions.
- Pour TRF2, WRN\_949-1079 et le domaine Myb-court de TRF2, elle est complétée par une étape de chromatographie d'affinité pour l'héparine.

En fin de purification les solutions protéiques sont toutes additionnées de 5 % v/v de glycérol puis congelées rapidement dans l'azote liquide et enfin stockées à -80 °C.

Avant toute expérience, les protéines décongelées sont soit dialysées dans le tampon de l'expérience, soit placées dans le tampon de l'expérience par chromatographie d'exclusion stérique sur une colonne Superdex 200 10/300 GL (GE Healthcare) à 0,5 mL/min.

### 5.3.2 Détails spécifiques à RAP1 et RAP1-ARAGA

#### 5.3.2.1 Tampons

Le tampon NiB a une concentration en NaCl de 1 M pour RAP1 et RAP1-ARAGA. Ceci permet d'élué complètement la protéine, qui est aussi retenue aux ions Ni<sub>2+</sub> par une grande région riche en résidus acides (chargés négativement au pH de la purification) qui confère à la protéine son pI très bas.

Pour la même raison, le tampon Ni ap TEV a une concentration en NaCl de 400 mM.

- **tampon sans sel**

- Tris-HCl 20 mM pH = 7,5
- 2-mercaptoethanol 5 mM
- **tampon ResQ-A**
  - Tris-HCl 20 mM pH = 7,5
  - NaCl 100 mM
  - 2-mercaptoethanol 5 mM
- **tampon ResQ-B**
  - Tris-HCl 20 mM pH = 7,5
  - NaCl 2 M
  - 2-mercaptoethanol 5 mM

### 5.3.2.2 Protocole

Après analyse par SDS-PAGE de l'étape d'HisTrap "inverse", les fractions d'intérêt sont réunies et la solution protéique est diluée avec le tampon sans sel pour l'ajuster à 100 mM en NaCl. La solution protéique est alors chargée à 2 mL/min sur une colonne ResourceQ 6 mL (GE Healthcare) préalablement équilibrée en tampon ResQ-A. La protéine est éluée par un gradient de 0 à 100 % v/v de tampon ResQ-B en 15 CV.

Après une dernière analyse par SDS-PAGE, les fractions d'intérêt finales sont réunies et stockées.

## 5.3.3 Détails spécifiques à TRF2, WRN\_979-1049 et le domaine Myb-court de TRF2

### 5.3.3.1 Tampons

- **tampon sans sel**
  - Tris-HCl 20 mM pH = 7,5
  - 2-mercaptoethanol 5 mM
- **tampon HepA**
  - Tris-HCl 20 mM pH = 7,5
  - NaCl 100 mM
  - 2-mercaptoethanol 5 mM
- **tampon HepB**
  - Tris-HCl 20 mM pH = 7,5
  - NaCl 2 M
  - 2-mercaptoethanol 5 mM

### 5.3.3.2 Protocole

Après analyse par SDS-PAGE de l'étape d'HisTrap "inverse", les fractions d'intérêt sont réunies et la solution protéique est diluée avec le tampon sans sel pour l'ajuster à 100 mM en NaCl. La solution protéique est alors chargée à 2 mL/min sur une colonne HiTrap Heparin HP 5 mL (GE Healthcare) préalablement équilibrée en tampon HepA. La protéine est éluée par un gradient de 0 à 100 % v/v de tampon HepB en 15 CV.

Après une dernière analyse par SDS-PAGE, les fractions d'intérêt finales sont réunies et stockées.

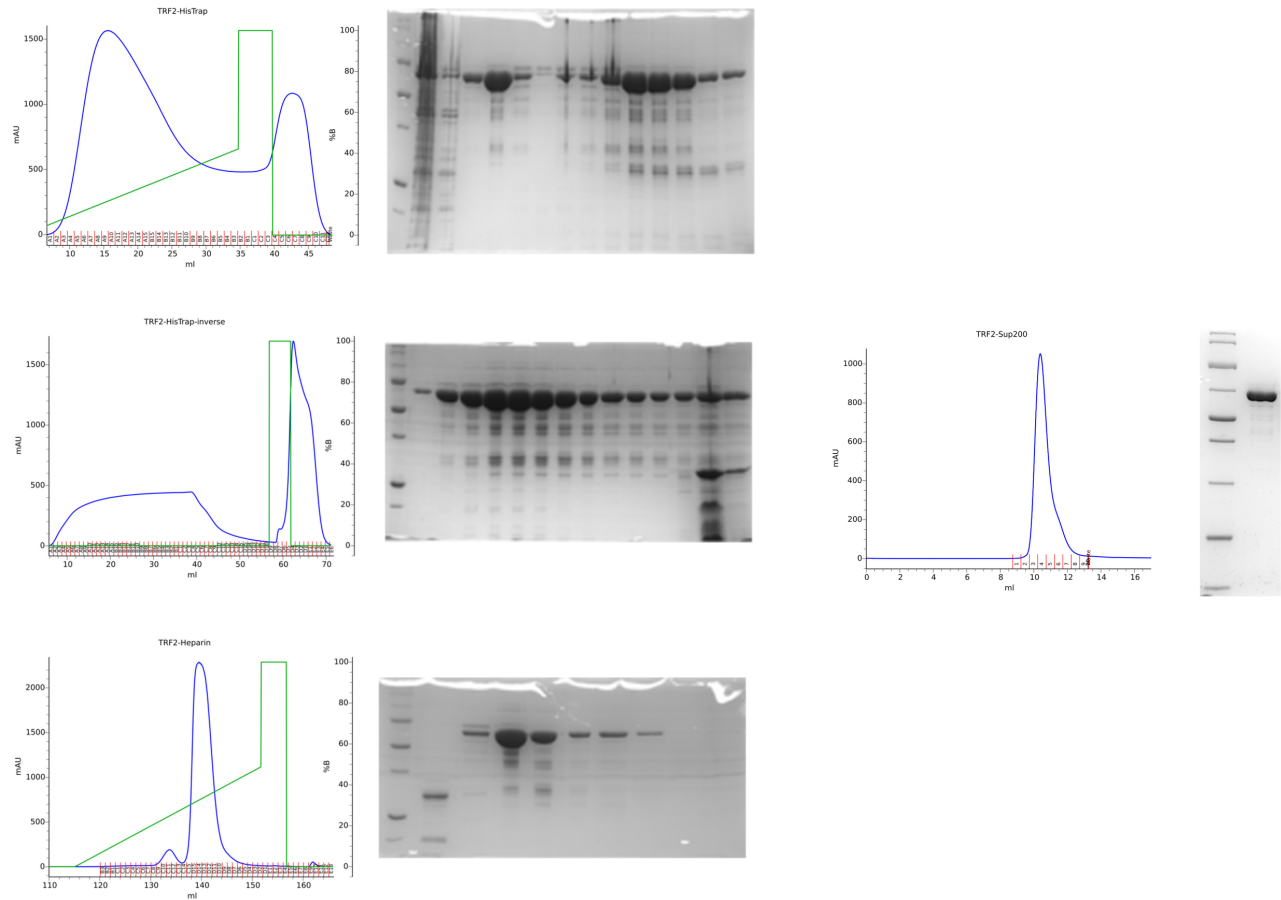
### 5.3.4 Résultats des purifications

Les résultats obtenus pour les purifications des protéines sont détaillés ci-dessous.

#### 5.3.4.1 TRF2

La purification de TRF2 est réalisée en quatre étapes : chromatographie d'affinité pour le nickel, clivage de l'étiquette histidine et séparation par une seconde chromatographie d'affinité pour le nickel, chromatographie d'affinité pour l'héparine, et finalement chromatographie d'exclusion stérique.

Les chromatogrammes et l'analyse des fractions par SDS-PAGE pour la purification de TRF2 sont présentés dans la figure suivante :



**Figure 75 : Purification de TRF2**

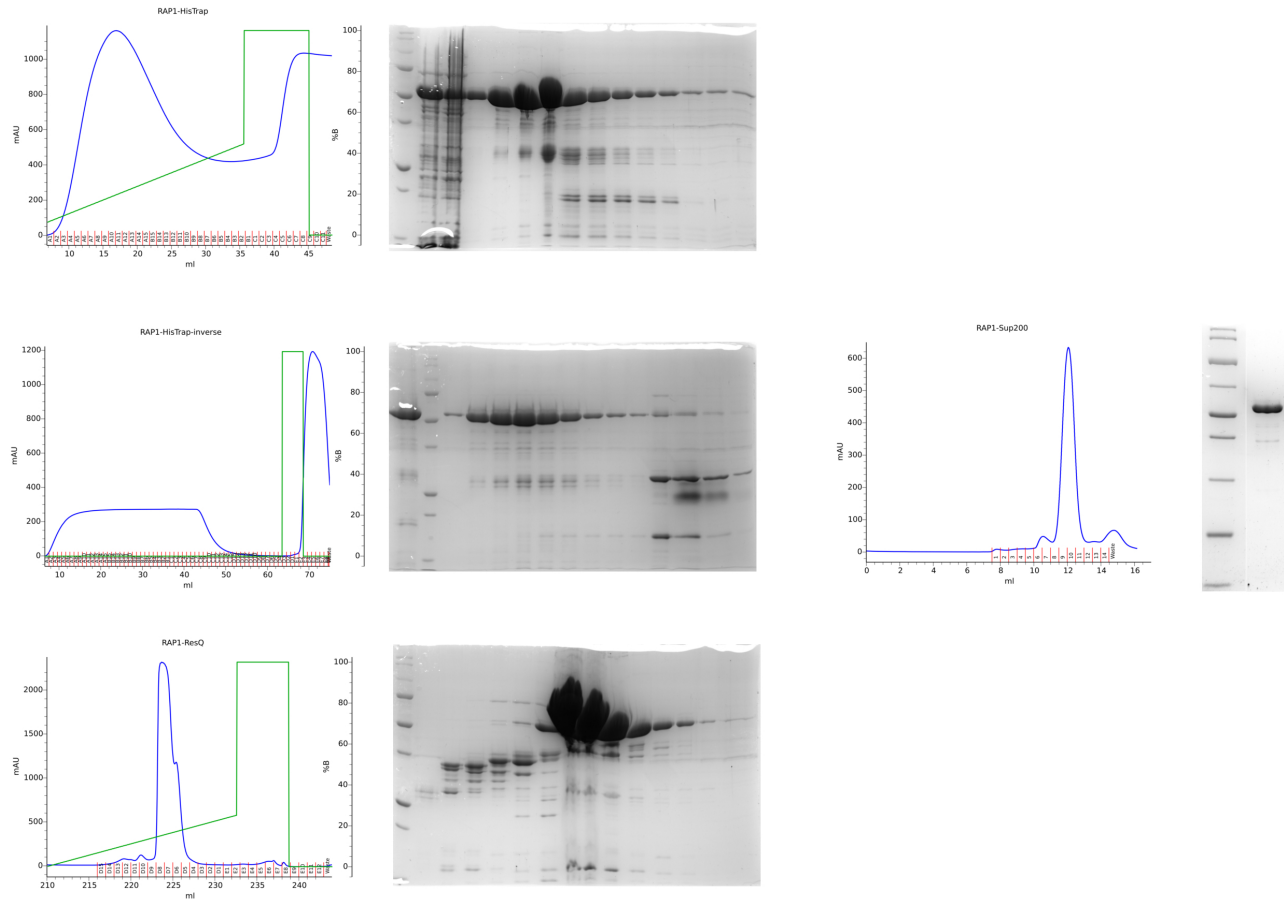
Le rendement est d'environ 7,5 mg de protéine par culot bactérien.

#### 5.3.4.2 RAP1

La purification de RAP1 est réalisée en quatre étapes : chromatographie d'affinité pour le nickel, clivage de l'étiquette histidine et séparation par une seconde chromatographie d'affinité pour le nickel, chromatographie d'échange d'anions, et finalement chromatographie d'exclusion stérique.

Les chromatogrammes et l'analyse des fractions par SDS-PAGE pour la purification de RAP1 sont présentés dans la figure suivante :





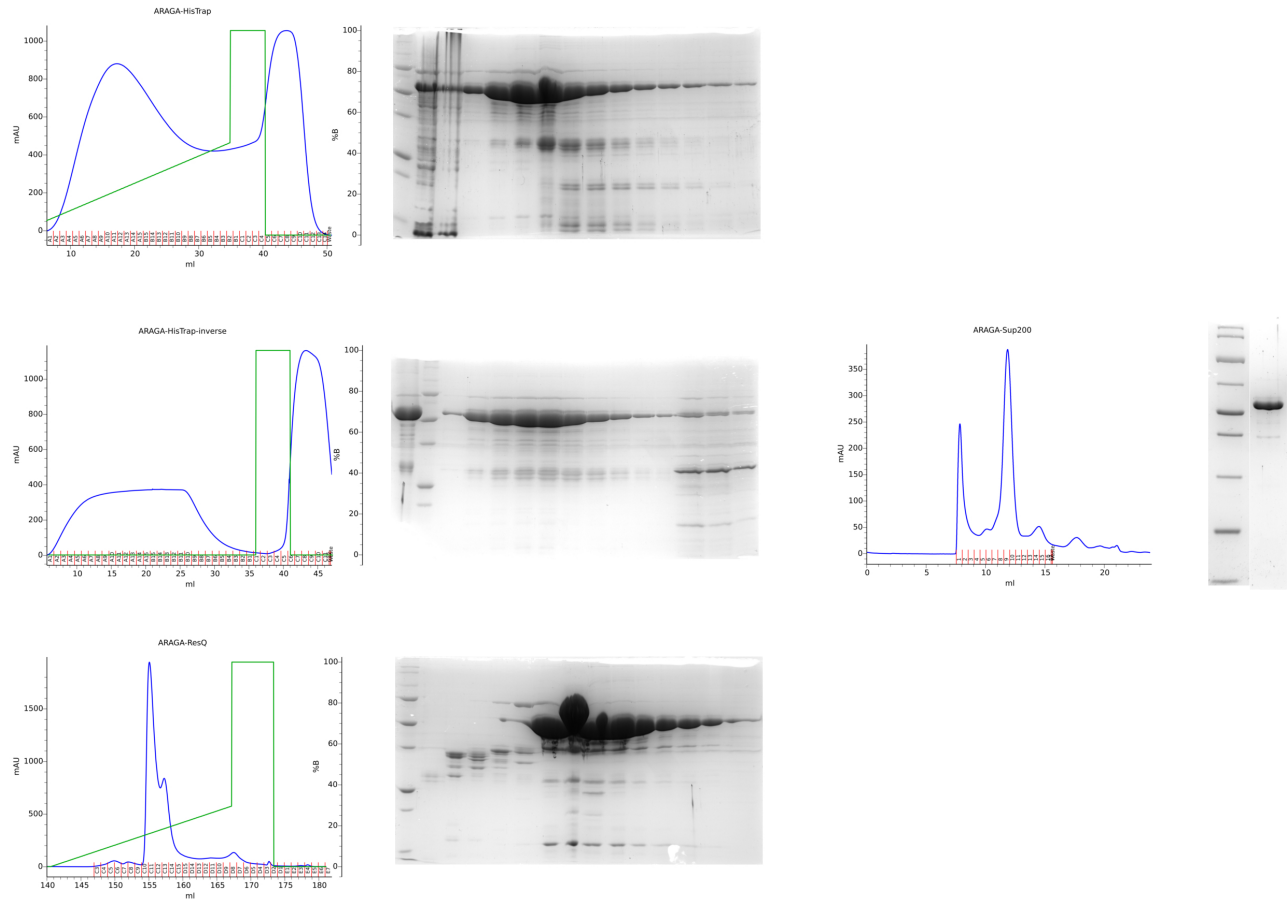
**Figure 76 : Purification de RAP1**

Le rendement est d'environ 8 mg de protéine par culot bactérien.

### 5.3.4.3 RAP1-ARAGA

La purification de RAP1-ARAGA est réalisée en quatre étapes : chromatographie d'affinité pour le nickel, clivage de l'étiquette histidine et séparation par une seconde chromatographie d'affinité pour le nickel, chromatographie d'affinité pour l'héparine, et finalement chromatographie d'exclusion stérique.

Les chromatogrammes et l'analyse des fractions par SDS-PAGE pour la purification de RAP1-ARAGA sont présentés dans la figure suivante :



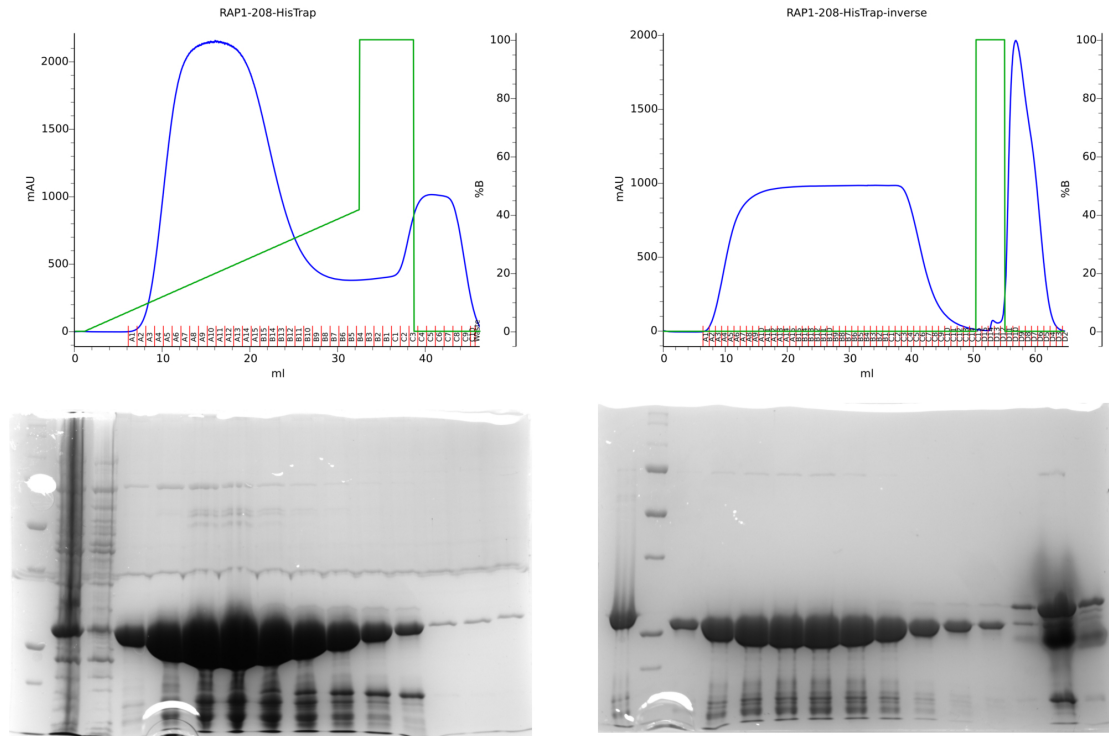
**Figure 77 : Purification de RAP1-ARAGA**

Le rendement est d'environ 6 mg de protéine par culot bactérien.

#### 5.3.4.4 RAP1<sub>1-208</sub>

La purification de RAP1<sub>1-208</sub> est réalisée en deux étapes : chromatographie d'affinité pour le nickel, clivage de l'étiquette histidine et séparation par une seconde chromatographie d'affinité pour le nickel.

Les chromatogrammes et l'analyse des fractions par SDS-PAGE pour la purification de RAP1<sub>1-208</sub> sont présentés dans la figure suivante :



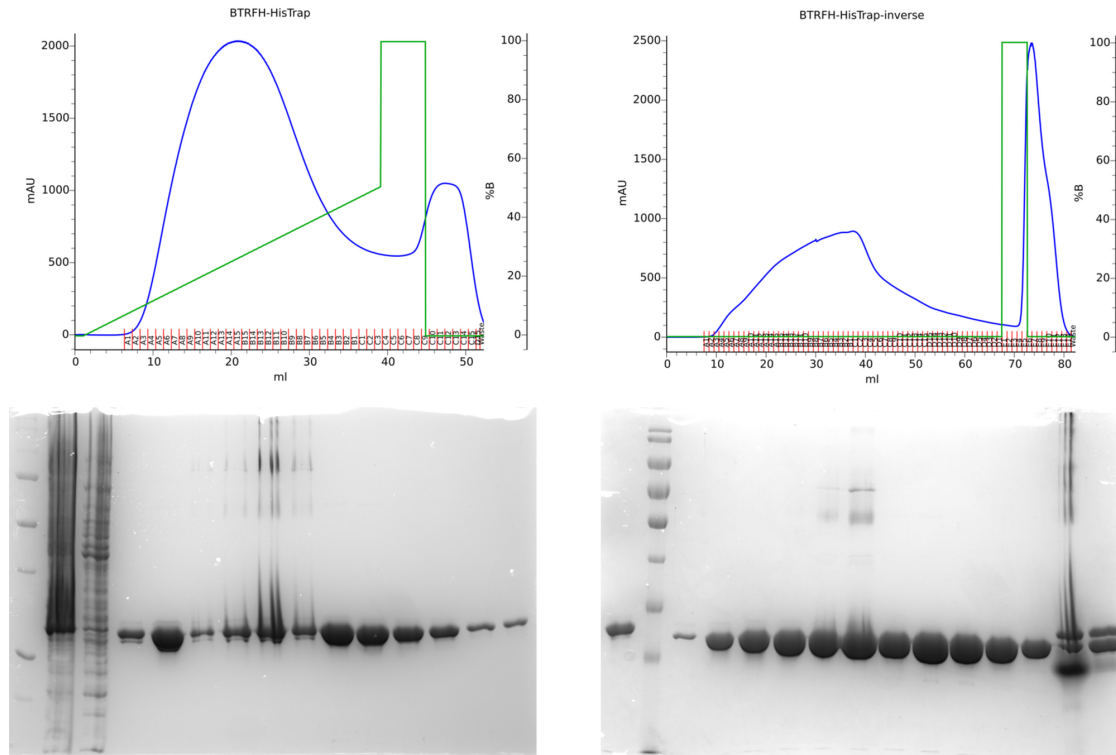
**Figure 78 : Purification de RAP1\_1-208**

Le rendement est d'environ 20,9 mg de protéine par culot bactérien.

#### 5.3.4.5 TRFH et B-TRFH

La purification de TRFH et B-TRFH est réalisée en deux étapes : chromatographie d'affinité pour le nickel, clivage de l'étiquette histidine et séparation par une seconde chromatographie d'affinité pour le nickel.

Les chromatogrammes et l'analyse des fractions par SDS-PAGE pour la purification de B-TRFH sont présentés dans la figure suivante :



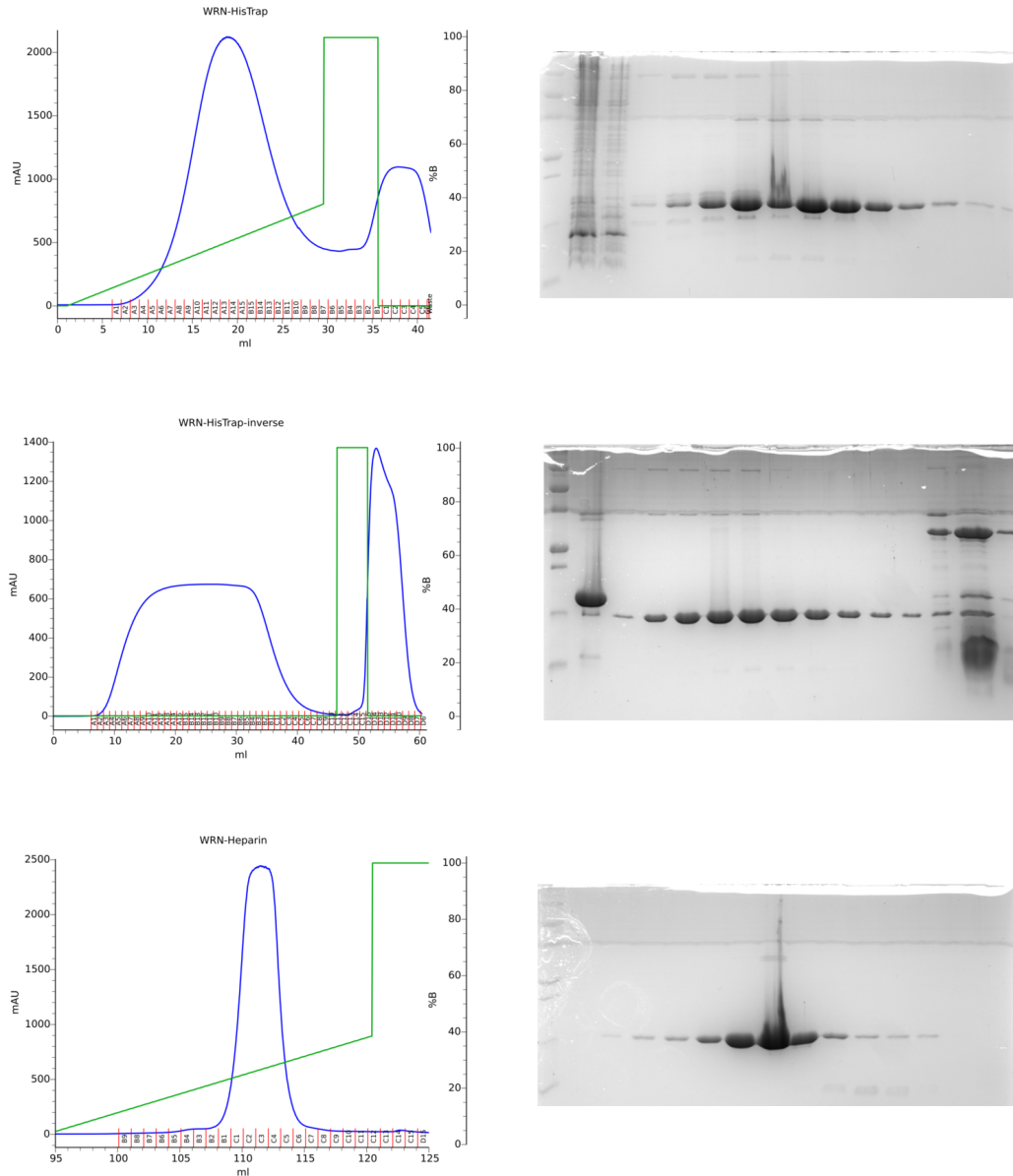
**Figure 79 : Purification de B-TRFH**

Le rendement est d'environ 6 mg de protéine par culot bactérien. Les résultats sont comparables pour TRFH, mais ne sont pas présentés car cette purification a été réalisée par une autre personne de l'équipe.

#### 5.3.4.6 WRN\_949-1079

La purification de WRN\_949-1079 est réalisée en trois étapes : chromatographie d'affinité pour le nickel, clivage de l'étiquette histidine et séparation par une seconde chromatographie d'affinité pour le nickel, et chromatographie d'affinité pour l'héparine.

Les chromatogrammes et l'analyse des fractions par SDS-PAGE pour la purification de WRN\_949-1079 sont présentés dans la figure suivante :



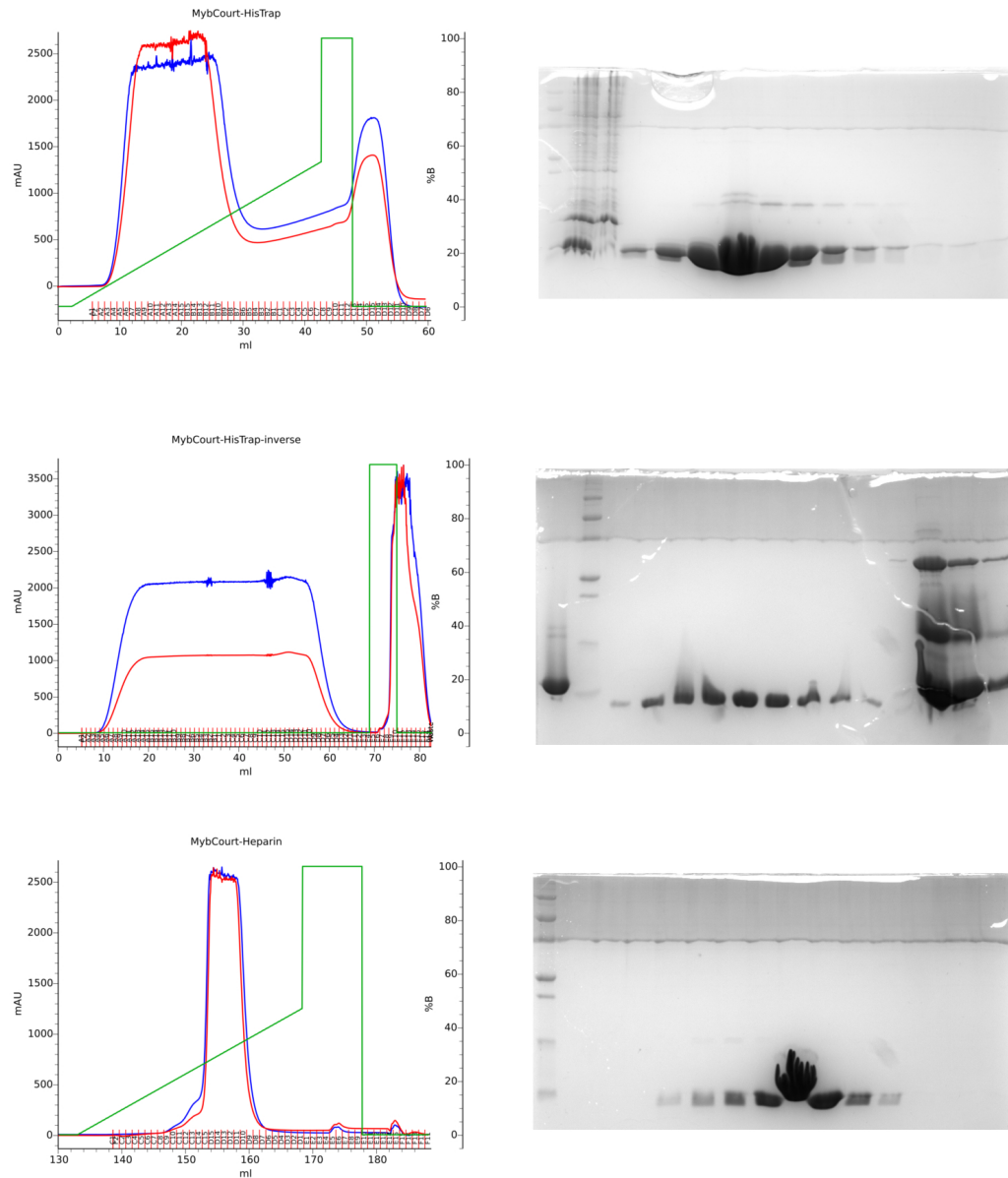
**Figure 80 : Purification de WRN\_949-1079**

Le rendement est d'environ 3,2 mg de protéine par culot bactérien.

#### 5.3.4.7 Myb-court

La purification du Myb-court est réalisée en trois étapes : chromatographie d'affinité pour le nickel, clivage de l'étiquette histidine et séparation par une seconde chromatographie d'affinité pour le nickel, et chromatographie d'affinité pour l'héparine.

Les chromatogrammes et l'analyse des fractions par SDS-PAGE pour la purification du Myb-court sont présentés dans la figure suivante :



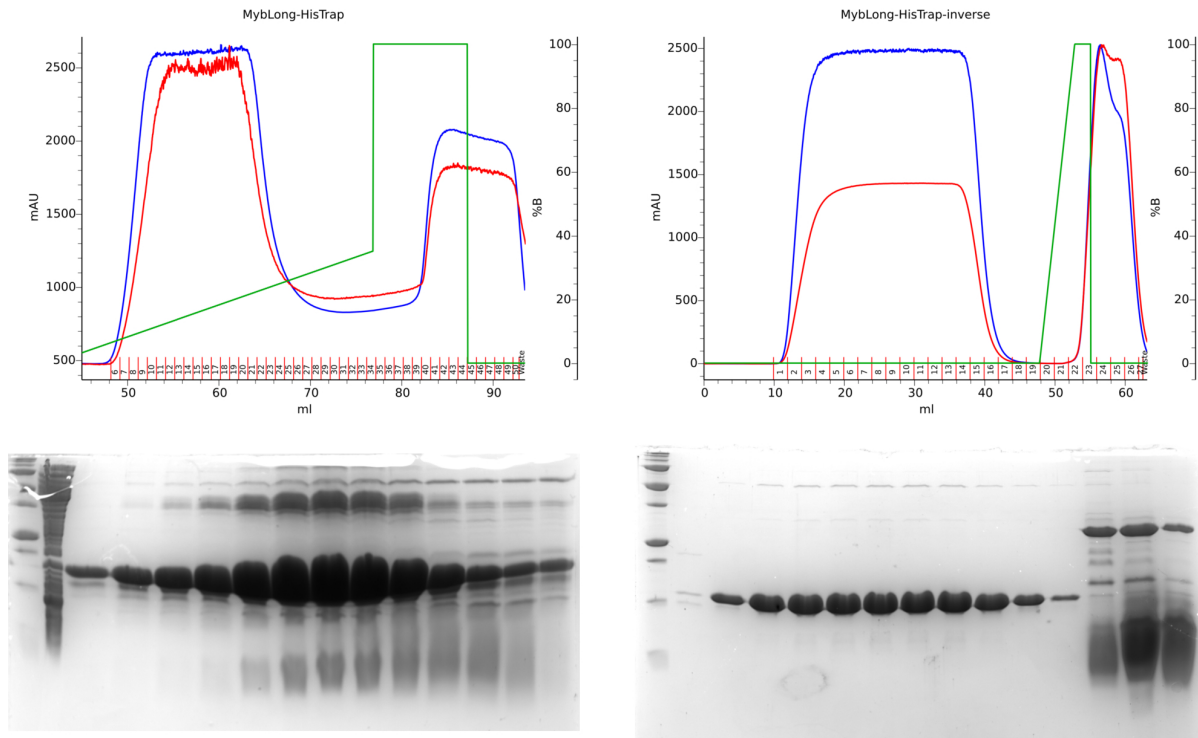
**Figure 81 : Purification du Myb-court**

Le rendement est d'environ 13,8 mg de protéine par culot bactérien.

#### 5.3.4.8 Myb-long

La purification du Myb-long est réalisée en deux étapes : chromatographie d'affinité pour le nickel, clivage de l'étiquette histidine et séparation par une seconde chromatographie d'affinité pour le nickel.

Les chromatogrammes et l'analyse des fractions par SDS-PAGE pour la purification du Myb-long sont présentés dans la figure suivante :



**Figure 82 : Purification du Myb-long**

Le rendement est d'environ 6,3 mg de protéine par culot bactérien.

## 5.4 Dichroïsme circulaire

### 5.4.1 Protocole

Les protéines sont dialysées dans un tampon Tris-HCl 9 mM pH = 7,5, NaF 300 mM. Pour chaque protéine, l'acquisition des spectres de dichroïsme circulaire est faite sur une série de concentrations : 3, 6 et 15  $\mu\text{M}$  en protéine (concentration exprimée en monomères pour TRF2). Les spectres sont mesurés dans une cuve en quartz de 200  $\mu\text{L}$  et 1 mm de trajet optique, avec un appareil Jasco J-815 CD Spectrometer. Chaque spectre est obtenu par l'accumulation de 20 acquisitions de 250 à 190 nm de longueur d'onde par pas de 0,1 nm et avec une largeur de bande de 2 nm. Les spectres finaux des protéines sont obtenus en soustrayant le spectre du tampon aux spectres initiaux. Pour pouvoir comparer les spectres entre eux, ils sont finalement exprimés en ellipticité molaire résiduelle selon la formule suivante :

$$[\theta] = \frac{\theta}{10 \times c \times L \times n}$$

où  $[\theta]$  est l'ellipticité molaire résiduelle (en  $\text{deg.cm}^2.\text{dmol}^{-1}.\text{résidu}^{-1}$ ),  $\theta$  est le signal de dichroïsme circulaire brut tel qu'enregistré par le spectromètre (en mdeg),  $c$  est la concentration molaire en protéine (en  $\text{mol.L}^{-1}$ ),  $L$  est le trajet optique (en cm) et  $n$  est le nombre de résidus dans la protéine.

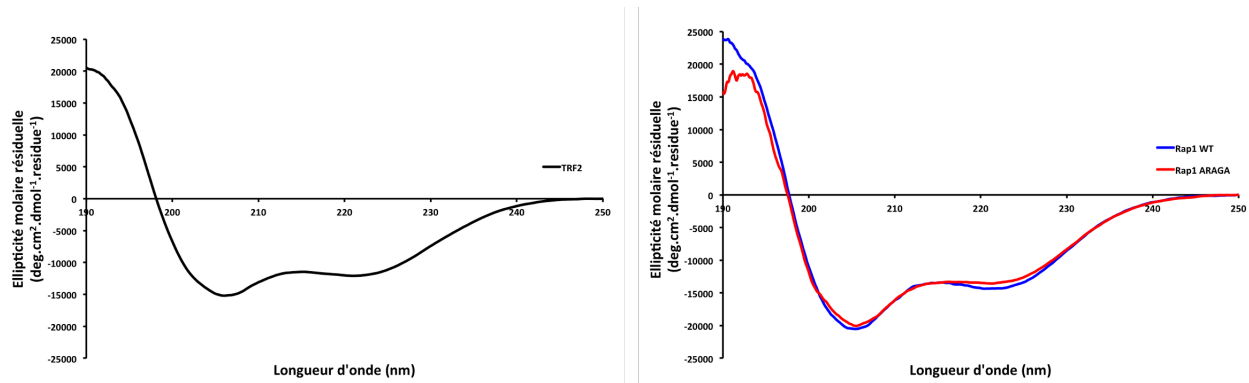
Pour une comparaison directe entre RAP1 et ARAGA, les spectres sont aussi mis à l'échelle pour que leur valeur à 209,5 nm (point d'inflexion) coïncide : ceci permet de s'affranchir des différences d'amplitude dues à l'incertitude sur la détermination de la concentration en protéine (estimée par l'absorbance à 280 nm mesurée avec un appareil NanoDrop, moins sensible que le spectromètre de dichroïsme circulaire).

La dénaturation thermique des protéines est suivie par la mesure du signal de dichroïsme circulaire à 222 nm (principal signal des hélices  $\alpha$ ) pendant que l'échantillon est chauffé de 10 à 95  $^{\circ}\text{C}$  par pas de 1  $^{\circ}\text{C}$ . Les données de dénaturation thermique sont modélisées avec une courbe sigmoïde à l'aide du logiciel SciDAVis (<http://scidavis.sourceforge.net>) afin d'estimer la valeur de la température de demi-dénaturation  $T_m$ .

### 5.4.2 Résultats

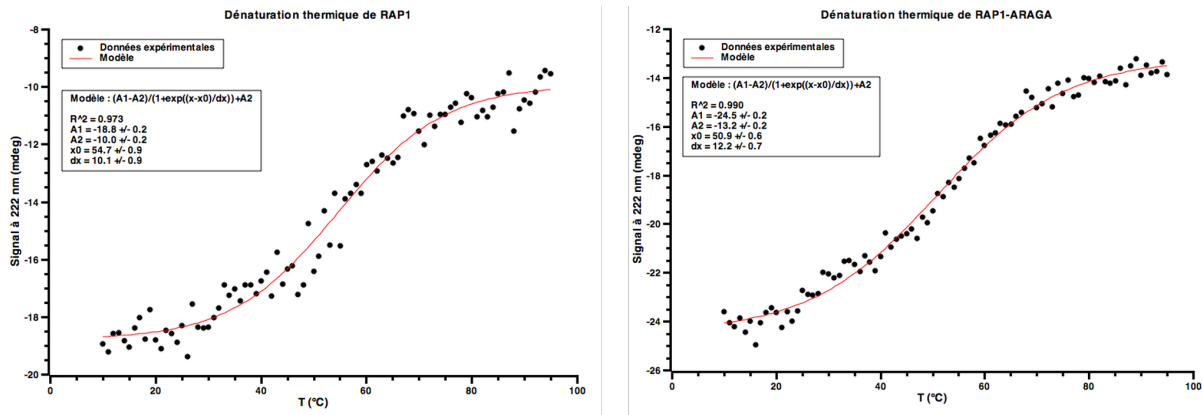
Les spectres de dichroïsme circulaire de TRF2 et de RAP1 et RAP1-ARAGA sont présentés dans la figure suivante :





**Figure 83 : Spectres de dichroïsme circulaire de TRF2 (à gauche) et de RAP1 et RAP1-ARAGA (à droite).**

Les spectres superposables de RAP1 et RAP1-ARAGA indiquent que la mutation n'a pas altéré le repliement de la protéine. Pour s'en assurer nous avons aussi suivi la dénaturation thermique des deux protéines, présentée dans la figure suivante :



**Figure 84 : Dénaturation thermique de RAP1 (à gauche) et RAP1-ARAGA (à droite).**

La modélisation des données de dénaturation thermique par une courbe sigmoïde indique les températures de demi-dénaturation suivantes :

- RAP1 :  $T_m \approx 54,7 \pm 0,9 \text{ } ^\circ\text{C}$
- RAP1-ARAGA :  $T_m \approx 50,9 \pm 0,6 \text{ } ^\circ\text{C}$

Ceci indique que la structure de RAP1-ARAGA est légèrement moins stable que celle de RAP1, néanmoins cette faible différence indique que la mutation n'affecte pas la stabilité du repliement de façon significative.

## 5.5 Titrations calorimétriques

Les titrations calorimétriques ont été réalisées avec un appareil VP-ITC ou iTC200 (au laboratoire, ou à la plateforme de calorimétrie de l'IBBMC à Orsay). Les protéines testées sont toujours dialysées **dans le même bain** au moins deux fois 2h contre au moins 100 volumes du tampon de l'expérience, à 4 °C. Les oligonucléotides et peptides ne sont jamais dialysés car trop petits, à la place un stock concentré est simplement dilué avec le tampon de l'expérience pour obtenir la concentration souhaitée. Avant chaque expérience, toutes les solutions sont dégazées et leur température est ajustée à celle de l'expérience.

Les conditions expérimentales de chaque expérience sont détaillées dans le tableau suivant (les concentrations en TRF2, TRFH, B-TRFH et complexe TRF2/RAP1 sont toujours exprimées en termes de monomères, les concentrations en Ku sont toujours exprimées en termes de l'hétérodimère).

Tampons utilisés :

- **Tampon A :**
  - Tris-HCl 50 mM, pH = 7,5
  - NaCl 150 mM
  - 2-mercaptoethanol 10 mM
- **Tampon B :**
  - Tris-HCl 25 mM, pH = 7,5
  - NaCl 100 mM
  - 2-mercaptoethanol 10 mM
- **Tampon C :**
  - Tris-HCl 20 mM, pH = 7,5
  - NaCl 150 mM
  - 2-mercaptoethanol 5 mM
- **Tampon D :**
  - Tris-HCl 20 mM, pH = 8,0
  - NaCl 150 mM
  - 2-mercaptoethanol 5 mM

**Tableau 10 : Conditions expérimentales des titrations calorimétriques.**

<b>Cuve</b>	<b>Seringue</b>	<b>Tampon</b>	<b>Température (K)</b>	<b>Volume d'injection (μL)</b>
TRF2 13 μM	RAP1 130 μM	A	283	10
TRF2 10 μM	RAP1-RCT 91 μM	A	283	10
TRF2 13 μM	RAP1-ARAGA 130 μM	A	283	10
TRF2 13 μM	Apollo-TBM 130 μM	A	283	10
TRFH 25 μM	RAP1-TBM 250 μM	B	279, 283, 293	10
TRF2/RAP1 13 μM	Apollo-TBM 130 μM	A	283	10
Apollo-TBM 25 μM	TRFH 250 μM	B	293	10
TRF2 27 μM	TIN2-TBM 270 μM	A	283	1,5 (iTC200)
TRF2/RAP1 27 μM	TIN2-TBM 270 μM	A	283	1,5 (iTC200)
B-TRFH 30 μM	RAP1_1-208 300 μM	C	283, 293, 303	10
WRN-RQC 9 μM	T2-S5S6 148 μM	C	293	10
TRFH 20 μM	WRN-RQC 200 μM	C	283, 293, 303	10
TRF2 13,46 μM	WRN-RQC 134,59 μM	C	293	10
RAP1 13,46 μM	WRN-RQC 134,59 μM	C	293	10
TRF2/RAP1 13,46 μM	WRN-RQC 134,59 μM	C	293	10
Ku 11,16 μM	TRFH 280 μM	D	298	6
Ku 9,04 μM	TRF2 180,8 μM	D	298	6
Myb-court 20 μM	T2-S5S6 100 μM	C	283	10
Myb-long 20 μM	T2-S5S6 100 μM	C	283	10

## 5.6 SAXS

### 5.6.1 Préparation des échantillons

Les protéines (les complexes préformés dans le cas des Mybs/ADN) sont tout d'abord injectées sur une colonne Superdex 200 10/300 GL (GE Healthcare) pour passer en tampon Tris-HCl 20 mM pH = 7,5, NaCl 150 mM, 2-mercaptoethanol 5 mM et éliminer les éventuels fragments de dégradation encore restants. Les protéines sont ensuite concentrées et les échantillons suivants sont préparés (concentration exprimée en monomères pour TRF2) :

- TRF2 seule à 215,5  $\mu$ M ;
- RAP1 seule à 326,7  $\mu$ M ;
- RAP1-RCT seul à 1015,7  $\mu$ M ;
- TRF2 à 120  $\mu$ M + RAP1 à 144  $\mu$ M ;
- TRF2 à 171,5  $\mu$ M + RAP1-RCT à 205,8  $\mu$ M ;
- Myb-long + ADN T2S5S8 ;
- Myb-court + ADN T2S5S8.

### 5.6.2 Acquisition des courbes de diffusion

Les acquisitions sont réalisées sur la ligne de lumière SWING du synchrotron SOLEIL, en utilisant le passeur d'échantillons pour RAP1 et la chromatographie en ligne (David & Pérez, 2009) pour tous les autres échantillons.

Tous les échantillons circulent à 0,2 mL/min, dans le tampon indiqué dans le paragraphe précédent. La chromatographie est réalisée sur une colonne d'exclusion stérique Bio-SEC3 (Agilent), avec des injections de 40 à 50  $\mu$ L. La circulation des échantillons assure que chaque acquisition est faite sur une partie de la solution qui n'a pas encore été irradiée, ce qui permet de s'affranchir de potentiels problèmes de radiolyse. Pour chaque acquisition, une série de courbes est également enregistrée sur le tampon de l'expérience.

Les restes des échantillons sont analysés par SDS-PAGE pour s'assurer qu'aucune dégradation ne s'est produite pendant le temps des acquisitions.

### 5.6.3 Analyse des données

#### 5.6.3.1 Traitement des données brutes

Les courbes de diffusion sont obtenues directement à la ligne de lumière, par les opérations suivantes à l'aide du logiciel FoxTrot (développé par les scientifiques de la ligne de lumière SWING) :

1. on dessine un masque pour délimiter la région utile du détecteur (il faut masquer le *beam stop* et son ombre),

2. on applique le masque sur les images de diffusion et le logiciel calcule automatiquement les courbes brutes (traitement par lot sur l'ensemble des images),
3. parmi les courbes obtenues sur le tampon, on supprime celles qui comportent des sauts brusques ou autres irrégularités,
4. on fait la moyenne des courbes restantes pour obtenir la courbe finale du tampon (CT),
5. on soustrait CT de chaque courbe obtenue sur l'échantillon,
6. on fait l'analyse de Guinier (approximation linéaire de la portion de courbe aux plus petits angles, qui permet de déterminer les paramètres  $I_0$  et  $R_g$ ) sur la courbe la plus haute pour déterminer une gamme d'angles  $Q_{min}$ - $Q_{max}$  telle que  $Q_{max}.R_g < 1,3$  (ou 1,2 dans le cas de protéines flexibles),
7. on applique l'analyse de Guinier sur toutes les autres courbes (traitement par lot) dans cette gamme d'angles pour obtenir les valeurs de  $I_0$  et  $R_g$  tout le long du chromatogramme (la valeur de  $I_0$  reproduit l'allure du chromatogramme obtenu avec l'absorbance à 280 nm car elle est proportionnelle à la concentration en protéine),
8. on choisit les courbes correspondant à une valeur de  $R_g$  constante sous le pic d'intérêt du chromatogramme et on calcule leur moyenne (à partir des courbes non soustraites du tampon) pour obtenir la courbe moyenne de l'échantillon (CME),
9. on effectue la soustraction CME - CT pour obtenir la courbe finale de l'échantillon.

En utilisant l'HPLC, la courbe finale de l'échantillon est typiquement obtenue à partir d'une moyenne de 10 à 15 courbes. Les courbes finales obtenues sont ainsi plus représentatives puisque la molécule est soumise à une dernière étape de purification sur place. D'autre part le rapport signal sur bruit aux grands angles est meilleur grâce à la redondance des données permise par ce mode d'acquisition. En utilisant le passeur d'échantillons, le nombre de courbes exploitables est limité par le nombre de réplicats que l'on peut mesurer avec la quantité d'échantillon disponible, et on utilise typiquement moins de courbes.

### 5.6.3.2 Calcul des enveloppes

Une fois en possession de la courbe de diffusion finale, nous utilisons le logiciel GNOM (Semenyuk & Svergun, 1991) pour calculer la fonction de distribution des distances  $P(r)$  qui est le résultat de la transformée de Fourier de la courbe de diffusion.

Cette fonction  $P(r)$  est utilisée en entrée du logiciel GASBOR (Svergun *et al*, 2001) pour générer automatiquement 100 enveloppes pour TRF2, RAP1, RAP1-RCT et les complexes TRF2/RAP1 et TRF2/RAP1-RCT, en renseignant un nombre de pseudo-résidus égal au nombre de résidus effectivement présents dans la protéine ou le complexe dont on calcule l'enveloppe. Nous n'imposons pas de symétrie pour ces calculs. Le programme utilise une approche *ab initio* pour construire l'enveloppe : il agence les pseudo-résidus dans l'espace de façon à minimiser l'écart entre la  $P(r)$  calculée à partir du modèle et la  $P(r)$  obtenue à

partir de la courbe de diffusion expérimentale, tout en conservant autant que possible un agencement en chaîne des pseudo-résidus.

Une fois que les 100 enveloppes issues de calculs indépendants sont prêtes, nous sélectionnons les meilleures selon les deux critères suivants :

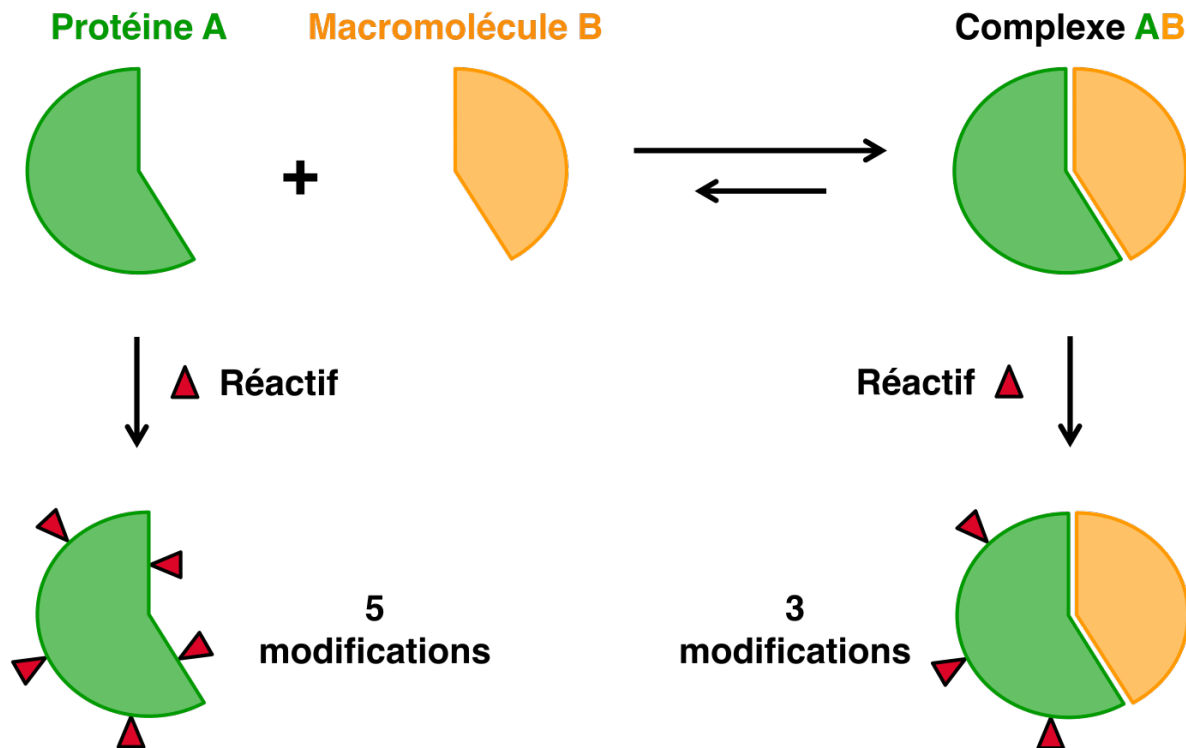
- le facteur Chi (qui mesure l'accord avec les données expérimentales) ne dépasse pas Chi-minimum + écart-standard(Chis),
- et le facteur NSD (*normalized spatial discrepancy*, vaut 0 pour deux objets identiques parfaitement superposés) calculé lors de la superposition des enveloppes sur l'enveloppe de meilleur Chi ne dépasse pas NSD-minimum + déviation-moyenne(NSDs).

Les modèles ainsi sélectionnés (typiquement une dizaine) sont finalement combinés avec le logiciel DAMAVER pour obtenir l'enveloppe finale.

## 5.7 Empreinte protéique

### 5.7.1 Principe et mise au point

Une expérience d'empreinte protéique (*protein footprinting*) consiste à mettre une protéine en présence d'un réactif capable de produire un adduit covalent sur certaines chaînes latérales. De nombreux réactifs sont disponibles pour cibler les différentes chaînes latérales, selon des mécanismes propres à chaque type de chaîne latérale (acide carboxylique, amine, alcool, thiol) (Mendoza & Vachet, 2009). Le principe de la technique est schématisé dans la figure suivante.



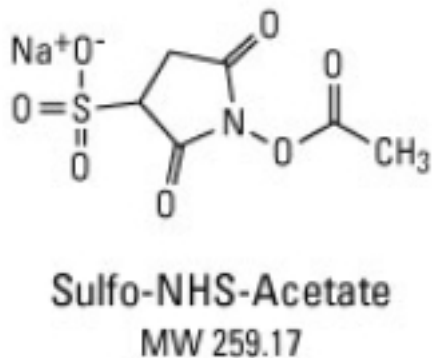
**Figure 85 : Principe de l'expérience d'empreinte protéique**

La méthode repose sur la possibilité d'identifier par spectrométrie de masse les résidus modifiés. En comparant les spectres de masse de la protéine traitée et non traitée par le réactif, nous pouvons identifier les résidus modifiés, c'est-à-dire ceux qui sont accessibles au réactif, donc exposés en surface de la protéine. Le même traitement est ensuite réalisé sur la protéine en complexe avec son partenaire, et permet d'identifier les résidus non modifiés (ou moins modifiés) en présence du partenaire, c'est-à-dire ceux qui sont protégés par la présence du partenaire, donc engagés dans l'interaction.

Cette méthode a été utilisée avec succès sur des complexes protéine/ADN et est complémentaire des techniques d'empreinte sur l'ADN (par exemple l'empreinte à la DNase I) puisqu'elle indique l'empreinte de l'ADN sur la protéine. Dans l'étude de la rétro-

transcriptase du VIH, par exemple, elle a pu apporter des informations complémentaires à celles obtenues avec la structure cristallographique (Kvaratskhelia *et al*, 2002).

Pour l'étude de l'interaction entre TRF2 et RAP1, nous avons utilisé l'acétylation des lysines comme réaction d'empreinte. Le réactif permettant cette réaction est le sulfo-NHS-acétate, dont la structure est représentée dans la figure suivante.



**Figure 86 : Structure du sulfo-NHS-acetate**

Puisque les deux molécules de notre complexe sont des protéines, nous avons dû mettre au point le protocole en tenant compte de deux contraintes :

1. pour s'assurer que la protéine que nous étudions est bien engagée à 100 % dans des complexes, il faut un léger excès de l'autre protéine (un excès molaire de 1,5 est suffisant du fait de la forte affinité de RAP1 pour TRF2 et des concentrations auxquelles nous travaillons) ;
2. le nombre total de lysines dans la réaction n'est pas constant d'une condition à une autre, puisque dans un cas nous avons un excès de TRF2 et dans l'autre un excès de RAP1 (et les deux protéines ne contiennent pas le même nombre de lysines) : par conséquent nous devons choisir une concentration en sulfo-NHS-acetate suffisante pour être en excès molaire par rapport au nombre total de lysines (si le réactif n'est pas en excès, une lysine non acétylée peut signifier qu'elle est inaccessible au réactif ou bien que tout le réactif disponible a déjà été consommé).

## 5.7.2 Protocole

### 5.7.2.1 Réaction

Les protéines, initialement stockées à -80 °C en tampon Tris (voir le protocole de purification), sont décongelées puis passées en tampon HEPES 20 mM pH = 7,5, NaCl 150 mM, 2-mercaptoethanol 5 mM par chromatographie d'exclusion stérique sur colonne Superdex 200 10/300 GL (GE Healthcare). Cette étape est nécessaire car le tampon Tris contient une amine susceptible de réagir avec le sulfo-NHS-acetate (ce tampon est d'ailleurs



utilisé pour arrêter la réaction). Elle permet également d'éliminer les quelques fragments de dégradation encore présents dans les préparations de protéines. Les protéines sont finalement concentrées par ultrafiltration jusqu'à une concentration suffisante pour pouvoir préparer les mélanges réactionnels.

La solution mère de sulfo-NHS-acetate est préparée immédiatement avant son utilisation en dissolvant la poudre dans de l'eau distillée. Les restes ne sont pas conservés car cette molécule s'hydrolyse spontanément.

Les mélanges suivants sont préparés dans un volume final de 20  $\mu$ L (les concentrations en TRF2 sont exprimées en termes de monomères) :

1. protéines libres :

- TRF2 à 20  $\mu$ M +/- sulfo-NHS-acetate à 4,075 mM
- RAP1 à 20  $\mu$ M +/- sulfo-NHS-acetate à 7,85 mM
- ARAGA à 20  $\mu$ M +/- sulfo-NHS-acetate à 7,85 mM

2. complexes :

- TRF2 à 20  $\mu$ M + RAP1 à 30  $\mu$ M + sulfo-NHS-acetate à 4,075 mM
- TRF2 à 20  $\mu$ M + ARAGA à 30  $\mu$ M + sulfo-NHS-acetate à 4,075 mM
- RAP1 à 20  $\mu$ M + TRF2 à 30  $\mu$ M + sulfo-NHS-acetate à 7,85 mM
- ARAGA à 20  $\mu$ M + TRF2 à 30  $\mu$ M + sulfo-NHS-acetate à 7,85 mM

Dans le cas des complexes, le partenaire est en excès molaire 1,5 face à la protéine que l'on souhaite étudier, et le sulfo-NHS-acetate est en excès molaire 2,5 face à la concentration totale de lysines présentes dans la réaction. Dans le cas des contrôles, la concentration en sulfo-NHS-acetate est identique à celle utilisée pour analyser la protéine en complexe.

Avant l'ajout du sulfo-NHS-acetate, les mélanges réactionnels sont incubés 10 min à 20 °C pour permettre la formation des complexes. Une fois le réactif ajouté, les réactions sont incubées 30 min à 20 °C puis finalement arrêtées par ajout de 2  $\mu$ L de tampon Tris 1 M pH = 7,5 (qui consomme l'excès de sulfo-NHS-acetate). Les produits d'acétylation sont enfin analysés par SDS-PAGE et spectrométrie de masse.

### 5.7.2.2 Spectrométrie de masse

Les bandes protéiques découpées dans les gels SDS-PAGE sont rincées deux fois avec 500  $\mu$ L de bicarbonate d'ammonium 100 mM pH = 8, puis deux fois avec 500  $\mu$ L de bicarbonate d'ammonium 50 mM pH = 8 + acétonitrile 50 % v/v. Les bandes sont alors partiellement séchées par lyophilisation puis réhydratées avec 20  $\mu$ L de bicarbonate d'ammonium 100 mM pH = 8.

Toutes les bandes sont digérées par la trypsine à 10 ng/ $\mu$ L dans le tampon bicarbonate d'ammonium 100 mM pH = 8, par une incubation à 50 °C pendant 2h sous agitation. Des

réplicats des bandes de RAP1 et ARAGA sont également digérés par la protéase V8 à 10 ng/ $\mu$ L dans le même tampon, par une incubation à 30 °C pendant une nuit sous agitation. Les digestions sont arrêtées par ajout de TFA qsp 1 % v/v final.

Pour chaque échantillon, on dépose 0,5  $\mu$ L directement sur une plaque MALDI avec 0,5  $\mu$ L de matrice 4-HCCA à 10 mg/mL dans un mélange acétonitrile 50 % v/v + eau 49,9 % v/v + TFA 0,1 % v/v et on laisse sécher. Pour les acquisitions de spectres MS/MS nécessitant plus de matériel, le reste d'échantillon est concentré par MicroZipTip C18 avant d'être déposé sur la plaque avec la matrice et séché.

L'acquisition des spectres MS est faite avec un appareil 4800 MALDI-TOF/TOF (ABSciex, Foster City, USA). Les spectres sont analysés avec le logiciel Mascot (<http://www.matrixscience.com>) contre la base de données de protéines non redondante du NCBI afin de déterminer la couverture de séquence, les fragments contenant des lysines, et les potentielles lysines acétylées. Les pics d'intérêt sont aussi identifiés par comparaison visuelle des spectres obtenus sur les protéines libres traitées ou non traitées par le sulfo-NHS-acetate. L'identité des fragments contenant des lysines a été confirmée expérimentalement par MS/MS.

### 5.7.2.3 Analyse

À l'aide du logiciel mMass (Strohalm *et al*, 2010), les spectres sont normalisés relativement au pic le plus haut :

- le pic à  $m/z = 1064$  pour RAP1 et ARAGA digérées par la trypsine,
- le pic à  $m/z = 1483$  pour RAP1 et ARAGA digérées par la protéase V8,
- le pic à  $m/z = 909$  pour TRF2 digérée par la trypsine.

Cette échelle relative permet la comparaison de spectres obtenus à partir d'acquisitions et d'échantillons différents, qui ont des amplitudes différentes si on les trace avec leur échelle absolue.

Une fois les spectres mis sur cette échelle relative, pour chaque fragment contenant une lysine (ou plusieurs lysines) nous comparons les hauteurs des pics entre les spectres de la protéine libre et de la protéine en complexe.

Considérons un exemple simple d'un fragment contenant une lysine acétylée, avec le pic correspondant bien identifié dans les spectres. Si ce pic est plus bas dans le spectre de la protéine en complexe que dans celui de la protéine libre, cela signifie que ce fragment acétylé est moins abondant dans cet échantillon, et donc que cette lysine est moins accessible au sulfo-NHS-acetate lorsque la protéine est en complexe.

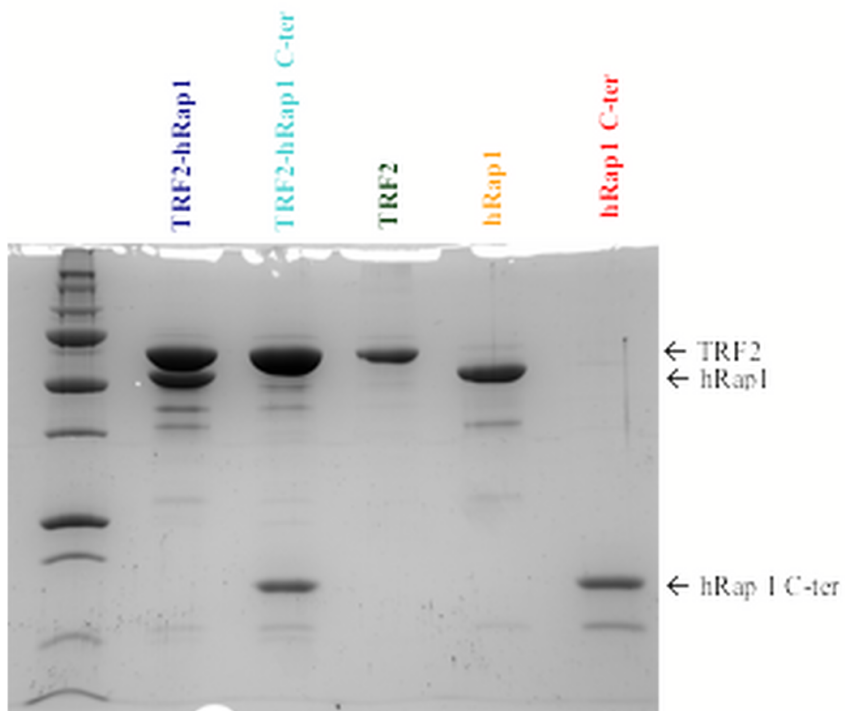
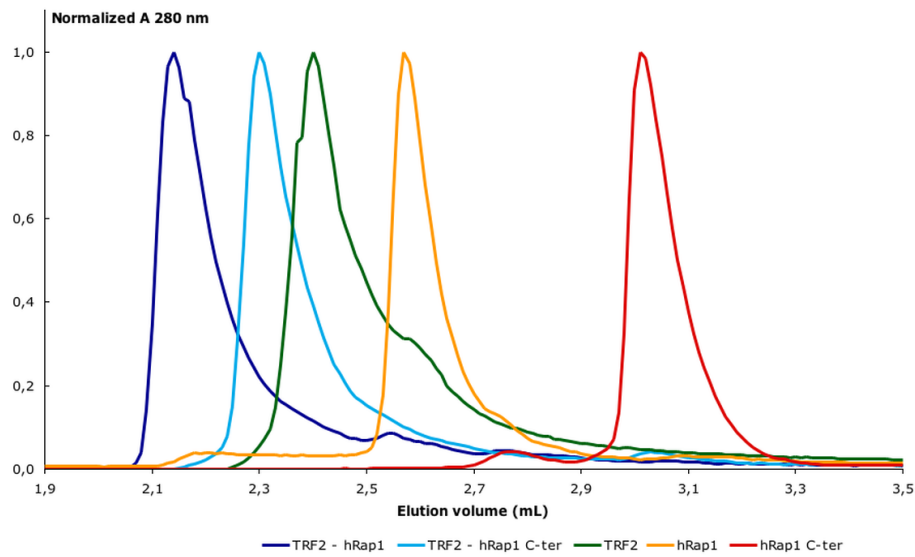
## 5.8 Tests d'interaction par chromatographie d'exclusion stérique

Les tests d'interaction par chromatographie d'exclusion stérique sont réalisés sur une colonne BIO SEC-3 (Agilent) ou Superdex 200 10/300 GL (GE Healthcare). Tous les tests sont réalisés en tampon Tris-HCl 20 mM, pH = 7,5, NaCl 150 mM, 2-mercaptoethanol 5 mM. Les conditions expérimentales des tests sont listées dans le tableau suivant.

**Tableau 11 : Tests d'interaction par chromatographie d'exclusion stérique.**

Mélange	Volume injecté ( $\mu\text{L}$ )	Colonne	Débit ( $\text{mL}/\text{min}$ )
TRF2 17,7 $\mu\text{M}$	50	SEC-3	0,15
RAP1 24,2 $\mu\text{M}$	50	SEC-3	0,15
RAP1-RCT 203 $\mu\text{M}$	50	SEC-3	0,15
TRF2/RAP1 10 $\mu\text{M}$	50	SEC-3	0,15
TRF2/RAP1-RCT 16,2 $\mu\text{M}$	50	SEC-3	0,15
TRFH 256,6 $\mu\text{M}$ (dimères)	50	SEC-3	0,2
Ku 94 $\mu\text{M}$ (dimères)	50	SEC-3	0,2
TRFH 50 $\mu\text{M}$ + Ku 25 $\mu\text{M}$	50	SEC-3	0,2
Ku 86,1 $\mu\text{M}$	100	Sup200	0,5
TRFH 116,7 $\mu\text{M}$ + Ku 86,1 $\mu\text{M}$	100	Sup200	0,5

Les chromatogrammes des tests d'interaction entre TRF2 et RAP1 et TRF2 et RAP1-RCT sont présentés dans la figure suivante :



**Figure 87 : Chromatogrammes des tests d'interaction entre TRF2 et RAP1 et TRF2 et RAP1-RCT (en haut). Analyse par SDS-PAGE des contenus des pics des chromatogrammes (en bas).**

## 5.9 Test d'interaction entre TRFH et Ku par rétention sur résine de nickel

L'expérience est réalisée en tampon Tris-HCl 20 mM, pH = 7,5, NaCl 150 mM, imidazole 10 mM, 2-mercaptoethanol 5 mM. Le tampon d'élution a la même composition sauf pour l'imidazole qui est à 700 mM. Toutes les étapes sont réalisées à 4 °C. Deux expériences sont réalisées en parallèle : une en présence et une en absence de Ku (comportant une étiquette 10 His) immobilisé sur la résine.

1. Placer 150 µL de résine Ni-NTA neuve dans une tube micro-spin.
2. Laver avec 2x 500 µL de tampon : **fraction L1** pour l'analyse.
3. Centrifuger brièvement pour enlever l'excès de liquide.
4. Fixer sur la résine 500 µg de 10His-Ku en incubant à 4° C pendant 30 min.
5. Centrifuger brièvement pour enlever l'excès de liquide : **fraction FT1** pour l'analyse (excès de Ku non fixé sur la résine).
6. Laver avec 500 µL de tampon : **fraction L2** pour l'analyse (excès de Ku faiblement fixé ou présent dans le volume mort).
7. Centrifuger brièvement pour enlever l'excès de liquide.
8. Suspendre la résine avec 100 µL de tampon et conserver 10 µL de suspension : **fraction R1** pour l'analyse (pour vérifier la fixation de Ku sur la résine).
9. Centrifuger brièvement pour enlever l'excès de liquide.
10. Mettre TRFH en contact avec Ku immobilisé (ou avec la résine en absence de Ku pour l'expérience contrôle) : ajouter 600 µg de TRFH (soit un excès molaire d'environ 3,9 en dimères de TRFH versus dimères de Ku) et incuber à 4 °C pendant 30 min.
11. Centrifuger brièvement pour enlever l'excès de liquide : **fraction FT2** pour l'analyse (TRFH non retenu par Ku ou par la résine de l'expérience contrôle).
12. Laver avec 500 µL de tampon : **fraction L3** pour l'analyse (TRFH faiblement retenu ou présent dans le volume mort).
13. Suspendre la résine avec 100 µL de tampon et conserver 10 µL de suspension : **fraction R2** pour l'analyse.
14. Centrifuger brièvement pour enlever l'excès de liquide.
15. Éluer avec 2x 100 µL de tampon d'élution : **fractions E1 et E2** pour l'analyse.
16. Laver avec 500 µL de tampon.
17. Suspendre la résine avec 100 µL de tampon et garder 10 µL de suspension : **fraction R3** pour l'analyse (fraction non éluee par 700 mM d'imidazole).
18. Analyser les fractions par SDS-PAGE.

## 5.10 Culture de cellules HeLa

### 5.10.1 Matériel

- milieu DMEM
- sérum de veau fœtal (SVF)
- pénicilline/streptomycine mix
- DMSO (cell culture grade)
- trypsin/EDTA 10X
- PBS 1X sans calcium et magnésium
- stock de cellules (ampoule congelée)
- incubateur à CO<sub>2</sub>
- bain à 37 °C
- hotte de culture
- cryo-tubes stériles (pour congeler des aliquots de cellules)
- **working medium**: DMEM + SVF 10 % v/v + pénicilline/streptomycine qsp 1X final
- **trypsin/EDTA 2X**: diluer la solution stock avec du PBS qsp 2X final
- **cryo medium**: working medium + SVF qsp 20 % v/v final + DMSO qsp 10 % v/v final

### 5.10.2 Protocoles

**Toutes les solutions doivent être préparées en milieu stérile.**

#### 5.10.2.1 Démarrer une nouvelle culture

1. Préchauffer le *working medium* dans le bain à 37 °C
2. Sortir un tube de cellules congelées, dévisser légèrement le bouchon pour éviter toute surpression à l'intérieur
3. Décongeler le tube dans le bain à 37 °C
4. Transférer les cellules dans un tube 14 mL et ajouter du milieu **goutte à goutte** qsp 14 mL
5. Centrifuger 5 min à ~300 g à température ambiante
6. Enlever le surnageant et resuspendre dans 10 mL de milieu neuf
7. Inoculation d'une boîte de culture :
  - 3 mL de suspension de cellules dans une boîte de 10 cm de diamètre
  - 7 mL de suspension de cellules dans une boîte de 15 cm de diamètre
8. Ajouter du **working medium** jusqu'à un volume final de :
  - 10 mL dans une boîte de 10 cm de diamètre
  - 20 mL dans une boîte de 15 cm de diamètre

### 5.10.2.2 Passage des cellules

Les cellules ont besoin d'être diluées quand elles atteignent 80-90% de confluence (contrôler au microscope). Il faut veiller à ne pas diluer trop car si les sous-cultures se retrouvent à moins de 10% de confluence les cellules peuvent ne pas se remettre à pousser. Les cellules HeLa sont assez tolérantes et sont toujours capables de pousser (c'est une lignée cancéreuse) même si on les dilue un peu trop, ou même si on les dilue trop tard quand elles sont à confluence. Les cellules HeLa doublent à peu près toutes les 24h, il faut en tenir compte si on sait qu'on ne pourra pas les passer tous les jours (p. ex. on peut diluer plus avant le week-end pour les retrouver à 80-90% de confluence le lundi).

La procédure suivante suppose que la culture est dans des boîtes de 15 cm de diamètre.

1. Préchauffer le *working medium*, le PBS et la trypsin/EDTA 2X à 37 °C
2. Enlever le milieu qui couvre la boîte
3. Rincer deux fois avec 10 mL de PBS
4. Ajouter 2.5 mL de trypsin/EDTA 2X pour couvrir toute la boîte
5. Incuber 5 min à 37 °C
6. Tapoter les bords de la boîte pour bien remettre en suspension toutes les cellules encore adhérentes
7. Vérifier au microscope que toutes les cellules sont bien resuspendues
8. Ajouter 10 mL de milieu et faire deux allers-retours à la pipette pour bien suspendre toutes les cellules (il peut rester des amas de cellules sinon). Le SVF contient des inhibiteurs de la trypsine, donc pas besoin de rincer la solution de trypsine.
9. Inoculer des nouvelles boîtes avec un volume approprié de suspension de cellules
10. Ajouter du milieu jusqu'à un volume final de 20 mL
11. Remettre les boîtes dans l'incubateur

### 5.10.2.3 Stockage d'aliqots de cellules

La procédure suivante suppose que la culture est dans une boîte de 15 cm de diamètre et à quasi-confluence.

**Les cellules doivent être refroidies lentement (d'abord à 4 °C) puis congelées lentement, surtout pas de flash freezing dans l'azote !**

1. Préchauffer le *working medium*, le PBS et la trypsin/EDTA 2X à 37 °C
2. Trypsiniser les cellules comme décrit dans le protocole précédent (étapes 2-7)
3. Pendant l'incubation, préparer le *cryo medium* (en conditions stériles)
4. Suspendre les cellules dans 18 mL de milieu
5. Centrifuger 5 min à ~300 g à température ambiante

6. Pendant la centrifugation, préchauffer le *cryo medium* à 37 °C
7. Enlever le surnageant complètement (pipeter l'excédent si nécessaire)
8. Resuspendre les cellules dans ~8 mL de *cryo medium*
9. Distribuer des aliquots de 1.5 mL dans des cryo tubes
10. Garder 2x 25 µL de suspension de cellules pour faire un comptage
11. Incuber ~10 min à 4 °C
12. Stocker à -80 °C dans une boîte en polystyrène (pour que le refroidissement soit lent) pour une conservation à moyen terme (quelques mois), ou stocker dans l'azote liquide pour une conservation à très long terme.

### 5.11 Préparation des extraits nucléaires

Les extraits nucléaires sont préparés selon le protocole décrit par Abmayr *et al* (2001) avec les modifications mineures suivantes :

- tous les tampons contiennent du 2-mercaptoethanol à 5 mM à la place du DTT ;
- la concentration en PMSF dans tous les tampons est de 1 mM au lieu de 0,2 mM ;
- la conductivité des extraits nucléaires n'est pas suivie pendant la dialyse finale, celle-ci est faite contre 100 volumes de tampon de dialyse (avec une membrane retenant les protéines de 6-8 kDa et plus) pendant 1h30 ;
- la centrifugation finale est faite à 16 100 g au lieu de 25 000 g.

Le dosage par la méthode de Bradford indique que les extraits nucléaires contiennent de 3 à 5 mg/mL de protéines totales, avec une petite variabilité d'une préparation à une autre.



## 5.12 Pull down streptavidine-ADN

### 5.12.1 Principe

Afin de tester les propriétés de liaison de TRF2 comparées à celles des complexes TRF2-RAP1 et TRF2-ARAGA dans un contexte où TRF2 est lié à l'ADN, nous avons mis au point une expérience de *pull down* à base d'ADN. Le principe est d'immobiliser un ADN biotinylé sur un support exposant la streptavidine. Cet ADN contient un site de liaison de TRF2 (deux motifs TTAGGG), nous pouvons donc fixer dessus la protéine TRF2 purifiée, seule ou en présence de RAP1 (ou ARAGA). Nous pouvons ensuite introduire un partenaire candidat ou même un mélange complexe comme un extrait nucléaire. Après rinçage et élution des protéines fixées il est possible de déterminer si le partenaire candidat se lie à TRF2 ou TRF2-RAP1 ou TRF2-ARAGA, ou d'identifier des protéines du mélange retenues par le système.

### 5.12.2 Protocole

Les oligonucléotides T2-S5-biotinylé et T2-S8 sont hybridés pour obtenir un ADN db biotinylé avec une base sortante en 5' à chaque extrémité et contenant un site de liaison de TRF2 (deux motifs TAGGG en tandem) :

```
biotine-5' -TACTAGGGTTAGGGTTAGCA -3'  
3' - TGATCCCAATCCCAATCGTA-5'
```

Une quantité de 9 nmoles de cet ADN est chargée à 0,1 mL/min sur une colonne HiTrap Streptavidin HP 1 mL (GE Healthcare) préalablement équilibrée en tampon Tris-HCl 20 mM pH = 7,5, NaCl 150 mM, 2-mercaptoethanol 5 mM, EDTA 0,2 mM, PMSF 0,1 mM. Un pic d'absorbance à 260 nm en sortie de colonne indique qu'on a bien saturé la colonne avec l'ADN.

La colonne est chargée à 0,2 mL/min avec une quantité saturante de TRF2, de complexe TRF2-RAP1 préformé ou de complexe TRF2-ARAGA préformé. Dans chaque cas, une absorbance à 280 nm non nulle en sortie de colonne indique qu'on a bien saturé la colonne avec TRF2 ou les complexes. La colonne est rincée avec du tampon jusqu'à ce que l'absorbance à 280 nm revienne à une ligne de base stable.

La colonne est alors chargée à 0,2 mL/min avec environ 490 µg d'extraits nucléaires, et rincée avec du tampon jusqu'à ce que l'absorbance à 280 nm revienne à une ligne de base stable.

Les protéines liées à la colonne sont alors éluées à 0,5 mL/min avec un tampon d'élution identique au précédent sauf pour la concentration en NaCl qui est ici de 1 M. Les protéines éluées sont analysées par SDS-PAGE et spectrométrie de masse.

Après chaque expérience la colonne est lavée avec le tampon d'éluion jusqu'à ce que l'absorbance à 280 revienne à une ligne de base stable, et équilibrée à nouveau avec le tampon de départ.

## 5.13 Hybridation des oligonucléotides

### 5.13.1 Matériel

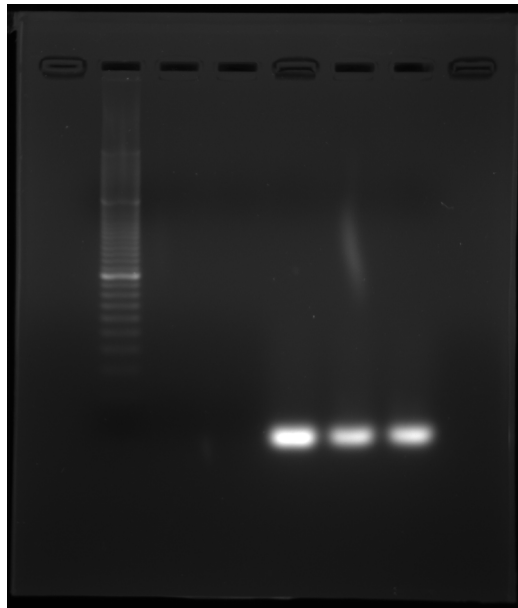
- Deux oligonucléotides simple-brin complémentaires
- Un grand bécher (2 à 5 L)
- Une plaque chauffante
- Un agitateur magnétique
- Un évaporateur “speedvac”
- Eau ultrapure (milliQ)
- Tampon d’hybridation : cacodylate de sodium 10 mM, pH=7.0; NaCl 150 mM
- Tubes de 1 mL avec un bouchon vissable étanche
- Agarose
- Tampon TAE 0,5 X (Tris-acétate 20 mM, EDTA 0,5 mM)
- Appareil d’électrophorèse

### 5.13.2 Protocole

1. Mettre les oligonucléotides simple-brin en solution dans l’eau ultrapure selon le protocole du fournisseur (viser une concentration finale d’environ 1 mM)
2. Déterminer précisément leur concentration par  $A_{260\text{ nm}}$
3. Mettre dans un tube des quantités égales de chaque brin (par exemple 500 nmol)
4. Sécher au speedvac
5. Pendant le séchage, faire bouillir de l’eau dans le grand bécher sur la plaque chauffante
6. Une fois les oligonucléotides simple-brin séchés, ajouter un volume adéquat de tampon d’hybridation (viser une concentration finale de 500  $\mu\text{M}$  à 2 mM)
7. Fermer le tube de façon bien étanche et agiter pour remettre les oligonucléotides en solution
8. Quand l’eau bout, placer le tube fermé dans l’eau (maintenu en surface par un flotteur), et après 5 min transférer le bécher sur un agitateur
9. Laisser refroidir passivement à température ambiante, en agitant à vitesse moyenne
10. Vers environ 30 °C, placer le bécher en chambre froide
11. Laisser refroidir passivement jusqu’à 4 °C pendant la nuit
12. Agiter le tube pour bien homogénéiser, et centrifuger brièvement pour ramener tout le liquide au fond du tube
13. Déterminer précisément la concentration de l’oligonucléotide simple-brin par  $A_{260\text{ nm}}$
14. Contrôler l’hybridation par électrophorèse sur gel d’agarose

15. Stocker à -20 °C (l'oligonucléotide est stable pendant plusieurs mois), éventuellement sous forme d'aliqots de plus petit volume pour éviter les décongélations et recongélations trop fréquentes

Un contrôle d'hybridation typiquement obtenu est présenté ci-dessous : 1,5 µg des oligonucléotides indiqués ont été déposés ; migration à 100 V constants dans un gel d'agarose 2,2 % m/v en tampon TAE 0,5 X ; coloration au BEt.



**Figure 88 : Contrôle de l'hybridation des oligonucléotides T2-S5 et T2-S7. De gauche à droite : marqueur de taille (PCR Low Ladder), simple-brin T2-S5, simple-brin T2-S7, double-brin T2-S5S7 en trois réplicats.**

Le BEt ne révèle pas les oligonucléotides simple-brin. L'apparition d'une bande de taille apparente de 20 pb confirme l'hybridation correcte. Nous ne pouvons pas détecter la présence d'un éventuel excès d'un des deux brins dans les produits d'hybridation, mais cela n'est pas problématique pour nos expériences car le domaine Myb de TRF2 n'interagit qu'avec les répétitions TTAGGG double-brin.

### 5.13.3 Remarques

La détermination précise de la concentration des oligonucléotides simple-brin est critique pour pouvoir réaliser un mélange le plus proche possible d'un rapport molaire 1:1. Utiliser les coefficients d'extinction molaire indiqués par le fournisseur, ou bien ceux obtenus avec l'un des outils suivants :

- <http://biophysics.idtdna.com/UVSpectrum.html>
- <http://www.basic.northwestern.edu/biotools/oligocalc.html>

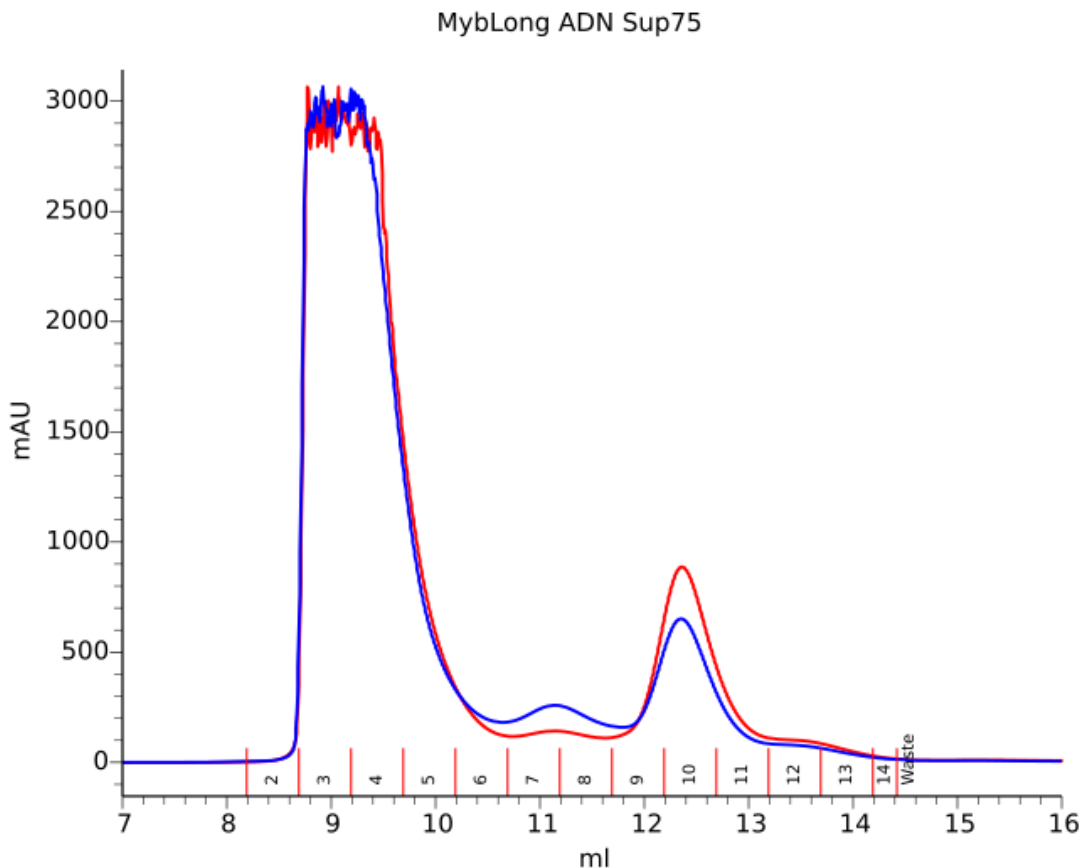
Un thermocycleur peut aussi être utilisé à la place de l'eau bouillante refroidissant passivement. Une descente en température finement contrôlée peut être nécessaire dans le cas d'oligonucléotides délicats à hybrider (plus longs, ou contenant des séquences avec lesquelles on souhaite former des tiges-boucles).

## 5.14 Résolution de la structure du complexe Myb-long/ADN

### 5.14.1 Préparation du complexe

La préparation du complexe Myb-long/ADN est réalisée en mélangeant la protéine avec un excès molaire en sites de liaison de 1,2. Ce mélange est concentré jusqu'à environ 100  $\mu$ L et injecté à 0,5 mL/min sur une colonne d'exclusion stérique Superdex75 10/300 GL (GE Healthcare) équilibrée en tampon Tris-HCl 20 mM, pH = 7,5, NaCl 150 mM, 2-mercaptoethanol 5 mM. Cette étape permet de séparer l'excès d'ADN du complexe, et nous assure que l'échantillon contient uniquement le complexe d'intérêt (les éventuels complexes où une seule protéine s'est fixée sur un oligonucléotide à deux sites sont aussi séparés). Les fractions d'intérêt sont réunies et concentrées jusqu'à environ 22 mg/mL en complexe.

Un chromatogramme typique obtenu en préparant le complexe Myb-long/T2-S5S7 est présenté dans la figure suivante :



**Figure 89 : Chromatogramme d'exclusion stérique du complexe Myb-long/ADN. Absorbance à 280 nm en bleu, absorbance à 260 nm en rouge.**

Les fractions 3, 4 et 5 sont réunies et concentrées à environ 23 mg/mL en complexe pour obtenir l'échantillon final (les fractions 7 et 8 contiennent probablement du complexe à une seule protéine, et les fractions 9, 10 et 11 contiennent l'excès d'oligonucléotide libre).

#### 5.14.1.1 Propriétés du complexe et de ses composants

Pour la protéine,  $M_W$  et  $\epsilon_{280\text{ nm}}$  sont obtenus avec l'outil ProtParam (<http://web.expasy.org/protparam/>) et  $\epsilon_{260\text{ nm}}$  est estimé par la mesure d'un spectre d'absorption UV au Nanodrop. Il s'obtient comme suit avec les valeurs d'absorbance lues :

$$\epsilon_{260\text{ nm}} = \epsilon_{280\text{ nm}} \times \frac{A_{260\text{ nm}}}{A_{280\text{ nm}}}$$

Pour l'oligonucléotide,  $M_W$  est connue car on connaît les masses molaires des deux brins hybridés (ces masses sont données par le fournisseur, ici Eurogentec),  $\epsilon_{260\text{ nm}}$  et  $\epsilon_{280\text{ nm}}$  sont déterminés avec l'outil IDT Biophysics (<http://biophysics.idtdna.com/UVSpectrum.html>).

Les propriétés du complexe sont déterminées en additionnant celles des composants en tenant compte de la stoechiométrie (deux protéines liées à un oligonucléotide).

- Myb-long
  - $M_W = 11691,1\text{ Da}$
  - $\epsilon_{280\text{ nm}} = 35980\text{ M}^{-1}\cdot\text{cm}^{-1}$
  - $\epsilon_{260\text{ nm}} \simeq 22264,85\text{ M}^{-1}\cdot\text{cm}^{-1}$
- oligonucléotide T2-S5S7
  - $M_W = 12233,1\text{ Da}$
  - $\epsilon_{280\text{ nm}} = 169099,5\text{ M}^{-1}\cdot\text{cm}^{-1}$
  - $\epsilon_{260\text{ nm}} = 320693,9\text{ M}^{-1}\cdot\text{cm}^{-1}$
- complexe 1 oligonucléotide + 2 Myb-long
  - $M_W = 35615,3\text{ Da}$
  - $\epsilon_{280\text{ nm}} \simeq 264786\text{ M}^{-1}\cdot\text{cm}^{-1}$
  - $\epsilon_{260\text{ nm}} \simeq 365223,7\text{ M}^{-1}\cdot\text{cm}^{-1}$

#### 5.14.2 Cristallogenèse

Toutes les expériences de cristallogenèse sont réalisées par la méthode de diffusion de vapeur. Les premiers essais de cristallisation sont réalisés par des cribles de 7x 96 conditions en gouttes assises de 0,2  $\mu\text{L}$  préparées par un robot Mosquito à la plateforme de cristallogenèse "PF6" de l'Institut Pasteur (Paris). Ces conditions de cristallisation viennent des kits commerciaux suivants :

- Crystal Screen 1 & 2 (Hampton Research)
- Wizard 1 & 2 (Emerald Biosystems)

- Structure Screen 1 & 2 (Molecular Dimensions Ltd)
- JBS 1-4 (Jena Bioscience)
- JBS 5-8 (Jena Bioscience)
- PEGion/Cryo (Hampton Research)
- Salt RX (Hampton Research)

Lorsque des conditions favorables sont trouvées dans ces cribles, elles sont reproduites manuellement au laboratoire en gouttes suspendues de 1  $\mu$ L dans des plaques de 24 puits.

Les 2x 48 conditions décrites dans (Pryor *et al*, 2012; Le Bihan *et al*, 2013) ont été préparées à l'aide du robot pipeteur MatrixMaker de la plateforme de cristallogénèse de l'Institut Pasteur. Ces solutions ont ensuite été utilisées pour préparer manuellement des gouttes suspendues en plaques de 24 puits.

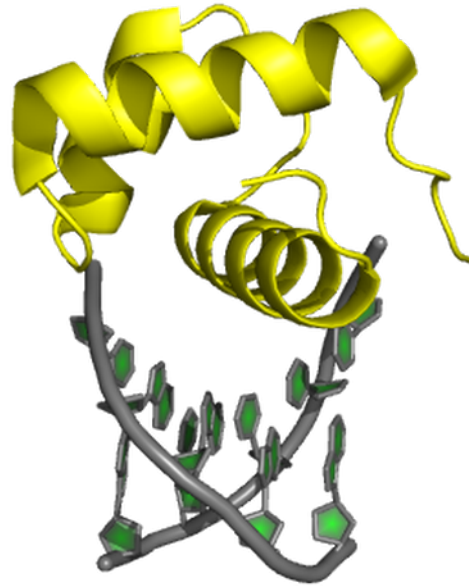
### 5.14.3 Collecte et traitement des données de diffraction

Toutes les expériences de diffraction des rayons X ont été faites au synchrotron SOLEIL, sur les lignes de lumière PROXIMA 1 et 2. Les images de diffraction ont été indexées et intégrées à l'aide du logiciel XDS (Kabsch, 2010).

### 5.14.4 Détermination des phases

Les phases sont déterminées par remplacement moléculaire avec le logiciel PHASER (McCoy *et al*, 2007) utilisé à travers l'interface de PHENIX (Adams *et al*, 2010). Le modèle a été préparé avec le logiciel PyMOL (Schrödinger LLC, 2010) à partir de l'entrée PDB 1W0U et contient un "mini-complexe" composé d'un Myb-court lié à un motif TAGGG double-brin. Ce modèle a 100 % d'identité de séquence avec notre complexe, et est représenté dans la figure suivante :





**Figure 90 : Modèle de remplacement moléculaire utilisé pour résoudre la structure du complexe Myb-long/ADN. La protéine est colorée en jaune et l'ADN en gris.**

Le coefficient de Matthews (Matthews, 1968; Kantardjieff & Rupp, 2003) est calculé avec l'outil Matthews de la suite CCP4 (Winn *et al*, 2011) et suggère que l'unité asymétrique contient deux copies du complexe entier (constitué d'un oligonucléotide à deux sites lié par deux molécules de Myb-long). Nous cherchons donc quatre copies de ce modèle à un Myb dans l'unité asymétrique.

#### 5.14.5 Affinement des modèles

Les affinements sont réalisés avec le logiciel phenix.refine (Afonine *et al*, 2012; Adams *et al*, 2010). Le premier tour d'affinement est toujours effectué en corps rigide pour ajuster le modèle dans la maille du cristal correspondant au jeu de données utilisées. Les affinements suivants agissent sur les coordonnées individuelles des atomes et les facteurs B.

## 5.15 Alignements de séquences

Les alignements présentés sont réalisés à l'aide du serveur Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo> ; Sievers *et al*, 2011; Li *et al*, 2015) ou manuellement dans le cas des motifs TBM courts. La coloration et l'annotation des alignements est réalisée avec le serveur ESPript (<http://esprict.ibcp.fr/ESPript/ESPript/> ; Robert & Gouet, 2014).

Pour l'analyse de conservation de TRF2 et RAP1, nous avons construit les alignements en utilisant les séquences issues des mêmes espèces de vertébrés pour les deux protéines. Selon ce critère, et en privilégiant également les entrées bien annotées de la base de données Uniprot, nous avons retenu 14 espèces.

Les séquences de TRF2 utilisées pour l'alignement sont les suivantes (identifiants Uniprot) :

1. Q15554 TERF2\_HUMAN
2. Q9PU53 TERF2\_CHICK
3. Q1WM12 Q1WM12\_XENLA
4. B8QB44 B8QB44\_XENTR
5. F6V5V7 F6V5V7\_CALJA
6. G1TE57 G1TE57\_RABIT
7. F1MMS4 F1MMS4\_BOVIN
8. D3ZJF7 D3ZJF7\_RAT
9. O35144 TERF2\_MOUSE
10. I2CTL7 I2CTL7\_MACMU
11. F7FBA4 F7FBA4\_MONDO
12. F1PBX6 F1PBX6\_CANFA
13. H0ZCM5 H0ZCM5\_TAEGU
14. Q8JGS4 Q8JGS4\_DANRE

Les séquences de RAP1 utilisées pour l'alignement sont les suivantes (identifiants Uniprot) :

1. Q9NYB0 TE2IP\_HUMAN
2. Q7TOL4 TE2IP\_CHICK
3. Q71M44 TE2IP\_XENLA
4. B8QB46 TE2IP\_XENTR
5. F7IE94 F7IE94\_CALJA
6. G1T9Q1 G1T9Q1\_RABIT
7. Q0VCT3 TE2IP\_BOVIN
8. Q5EAN7 TE2IP\_RAT

9. Q91VL8 TE2IP\_MOUSE
10. H9FW08 H9FW08\_MACMU
11. F7DF74 F7DF74\_MONDO
12. E2RSQ7 E2RSQ7\_CANFA
13. H0ZDI5 H0ZDI5\_TAEGU
14. Q6NYJ3 TE2IP\_DANRE

## 6 Références

- Abmayr SM, Yao T, Parmely T & Workman JL (2001) Preparation of Nuclear and Cytoplasmic Extracts from Mammalian Cells. In *Current Protocols in Molecular Biology* John Wiley & Sons, Inc.
- Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**: 213–221
- Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D Biological Crystallography* **68**: 352–367
- Aihara H, Huang WM & Ellenberger T (2007) An Interlocked Dimer of the Protelomerase TelK Distorts DNA Structure for the Formation of Hairpin Telomeres. *Molecular Cell* **27**: 901–913
- Amiard S, Doudeau M, Pinte S, Poulet A, Lenain C, Faivre-Moskalenko C, Angelov D, Hug N, Vindigni A, Bouvet P, Paoletti J, Gilson E & Giraud-Panis M (2007) A topological mechanism for TRF2-enhanced strand invasion. *Nature Structural & Molecular Biology* **14**: 147–154
- Ancelin K, Brunori M, Bauwens S, Koering C, Brun C, Ricoul M, Pommier J, Sabatier L & Gilson E (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. *Molecular and Cellular Biology* **22**: 3474–3487
- Bae NS & Baumann P (2007) A RAP1/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends. *Molecular Cell* **26**: 323–334
- Baumann P & Cech TR (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**: 1171–1175
- Bianchi A & Shore D (2008) How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex. *Molecular Cell* **31**: 153–165
- Bilaud T, Brun C, Ancelin K, Koering CE, Laroche T & Gilson E (1997) Telomeric localization of TRF2, a novel human telobox protein. *Nature Genetics* **17**: 236–239
- Bilaud T, Koering CE, Binet-Brasselet E, Ancelin K, Pollice A, Gasser SM & Gilson E (1996) The Telobox, a Myb-Related Telomeric DNA Binding Motif Found in Proteins from Yeast, Plants and Human. *Nucleic Acids Research* **24**: 1294–1303

- Blackburn EH, Greider CW & Szostak JW (2006) Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nature Medicine* **12**: 1133–1138
- Brown DG & Freemont PS (1996) Crystallography in the study of protein-DNA interactions. In *Crystallographic Methods and Protocols* p 293318. Springer
- Bryan TM, Englezou A, Dalla-Pozza L, Dunham MA & Reddel RR (1997) Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nature Medicine* **3**: 1271–1274
- Bryan TM, Englezou A, Gupta J, Bacchetti S & Reddel RR (1995) Telomere elongation in immortal human cells without detectable telomerase activity. *The EMBO Journal* **14**: 4240–4248
- Budd A (2012) Diversity of Genome Organisation. In *Evolutionary Genomics: Statistical and Computational Methods*, Anisimova M (ed) Humana Press
- Callebaut I & Mornon JP (1997) From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Letters* **400**: 25–30
- Celli GB & de Lange T (2005) DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature cell biology* **7**: 712–718
- Cesare AJ & Reddel RR (2010) Alternative lengthening of telomeres: models, mechanisms and implications. *Nature Reviews. Genetics* **11**: 319–330
- Chaconas G & Kobryn K (2010) Structure, Function, and Evolution of Linear Replicons in Borrelia. *Annual Review of Microbiology* **64**: 185–202
- Chen CW (2007) Streptomyces Linear Plasmids: Replication and Telomeres. In *Microbial Linear Plasmids*, Meinhardt F & Klassen R (eds) pp 33–61. Springer Berlin Heidelberg
- Chen Y, Rai R, Zhou Z, Kanoh J, Ribeyre C, Yang Y, Zheng H, Damay P, Wang F, Tsujii H, Hiraoka Y, Shore D, Hu H, Chang S & Lei M (2011) A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms. *Nature Structural & Molecular Biology* **18**: 213–221
- Chen Y, Yang Y, van Overbeek M, Donigian JR, Baciu P, de Lange T & Lei M (2008) A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science (New York, N.Y.)* **319**: 1092–1096
- Choi KH, Farrell AS, Lakamp AS & Ouellette MM (2011) Characterization of the DNA binding specificity of Shelterin complexes. *Nucleic Acids Research* **39**: 9206–9223
- Chong L, van Steensel B, Broccoli D, Erdjument-Bromage H, Hanish J, Tempst P & de Lange T (1995) A Human Telomeric Protein. *Science* **270**: 1663–1667
- Collins K (2011) Single-stranded DNA repeat synthesis by telomerase. *Current Opinion in Chemical Biology* **15**: 643–648

- Cong Y, Wright WE & Shay JW (2002) Human Telomerase and Its Regulation. *Microbiology and Molecular Biology Reviews* **66**: 407–425
- Court R, Chapman L, Fairall L & Rhodes D (2005) How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Reports* **6**: 39–45
- Crabbé L & Karlseder J (2010) Mammalian Rap1 widens its impact. *Nature cell biology* **12**: 733–735
- Cui T, Moro-oka N, Ohsumi K, Kodama K, Ohshima T, Ogasawara N, Mori H, Wanner B, Niki H & Horiuchi T (2007) Escherichia coli with a linear genome. *EMBO reports* **8**: 181–187
- David G & Pérez J (2009) Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *Journal of Applied Crystallography* **42**: 892–900
- de Lange T (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & Development* **19**: 2100–2110
- Diederichs K & Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural & Molecular Biology* **4**: 269–275
- Doksani Y, Wu JY, de Lange T & Zhuang X (2013) Super-Resolution Fluorescence Imaging of Telomeres Reveals TRF2-Dependent T-loop Formation. *Cell* **155**: 345–356
- Durand D, Vivès C, Cannella D, Pérez J, Pebay-Peyroula E, Vachette P & Fieschi F (2010) NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of Structural Biology* **169**: 45–53
- Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta crystallographica. Section D, Biological crystallography* **67**: 282–292
- Fairall L, Chapman L, Moss H, de Lange T & Rhodes D (2001) Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Molecular Cell* **8**: 351–361
- Fang G & Cech TR (1993) Oxytricha telomere-binding protein: DNA-dependent dimerization of the alpha and beta subunits. *Proceedings of the National Academy of Sciences* **90**: 6056–6060
- Fell VL & Schild-Poulter C (2015) The Ku heterodimer: Function in DNA repair and beyond. *Mutation Research/Reviews in Mutation Research* **763**: 15–29
- Fischer H, Oliveira Neto M de, Napolitano HB, Polikarpov I & Craievich AF (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography* **43**: 101–109

- Fouché N, Cesare AJ, Willcox S, Ozgür S, Compton SA & Griffith JD (2006) The basic domain of TRF2 directs binding to DNA junctions irrespective of the presence of TTAGGG repeats. *The Journal of Biological Chemistry* **281**: 37486–37495
- Garavís M, González C & Villasante A (2013) On the Origin of the Eukaryotic Chromosome: The Role of Noncanonical DNA Structures in Telomere Evolution. *Genome Biology and Evolution* **5**: 1142–1150
- Ghosh A & Tergaonkar V (2012) Extra-Telomeric Roles of Telomeric Proteins. In *Reviews on Selected Topics of Telomere Biology*, Li B (ed) InTech
- Gillis AJ, Schuller AP & Skordalakes E (2008) Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* **455**: 633–637
- Gilson E & Géli V (2007) How telomeres are replicated. *Nature Reviews. Molecular Cell Biology* **8**: 825–838
- Giraud-Panis M, Pisano S, Benarroch-Popivker D, Pei B, Le Du MH & Gilson E (2013) One identity or more for telomeres? *Frontiers in oncology* **3**: 48
- Giraud-Panis M, Teixeira MT, Géli V & Gilson E (2010) CST Meets Shelterin to Keep Telomeres in Check. *Molecular Cell* **39**: 665–676
- Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E & Ben-Tal N (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **19**: 163–164
- Gobbini E, Cesena D, Galbiati A, Lockhart A & Longhese MP (2013) Interplays between ATM/Tel1 and ATR/Mec1 in sensing and signaling DNA double-strand breaks. *DNA Repair* **12**: 791–799
- Gomez D (2012) Telomere structure and telomerase in health and disease (Review). *International Journal of Oncology*
- Gottschling DE & Zakian VA (1986) Telomere proteins: Specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell* **47**: 195–205
- Greider CW & Blackburn EH (1985) Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**: 405–413
- Greider CW & Blackburn EH (1987) The telomere terminal transferase of *tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**: 887–898
- Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H & de Lange T (1999) Mammalian telomeres end in a large duplex loop. *Cell* **97**: 503–514
- Grundy GJ, Moulding HA, Caldecott KW & Rulten SL (2014) One ring to bring them all The role of Ku in mammalian non-homologous end joining. *DNA Repair* **17**: 30–38

Guinier A (1939) La Diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques.

Hanaoka S, Nagadoi A & Nishimura Y (2005) Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. *Protein Science: A Publication of the Protein Society* **14**: 119–130

Hanaoka S, Nagadoi A, Yoshimura S, Aimoto S, Li B, de Lange T & Nishimura Y (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. *Journal of Molecular Biology* **312**: 167–175

Hayflick L (1965) The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research* **37**: 614–636

Hiraoka Y (1998) Meiotic telomeres: a matchmaker for homologous chromosomes. *Genes to Cells* **3**: 405–413

Hol W (2011) Protein-DNA complexes.  
<http://bstr521.biostr.washington.edu/PDF/2011%20Protein-DNA%20ComplexesV06.pdf>

Hollis T (2007) Crystallization of protein-DNA complexes. *Methods in molecular biology (Clifton, N.J.)* **363**: 225–237

Horvath MP, Schweiker VL, Bevilacqua JM, Ruggles JA & Schultz SC (1998) Crystal Structure of the Oxytricha nova Telomere End Binding Protein Complexed with Single Strand DNA. *Cell* **95**: 963–974

Houghtaling BR, Cuttonaro L, Chang W & Smith S (2004) A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Current biology* **14**: 1621–1631

Hsu H, Gilley D, Galande SA, Hande MP, Allen B, Kim S, Li GC, Campisi J, Kohwi-Shigematsu T & Chen DJ (2000) Ku acts in a unique way at the mammalian telomere to prevent end joining. *Genes & Development* **14**: 2807–2812

Hu J, Feng H, Zeng W, Lin G-x & Xi XG (2005) Solution structure of a multifunctional DNA- and protein-binding motif of human Werner syndrome protein. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 18379–18384

Huang WM, DaGloria J, Fox H, Ruan Q, Tillou J, Shi K, Aihara H, Aron J & Casjens S (2012) Linear Chromosome-generating System of *Agrobacterium tumefaciens* C58 : protelomerase generates and protects hairpin ends. *Journal of Biological Chemistry* **287**: 25551–25563

Ishikawa F & Naito T (1999) Why do we have linear chromosomes? A matter of Adam and Eve. *Mutation Research/DNA Repair* **434**: 99–107



- Janoušková E, Nečasová I, Pavloušková J, Zimmermann M, Hluchý M, Marini V, Nováková M & Hofr C (2015) Human Rap1 modulates TRF2 attraction to telomeric DNA. *Nucleic Acids Research* **43**: 2691–2700
- Kabir S, Hockemeyer D & de Lange T (2014) TALEN Gene Knockouts Reveal No Requirement for the Conserved Human Shelterin Protein Rap1 in Telomere Protection and Length Regulation. *Cell Reports* **9**: 1273–1280
- Kabir S, Sfeir A & de Lange T (2010) Taking apart Rap1: an adaptor protein with telomeric and non-telomeric functions. *Cell cycle (Georgetown, Tex.)* **9**: 4061–4067
- Kabsch W (2010) XDS. *Acta Crystallographica Section D Biological Crystallography* **66**: 125–132
- Kantardjieff KA & Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein science: a publication of the Protein Society* **12**: 1865–1871
- Karplus PA & Diederichs K (2012) Linking Crystallographic Model and Data Quality. *Science* **336**: 1030–1033
- Kim H, Lee O, Xin H, Chen L, Qin J, Chae HK, Lin S, Safari A, Liu D & Songyang Z (2009) TRF2 functions as a protein hub and regulates telomere maintenance by recognizing specific peptide motifs. *Nature Structural & Molecular Biology* **16**: 372–379
- Kim S-h, Kaminker P & Campisi J (1999) TIN2, a new regulator of telomere length in human cells. *Nature Genetics* **23**: 405–412
- Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, Harley CB & Aviv A (2010) Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nature Protocols* **5**: 1596–1607
- Kitano K, Kim S & Hakoshima T (2010) Structural Basis for DNA Strand Separation by the Unconventional Winged-Helix Domain of RecQ Helicase WRN. *Structure* **18**: 177–187
- Kobryn K & Chaconas G (2002) ResT, a Telomere Resolvase Encoded by the Lyme Disease Spirochete. *Molecular Cell* **9**: 195–201
- König P, Giraldo R, Chapman L & Rhodes D (1996) The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell* **85**: 125–136
- Kupiec M (2014) Biology of telomeres: lessons from budding yeast. *FEMS Microbiology Reviews* **38**: 144–171
- Kvaratskhelia M, Miller JT, Budihas SR, Pannell LK & Grice SFJL (2002) Identification of specific HIV-1 reverse transcriptase contacts to the viral RNA:tRNA complex by mass spectrometry and a primary amine selective reagent. *Proceedings of the National Academy of Sciences* **99**: 15988–15993

- Kyrion G, Liu K, Liu C & Lustig AJ (1993) RAP1 and telomere structure regulate telomere position effects in *Saccharomyces cerevisiae*. *Genes & Development* **7**: 1146–1159
- Lamarche BJ, Orazio NI & Weitzman MD (2010) The MRN complex in double-strand break repair and telomere maintenance. *FEBS letters* **584**: 3682–3695
- Leavitt S & Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Current Opinion in Structural Biology* **11**: 560–566
- Le Bihan YV, Matot B, Pietrement O, Giraud-Panis MJ, Gasparini S, Le Cam E, Gilson E, Sclavi B, Miron S & Le Du MH (2013) Effect of Rap1 binding on DNA distortion and potassium permanganate hypersensitivity. *Acta crystallographica. Section D, Biological crystallography* **69**: 409–419
- Lee O, Kim H, He Q, Baek HJ, Yang D, Chen L, Liang J, Chae HK, Safari A, Liu D & Songyang Z (2011) Genome-wide YFP fluorescence complementation screen identifies new regulators for telomere signaling in human cells. *Molecular & Cellular Proteomics: MCP* **10**: M110.001628
- Lei M, Zaug AJ, Podell ER & Cech TR (2005) Switching Human Telomerase On and Off with hPOT1 Protein in Vitro. *Journal of Biological Chemistry* **280**: 20449–20456
- Lenain C, Bauwens S, Amiard S, Brunori M, Giraud-Panis M & Gilson E (2006) The Apollo 5' Exonuclease Functions Together with TRF2 to Protect Telomeres from DNA Repair. *Current Biology* **16**: 1303–1310
- Li B & de Lange T (2003) Rap1 affects the length and heterogeneity of human telomeres. *Molecular Biology of the Cell* **14**: 5060–5068
- Li B, Oestreich S & de Lange T (2000) Identification of human Rap1: implications for telomere evolution. *Cell* **101**: 471–483
- Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N & Lopez R (2015) The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*
- Liu D, O'Connor MS, Qin J & Songyang Z (2004) Telosome, a Mammalian Telomere-associated Complex Formed by Multiple Telomeric Proteins. *Journal of Biological Chemistry* **279**: 51338–51342
- Loayza D & de Lange T (2003) POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **423**: 1013–1018
- Lundblad V & Blackburn EH (1993) An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* **73**: 347–360
- Luo Z, Dai Z, Xie X, Feng X, Liu D, Songyang Z & Xiong Y (2015) TeloPIN: a database of telomeric proteins interaction network in mammalian cells. *Database* **2015**: bav018

- Makarov VL, Hirose Y & Langmore JP (1997) Long G Tails at Both Ends of Human Chromosomes Suggest a C Strand Degradation Mechanism for Telomere Shortening. *Cell* **88**: 657–666
- Manke IA, Lowery DM, Nguyen A & Yaffe MB (2003) BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science (New York, N.Y.)* **302**: 636–639
- Marcand S, Gilson E & Shore D (1997) A Protein-Counting Mechanism for Telomere Length Regulation in Yeast. *Science* **275**: 986–990
- Martínez P & Blasco MA (2011) Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. *Nature Reviews. Cancer* **11**: 161–176
- Matot B (2010) Etude structurale de la protéine Rap1 impliquée dans le maintien de la stabilité des télomères chez la levure *Saccharomyces cerevisiae*.
- Matot B, Le Bihan Y, Lescasse R, Pérez J, Miron S, David G, Castaing B, Weber P, Raynal B, Zinn-Justin S, Gasparini S & Le Du MH (2012) The orientation of the C-terminal domain of the *Saccharomyces cerevisiae* Rap1 protein is determined by its binding to DNA. *Nucleic Acids Research* **40**: 3197–3207
- Matthews BW (1968) Solvent content of protein crystals. *Journal of molecular biology* **33**: 491–497
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *Journal of applied crystallography* **40**: 658–674
- Mendoza VL & Vachet RW (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrometry Reviews* **28**: 785815
- Milev S (2015) Don't throw away your 'bad' N-value ITC data. <http://www.materials-talks.com/blog/2015/04/23/dont-throw-away-your-bad-n-value-itc-data>
- Mitchell M, Gillis A, Futahashi M, Fujiwara H & Skordalakes E (2010) Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nature Structural & Molecular Biology* **17**: 513–518
- Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL & Wu JR (1988) A highly conserved repetitive DNA sequence, (TTAGGG)<sub>n</sub>, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 6622–6626
- Murti KG & Prescott DM (2002) Topological organization of DNA molecules in the macronucleus of hypotrichous ciliated protozoa. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* **10**: 165–173
- Mylonas E & Svergun DI (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of Applied Crystallography* **40**: s245–s249

- Nair SK, Sliverman SK, Chen JH & Xiao Y (2011) Crystal Structure Analysis of TRF2-Dbp-DNA complex. *To be Published*: null-null
- Nora GJ, Buncher NA & Opresko PL (2010) Telomeric protein TRF2 protects Holliday junctions with telomeric arms from displacement by the Werner syndrome helicase. *Nucleic Acids Research* **38**: 3984–3998
- Nosek J, Kosa P & Tomaska L (2006) On the origin of telomeres: a glimpse at the pre-telomerase world. *BioEssays* **28**: 182–190
- Opresko PL, Kobbe C von, Laine J, Harrigan J, Hickson ID & Bohr VA (2002) Telomere-binding Protein TRF2 Binds to and Stimulates the Werner and Bloom Syndrome Helicases. *Journal of Biological Chemistry* **277**: 41110–41119
- O'Connor MS, Safari A, Liu D, Qin J & Songyang Z (2004) The human Rap1 protein complex and modulation of telomere length. *The Journal of Biological Chemistry* **279**: 28585–28591
- Palm W & de Lange T (2008) How shelterin protects mammalian telomeres. *Annual Review of Genetics* **42**: 301–334
- Peersen OB, Ruggles JA & Schultz SC (2002) Dimeric structure of the Oxytricha nova telomere end-binding protein  $\alpha$ -subunit bound to ssDNA. *Nature Structural & Molecular Biology* **9**: 182–187
- Pérez J & Koutsioubas A (2015) Memprot: a program to model the detergent corona around a membrane protein based on SECSAXS data. *Acta Crystallographica Section D Biological Crystallography* **71**: 86–93
- Pickett HA & Reddel RR (2009) Alternative Lengthening of Telomeres in Human Cells. In *Telomeres and Telomerase in Cancer*, Hiyama K (ed) pp 127–148. Humana Press
- Podlevsky JD, Bley CJ, Omana RV, Qi X & Chen JJ (2008) The Telomerase Database. *Nucleic Acids Research* **36**: D339–D343
- Poulet A, Buisson R, Faivre-Moskalenko C, Koelblen M, Amiard S, Montel F, Cuesta-Lopez S, Bornet O, Guerlesquin F, Godet T, Moukhtar J, Argoul F, Déclais A, Lilley DMJ, Ip SCY, West SC, Gilson E & Giraud-Panis M (2009) TRF2 promotes, remodels and protects telomeric Holliday junctions. *The EMBO Journal* **28**: 641–651
- Poulet A, Pisano S, Faivre-Moskalenko C, Pei B, Tauran Y, Haftek-Terreau Z, Brunet F, Le Bihan Y, Le Du MH, Montel F, Hugo N, Amiard S, Argoul F, Chaboud A, Gilson E & Giraud-Panis M (2012) The N-terminal domains of TRF1 and TRF2 regulate their ability to condense telomeric DNA. *Nucleic Acids Research* **40**: 2566–2576
- Premkumar VL, Cranert S, Linger BR, Morin GB, Minium S & Price C (2014) The 3' Overhangs at Tetrahymena thermophila Telomeres Are Packaged by Four Proteins, Pot1a, Tpt1, Pat1, and Pat2. *Eukaryotic Cell* **13**: 240–245

- Pryor J Edward E, Wozniak DJ & Hollis T (2012) Crystallization of *Pseudomonas aeruginosa* AmrZ protein: development of a comprehensive method for obtaining and optimization of protein-DNA crystals. *Acta crystallographica. Section F, Structural biology and crystallization communications* **68**: 985–993
- Qin J & Gronenborn AM (2014) Weak protein complexes: challenging to study but essential for life. *FEBS Journal* **281**: 1948–1949
- Raffa GD, Cenci G, Ciapponi L & Gatti M (2013) Organization and Evolution of *Drosophila* Terminin: Similarities and Differences between *Drosophila* and Human Telomeres. *Frontiers in oncology* **3**: 112
- Rambo RP & Tainer JA (2010) Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Current Opinion in Structural Biology* **20**: 128–137
- Receveur-Brechot V & Durand D (2012) How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Current Protein & Peptide Science* **13**: 55–75
- Ribes-Zamora A, Indiviglio SM, Mihalek I, Williams CL & Bertuch AA (2013) TRF2 Interaction with Ku Heterotetramerization Interface Gives Insight into c-NHEJ Prevention at Human Telomeres. *Cell Reports* **5**: 194–206
- Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**: W320–W324
- Rodriguez M, Yu X, Chen J & Songyang Z (2003) Phosphopeptide Binding Specificities of BRCA1 COOH-terminal (BRCT) Domains. *Journal of Biological Chemistry* **278**: 52914–52918
- San Filippo J, Sung P & Klein H (2008) Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry* **77**: 229–257
- Sarthy J, Bae NS, Scrafford J & Baumann P (2009) Human RAP1 inhibits non-homologous end joining at telomeres. *The EMBO Journal* **28**: 3390–3399
- Scherthan H (2007) Telomeres and meiosis in health and disease. *Cellular and Molecular Life Sciences* **64**: 117–124
- Schrödinger LLC (2010) The PyMOL Molecular Graphics System, Version 1.7.4.
- Semenyuk AV & Svergun DI (1991) GNOM a program package for small-angle scattering data processing. *Journal of Applied Crystallography* **24**: 537–540
- Sfeir A & de Lange T (2012) Removal of shelterin reveals the telomere end-protection problem. *Science (New York, N.Y.)* **336**: 593–597
- Shi K, Huang WM & Aihara H (2013) An Enzyme-Catalyzed Multistep DNA Refolding Mechanism in Hairpin Telomere Formation. *PLoS Biol* **11**: e1001472

- Siderakis M & Tarsounas M (2007) Telomere regulation and function during meiosis. *Chromosome Research* **15**: 667–679
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**:
- Song K, Jung D, Jung Y, Lee S & Lee I (2000) Interaction of human Ku70 with TRF2. *FEBS Letters* **481**: 81–85
- Strohalm M, Kavan D, Nova?k P, Volny? M & Havli?c?ek V (2010) mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Analytical Chemistry* **82**: 4648–4651
- Svergun DI, Petoukhov MV & Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* **80**: 2946–2953
- Takai H, Smogorzewska A & de Lange T (2003) DNA Damage Foci at Dysfunctional Telomeres. *Current Biology* **13**: 1549–1556
- Takai KK, Hooper S, Blackwood S, Gandhi R & de Lange T (2010) In vivo stoichiometry of shelterin components. *The Journal of biological chemistry* **285**: 1457–1467
- Teo H, Ghosh S, Luesch H, Ghosh A, Wong ET, Malik N, Orth A, Jesus P de, Perry AS, Oliver JD, Tran NL, Speiser LJ, Wong M, Saez E, Schultz P, Chanda SK, Verma IM & Tergaonkar V (2010) Telomere-independent Rap1 is an IKK adaptor and regulates NF-kappaB-dependent gene expression. *Nature cell biology* **12**: 758–767
- van Overbeek M & de Lange T (2006) Apollo, an Artemis-Related Nuclease, Interacts with TRF2 and Protects Human Telomeres in S Phase. *Current Biology* **16**: 1295–1302
- van Steensel B & de Lange T (1997) Control of telomere length by the human telomeric protein TRF1. *Nature* **385**: 740–743
- Walker JR & Zhu X (2012) Post-translational modifications of TRF1 and TRF2 and their roles in telomere maintenance. *Mechanisms of ageing and development* **133**: 421–434
- Walker JR, Corpina RA & Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**: 607–614
- Wan B, Yin J, Horvath K, Sarkar J, Chen Y, Wu J, Wan K, Lu J, Gu P, Yu EY, Lue NF, Chang S, Liu Y & Lei M (2013) SLX4 Assembles a Telomere Maintenance Toolkit by Bridging Multiple Endonucleases with Telomeres. *Cell Reports* **4**: 861–869
- Wang F, Podell ER, Zaug AJ, Yang Y, Baciú P, Cech TR & Lei M (2007) The POT1TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**: 506–510
- Williams ES & Bailey SM (2009) Chromosome Orientation Fluorescence In Situ Hybridization (CO-FISH). *Cold Spring Harbor Protocols* **2009**: pdb.prot5269

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* **67**: 235–242

Wotton D & Shore D (1997) A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*. *Genes & Development* **11**: 748–760

Xin H, Liu D, Wan M, Safari A, Kim H, Sun W, O'Connor MS & Songyang Z (2007) TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature* **445**: 559–562

Ye JZ, Hockemeyer D, Krutchinsky AN, Loayza D, Hooper SM, Chait BT & de Lange T (2004) POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes & Development* **18**: 1649–1654

Yu EY, Yen W, Steinberg-Neifach O & Lue NF (2010) Rap1 in *Candida albicans*: an Unusual Structural Organization and a Critical Function in Suppressing Telomere Recombination. *Molecular and Cellular Biology* **30**: 1254–1268

Yu X, Chini CCS, He M, Mer G & Chen J (2003) The BRCT Domain Is a Phospho-Protein Binding Domain. *Science* **302**: 639–642

Zhong Z, Shiue L, Kaplan S & de Lange T (1992) A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Molecular and Cellular Biology* **12**: 4834–4843

## 7 Annexes

### 7.1 Alignement des séquences de RAP1

Les résidus identiques sont surlignés en rouge, les résidus similaires sont écrits en rouge et encadrés en bleu. Les résidus du motif TBM sont signalés par les triangles et l'étoile. Les lysines protégées dans le complexe TRF2/RAP1 sont signalées par des cercles (lysine protégée chez RAP1 sauvage et RAP1-ARAGA) et par un carré (lysine protégée uniquement chez RAP1 sauvage).







Homo_sapiens	357	.DGYPIWSRQDDIDLQKDD	DTREALVKKFGAQN	VARRIEFR	KK...
Gallus_gallus	287	.DGYPVWSREDDLELQKDD	DVRSKLIKFGAEN	VARRVAFR	KS...
Xenopus_laevis	471	.DGFPIWIRKDDLDLQKADA	ETQKRLIKYGSD	NVAKRVAFL	AS...
Xenopus_tropicalis	467	.DGYPIWVRKDDLDLQKDDA	ETLKRLIKYGAD	NVAKRVAFL	AS...
Callithrix_jacchus	357	.DGYPIWSRQDDIDLQKDD	DTREALVKKFGAQN	VARRIEFR	KK...
Oryctolagus_cuniculus		.....	.....	.....	.....
Bos_taurus	357	.DGYPIWSRQDDLDLQKDD	ATRDALVKKFGAQN	VARRIEFR	KK...
Rattus_norvegicus	351	.DGFPIWCRQDDLDLQKDD	DTRNALVKKYGAN	VARRIEFR	KK...
Mus_musculus	351	.DGYPIWCRQDDLDLQKDD	DTKNALVKKFGAN	VARRIEFR	KK...
Macaca_mulatta	357	.DGYPIWSRQDDIDLQKDD	DTREALVKKFGAQN	VARRIEFR	KK...
Monodelphis_domestica	345	.DGYPIWTRQDDVDLQKDD	ENTRSKLIKKFGAK	NVARRIEFR	SN...
Canis_familiaris	361	.DGYPIWSRQDDLDLQKDD	DARDALVKKFGAQN	VARRIEFR	KK...
Taeniopygia_guttata	264	.DGYPVWSREDDLELQKDD	HVRNKLVAKFGAG	NVEKRIVFR	ES...
Danio_rerio	570	E THGFLWTRLDDETLLAAD	EYELLEQLQS	KFGE	EEVTRRKSEFLATDVK

••

## 7.2 Alignement des séquences de TRF2

Notre construction de TRF2 recombinante est plus courte de 42 résidus que la séquence présente dans la base de données Uniprot. Pour cette raison, la numérotation de l'alignement est décalée de +42 par rapport à la numérotation utilisée dans toute la thèse (qui se rapporte toujours à notre construction de 500 résidus). La différence entre la séquence de notre TRF2 recombinante et la séquence présente dans Uniprot (code d'accès Q15554 TRF2\_HUMAN) est indiquée ci-dessous :

```

sp|Q15554|TRF2_HUMAN      MAAGAGTAGPASGPGVVRDPAASQPRKRPREGGEGARRSDTMAGGGGSSDGSGRAAGR 60
TRF2-recombinant        -----GAGGGGSSDGSGRAAGR 18
                          *****

sp|Q15554|TRF2_HUMAN      ASRSSGRARRGRHEPGLGGPAERGAGEARLEEAVNRWVLKIFYHEALRAFGRSRYGDFRQ 120
TRF2-recombinant        ASRSSGRARRGRHEPGLGGPAERGAGEARLEEAVNRWVLKIFYHEALRAFGRSRYGDFRQ 78
                          *****

sp|Q15554|TRF2_HUMAN      IRDIMQALLVRLPLGKHTVSRLLRVMQCLSRIEEENLDCSFDMEAEPLTESAINVLEM 180
TRF2-recombinant        IRDIMQALLVRLPLGKHTVSRLLRVMQCLSRIEEENLDCSFDMEAEPLTESAINVLEM 138
                          *****

sp|Q15554|TRF2_HUMAN      IKTEFTL TEAVVESSRKLKVEAAVVICIKNKEFEKASKILKKHMSKDPTTQKLRNDLLNI 240
TRF2-recombinant        IKTEFTL TEAVVESSRKLKVEAAVVICIKNKEFEKASKILKKHMSKDPTTQKLRNDLLNI 198
                          *****

sp|Q15554|TRF2_HUMAN      IREKNLAHPVIQNFYSYETFQKMLRFLESHLDDAEPYLLTMAKKALKSESAASSTGKEDK 300
TRF2-recombinant        IREKNLAHPVIQNFYSYETFQKMLRFLESHLDDAEPYLLTMAKKALKSESAASSTGKEDK 258
                          *****

sp|Q15554|TRF2_HUMAN      QPAPGPVEKPPREPARQLRNPPTTIGMMLTKAAFKTLSGAQDSEAAFAKLDQKDLVLPQTQ 360
TRF2-recombinant        QPAPGPVEKPPREPARQLRNPPTTIGMMLTKAAFKTLSGAQDSEAAFAKLDQKDLVLPQTQ 318
                          *****

sp|Q15554|TRF2_HUMAN      ALPASPALKNKRPRKDENEESAPADGEGGSELQPKNKRMTISRLVLEEDSQSTEPSAGLN 420
TRF2-recombinant        ALPASPALKNKRPRKDENEESAPADGEGGSELQPKNKRMTISRLVLEEDSQSTEPSAGLN 378
                          *****

sp|Q15554|TRF2_HUMAN      SSQEASAPPSKPTVLNQPLPGEKNPKVPGKWNSSNGVEEKETWVEEDELQVQAAPDE 480
TRF2-recombinant        SSQEASAPPSKPTVLNQPLPGEKNPKVPGKWNSSNGVEEKETWVEEDELQVQAAPDE 438
                          *****

sp|Q15554|TRF2_HUMAN      DSTTNITKKQKWTVEESEWVKAGVQKYGEGNWAAISKNYPFVNRTAVMIKDRWRMTKRLG 540
TRF2-recombinant        DSTTNITKKQKWTVEESEWVKAGVQKYGEGNWAAISKNYPFVNRTAVMIKDRWRMTKRLG 498
                          *****

sp|Q15554|TRF2_HUMAN      MN      542
TRF2-recombinant        MN      500
                          **

```

Dans l'alignement, les résidus identiques sont surlignés en rouge, les résidus similaires sont écrits en rouge et encadrés en bleu. Les lysines protégées dans le complexe TRF2/RAP1 sont signalées par des cercles (lysine protégée par RAP1 sauvage et RAP1-ARAGA) et par des carrés (lysines protégées uniquement par RAP1 sauvage).

Homo_sapiens	1	MAAGAGTAGPASGPGVVRDPAASQPRKRPGREGG...EGARRSDTMAGGGGSSDGSGRAAGRASRSSGRARRGRHEPG	LGGPAER	GAG	EAR	LE	AVNRWV	LK	FYF	HE	HAL
Gallus_gallus	1	.....	MAAKR..SRAAM	EEQEKTS	TRSDDR	EQAVNRWV	LK	FYF	HQ	AV	
Xenopus_laevis	1	.....	ME...SNST	LR.ECGSP	DP	CI	ERT	IN	OW	AF	DYYFHTAV
Xenopus_tropicalis	1	.....	ME...TNST	LE.ERRSF	DSCK	Q	ERT	IN	OW	FD	DYYFHTAV
Callithrix_jacchus	1	.....	GLGDVDFPK	QEAELRPAQ	RELRRFR	RRGPGPPV	L	SLPEAR	GRL	EAR	LEAVNRWV
Oryctolagus_cuniculus	1	MAAGAGTAGPASGPGVVRDPAASQPRKRPSRGGG...EGARRSDAMAGGGGSSDGSGR.AGRASRS	GGRRARRGRHEPG	LGGPDR	GAG	EAR	LE	AVNRWV	LK	FYF	HEHAL
Bos_taurus	1	MAAGAGTAGSASGPGVVRDPAASQSRKRPGREGG...EGARRSDAMAGGGGSSDGSRAAGRASRS	GGRRARRGRHAPG	LGGAAER	GAG	EV	RT	EA	AVNRWV	LK	FYFHEHAL
Rattus_norvegicus	1	.....	MAGGGGSSDNG	GRAANRRAS	RS	GGRRARRGRHEPG	LGGAAER	GAG	EAR	LE	AVNRWV
Mus_musculus	1	MAAGAGTAGPASGPGVVRDPMASQPRKRPSREGGEGGEGERRSNTMAGGGGSSDSSGRAASRRAS	RS	GGRRARRGRHEPG	LGGAAER	GAG	EAR	LE	AVNRWV	LK	FYFHEHAL
Macaca_mulatta	1	.....	MAGGGGSSDGS	GRAAGRASRS	SGRARRGRHEPG	LGGPAER	GAG	EAR	LE	AVNRWV	LK
Monodelphis_domestica	1	.....	.....	HETG	LGGDGER	GLG	EAR	LE	AVNRWV	LK	FYFHEHAL
Canis_familiaris	1	.....	MAGGGGSSDGS	GRAAGRASRS	GGRRARRGRQEPG	LGGATER	GAG	EAR	LE	AVNRWV	LK
Taeniopygia_guttata	1	.....	.....	GAG	MAALP	E	K	M	V	N	G
Danio_erio	1	.....	.....	MSD	K	P	C	E	P	S	W

Homo_sapiens	108	RAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Gallus_gallus	41	AAVRAGRNRFQRLRDVMQALLVRPLEREPVAQAQLRVMQFLSRIEEENLDCSFDKETEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Xenopus_laevis	37	EAFRSDRHEDFQEIRDIVSVLLLRPLETKAQNDCCKLRIMQCLSRIEEENLDCSFDNENNE	TPLES	AVGL	LD	IQ	AK	EM	PL	DE	ED
Xenopus_tropicalis	37	EAFRSDRRKDFQEIRDIVSVLLLRPLETKAQNDCCKLRIMQCLSRIEEENLDCSFDNENNE	TPLES	AVGL	LD	IQ	AK	EM	PL	DE	ED
Callithrix_jacchus	63	RAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Oryctolagus_cuniculus	107	RAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Bos_taurus	108	RAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Rattus_norvegicus	66	RAFRSSRYRDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Mus_musculus	111	RAFRSSRYRDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Macaca_mulatta	66	RAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Monodelphis_domestica	36	RAFRGGRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRIMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Canis_familiaris	66	RAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEAEELTPLESAINVLEMIKT	EFTL	TEAVV	ESSRKL	VEAAV	IT	CI	KNKE	FE	KAS
Taeniopygia_guttata	26	EAFRSGRNRFQFRDIMQALLVRPLDRPVDVTQMLRIMQLSRVEEENLDCSFDKLELELTPLESAMLVDFTR	EFSV	ADRT	ME	TV	Q	M	V	R	EAAV
Danio_erio	28	SAFRNGDYTAFTHFINIIESVVVRVVDGHTDMIPKLRLLMQLSRINNGDKLDTFEE...PKTPLESALD	LES	IC	R	D	M	S	V	P	H

Homo_sapiens	218	KILKKHMSKDP	TQKLRNDLLNI	IREKNLAHP	VIQNF	SYETFOQ	KMLRF	FLESHLDDA	EPYLL	TMAKKA	LKSE	SAA..S.S	S	.....	T	G	.....	
Gallus_gallus	151	AVLKRHMEKDT	GNQKKRTELQTI	IREKNRAHP	IIRDF	SYVNF	QOYMF	FLKTYVDT	SEPL	LV	TMMKSL	NSE	RAE	E	P	K	R	S
Xenopus_laevis	147	RILKKYISNY	.SKRKLRAELHTI	IREKNQH	PLITN	FSFTTI	IKKMIY	QLFEKHIDV	DP	ILL	TI	AKKE	Q	S	C	L	S	K
Xenopus_tropicalis	147	RILKKYISNN	.NKRKLRSDLQTI	IREKNQH	PLIAN	FSFTTI	IKKMIY	QLFEKHIEL	TA	PV	L	T	MAKKE	Q	S	C	L	S
Callithrix_jacchus	173	KILKKHMSKDP	TTQKLRNDLLNI	IREKNLAHP	VIQNF	SYETFOQ	KMLRF	FLESHLDDA	EPYLL	TMAKKA	LKSE	SAA..S.S	S	.....	T	V	.....	
Oryctolagus_cuniculus	217	KILKKHMSKDP	TTQKMKNDLLNI	IREKNVAHP	VIQNF	SYETFOQ	KMLRF	FLESHLDDT	EPYLL	TMAKKA	LKSE	SAA..S.S	S	.....	T	V	.....	
Bos_taurus	218	KILKKHMSKDP	TTQKLRNDLLNI	IREKNLAHP	VIQNF	SYETFOQ	KMLRF	FLESHLDDA	EPYLL	TSLPKA	LKSE	SAA..S.A	S	.....	T	V	.....	
Rattus_norvegicus	176	KILKKYMSKDP	TTQKVRTDLLNI	IREKNLAHP	VIQNF	SYETFOQ	KMLRF	FLESHLDDT	EPYLL	TMAKKA	LKSE	SAA..S.S	S	.....	T	M	.....	
Mus_musculus	221	KILKKYMSKDP	TTQKLRDLDLLNI	IREKNLAHP	VIQNF	SYETFOQ	KMLRF	FLESHLDDT	EPYLL	TMAKKA	LKSE	SAA..S.S	S	.....	T	M	.....	
Macaca_mulatta	176	KILKKHMSKDP	TTQKLRNDLLNI	IREKNLAHP	VIQNF	SYETFOQ	KMLRF	FLESHLDDA	EPYLL	TMAKKA	LKSE	SAA..S.S	S	.....	P	G	.....	
Monodelphis_domestica	146	KILKKHMSKDP	TTQKLRNDLLGI	IREKNITH	VIQNF	SYETFOQ	KMLRF	FLESHMDDA	EPYLL	TMAKRA	LKSE	SAS..T.S	S	.....	S	.....		
Canis_familiaris	176	KILKKHMSKDP	TTQKLRNDLLNI	IREKNLAHP	VIQNF	SYETFOQ	KMLRF	FLESHLDDA	EPYLL	TMARKA	LKSE	TAA..S.S	S	.....	T	V	.....	
Taeniopygia_guttata	136	DIVKHF	FGKEPRNQKRNELAV	IREKNPS	HPKVN	KSYED	FQOS	IF	FL	KGY	MDD	S	E	P	A	L	L	
Danio_erio	136	EMLKHF	SSSS.DSAGK	KKLLHL	IR	CR	SS	H	T	V	L	Q	L	S	S	Y	S	H

Homo_sapiens	297	.....KE.DK.....QPAP.....
Gallus_gallus	247	EPL.....RRVKHSGVTLRRAETAGGVAGAPSCPEMAKDPTGAPEHVGTVKDAVRAPCPAESTEDSQGTPRCAETARDVMGAPSPSEMTRDLL
Xenopus_laevis	230	.....QE.....
Xenopus_tropicalis	230	.....QE.....
Callithrix_jacchus	252	.....KE.DK.....QPAP.....
Oryctolagus_cuniculus	296	.....KE.DK.....QPAP.....
Bos_taurus	297	.....KEPQP.....QPAP.....
Rattus_norvegicus	255	.....KEE.....KHT.....
Mus_musculus	300	.....REE.....KHP.....
Macaca_mulatta	255	.....KE.DK.....QPAP.....
Monodelphis_domestica	224	.....QLAL.....
Canis_familiaris	255	.....KE.DK.....QPAP.....
Taeniopygia_guttata	207	.....EGAASPAK.....
Danio_rerio	245	SAVSRQGLQQVFDLLSEQYGVGISFFQLQDSVEAEAQQQQEGGVAGPELCLTLSETPMQTVILSDT.....

Homo_sapiens	305	.....GVEKP.....PRE.P.....
Gallus_gallus	335	GAPKCTETARDVVRAPSPAESTKD.PVGT.PGHAETARDVARAPSPAETTKNLPGAPECADTVKNTVRAPSPAERRKDLVRA.PKRAETARDVVRAPSPAERVKDTAGASEPM
Xenopus_laevis	232	.....QITLH.....L.....QNE.S.P.R.S.....
Xenopus_tropicalis	232	.....RTTSQ.....L.....QNE.S.P.R.S.....
Callithrix_jacchus	260	.....GVEKP.....PRE.P.....
Oryctolagus_cuniculus	304	.....E.V.A.K.P.....PRE.P.....
Bos_taurus	306	.....E.V.E.K.P.....LRE.P.....
Rattus_norvegicus	261	.....E.V.E.K.P.....LRE.P.....
Mus_musculus	306	.....E.V.E.K.P.....LRE.P.....
Macaca_mulatta	263	.....G.V.E.K.P.....PRE.P.....
Monodelphis_domestica	228	.....E.P.D.K.P.....LQ.E.P.....
Canis_familiaris	263	.....E.V.Q.K.P.....LKE.P.....
Taeniopygia_guttata	215	.....R.P.A.T.....S.P.E.A.R.H.R.A.L.A.E.L.....
Danio_rerio	311	.....E.....E.Q.D.E.P.R.R.Y.....

Homo_sapiens	315	.....A.RQL.RNPP.TT.I.G.M.M.T.L.K.A.A.F.K.T.L.S.G.A.O.D.S.E.A.A.F.A.K.L.D.Q.K.D.L.V.L.P.T.Q.A.L.P.A.S.P.A.L.K.N.K.R.P.R.K
Gallus_gallus	445	KSASYPTASQPRIAAVKSSKVFSVPVEISEQP..AAAAPVHAGVSSRDLE.R.T.P.F.R.T.V.T.T.Y.G.I.S.V.L.R.E.A.F.K.M.L.S.N.S.P.D.S.D.A.L.F.N.K.L.D.E.T.D.L.P.S.P.Q.M.S.P.S.V.S.H.R.T.K.R.R.K.E
Xenopus_laevis	245	.....P.Q.V.V.S.E.A.D.C.V.I.S.Y.G.L.S.V.L.K.S.K.F.K.S.L.C.Q.E.E.D.P.D.T.I.F.R.G.L.C.E.M.D.F.C.R.O.D.T.P.N.D.I.P.L.E.K.A.....
Xenopus_tropicalis	245	.....S.Q.V.V.S.E.A.D.C.V.I.S.Y.G.L.S.V.L.K.S.K.F.K.S.L.C.Q.E.E.D.T.D.T.M.F.R.K.L.C.E.T.D.F.C.R.K.E.T.P.I.D.I.P.V.E.K.A.....
Callithrix_jacchus	270	.....T.S.R.S.A.A.S.D.P.T.T.F.G.M.A.T.L.K.A.A.F.K.T.L.S.G.A.O.D.S.E.A.A.F.A.K.L.D.Q.K.D.L.V.L.P.T.Q.A.P.S.T.S.P.A.L.K.N.K.R.P.R.I
Oryctolagus_cuniculus	314	.....G.R.Q..L.Q.N.T.P.T.T.I.G.I.T.T.L.K.A.A.F.K.A.L.S.G.A.O.D.S.D.A.A.F.S.K.L.D.Q.K.D.L.V.L.P.N.Q.V.T.P.P.S.P.A.L.K.N.K.R.P.R.K
Bos_taurus	316	.....A.S.R.Q.L.Q.S.T.P.T.T.I.G.I.R.T.L.K.A.A.F.K.T.L.S.S.A.O.D.S.E.A.A.F.S.K.L.D.Q.K.D.M.V.F.P.N.K.V.C.P.P.S.P.A.L.K.N.K.R.P.R.K
Rattus_norvegicus	271	.....P.R.Q.P.R.N.P.P.T.T.I.G.L.R.T.V.K.A.A.F.K.A.L.S.T.A.Q.D.S.E.A.A.F.A.K.L.D.Q.K.D.L.V.L.P.N.L.A.S.P.S.P.T.H.K.N.K.R.P.R.K
Mus_musculus	316	.....P.S.R.Q.P.Q.N.P.P.A.T.I.G.I.R.T.L.K.A.A.F.K.A.L.S.T.A.Q.D.S.E.A.A.F.A.K.L.D.Q.K.D.L.V.L.A.N.L.A.S.P.S.P.A.H.H.K.H.K.R.P.R.K
Macaca_mulatta	273	.....T.R.Q.L.R.N.P.P.T.T.I.G.M.M.T.L.K.A.A.F.K.T.L.S.G.A.O.D.S.E.A.A.F.A.K.L.D.Q.K.D.L.V.L.P.T.Q.A.S.P.A.S.P.A.L.K.N.K.R.P.R.K
Monodelphis_domestica	238	.....M.R..P.L.P.E.S.P.A.T.F.G.I.S.T.L.K.K.A.F.K.S.L.S.D.S.O.D.P.D.A.A.F.T.K.L.D.Q.K.D.V.I.L.A.R.P.V.S.P.S.A.A.L.K.N.K.R.Q.R.Q
Canis_familiaris	273	.....V.R.Q.L.R.S.T.P.T.T.I.G.I.M.T.L.K.A.A.F.K.T.L.S.S.A.O.D.S.E.A.A.F.S.K.L.D.Q.K.D.L.V.L.P.N.Q.V.S.M.P.S.P.A.L.K.N.K.R.P.R.K
Taeniopygia_guttata	231	.....P.R.Q.Q.P.P.G.A.V.T.S.C.G.I.S.V.L.R.E.A.F.K.I.L.S.D.S.G.D.P.D.T.L.F.T.K.L.D.E.T.D.L.P.C.P.K.Q.L.S.P.S.V.S.H.R.T.K.R.C.R.E
Danio_rerio	320	.....A.G.M.T.I.S.R.L.V.Q.E.E.D.S.Q.I.S.A.E.H.T.H.I.E.D.E.E.E.E.E.E.E.E.Q.T.H.T.H.R.G.G.D..C.S.L.S.E.P.Q.C.S.S.T...Q.R.S.T.P.A.R.L.C.K.....L.R.S.N.T.R.V.Q.Q.S.E.S.E.S.Q.S.S.P.A.H.R.S.T.P.A.R.A

Homo_sapiens	376	.....D.....	ENESAPAD	GEGGSELQP	.KNKRMTISRLVLE	EDSQSTEP	SAGLN	SSQEAASA	...PP	SKPTV	...L	NQPLPGE	KNP	KVP	KGKW	
Gallus_gallus	552	.....E.....	KNQGS	SET...LDSP	EIPHKSKR	LFTISKMIMD	QGSQSKS	SESPD	SSQER	VVS...SA	YRPVQ	ELPD	QPVST	...KR	SSQQRW	
Xenopus_laevis	303	.....G.....	TDTEES	...ILGR	HRK	KLVS	LKQVVID	KDSQC	HLSEN	SNVSS	DR	NQAAPSA	QPKN	...Q	ITPKK	
Xenopus_tropicalis	303	.....G.....	HTHEE	PDSLIS	ERKGG	KKLVS	SLQQLVMKK	DNQC	INLSEN	ANISH	DS	AHQNVSSA	QKPN	...P	ITPKK	
Callithrix_jacchus	333	.....D.....	ENGSAPAE	GEGGSELQL	.KNKRMTISRLVLE	KDSQVTEP	STGLT	SSQEA	VPA..PP	SKPTV	...L	NQPLPGE	KNP	KIP	KGKW	
Oryctolagus_cuniculus	375	.....D.....	ENESAPAE	GEGGSELQP	.KSKRMTISRLILE	EDSQSTEP	STGLT	SSQEA	VPA...SP	PKPTA	...L	NQPRPRE	ISL	RVP	KSKW	
Bos_taurus	378	.....D.....	ENESAPTE	GEGGSELQP	.KTKRMTISRLVLE	EDSQSTEP	SSGLD	SSQEV	TPA...SP	SKPTV	...L	NQLPGE	KNP	KVT	KGKW	
Rattus_norvegicus	332	.....D.....	ENESAPAE	GEGGSE	ERQP	RNNRMTISRLVLE	EDSQSE	EP	SPDLN	SSRQA	LSA...SP	SKPRA	...L	NQPH	TGEP	
Mus_musculus	378	.....D.....	EHESA	APAE	GEGGSD	RQP	RNSP	MTISRLLE	EDSQSTEP	SPGLN	SSHKA	MS...A	SKPRA	...L	NQHP	GEP
Macaca_mulatta	334	.....D.....	ENESAPAE	GEGGSELQP	.KNKRMTISRLVLE	EDSQSTEP	SAGLN	SSQEA	APA...PP	SKPTV	...L	NQPLPGE	KNP	KVP	KGKW	
Monodelphis_domestica	298	.....E.....	ESSPS	SADNSELQP	.KNKRMTISRLVLE	HDSQISEQ	S	SDSPD	SSQET	IVS...SA	NS	TVL	KL	HN	QPLS	
Canis_familiaris	334	.....D.....	ENESAPAE	GEGGSELQP	.KNKRMTISRLVLE	EDSQSTEP	SASLD	SSQEV	VPA...SP	SKPTV	...L	NQPLPGE	KNP	KVP	KGKW	
Taeniopygia_guttata	292	.....V.....	ENQAS	SET...SQPP	EMPHNI	KNLFS	ISN	LIME	LSSSS	SKS	SECPD	SSQEH	VVS...SA	SKPAP	KLPEEP	
Danio_rerio	407	SHP	SNPR	SSTR	KKTP	SRRI	I	ESSD	EN	EQTTAA	AATH	TPR	GR	QR	QR	

Homo_sapiens	454	NSSNGVEE	.KETWV	EEDELF	QVQA...AP	DEDS	TNTI	TKKQ	KTVEE	SEWVKA	GVQK	YEGN	WAA	ISKNY	PFV	NRTAV	MIKDR	RTM	KRL	GMN...																																																	
Gallus_gallus	627	NSSYGEE	.KDSWS	DEDELF	TDAA	L	TETSS	NNS	IVYGS	KKQK	WTVEE	SEWIKD	GVRKY	YEGR	WKI	ISEK	YPF	QNR	TSV	QIKDR	RTM	KKL	GIA...																																														
Xenopus_laevis	374	DRS.DIVE	D	RDNWI	EAD	ELF	GSKRE	ENGK	TSNNT	STNS	KKQK	WTEDE	TEWIK	GVRKY	YEGN	WKD	IMK	YF	LN	R	TSV	MIKDR	RTM	KKL	GIV																																												
Xenopus_tropicalis	377	DGS.DIVE	E	QDTWS	EAD	ELF	KSKRE	ENK	TNN	TT	NG	KKQK	WTEDE	TEWIK	GVRKY	YEGN	WKD	ILK	K	PF	QNR	TSV	MIKDR	RTM	KKL	GIV																																											
Callithrix_jacchus	411	NSSNGVEE	.KETWV	EEDELF	PVQA...AP	DEDS	TNTI	TKKQ	KTVEE	SEWVKA	GVQK	YEGN	WAA	ISKNY	PFV	NRTAV	MIKDR	RTM	KRL	GMN...																																																	
Oryctolagus_cuniculus	453	NSSNGVEE	.KETWV	EEDELF	QVQGE...P	WNRL	GAS	LIL	LQK	WTVEE	SEWVKA	GVQK	YEGN	WAA	ISKNY	PFV	NRTAV	MIKDR	RTM	KKL	GMN...																																																
Bos_taurus	456	NSSNGVEE	.KETWV	EEDELF	QVQA...AP	DEES	ATNT	TKKQ	KTVEE	SEWVKA	GVQK	YEGN	WAA	ISKNY	PFV	NRTAV	MIKDR	RTM	KRL	GMN...																																																	
Rattus_norvegicus	410	NSPNGLEE	.KEVWS	EEDOLF	EVRA...P	G	EDK	SSN	L	ARKQ	KTVEE	SEWVKD	GVQK	YEGN	WVA	ISK	Y	PF	V	N	R	T	A	V	M	I	K	D	R	R	T	M	K	K	L	G	L	N...																															
Mus_musculus	454	NSPNGLEE	.KEVWL	EEDOLF	EVQA...P	G	EDR	SSS	L	R	KQ	KTVEE	SEWVKD	GVRKY	YEGN	WAA	ISK	Y	PF	V	N	R	T	A	V	M	I	K	D	R	R	T	M	K	K	L	G	M	N...																														
Macaca_mulatta	412	NSSNGVEE	.KETWV	EEDELF	QVQA...AP	DEDS	TNTI	TKKQ	KTVEE	SEWVKA	GVQK	YEGN	WAA	ISKNY	PFV	NRTAV	MIKDR	RTM	KRL	GMN...																																																	
Monodelphis_domestica	376	NSSNGLEE	.KDTWL	EEEELF	HVQ	V	PT	SEW	STH	TN	PL	T	L	KQ	KTVEE	SEWIK	GVQK	YEGN	WAA	ISK	N	Y	P	F	V	N	R	T	A	V	M	I	K	D	R	R	T	M	K	K	L	G	M	N...																									
Canis_familiaris	412	DSSNGAEE	KRHTWL	EEEELF	QVQ	G	K	S	A	P	D	E	S	T	S	I	T	I	T	K	K	Q	W	T	V	E	E	S	E	W	V	K	A	G	V	Q	K	Y	E	G	N	W	A	A	ISK	N	Y	P	F	V	N	R	T	A	V	M	I	K	D	R	R	T	M	K	K	L	G	M	N...
Taeniopygia_guttata	363	NSSCGTEE	.KDSWS	DEDELF	ANA	E	P	G	L	P	N	S...P	H	A	G	Q	V	L	K	W	T	I	Q	E	S	E	W	I	K	E	G	V	K	K	Y	E	G	K	W	K	S	I	C	L	K	Y	P	F	R	N	R	T	A	V	M	I	K	D	R	R	T	M	K	K	L	G	M	L...	
Danio_rerio	494	LDVSGI..	QDNWS	DEDSLE	H	T	S	T	A	P	A...K	K	Y	T	R	M	W	S	V	Q	E	S	E	W	L	K	Q	G	V	Y	G	V	G	H	E	R	I	R	S	A	E	P	F	A	G	R	T	A	V	N	L	K	D	R	R	T	M	V	K	L	K	M	V...						

### 7.3 Publication

Cet article est actuellement en cours de finalisation pour être soumis à publication. Les figures supplémentaires ne sont pas présentées afin de ne pas alourdir les annexes avec des informations redondantes déjà présentées dans la thèse. Elles sont néanmoins disponibles dans la version numérique de cette thèse.



1 **A higher-order entity formed by the flexible assembly of TRF2 and RAP1**

2 Guillaume Gaullier<sup>1,#</sup>, Simona Miron<sup>1,7,#</sup>, Sabrina Pisano<sup>2</sup>, Rémi Buisson<sup>2,8</sup>, Yann-Vaï Le  
3 Bihan<sup>1,9</sup>, Carine Tellier-Lebègue<sup>1</sup>, Wala Messaoud<sup>1</sup>, Pierre Roblin<sup>3,4</sup>, Beatriz G. Guimarães<sup>3</sup>,  
4 Robert Thai<sup>5</sup>, Marie-Josèphe Giraud-Panis<sup>2</sup>, Eric Gilson<sup>2,6</sup> and Marie-Hélène Le Du<sup>§1</sup>.

5 **Addresses:**

6 <sup>1</sup> Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud,  
7 Batiment 144, CEA Saclay, Gif-sur-Yvette, F-91191, France.

8 <sup>2</sup> Institute for Research on Cancer and Aging, Nice (IRCAN); CNRS UMR7284/INSERM  
9 U1081; Faculty of Medicine; Nice, France; Laboratoire de Biologie Moléculaire de la Cellule;  
10 CNRS UMR5239; Ecole Normale Supérieure de Lyon; UCBL1; IFR128; Lyon, France.

11 <sup>3</sup> Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91192 GIF-SUR-  
12 YVETTE Cedex, France

13 <sup>4</sup> Institut National de la Recherche Agronomique, Unité Biopolymères, Interactions,  
14 Assemblages, 44316 Nantes, France

15 <sup>5</sup> CEA, iBiTecS, F-91191 Gif-sur-Yvette, France

16 <sup>6</sup> Department of Genetics, CHU; Nice, France;

17 **Contact** : § : [marie-helene.ledu@cea.fr](mailto:marie-helene.ledu@cea.fr)

18 **Additional footnotes:**

19 # : These two authors equally contributed

20 <sup>7</sup> present address: Institute of Biochemistry, Splaiul Independentei 296, 060031, Bucharest 17,  
21 Romania

22 <sup>8</sup> present address: Massachusetts General Hospital Cancer Center, Harvard Medical School,  
23 Boston, MA 02129, USA

24 <sup>9</sup> present address: Division of Cancer Therapeutics, The Institute of Cancer Research,  
25 London, UK.

26

1 **ABSTRACT**

2 Telomere integrity is essential to maintain genome stability, and telomeric dysfunctions are  
3 associated with cancer and aging pathologies. In human, the shelterin complex binds  
4 TTAGGG DNA repeats and provides capping to chromosome ends. Within shelterin, TRF2 is  
5 required for telomere protection through a network of nucleic acid and protein interactions.  
6 One unresolved question is how TRF2 partners influence its properties toward DNA and  
7 regulate its capacity to bind multiple proteins. Through a combination of biochemical,  
8 biophysical and structural approaches, we unveiled a unique mode of assembly between  
9 TRF2 and its shelterin partner RAP1. The complete interaction scheme between the full-  
10 length proteins involves a primary site of RAP1 and a secondary interaction that directly  
11 affects the binding properties of the assembly. These results reveal how a non-DNA binding  
12 protein can influence the properties of a DNA-binding partner by mutual conformational  
13 adjustments.

14

1 Telomeres are specific nucleoprotein complexes that protect the chromosomes termini against  
2 degradation, checkpoint activation and illicit repair. Shelterin is a six-protein complex that  
3 constitutes the core of the telomere structure by binding double-stranded DNA through TRF1  
4 and TRF2 (Telomeric Repeat-binding Factors 1 and 2) and single-stranded DNA through  
5 POT1. TIN2 bridges TRF1 and TRF2, whereas TPP1 links double-stranded DNA and single-  
6 stranded DNA through its interaction with POT1 and TIN2, and RAP1 tightly interacts with  
7 TRF2 (1, 2). Shelterin functions depend on the recruitment of many associated factors (2, 3),  
8 and on the formation of specific assemblies with DNA (4). Systematic study of the telomeric  
9 proteins interaction network reveals more than one thousand protein-protein interactions and  
10 telomere-associated proteins (5). The individual implication of these proteins in shelterin  
11 functions has been extensively studied, but the functional synergies upon complex formation  
12 remain poorly understood. Telomere repeats induce a subcellular compartments significantly  
13 enriched in shelterin proteins. The local high concentration of shelterin proteins implies that  
14 the description of high affinity interactions is not sufficient to properly understand the  
15 regulatory processes involved in telomere maintenance. It is known that weak-affinity ( $K_D >$   
16  $10^{-4}$  M) and transient interactions are equally important than high-affinity interactions ( $K_D <$   
17  $10^{-6}$  M) in the regulation of many cellular pathways (6). However, our understanding of these  
18 regulatory processes is limited by our ability to study short-lived interactions. Weak and  
19 transient interactions are particularly difficult to study in multifunctional proteins, which  
20 interact with numerous partners often through a common or similar interface.

21 TRF2 is a multifunctional protein that prevents Non Homologous End Joining (NHEJ) and  
22 ATM-dependent checkpoint as well as DNA replication (2, 3, 7-9). Moreover, TRF2 binds  
23 several extra-telomeric sites where it regulates gene expression (10, 11). TRF2 is composed  
24 of a TRFH domain involved in homodimer assembly and in recruitment of several proteins  
25 (12-17), and a C-terminal Myb/telobox domain that specifically binds double-stranded DNA  
26 (18, 19). The basic N-terminal region of TRF2 hosts binding sites for several telomere-  
27 interacting molecules including the telomeric RNA TERRA, provides a specific ability to  
28 bind Holliday junctions and to protect them from resolvase cleavage, and helps TRF2 to  
29 condense DNA and chromatin (20-24). The long linker of TRF2 contains the binding sites for  
30 the two shelterin proteins RAP1 (Repressor Activator Protein 1) and TIN2 (25, 26).  
31 TRF2/RAP1 complex is involved in NHEJ protection in an *in vitro* assay (27), but RAP1  
32 deletion does not seem to affect cell viability or telomere protection *in vivo* (28, 29).  
33 Moreover, ChIP-seq analyses have shown that RAP1 can be found at chromatin sites in the

1 absence of TRF2 (30, 31). Cytoplasmic localization of RAP1 has also been reported, as well  
2 as its involvement in NF- $\kappa$ B signaling (32, 33). The interaction of RAP1 with TRF2 at  
3 telomeres could therefore correspond to a reservoir of RAP1. However, a recent study shows  
4 that RAP1 can affect TRF2 interaction properties toward duplex DNA (34). In TeloPIN  
5 database (5), TRF2 and RAP1 interact respectively with 247 and 131 partners, among whom  
6 32 are common partners between TRF2, RAP1 and other shelterin proteins, and 12 are  
7 common partners exclusively between TRF2 and RAP1. Therefore, whether or not  
8 TRF2/RAP1 interaction may affect/reinforce their respective properties and how DNA/TRF2  
9 complex recruits its numerous proteins partners compared to RAP1/TRF2/DNA complex  
10 remain to be fully characterized.

11 In order to address this central question, we initiated a comprehensive study of the parameters  
12 that describe the assembly of TRF2 and RAP1. We analyzed the TRF2/RAP1 complex both  
13 in terms of structure and binding properties. We observed that RAP1 affects TRF2 behavior  
14 *in vitro* toward particular DNA structures. Through a combination of SAXS, ITC,  
15 crystallography and biochemical approaches, we were able to draw the complete interaction  
16 scheme between the full-length proteins. Finally, we confirmed that a construct of RAP1  
17 deleted for its C-terminal high-affinity binding site is sufficient to modulate TRF2 properties  
18 toward Holliday junction, and observed that the binding properties of TRF2/DNA or of  
19 RAP1/TRF2/DNA toward proteins from HeLa cells nuclear extracts are different.

20

## 1 RESULTS

### 2 RAP1 affects TRF2 binding properties toward DNA

3 RAP1 is a key partner of TRF2 and is thus probably implicated in TRF2 multi-functionality.  
4 We addressed the effect of this main partner on TRF2 ability to interact with various DNA  
5 structures. In addition to linear double stranded TTAGGG (T<sub>2</sub>AG<sub>3</sub>) DNA, TRF2 recognizes  
6 branched DNA structures like Holliday junctions (HJ), the four-stranded DNA intermediates  
7 of homologous recombination (21). ITC (Isothermal Titration Calorimetry) and EMSA  
8 (Electrophoretic Mobility Shift Assay) experiments confirm that RAP1 does not bind linear  
9 double stranded TTAGGG (T<sub>2</sub>AG<sub>3</sub>) DNA, and does not prevent TRF2/DNA or TRF2/HJ  
10 binding (Supplementary Figure 1). Binding of TRF2 on HJ has been shown to protect this  
11 structure against the cleavage of various archetypical resolvases: human GEN1, T7  
12 Endonuclease I, yeast mitochondrial CCEI and the *E. coli* RusA enzyme. Interestingly, we  
13 observe that RAP1 inhibits TRF2 protection property against these enzymes in a  
14 concentration dependent manner (Figure 1A, Supplementary Figure 2). On the other hand, we  
15 do not observe a direct effect of RAP1 on the resolvases (Figure 1A, Supplementary Figures 2  
16 and 3), nor an increase in cleavage when we used the protection-deficient mutant of TRF2,  
17 TRF2<sup>AB</sup>, deleted for the N-terminus B-domain (Supplementary Figure 3) (21). These results  
18 suggest that this “recovery” of resolvase activity is not caused by a RAP1-dependent increase  
19 in the enzymes activity but rather by a decrease in TRF2-dependent protection. This effect of  
20 RAP1 was not expected since TRF2 HJ protection originates from the binding of the N-  
21 terminal basic domain of TRF2 on the center of the HJ (21), and the known RAP1 Binding  
22 Motif (TRF2-RBM) is located at residue 275-316 (25). Intriguingly, when we use a TRF2  
23 mutant lacking the whole linker (or hinge) domain and thus the TRF2-RBM (TRF2<sup>ΔL</sup>, Figure  
24 1B), we still observe this RAP1 effect of “anti-protection” for CCE1 and RusA, although  
25 higher concentration of RAP1 would be needed (Figure 1C). Therefore, we conclude that  
26 RAP1 does not directly interact with DNA as linear double stranded form or as HJ, and does  
27 not modify the capacity of TRF2 to interact with linear or branched DNA. But interestingly,  
28 the presence of RAP1 prevents TRF2 protection of HJ independently of the presence of  
29 TRF2-RBM.

30 In addition to its ability to bind particular DNA structures, TRF2 has been shown to condense  
31 linear DNA by promoting positive supercoils in the bound DNA (24, 35). As previously  
32 described (24), we drew 2D-probability maps giving the probability of any TRF2/DNA  
33 complex to exhibit a given volume in nm<sup>3</sup> for a given contour length in nm as measured by

1 Atomic Force Microscopy (AFM). As previously observed (24), TRF2 shortens DNA and  
2 drives the formation of complexes that become more and more condensed when the quantity  
3 of TRF2 increases (Figure 1D, top). In the presence of a small excess of full-length RAP1,  
4 DNA is still condensed in the complex compared to unbound DNA, but the highly condensed  
5 complexes largely disappear (Figure 1D, middle). Interestingly, the use of the known RAP1  
6 C-terminal TRF2 high-affinity binding domain (RAP1-RCT) instead of full-length RAP1  
7 leads to an increased condensation of DNA (Figure 1D, bottom). Therefore, RAP1 does not  
8 prevent TRF2 to condense linear DNA, but alters the level of condensation by blocking high  
9 degrees of oligomerization in the TRF2/DNA complexes. But interestingly, RAP1-RCT  
10 domain, which tightly binds TRF2, is not sufficient to block this oligomerization.

11 Altogether, these data show that RAP1 does not prevent TRF2 interaction with linear or  
12 branched DNA, but it clearly alters some of the properties of TRF2 toward DNA in particular  
13 by relieving the protective action of TRF2 on HJ cleavage by resolvase enzymes. Strikingly,  
14 this effect of RAP1 is maintained despite the deletion of the high affinity RAP1-RCT binding-  
15 site of TRF2. In the same way, the isolated RAP1-RCT domain, which contains the high  
16 affinity TRF2 binding-site, is not sufficient to prevent TRF2 to form highly condensed  
17 complexes with linear DNA. Therefore, these properties of RAP1 necessarily involve  
18 additional regions than the described TRF2 high affinity binding area.

### 19 **Interaction of RAP1 with TRF2 is associated to conformational adjustment**

20 The effect of RAP1 on TRF2 properties toward DNA raised questions on the structure of the  
21 TRF2/RAP1 complex. In particular, we wanted to investigate whether the modified behavior  
22 of TRF2 in the presence of RAP1 resulted from a conformational adjustment.

23 TRF2 and RAP1 are each made of globular domains linked by long unstructured regions,  
24 suggesting a high flexibility of the molecules (Figure 2A), not compatible with X-ray  
25 crystallography or electron microscopy approach, and the size of the complex (around 200  
26 kDa) is not compatible with an NMR study. On the opposite, Small Angle X-ray Scattering  
27 (SAXS) analysis is particularly appropriate to gain information about the average  
28 conformation of proteins and their assembly in solution (36). After careful control of the high  
29 purity level (>95%), and structural integrity of our tag-free samples (Supplementary Figure  
30 4), we collected SAXS curves on SWING beamline at SOLEIL synchrotron, using the online  
31 HPLC to avoid any sample aggregation and calculated *ab initio* envelopes for each of the  
32 isolated full length TRF2, full length RAP1 and RAP1/TRF2 complex, a RAP1-RCT  
33 construct containing residues 270-399, and RAP1-RCT/TRF2 complexes. The experimental

1 molecular weight derived from the SAXS curves revealed that TRF2 is dimer, RAP1 is  
2 monomer and that RAP1/TRF2 contains one TRF2 dimer and two RAP1 monomers (Figure  
3 2B). The p(R) functions and normalized Kratky plots (37) show that proteins alone or in  
4 complex behave as elongated molecules and adopt pearls necklace like conformation  
5 (Supplementary Figure 5).

6 TRF2 dimer is highly extended with a high degree of flexibility and a distance between the  
7 Myb DNA-binding domains of each monomer that can reach 230 Å (Figure 2C). TRF2,  
8 TRF2/RAP1 and TRF2/RAP1-RCT complexes both display a pseudo two-fold symmetry  
9 although no symmetry constraint was used during calculation process (Figures 2E and  
10 Supplementary 5G). In the case of the short construct RAP1-RCT, the complex formed with  
11 TRF2 does not involve particular conformational adjustment (Supplementary Figure 6A). On  
12 the contrary, the superimposition of TRF2 and RAP1/TRF2 envelopes shows a spreading for  
13 the complex compared to TRF2 alone, in agreement with the increased value of Dmax  
14 between the two p(R) functions (Supplementary Figure 6B). In addition, the shape of  
15 RAP1/TRF2 envelope is not compatible with a complete superimposition with TRF2  
16 envelope and two RAP1 envelopes (Supplementary Figures 6C and 6D). Therefore, the  
17 formation of TRF2/RAP1 complex involves a major conformational adjustment of both  
18 proteins.

19 In order to further characterize and quantify TRF2/RAP1 mode of assembly, we measured the  
20 interaction between full-length RAP1 and TRF2 by ITC. As compare to previous studies (34,  
21 Supplementary Figure 7A), in our condition we were able to observe a reproducible biphasic  
22 isotherm. The nonlinear least squares fit of the data to a two-independent sites binding model  
23 gives a dissociation constant of less than 1 nM for the high affinity signal, and a dissociation  
24 constant of 0.5 µM for the lower affinity signal (Figure 2F). The high affinity signal observed  
25 in the biphasic curve of the full-length proteins is equivalent to the one measured between  
26 RAP1-RCT and TRF2 (Figure 2G), suggesting that it corresponds to the known interaction  
27 between RAP1-RCT and TRF2-RBM (25). The presence of a second signal also suggests that  
28 the interaction between RAP1 and TRF2 full length involves additional regions from both  
29 partners than the previously described RAP1-RCT and TRF2-RBM domains (25).

### 30 **Complete interaction of RAP1 with TRF2 extends from BRCT to C-terminal domain**

31 Previous structural studies have shown that TRF2 dimerization domain (TRFH) specifically  
32 interacts with proteins containing the consensus sequence **YxLxP** (17, 25). This motif belongs

1 to numerous TRF2 partners among which the protein Apollo. Indeed a peptide containing the  
2 consensus sequence, Apollo-YLP (<sup>494</sup>ATEFRGLALKYLLTPVNF<sup>511</sup>), interacts with the TRFH  
3 domain with an affinity of 120 nM (15, 17), or 59 nM in our hands (Supplementary Figure  
4 7B). This motif is present in RAP1 sequence at position <sup>99</sup>YRLGP<sup>103</sup>, located at the C-  
5 terminus of the BRCT domain (Figure 2A). The corresponding peptide  
6 <sup>89</sup>ENRERLELEAYRLGPASA<sup>106</sup> (RAP1-YLP) interacts with TRFH with a dissociation constant  
7 of 13.0 μM (Supplementary Figure 7C). In order to delineate the structural determinants of  
8 this interaction we solved the crystal structure of the RAP1-YLP/TRFH complex at 2.4Å  
9 resolution (Figure 3A). The superimposition of our structure with this of TRFH in complex  
10 with the Apollo-YLP peptide (PDB entry 3BUA, 15) shows that the structures are similar,  
11 although the N-terminal part of Apollo-YLP and RAP1-YLP peptides are oriented in opposite  
12 directions (Figure 3B). In the case of the Apollo peptide, substitution of Tyr, Leu or Pro into  
13 Ala affects the interaction with the TRFH domain, although the effect was not equivalent for  
14 each position (17). We mutated the region <sup>99</sup>YRLGP<sup>103</sup> into <sup>99</sup>ARAGA<sup>103</sup> in RAP1 full length  
15 and, after checking its structural integrity by circular dichroism (Supplementary Figure 4C-E),  
16 we measured its interaction with TRF2. The high affinity ITC transition is maintained with  
17 the same dissociation constant than for wild-type RAP1, but the low affinity signal becomes  
18 not significant (Figure 3C), as in the case of RAP1-RCT (Figure 2G). Altogether, our data  
19 indicate that the sequence YRLGP from RAP1 interacts with the TRFH domain, is essential  
20 for the secondary interaction to occur with TRF2, and is necessary for a complete interaction  
21 of RAP1 with TRF2, although the residual secondary signal suggests that additional region  
22 may participate.

### 23 **RAP1 stretches along TRF2 dimerization domain**

24 In order to recover the protein surfaces involved in TRF2/RAP1 assembly, we performed  
25 protein footprinting using *in vitro* lysine acetylation by N-hydroxysulfosuccinimide acetate.  
26 This compound can acetylate lysines that are solvent-accessible but neither those buried in the  
27 structure nor those protected by interactions with another molecule. Through analysis of  
28 acetylation profiles using mass spectrometry, it is thus possible to determine which lysines are  
29 protected/buried in a protein. TRF2 and RAP1 contain respectively 44 and 25 lysines among  
30 which we observed acetylation for 24 lysines from TRF2 and 14 from RAP1 (Supplementary  
31 Figures 8 and 9). We compared the lysine acetylation profiles of TRF2, RAP1 and RAP1-  
32 ARAGA alone, from these of TRF2/RAP1 and TRF2/RAP1-ARAGA complexes. The  
33 comparison of lysine acetylation profiles between protein alone or in complex highlights



1 various lysine populations that are: 1) acetylated in all conditions, indicating constantly  
2 accessible regions that are therefore not involved in the interactions; 2) acetylated in the  
3 proteins alone and not in the complex with RAP1 wild type or with RAP1-ARAGA,  
4 indicating regions that participate in the interaction; or 3) acetylated in the proteins alone and  
5 in the complex with RAP1-ARAGA but not with wild-type RAP1, showing that the mutation  
6 ARAGA affects the interaction in this region; 4) not acetylated in the proteins alone and  
7 acetylated in either both complexes or in TRF2/RAP1 complex only, pointing regions that  
8 become accessible upon complex formation, probably due to an opening of the corresponding  
9 region (Supplementary information, Supplementary Figure 8 and 9). Taken together, the  
10 modifications in acetylation profiles allowed us to map the region of TRFH domain and  
11 RAP1 N-terminal region involved or not in TRF2/RAP1 complex and affected or not by the  
12 ARAGA mutation (Figure 4A and B, and supplementary figure 10)

13 Altogether, our results not only confirm the interaction of the N-terminal moiety of RAP1  
14 with TRFH, but also enable us to draw an interaction path along TRFH and RAP1-Nter, and  
15 to propose an interaction scheme (Figure 4C, and supplementary video). When the TRFH  
16 dimer is oriented with the inner side on top, RAP1 interaction path goes through the area  
17 facing us of the right monomer, the symmetrical surface still being accessible for the binding  
18 of a second RAP1 molecule without any steric hindrance, and ends at both the N- and C-  
19 terminal ends at the bottom of the TRFH domain (Figure 4C, left). Interestingly, the location  
20 of the **YxLxP** peptide at one extremity of RAP1 BRCT domain implies that RAP1/TRF2  
21 assembly induces a dimerization of the BRCT domain (Figure 4C, right). Therefore, RAP1  
22 complete interaction with TRF2 anchors both at the C-terminal domain and at the **YxLxP**  
23 sequence from the BRCT domain, and involves large additional regions from both proteins.  
24 Conservation analysis of TRF2 and RAP1 among vertebrates shows that the region of the  
25 TRFH domain from TRF2 that binds the motif **YxLxP** is highly conserved in agreement with  
26 its known crucial implication in TRF2 function (Figure 4D, left). In the case of RAP1, the  
27 surface from the N-terminal moiety involved in the interaction with TRFH domain from  
28 TRF2 is also highly conserved, suggesting that this interaction is also important for RAP1  
29 function among vertebrates (Figure 4D, right).

### 30 **Effect of Rap1[1-208] on TRF2 HJ protection**

31 Acetylation profile shows that both RAP1-wt and RAP1-ARAGA protects acetylation of  
32 lysines located near the N-terminal and region of TRF2: Lys58, Lys93, Lys242 and Lys245.  
33 On the contrary, RAP1-wt protects Lys184 and Lys190 but RAP1-ARAGA does not (Figure

1 4A and 4C, left). It has been previously observed that the HJ protection involves the N-  
2 terminal region of TRF2 (21), located close to the region protected both by RAP1-wt and by  
3 RAP1-ARAGA, but on the opposite side than the region that binds YxLxP from RAP1 or  
4 than the region protected by RAP1-wt only. We therefore expect that HJ protection by TRF2  
5 should be affected in the same way by RAP1 or by RAP1-ARAGA. On the contrary, we  
6 expect that a construct of RAP1 lacking the high-affinity C-terminal TRF2 binding site should  
7 have similar effect on HJ protection by TRF2 than this observed with RAP1 on TRF2  $\Delta$ L. We  
8 therefore tested the effect of RAP1-ARAGA and of RAP1[1-208] on HJ protection by TRF2.  
9 We observed that RAP1-ARAGA affects TRF2 HJ protection in a similar manner than RAP1-  
10 wt. Furthermore, the effect RAP1[1-208] on TRF2 HJ protection is similar to this observed  
11 with RAP1 full length on TRF2 $\Delta$ L mutant, higher quantity of RAP1[1-208] are necessary  
12 (Figure 5A) in the same way than higher quantity of RAP1 was necessary with TRF2 $\Delta$ L  
13 (Figure 1B and AC). Therefore, the interaction of RAP1[1-208] with a region of TRF2  
14 located at the bottom of the TRFH domain near the N-terminal region directly affects the  
15 property of TRF2 to protect HJ from resolvase. We conclude that this particular region is  
16 involved in HJ protection by TRF2 (Figure 4C).

### 17 **Protein binding properties of TRF2 versus TRF2/RAP1**

18 The partial overlap between RAP1-YLP and Apollo-YLP binding sites on the TRFH domain  
19 raises the question of a possible competition between both proteins for TRF2 binding. The  
20 strong difference in affinities for TRFH domain exhibited by RAP1-YLP and Apollo-YLP (13  
21  $\mu$ M versus 59 nM respectively) suggests that Apollo-YLP should be able to interact with  
22 TRF2 in the presence of RAP1. On the other hand, the high-affinity provided by the RAP1-  
23 RCT/ TRF2-RBM anchor (1 nM) brings the two molecules together thus increasing local  
24 concentration of the RAP-YLP. To choose between these two hypotheses, we measured both  
25 interactions of the Apollo-YLP peptide on a pre-formed TRF2-RAP1 complex and of RAP1  
26 on a pre-formed TRF2/Apollo-YLP peptide complex (Supplementary Figure 7C-7E). In the  
27 first case, we observed a binding of the Apollo-YLP peptide with an affinity of 47 nM, similar  
28 to this of 46 nM obtained in the absence of RAP1. In the case of pre-formed TRF2/Apollo-  
29 YLP complex, we still observed a biphasic isotherm, although the low affinity signal was too  
30 low to be quantified. This suggests that the formation of RAP1/TRF2 complex does not affect  
31 TRF2 binding properties toward Apollo protein, in agreement with the *in vivo* described  
32 capacity of Apollo to interact with TRF2 in the presence of RAP1 (13).

1 In order to verify on a general basis how the presence of RAP1 may affect or not the  
2 interaction between TRF2 and other protein partners, we designed an original streptavidin-  
3 DNA pull down experiment, using a biotinylated  $(T_2AG_3)_2$  oligo-nucleotide immobilized on a  
4 streptavidin column. After saturation of the column with TRF2, TRF2/RAP1 or TRF2/RAP1-  
5 ARAGA, we injected HeLa cells nuclear extract, and compared both chromatography elution  
6 and SDS-PAGE profiles. We first observe that the chromatography elution profiles are  
7 different in the case of TRF2, TRF2/RAP1 and TRF2/RAP1-ARAGA (Supplementary Figure  
8 12). We performed our comparison on the region of the gel corresponding to molecular  
9 weight higher than this of TRF2 or RAP1 in order to avoid contamination with degradation  
10 products of TRF2, RAP1 or RAP1-ARAGA. We first observed that a large number of  
11 proteins are retained by TRF2, TRF2/RAP1 or TRF2/RAP1-ARAGA (lanes 5, 8 and 11  
12 respectively in Figure 5B). Our comparison focuses on the presence or absence of specific  
13 bands, and not of their intensity, which would not be significant enough. We observed that the  
14 presence of RAP1 or RAP1-ARAGA is associated with supplementary bands compared to  
15 TRF2 alone, which agrees with TeloPIN. Interestingly, we also observed that the SDS-PAGE  
16 patterns are different between TRF2, TRF2/RAP1 and TRF2/RAP1-ARAGA (lanes 5, 8 and  
17 11 in Figure 5B). Therefore, the binding properties of TRF2, TRF2/RAP1 or TRF2/RAP1-  
18 ARAGA, loaded on  $(T_2AG_3)_2$  oligonucleotide, are all different from each other. We conclude  
19 that the mutation of the **YxLxP** motif associated to the complete interaction of RAP1 with  
20 TRF2 dimerization domain directly affects the binding properties of the assembly formed  
21 with TRF2.  
22

## 1 **DISCUSSION**

2 Our comprehensive study of TRF2, RAP1 and the effect of RAP1/TRF2 complex formation  
3 on their conformation and interaction properties reveal new elements that strongly deepen our  
4 knowledge of these two proteins and their inter-relation. We first observed that RAP1 affects  
5 both the ability of TRF2 to form condensed multimeric complexes with DNA and to protect  
6 Holliday junction from resolvases (Figure 1), that the formation of TRF2/RAP1 assembly is  
7 more complex than the known RAP1-RCT/TRF2-RBM interaction (15), and that the  
8 assembled molecules form a distinct entity. The combination of SAXS and ITC analysis  
9 confirms that the formation of TRF2/RAP1 complex involves a profound conformational  
10 adjustment (SAXS), associated with a two-transition mode of interaction (ITC) (Figure 2).  
11 Moreover, the isolated RAP1-RCT domain is not sufficient to induce a conformational  
12 adjustment, or a secondary ITC signal (Figure 2, Supplementary Figure 5G). A **YxLxP** motif,  
13 specific for the interaction with the TRF2 dimerization domain, TRFH (14), is located at the  
14 C-terminus of the BRCT domain at position 99-103. The crystal structure of TRFH in  
15 complex with RAP1 peptide containing the **YxLxP** motif show that the molecular  
16 determinants of the interaction involve the **YxLxP** motif and is similar to that previously  
17 observed between TRFH and Apollo-YLP peptide (Figure 3A-C). Moreover, the similar ITC  
18 interaction profiles between TRF2 and RAP1-ARAGA (Figure 3C) and between the pre-  
19 formed TRF2/Apollo-YLP complex and RAP1 (Supplementary Figure 6F) show that the  
20 Apollo-YLP peptide compete with RAP1 for binding on TRF2 TRFH domain and that this  
21 competition is centered on the **YxLxP** site. The residual secondary signal in both case also  
22 suggests that additional region of RAP1 are involved in the secondary binding site (Figure 3C  
23 and Supplementary figure 6F). Protein footprinting experiments allowed us to fulfill the  
24 TRF2/RAP1 interaction scheme and to draw a model of TRFH/RAP1 N-terminus complex. In  
25 this model, the N-terminus moiety of RAP1 follows one face of the TRFH domain, interacts  
26 through **YxLxP** sequence in the inner part of the TRFH, and ends at the external side of the  
27 domain (Figure 4C, left). Looking from the top of the TRFH, we observe that the interaction  
28 of the N-terminal region of RAP1 induces an unexpected dimerization of the BRCT domain  
29 (Figure 4C, right). The high conservation among vertebrates of the YLP binding site of TRF2  
30 and of the surface of RAP1 involved in the secondary interaction also reveals that this  
31 secondary interaction of RAP1 and the capacity of TRF2 to interact with YLP motif  
32 containing proteins are of equal importance.

1 The similar ITC interaction profiles between TRF2 and Apollo-YLP and between the  
2 preformed TRF2/RAP1 complex and Apollo-YLP also suggests that RAP1 interaction may  
3 probably not hinder Apollo interaction (Supplementary Figure 7C and 7D). Similar analysis  
4 could apply to SLX4, which contains a TRFH binding motif (RGLV**H**R**L**A**P**W), is associated  
5 to a dissociation constant of 750 nM with TRF2, and adopts the same conformation than  
6 Apollo-YLP peptide in complex with TRFH (PDB entry 4M7C, 15). In TRF1 dimerization  
7 domain, the region equivalent to this of TRF2 specifically interacts with the TIN2 protein. A  
8 TIN2 peptide<sup>248</sup>THPEPLAGR**H**F**N**L**A**P**L**GR<sup>265</sup>, specific for TRF1 dimerization domain, interacts  
9 *in vitro* with TRF2 TRFH with a dissociation constant of 6 μM (15), in the same order of  
10 magnitude than this of RAP1-YLP peptide on TRF2 (13 μM). The variable affinity measured  
11 for the **YXLXP**-containing peptides suggests that RAP1 secondary interaction might 1/  
12 impede inappropriate binding on the TRFH of TRF2 (in case of low affinity binding like for  
13 TIN2); 2/ vary depending on the chemical environment. In this way, our streptavidin-DNA  
14 pull down reveals that, when bound to a (T<sub>2</sub>AG<sub>3</sub>)<sub>2</sub> oligonucleotide, TRF2, TRF2/RAP1 or  
15 TRF2/RAP1-ARAGA do not retain the same proteins from a HeLa nuclear extract. In  
16 particular, the capacity of RAP1 to form or not a complete interaction with TRF2 through the  
17 **YXLXP** motif is associated with different binding properties (Figure 5B).

18 Based on our results, we propose an overall scheme of the formation of the TRF2/RAP1  
19 complex and the possible effect on their binding properties (Figure 6). In the absence of  
20 RAP1, TRF2 interacts with multiple partners, with no site discrimination in the case of TIN2  
21 (Figure 6, left). The RAP1/TRF2 complex can adopt alternative open or closed conformation  
22 depending on the presence or absence of TRFH high-affinity **YXLXP**-containing protein  
23 partners (such as Apollo or SLX4) that compete with RAP1 N-terminal region, or depending  
24 on the chemical environment that may favor or not secondary interaction (Figure 6, right).  
25 TRF2/RAP1 closed conformation inhibits additional illicit interaction with TRFH (such as the  
26 one with TIN2), induces RAP1 BRCT domain dimerization (Figure 4C), and forms a new  
27 entity that may interact with other specific telomeric or non-telomeric partners, whereas  
28 TRF2/RAP1 open conformation has different binding properties (Figure 6). Of note, TRF2  
29 alone, RAP1 alone and closed TRF2/RAP1 models are in agreement with our SAXS results.  
30 The open form of TRF2/RAP1 model was extrapolated from the ITC and streptavidin-DNA  
31 pull down experiments. Can this scheme of interaction explain the effect of RAP1 on the  
32 intrinsic properties of TRF2?

1 We have seen that RAP1 does not modify TRF2 property to condense DNA (Figure 1), but it  
2 does affect the oligomerization level that we observed in AFM experiments. This is a  
3 consequence of the complete RAP1 interaction since RAP1-RCT alone is not able to do so.  
4 The mechanism behind could involve steric hindrance or hiding of an oligomerization-prone  
5 domain of TRF2, located at or close to the TRFH domain. Moreover, how RAP1 affects the  
6 basic domain-dependent ability of TRF2 to protect HJ against resolvases remains to be  
7 elucidated. However, our data show that the interaction of RAP1 with the region of the TRFH  
8 domain of TRF2 is part of the mechanism behind the alteration in the protection of HJ against  
9 resolvases.

10 Our data not only show that RAP1 affects TRF2 conformation and function but also that  
11 TRF2 deeply affects RAP1 conformation. The induced dimerization of RAP1 BRCT domain  
12 upon full interaction with TRF2 is of particular interest (Figure 4C, right). The lysine  
13 acetylation profile of RAP1-ARAGA in the footprinting experiments shows that mutation of  
14 the **YxLxP** motif affects RAP1 interaction with TRFH, and particularly the induced  
15 dimerization of the BRCT domain. In addition, the streptavidin-DNA pull down shows that  
16 the capacity of RAP1 to form a complete interaction with TRF2 or not is associated with  
17 different binding properties. This provides both a clue about the functional role of RAP1  
18 BRCT domain, and an explanation of the high conservation of this domain among Rap1  
19 proteins from yeast to human. Indeed, BRCT domains are found in a wide array of  
20 prokaryotic and eukaryotic proteins. In human, at least 23 genes coding for proteins with  
21 BRCT domains have been listed, most of them implicated in DNA damage response. The  
22 induced dimerization mimics a BRCT tandem, known for its propensity to interact with  
23 phosphorylated proteins (38). The induced dimerization of RAP1 BRCT upon interaction with  
24 TRF2 suggests that RAP1 functions could be quite different whether RAP1 co-localizes with  
25 TRF2 or not.

26 Altogether, our data indicate that formation of the TRF2/RAP1 complex needs the presence of  
27 several domains of the proteins. Importantly, we show that the TRF2/RAP1 complex forms a  
28 structural entity with different competences than individual TRF2 and RAP1. We have shown  
29 that the available accessible surface of TRF2 or TRF2/RAP1 determines binding preferences  
30 for other partners of TRF2. Determining which proteins is more specific for TRF2 or for  
31 TRF2/RAP1 should provide a key to the regulation of TRF2 and RAP1 functions. In a general  
32 manner, one should consider TRF2, RAP1 and TRF2/RAP1 as three distinct, although related,  
33 entities.

## 1 **METHODS**

### 2 **Protein constructs used in this study**

3 See Supplementary Table 1 for a complete list of the constructs used in this study.

### 4 **HJ protection**

5 5 nM of tHJ labelled on one strand was incubated with different concentrations of proteins in  
6 10  $\mu$ l of binding buffer containing 10 mM MgAc<sub>2</sub> and digested for 60 min using 2 nM of T7  
7 Endonuclease I, MBP-CceI and MBP-RusA or 50 nM of GEN1<sup>1-527</sup>. Reactions were  
8 performed at 4°C for Endo I or 37°C for MBP-CceI, MBP-RusA and GEN1<sup>1-527</sup>. The reaction  
9 was stopped by addition of 6  $\mu$ g of proteinase K followed by 20 min incubation at 4°C or 15  
10 min at 37°C. Formamide was added to a final concentration of 60% (v/v) and the samples  
11 loaded on a 10 % denaturing acrylamide (19:1, bis:mono) 1xTBE gel. After migration in  
12 1xTBE, gels were analysed using a Phosphorimager TyphoonFLA9000 (GE Healthcare) and  
13 the Image Quant software (GE Healthcare).

### 14 **EMSA**

15 Proteins were incubated 15 min at 4°C with 5 nM DNA in 10  $\mu$ L of 20 mM Tris-Acetate pH8,  
16 0.1 mg ml<sup>-1</sup> BSA, 1 mM DTT, 50 mM KAc, 30 mM NaCl, 5% (v/v) glycerol (binding buffer).  
17 Gel electrophoresis was performed as published (34) and gels were analysed as above.

### 18 **AFM**

19 50 nM of TRF2 dimer and either 100 nM of full length Rap1 or Rap1-RCT were pre-  
20 incubated at 0°C for 15 min in the binding buffer (150 mM KCl, 5 mM HEPES pH 7.4).  
21 DNA (final concentration 10 nM) was then incubated with a 1:1 molar ratio of TRF2 with the  
22 preincubated proteins for 20 min at 25°C in the same binding buffer. Samples were  
23 crosslinked with glutaraldehyde (0.1% final) for 30 min on ice and applied on freshly cleaved  
24 mica surfaces treated with 10 mM MgCl<sub>2</sub>. After 2 min, mica was washed with 1 ml of  
25 deionized water and dried with a gentle flux of nitrogen. Imaging was performed on a  
26 Nanoscope IIIa equipped with E-scanner (Digital Instruments Inc., Santa Barbara, CA, USA),  
27 in air under Tapping Mode using silicon tips. Images were recorded at 1.5–2.0 Hz over scan  
28 areas 1  $\mu$  m wide (512  $\times$  512 pixels). Raw scanning force microscopy images were flattened  
29 using the manufacturer's software and converted into TIF files. Contour lengths (CLs) were  
30 measured by the read-through length method using SigmaScan Pro software (SPSS Inc.,  
31 Chicago, IL, USA). Volumes calculated as half-ellipsoids as published (35). Between 150 and

1 250 objects were scored for each condition. The 2D probability density map is a smoothed  
2 distribution that represents the convolution of the experimental distribution with a  $s_1 \times s_2$   
3 gaussian pulse. To construct the 2D map, each protein/DNA complex with a raw DNA length  
4  $L_{\text{raw}}$  and a complex volume  $\text{Vol}_{\text{complex}}$  contributes as gaussian distribution of respectively  $s_1$   
5 and  $s_2$  width in  $L_{\text{raw}}$  and  $\text{Vol}_{\text{complex}}$  directions. Each pixel of the 2D map is the sum of the  
6 various gaussian contributions from the experimental data points. The resulting 2D map  
7 represents a 100 by 100 pixel image of the probability to find a protein/DNA complex with  
8 given raw DNA length and complex volume, the intensity of each pixel is given by the  
9 associated probability density color scale.

## 10 **Protein expression and purification**

11 Protocols are detailed in supplementary information.

## 12 **SAXS**

13 SAXS experiments were performed at SWING beamline, synchrotron SOLEIL. TRF2, RAP1  
14 and RAP1-RCT were set in the buffer containing 20 mM Tris pH = 7.5, 150 mM NaCl, 5 mM  
15 2-mercaptoethanol and 0.2 mM PMSF. TRF2 was concentrated to 215  $\mu\text{M}$  (monomers),  
16 RAP1 to 326.7  $\mu\text{M}$ , RAP1-RCT to 1015.7 $\mu\text{M}$ , TRF2 and RAP1 were mixed to 120 and  
17 144 $\mu\text{M}$  respectively, and TRF2 and RAP1-RCT were mixed to 171.5 and 205.8  $\mu\text{M}$   
18 respectively. All data were collected at SWING beamline at SOLEIL synchrotron using either  
19 the sample changer (RAP1 full-length), or using the online HPLC system (39), equipped with  
20 a BIO-SEC3 column (Agilent) for all other proteins or complexes. After data collection, all  
21 leftovers from these preparations were analyzed by SDS-PAGE to ensure no degradation  
22 occurred during the experiments. The detailed analysis of the data is described in  
23 supplementary data.

## 24 **Crystallization, and structure determination**

25 Crystals of TRFH in complex with peptide RAP1-YLP <sup>89</sup>ENRERLELEAYRLGPASA<sup>106</sup>  
26 were obtained by hanging drop vapor diffusion at room temperature, with a concentration of  
27 TRFH of 34 mg/ml and a molar ratio of TRFH//Rap1-YLP of 1:3.5 in a buffer containing  
28 12.5mM Tris pH 8.0, 100mM NaCl and DTT 2.5mM. The most suitable crystals grew in 16%  
29 PEGmme 550, 50mM Hepes pH 8.0, 5mM MgCl<sub>2</sub>, 5mM DTT. Native diffraction data were  
30 collected on PX1 beamline at SOLEIL synchrotron, and reduced with XDS (40). The  
31 structure was solved by molecular replacement with PHASER (41), using one TRFH  
32 monomer from 3BUA pdb entry as model probe (14), and refined with BUSTER5 (42).



## 1 **Protein footprinting and Mass spectrometry analysis**

2 The setup of protein footprinting protocols and mass spectrometry analysis are detailed in  
3 supplementary informations.

## 4 **ITC**

5 Molecular interactions (see Supplementary Table 2) were investigated by Isothermal Titration  
6 Calorimetry using a VP-ITC and an iTC200 instruments (GE Healthcare). The solutions were  
7 degassed prior to use, at the same temperature as those used in the experiments. Integration of  
8 the peaks corresponding to each injection, correction for the baseline, and the fit were done  
9 using Origin-based software provided by the manufacturer and enabled to calculate the  
10 equilibrium dissociation constants of the interactions presented in this study. The  
11 experimental conditions of all ITC experiments are indicated in Supplementary Table 2.  
12 Blank experiments (heats of dilution from titrants into buffer) were also performed.

## 13 **SEC-MALS**

14 TRF2 protein at a concentration of 100  $\mu$ M (in a buffer containing 50 mM Tris, pH 7.5, 150  
15 mM NaCl, 10 mM 2-mercaptoethanol) was applied to Bio SEC-3 column (Agilent) on a  
16 HPLC Shimadzu system coupled to MALS/QELS/UV/RI (Wyatt Technology). Data were  
17 analysed using ASTRA software after calibration with a bovine serum albumin standard.

## 18 **Cell culture**

19 HeLa cells were kindly provided by Dr Carl Mann (I2BC), and grown in 15-cm dishes in  
20 DMEM medium supplemented with 10 % v/v fetal bovine serum and a mix of penicillin and  
21 streptomycin to prevent bacterial contamination. Cells were amplified to 30 15-cm dishes at  
22 80-90 % confluence, and finally harvested by scraping in cold PBS just prior to the  
23 preparation of nuclear extracts.

## 24 **Preparation of nuclear extracts**

25 Nuclear extracts were prepared following the protocol described in (43), with minor  
26 modifications detailed in supplementary informations.

## 27 **Streptavidin-DNA pull down experiments**

28 The following oligonucleotides were annealed to make a biotinylated duplex DNA molecule  
29 displaying a TRF2 binding site (two tandem TAGGG motifs): biotin-5'-  
30 TACTAGGGTTAGGGTTAGCA-3' and complementary 5'-

1 ATGCTAACCCCTAACCCCTAGT-3' yielding a single base 5'-overhang at each end. 9 nmoles  
2 of biotinylated DNA were loaded at 0.1 mL/min on a HiTrap Streptavidin HP 1 mL column  
3 (GE Healthcare) first equilibrated in binding buffer (Tris-HCl 20 mM pH=7.5; NaCl 150 mM;  
4 2-mercaptoethanol 5 mM; EDTA 0.2 mM; PMSF 0.1 mM). A peak in A260 indicated that the  
5 column was loaded with a saturating amount of biotinylated DNA. The column was then  
6 loaded at 0.2 mL/min with either TRF2, preformed TRF2/RAP1 complex, preformed  
7 TRF2/RAP1-ARAGA complex, or nothing (control experiment) and washed with binding  
8 buffer until A280 returned to a flat baseline. In each case, non-zero A280 of the flow through  
9 indicated that the column was loaded with a saturating amount of protein. 490 µg of nuclear  
10 extracts were loaded at 0.2 mL/min on the column. When A280 returned to a flat baseline,  
11 bound proteins were eluted at 0.5 mL/min with elution buffer (same as binding buffer except  
12 NaCl 1 M). Between each experiment, the column was washed with elution buffer until A280  
13 returned to a flat baseline, and equilibrated again in binding buffer. Eluted samples were  
14 analyzed by SDS-PAGE.

#### 15 **Model building and conservation analysis**

16 The BRCT domain of RAP1 was modeled using MODELLER software (44) and the structure  
17 BRCT domain of Rap1 from *S. cerevisiae* (45, entry 2L42). The resulting model and Myb  
18 domain of RAP1 (46) were then manually positioned along the TRFH domain of TRF2 with  
19 Pymol (47) based on the TRFH/RAP1-peptide crystal structure and the protein footprinting  
20 results. Conservation analysis was performed with CONSURF software (48) using 14  
21 sequences for both TRF2 and RAP1 from the same species of vertebrates.

22

1 **Aknowledgements :**

2 The authors wish to thank Stephen CY Ip, Stephen C West, Anne-Cécile Déclais and David  
3 MJ Lilley for kindly providing resolving enzymes or plasmids to produce them. We thank Dr  
4 Carl Mann for his help with cell culture and for the HeLa cells he provided. We are also  
5 grateful to Magali Aumont-Nicaise for her support in the use of ITC instruments at IBBMC,  
6 and to Valérie Campanacci for her help in SEC-MALS experiments. Initial discussions with  
7 Delphine Benarroch-Popivker and Bianca Sclavi were very useful for the design of the  
8 protein footprinting experiments. We are very grateful to PX1, PX2 and SWING beamlines  
9 local contacts from synchrotron SOLEIL for their help during data acquisition.

10 **Authors contributions:**

11 GG SM SP Y-VLB RT M-JG-P EG and M-HLD conceived and designed the experiments.  
12 GG SM SP RB Y-VLB CTL WM PR BG RT M-JG-P and M-HLD performed the  
13 experiments. GG SM SP RB Y-VLB CTL WM PR BG RT M-JG-P EG and M-HLD analyzed  
14 the data. GG SM SP M-JG-P EG and M-HLD wrote the manuscript.

15 **Funding**

16 This study was funded by the Agence Nationale de la Recherche (TELOLOOP grant), the  
17 Ligue contre cancer (EG équipe labellisée) the INCa (TELOCHROM grant), the Essonne  
18 committee of the Ligue contre le cancer, the foundation ARC pour la Recherche sur le  
19 Cancer, the Commissariat à l’Energie Atomique et aux Energies Alternatives (CEA), the  
20 French Infrastructure for Integrated Structural Biology (FRISBI), the Centre National de la  
21 Recherche Scientifique (CNRS), and the Institut National de la Santé et de la Recherche  
22 Médicale (INSERM).

23 **Abbreviations**

24 TRF2: Telomeric Repeat-binding Factor 2  
25 RAP1: Repressor Activator Protein 1  
26 SAXS: Small Angle X-ray Scattering  
27 ITC: Isothermal Titration Calorimetry  
28 EMSA: Electrophoretic Mobility Shift Assay  
29 SEC-MALS: Size Exclusion Chromatography coupled to Multi-Angle Light Scattering  
30 HJ: Holliday Junction  
31 TRF2-RBM: TRF2 RAP1 Binding Motif  
32 RAP1-RCT: RAP1 C-terminal TRF2 binding domain

- 1 RAP1-YLP: RAP1 peptide containing the sequence **YXLXP**
- 2 Apollo-YLP: Apollo peptide containing the sequence **YXLXP**
- 3 RAP1-ARAGA: RAP1 protein with sequence **YRLGP** mutated to **ARAGA**
- 4
- 5
- 6
- 7

1 **References :**

- 2 1- Liu D, O'Connor MS, Qin J, Songyang Z (2004) Telosome, a mammalian telomere-  
3 associated complex formed by multiple telomeric proteins. *J Biol Chem* **279**: 51338-51342
- 4 2- de Lange T (2005) Shelterin: the protein complex that shapes and safeguards human  
5 telomeres. *Genes Dev* **19**: 2100-2110
- 6 3- Songyang Z, Liu D (2006) Inside the mammalian telomere interactome: regulation and  
7 regulatory activities of telomeres. *Crit Rev Eukaryot Gene Expr* **16**: 103-18
- 8 4- Giraud-Panis M-J, Pisano S, Benarroch-Popivker D, Pei B, Le Du M-H, Gilson E (2013)  
9 One identity or more for telomeres? *Front. Oncol* **3**: 48
- 10 5- Luo Z, Dai Z, Xie X, Feng X, Liu D, Songyang Z, Xiong Y (2015) TeloPIN: a database of  
11 telomeric proteins interaction network in mammalian cells. Database (Oxford)
- 12 6- Qin, J. and Gronenborn, A.M. (2014) Weak protein complexes: challenges to study but  
13 essential for life. *FEBS Journal*, 281, 1948-1949
- 14 7- Gilson E, Géli V (2007) How telomeres are replicated. *Nature reviews. Mol Cell Biol* **8**:  
15 825–838
- 16 8- Verdun RE, Karlseder J (2007) Replication and protection of telomeres. *Nature* **447**: 924–  
17 931
- 18 9- Ye J, Lenain C, Bauwens S, Rizzo A, Saint-Léger A, Poulet A, Benarroch D, Magdinier F,  
19 Morere J, Amiard S, Verhoeyen E, Britton S, Calsou P, Salles B, Bizard A, Nadal M, Salvati  
20 E, Sabatier L, Wu Y, Biroccio A, Londoño-Vallejo A, Giraud-Panis MJ, Gilson E (2010)  
21 TRF2 and apollo cooperate with topoisomerase 2alpha to protect human telomeres from  
22 replicative damage. *Cell* **142**: 230-242
- 23 10- Biroccio A, Cherfils-Vicini J, Augereau A, Pinte S, Bauwens S, Ye J, Simonet T, Horard  
24 B, Jamet K, Cervera L, Mendez-Bermudez A, Poncet D, Grataroli R, de Rodenbeeke CT,  
25 Salvati E, Rizzo A, Zizza P, Ricoul M, Cognet C, Kuilman T, Duret H, Lépinasse F, Marvel J,  
26 Verhoeyen E, Cosset FL, Peeper D, Smyth MJ, Londoño-Vallejo A, Sabatier L, Picco V,  
27 Pages G, Scoazec JY, Stoppacciaro A, Leonetti C, Vivier E, Gilson E (2013) TRF2 inhibits a  
28 cell-extrinsic pathway through which natural killer cells eliminate cancer cells. *Nat Cell Biol*  
29 **15**: 818-828

- 1 11- Ye J, Renault VM, Jamet K, Gilson E (2014) Transcriptional outcome of telomere  
2 signalling. *Nat Rev Genet* **15**: 491-503
- 3 12- Lenain C, Bauwens S, Amiard S, Brunori M, Giraud-Panis MJ, Gilson E (2006) The  
4 Apollo 5' exonuclease functions together with TRF2 to protect telomeres from DNA repair.  
5 *Curr Biol* **16**: 1303-1310
- 6 13- van Overbeek M, de Lange T (2006) Apollo, an Artemis-Related Nuclease, Interacts with  
7 TRF2 and Protects Human Telomeres in S Phase. *Current Biology* **16**: 1295-1302
- 8 14- Fairall L, Chapman L, Moss H, de Lange T, Rhodes D (2001) Structure of the  
9 dimerization domain of the human telomeric proteins TRF1 and TRF2. *Mol Cell* **8**: 351-361
- 10 15- Chen Y, Yang Y, van Overbeek M, Donigian JR, Baciou P, de Lange T, Lei M (2008) A  
11 shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric  
12 proteins. *Science* **22**: 1092-1096
- 13 16- Wan B, Yin J, Horvath K, Sarkar J, Chen Y, Wu J, Wan K, Lu J, Gu P, Yu EY, Lue NF,  
14 Chang S, Liu Y, Lei M (2013) SLX4 assembles a telomere maintenance toolkit by bridging  
15 multiple endonucleases with telomeres. *Cell Rep* **4**: 861-869.
- 16 17- Kim H, Lee OH, Xin H, Chen LY, Qin J, Chae HK, Lin SY, Safari A, Liu D, Songyang Z  
17 (2009) TRF2 functions as a protein hub and regulates telomere maintenance by recognizing  
18 specific peptide motifs. *Nat Struct Mol Biol* **16**: 372-379
- 19 18- Billaud T, Brun C, Ancelin K, Koering CE, Laroche T, Gilson E (1997) Telomeric  
20 localization of TRF2, a novel human telobox protein. *Nat Genet* **17**: 236-239
- 21 19- Court R, Chapman L, Fairall L, Rhodes D (2005) How the human telomeric proteins  
22 TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures.  
23 *EMBO reports* **6**: 39-45
- 24 20- Palm W, de Lange T (2008) How shelterin protects mammalian telomeres. *Annu Rev*  
25 *Genet* **42**: 301-334
- 26 21- Poulet A, Buisson R, Faivre-Moskalenko C, Koelblen M, Amiard S, Montel F, Cuesta-  
27 Lopez S, Bornet O, Guerlesquin F, Godet T, Moukhtar J, Argoul F, Déclais AC, Lilley DM,  
28 Ip SC, West SC, Gilson E, Giraud-Panis MJ (2009) TRF2 promotes, remodels and protects  
29 telomeric Holliday junctions. *EMBO J* **28**: 641-651

- 1 22- Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM (2009) TERRA RNA  
2 binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres.  
3 *Mol Cell* **35**: 403-413
- 4 23- Baker AM, Fu Q, Hayward W, Victoria S, Pedroso IM, Lindsay SM, Fletcher TM (2011)  
5 The telomere binding protein TRF2 induces chromatin compaction. *PLoS One* **6**: e19124
- 6 24- Poulet A, Pisano S, Faivre-Moskalenko C, Pei B, Tauran Y, Haftek-Terreau Z, Brunet F,  
7 Le Bihan YV, Ledu MH, Montel F, Hugo N, Amiard S, Argoul F, Chaboud A, Gilson E,  
8 Giraud-Panis MJ (2012) The N-terminal domains of TRF1 and TRF2 regulate their ability to  
9 condense telomeric DNA. *Nucleic Acids Res* **40**: 2566-2576
- 10 25- Chen Y, Rai R, Zhou ZR, Kanoh J, Ribeyre C, Yang Y, Zheng H, Damay P, Wang F,  
11 Tsujii H, Hiraoka Y, Shore D, Hu HY, Chang S, Lei M (2011) A conserved motif within  
12 RAP1 has diversified roles in telomere protection and regulation in different organisms. *Nat*  
13 *Struct Mol Biol* **18**:213-221
- 14 26- Li B, Oestreich S, de Lange T (2000) Identification of human Rap1: implications for  
15 telomere evolution. *Cell* **101**: 471-483
- 16 27- Bae NS, Baumann P (2007) A RAP1/TRF2 complex inhibits nonhomologous end-joining  
17 at human telomeric DNA ends. *Mol Cell* **26**: 323–334
- 18 28- Kabir S, Sfeir A, de Lange T (2010) Taking apart Rap1: an adaptor protein with telomeric  
19 and non-telomeric functions. *Cell Cycle* **9**: 4061–4067
- 20 29- Kabir S, Hockemeyer D, de Lange T (2014) TALEN gene knockouts reveal no  
21 requirement for the conserved human shelterin protein Rap1 in telomere protection and length  
22 regulation. *Cell Rep* **9**: 1273–1280
- 23 30- Martinez P, Thanasoula M, Carlo AR, Gómez-López G, Tejera AM, Schoeftner S,  
24 Dominguez O, Pisano DG, Tarsounas M, Blasco MA (2010) Mammalian Rap1 controls  
25 telomere function and gene expression through binding to telomeric and extratelomeric sites.  
26 *Nature Cell Biol* **12**: 768-780
- 27 31- Yang D, Xiong Y, Kim H, He Q, Li Y, Chen R, Songyang Z (2011) Human telomeric  
28 proteins occupy selective interstitial sites. *Cell Research* **21**: 1013-1027
- 29 32- Teo H, Ghosh S, Luesch H, Ghosh A, Tsin Wong E, Malik N, Orth A, de Jesus A, Oliver  
30 JD, Tran NL, Speiser LJ, Wong M, Saez E, Schultz P, Chanda SK, Verma IM, nd Tergaonkar

- 1 V (2010) Telomere-independent Rap1 is an IKK adaptor and regulates NF -  $\kappa$  B - dependent  
2 gene expression. *Nature Cell Biol* **12**: 758-767
- 3 33- Takai KK, Hooper S, Blackwood S, Gandhi R, de Lange T (2010) In vivo stoichiometry  
4 of shelterin components. *J Biol Chem* **285**: 1457–1467
- 5 34- Janoušková E, Nečasová I, Pavloušková J, Zimmermann M, Hluchý M, Marini V,  
6 Nováková M, Hofr C (2015) Human Rap1 modulates TRF2 attraction to telomeric DNA.  
7 *Nucleic Acids Res* **43**: 2691-2700
- 8 35- Amiard S, Doudeau M, Pinte S, Poulet A, Lenain C, Faivre-Moskalenko C, Angelov D,  
9 Hug N, Vindigni A, Bouvet P, Paoletti J, Gilson E, Giraud-Panis MJ (2007) A topological  
10 mechanism for TRF2-enhanced strand invasion. *Nat Struct Mol Biol* **14**: 147-54
- 11 36- Rambo RP & Tainer JA (2010) Bridging the solution divide: comprehensive structural  
12 analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering.  
13 *Current Opinion in Structural Biology* **20**: 128–137
- 14 37- Durand D, Vivès C, Cannella D, Perez J, Pebay-Peyroula E, Vachette P, Fieschi F (2010)  
15 NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-  
16 flexible linkers. *J Struct Biol* **169**: 45-53
- 17 38- Gerloff DL, Woods NT, Farago A, Monteiro ANA (2012) BRCT domains: a little more  
18 than kin, and less than kind. *FEBS Lett* **586**: 2711-2716
- 19 39- David G, Pérez J (2009) Combined sampler robot and HPLC: a fully automated system  
20 for biological SAXS experiments at the Synchrotron SOLEIL Swing beamline. *J Appl Cryst*  
21 **42**: 892-900
- 22 40- Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta*  
23 *Crystallogr D Biol Crystallogr* **66**: 133-144
- 24 41- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007)  
25 Phaser crystallographic software. *J Appl Cryst* **40**: 658-674
- 26 42- Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Smart OS,  
27 Vonrhein C, Womack TO (2009) BUSTER, version 2.8.0. Cambridge, United Kingdom:  
28 Global Phasing Ltd



- 1 43- Abmayr SM, Yao T, Parmely T, Workman JL (2001) Preparation of Nuclear and  
2 Cytoplasmic Extracts from Mammalian Cells. In *Current Protocols in Molecular Biology*  
3 John Wiley & Sons, Inc
- 4 44- Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Protein structure modeling with  
5 MODELLER. *Methods Mol Biol* **426**: 145–159
- 6 45- Zhang W, Zhang J, Zhang X, Xu C, Tu X (2011) Solution structure of Rap1 BRCT  
7 domain from *Saccharomyces cerevisiae* reveals a novel fold. *Biochem Biophys Res Commun*  
8 **404**, 1055-1059.
- 9 46- Hanaoka S, Nagadoi A, Yoshimura S, Aimoto S, Li B, de Lange T, Nishimura Y (2001)  
10 NMR structure of the hRap1 Myb motif reveals a canonical three-helix bundle lacking the  
11 positive surface charge typical of Myb DNA-binding domains. *J Mol Biol* **312**: 167-175
- 12 47- DeLano W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San  
13 Carlos, CA, USA. <http://www.pymol.org>
- 14 48- Ashkenazy H., Erez E., Martz E., Pupko T. and Ben-Tal N. (2010) ConSurf 2010:  
15 calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.  
16 *Nucl. Acids Res.* Vol. 38, Web Server issue W529–W533.
- 17  
18  
19  
20

1 **Table 1: X-ray diffraction data collection and refinement statistics.**

2

Wavelength (Å)	0.98
Space-group	P2 <sub>1</sub>
Diffraction limits (last shell)	2.44 Å (2.63 Å – 2.44 Å)
Unit cells (axbxcx□x□x□)	53.8x104.8x85.3 90×94.4×90
R <sub>merge</sub>	0.068 (0.664)
Number of unique reflections	33979 (5718)
I/□	14,29 (2.26)
Completeness	0.965 (0.794)
Wilson B-factor (Å <sup>2</sup> )	51.6
<b>Molecular Replacement (LLG)</b>	1391
<b>Refinement</b>	
Resolution	2.44 Å
R <sub>work</sub>	0.193
R <sub>free</sub>	0.237
Number of non-solvent atoms	6938
Number of water molecules	94
Figure of merit	0.7813
RMSD bond lengths	0.004
RMSD bond angles	0.976
RMSD chirality	15.25
PDB entry	4RQI

3 LLG: Log of Likelihood Gain

4

1 **Figures legends**

2 **Figure 1: RAP1 affects TRF2 binding properties on linear and branched DNA.**

3 **A:** 5 nM of tHJ labelled on strand indicated at the top was incubated with 200 nM of TRF2  
4 with and without increasing amount of RAP1 prior to cleavage with GEN1[1-527].  
5 Concentrations of RAP1 used were 100, 200, 400, 1000, 1500 nM (experiment on the left) or  
6 at 1500 nM for the other three experiments.

7 **B:** 5 nM of tHJ labelled on strand 4 was incubated with increasing amounts of TRF2<sup>ΔL</sup> prior  
8 to cleavage with either Endonuclease I, CCE1 or RusA. Concentrations of TRF2<sup>ΔL</sup> used were  
9 10, 20, 50, 100, 200, 500, 1000 nM. In lanes 3, 12 and 21 the junction is incubated with 200  
10 nM of TRF2 prior to cleavage. In lanes 29 and 30 the junction is incubated with 1000 nM of  
11 TRF2 and TRF2<sup>ΔL</sup>.

12 **C:** 5 nM of tHJ labelled on strand 4 was incubated with 200 nM of TRF2<sup>ΔL</sup> with and without  
13 increasing amount of RAP1 prior to cleavage with either Endonuclease I, CCE1 or RusA.  
14 Concentrations of RAP1 used were 100, 200, 400, 1000, 1500 nM. In lanes 3, 12 and 21 the  
15 junction is incubated with 1500 nM of RAP1 prior to cleavage. In lanes 4, 13 and 22 the  
16 junction is incubated with 200 nM of TRF2 prior to cleavage. In lanes 29, 30 and 31 the  
17 junction is incubated with 1500 nM of RAP1, 200 nM of TRF2 and 200 nM of TRF2<sup>ΔL</sup>.

18 **D:** 2D probability maps calculated from AFM experiments showing the binding of TRF2 (10  
19 nM of dimers) on a 650 bp telomeric DNA (10 nM) either alone (top panel) or in the presence  
20 of 40 nM (monomer) of either full-length RAP1 (middle panel) or RAP1-RCT (bottom  
21 panel). Yellow dotted lines show the contour length of unbound DNA. Green and red dotted  
22 lines show the contour length and volume of the main TRF2-DNA complex respectively.  
23 Color scale of the probability on the right of the maps.

24 **Figure 2: Formation of TRF2 complex with RAP1 is associated to conformational**  
25 **adjustment and biphasic signal of interaction.**

26 **A:** Domains organisation of RAP1 and TRF2. The blue arrow shows the limit of RAP1-RCT  
27 construct.

28 **B:** Table of theoretical molecular weight (MW<sub>th</sub>), giration radius (R<sub>g</sub>), maximal distances  
29 (D<sub>max</sub>) and experimental molecular weight (MW<sub>exp</sub>) from Guinier calculation performed  
30 with each corresponding SAXS curve.

1 **C–E:** *Ab initio* averaged envelopes of TRF2 (**C**), RAP1 (**D**), and TRF2/RAP1 (**E**) with  
2 superimposed globular domains as references.

3 **F–G:** ITC characterization of the interaction between TRF2 and RAP1 (**F**) or RAP1-RCT (**G**).

4 **Figure 3: Complete interaction of RAP1 with TRF2 involves YXLXP motif and TRF2**  
5 **dimerization domain.**

6 **A:** Superimposition of the four TRFH monomers observed in the crystal asymmetric unit  
7 (from pale to dark yellow) and of the four corresponding RAP1-YLP peptides (from pale to  
8 dark cyan).

9 **B:** Superimposition of TRFH (yellow) in complex with Apollo-YLP (dark purple) or RAP1-  
10 YLP peptide (cyan). The region of TRFH involved in the interaction with Apollo-YLP is  
11 shown in violet, and with RAP1-YLP in light cyan. The rms deviation calculated on C-alpha  
12 of each of the four TRFH monomer / RAP1-YLP and of TRFH monomer / Apollo-YLP is  
13 between 0.54 and 0.88 Å.

14 **C:** ITC characterization of the interaction between TRF2 with RAP1-ARAGA.

15 **Figure 4: Complete interaction of RAP1 with TRF2 involves large area from RAP1 N-**  
16 **terminal moiety and TRF2 dimerization domain.**

17 **A–B:** Acetylation profile of lysine residues in TRF2, RAP1 or RAP1-ARAGA alone,  
18 compared to TRF2/RAP1 or TRF2/RAP1-ARAGA. Lysine residues are represented in small  
19 sphere and coloured differently if they are protected (blue, or light blue if the precise  
20 acetylation position was not determined), more accessible (red), or not affected (black) in  
21 both TRF2/RAP1 and TRF2/RAP1-ARAGA complex, or protected (green) or more  
22 accessible (orange) only in TRF2/RAP1 complex. Residues Y-L-P are shown as purple  
23 spheres. The inserted tables show the lysine number in the sequence, highlighted in blue in  
24 RAP1 globular domain, in yellow in TRF2 globular domain. The acetylation profile follows  
25 the same colour code as in the structure representation. **A:** TRFH/RAP1-YLP peptide crystal  
26 structure; **B:** Juxtaposition of RAP1[1-121] 3D model calculated with Modeller (44), and  
27 RAP1-Myb NMR structure (PDB entry 1FEX).

28 **C:** Positioning of RAP1 BRCT and Myb domains on TRFH surface based on combination of  
29 the crystal structure and footprinting experiments, showing on the right side the induced  
30 dimerization of BRCT domain (colour code identical than in figure 3).

1 **D:** Surface representation of RAP1[1-208] coloured upon the conservation score of  
2 CONSURF software (47) with the trace of the TRFH domain in black shadow (left) ; Surface  
3 representation of TRFH domain of TRF2 coloured upon the conservation score of CONSURF  
4 software (47) with the RAP1-YLP peptide in black (right). All structure representations were  
5 drawn with software Pymol (46).

6 **Figure 5: Interaction of RAP1 with TRF2 is associated to different binding properties.**

7 **A:** 5 nM of tHJ labelled on strand 4 was incubated with 200 nM of TRF2 with and without  
8 increasing amount of RAP1-ARAGA or RAP1[1-208] prior to cleavage with CCE1.  
9 Concentrations of RAP1-ARAGA used were 200, 400, 1000, 1500, 3000 nM (lanes 8-12, and  
10 3000 nM in lanes 4 and 20. Concentrations of RAP1[1-208] used were 100, 200, 400, 1000,  
11 1500 nM (lanes 13-17), and 1500 nM in lanes 5 and 21. Concentration of RAP1-wt in lanes 3,  
12 7 and 19 is 1500 nM. In lanes 3, 12 and 21 the junction is incubated with 1500 nM of RAP1  
13 prior to cleavage.

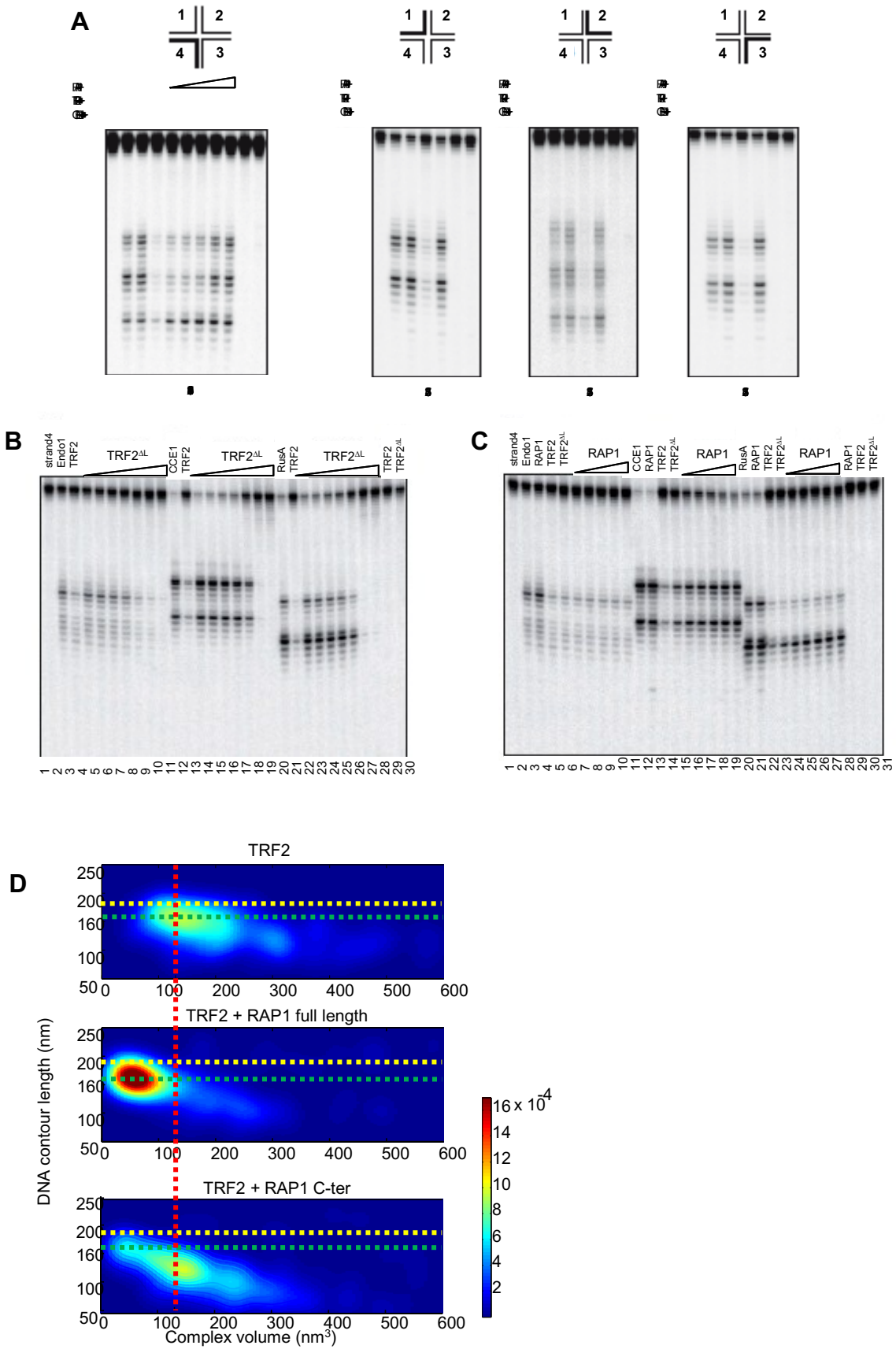
14 **B:** Left panel, SDS-PAGE analysis of a typical streptavidin-DNA pulldown experiment  
15 (Coomassie blue stained). Lanes contain: 1) input nuclear extract, 2) molecular weight marker  
16 (Weights from top to bottom: 250, 150, 100, 80, 60 kDa); 3) input purified TRF2, 4) nuclear  
17 extract flow through, 5) elution of TRF2-bound proteins; 6) input purified TRF2/RAP1  
18 complex, 7) nuclear extract flow through, 8) elution of TRF2/RAP1-bound proteins; 9) input  
19 purified TRF2/RAP1-ARAGA complex, 10) nuclear extract flow through, 11) elution of  
20 TRF2/RAP1-ARAGA-bound proteins; 12) nuclear extract flow through in absence of  
21 telomeric proteins, 13) elution of DNA-bound proteins (control). Right panel, enlarged view  
22 of lanes 5, 8 and 11 of the SDS-PAGE.

23 **Figure 6: Model of binding properties of TRF2/RAP1 assembly**

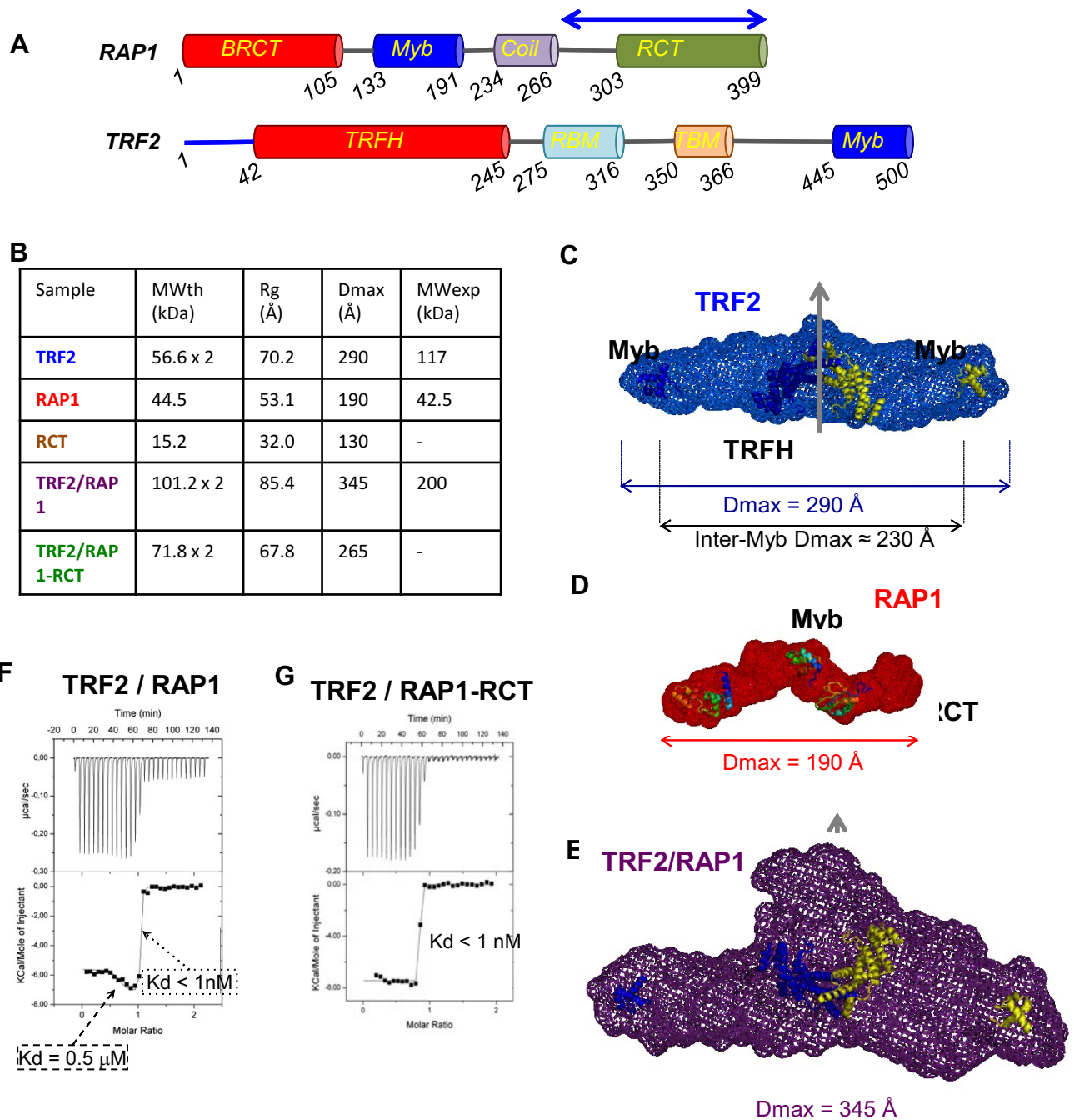
24 Proposed model of the assembly of TRF2/RAP1 and the effect of its malleability on its  
25 binding properties. The presence of RAP1 (cyan) prevents the binding of non-specific  
26 proteins like FLP containing proteins (orange) in the TBM area of TRF2 (yellow), but does  
27 not prevent the binding of high-affinity YLP containing proteins (purple). In the absence of  
28 YLP containing proteins, RAP1 complete interaction induces the formation of BRCT dimer  
29 (blue circle) that could recruit new phosphorylated partners (green).

30

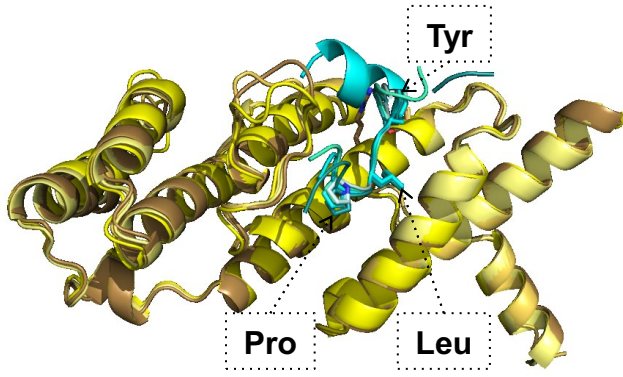
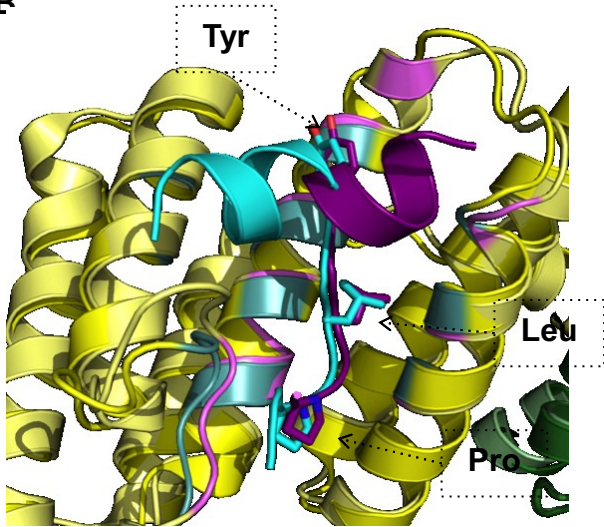
31



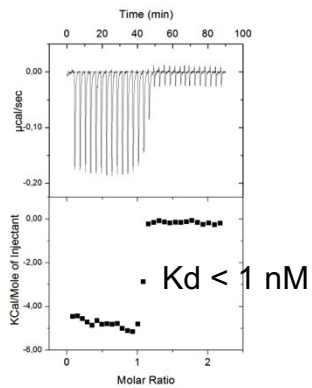
**Figure 1: RAP1 affects TRF2 binding properties on linear and branched DNA.**



**Figure 2: Formation of RAP1/TRF2 complex is associated to conformational adjustment and biphasic signal of interaction .**

**A****B****C**

TRF2 / RAP1-ARAGA



**Figure 3: Complete interaction of RAP1 with TRF2 involves YXLXP motif and TRF2 dimerization domain.**



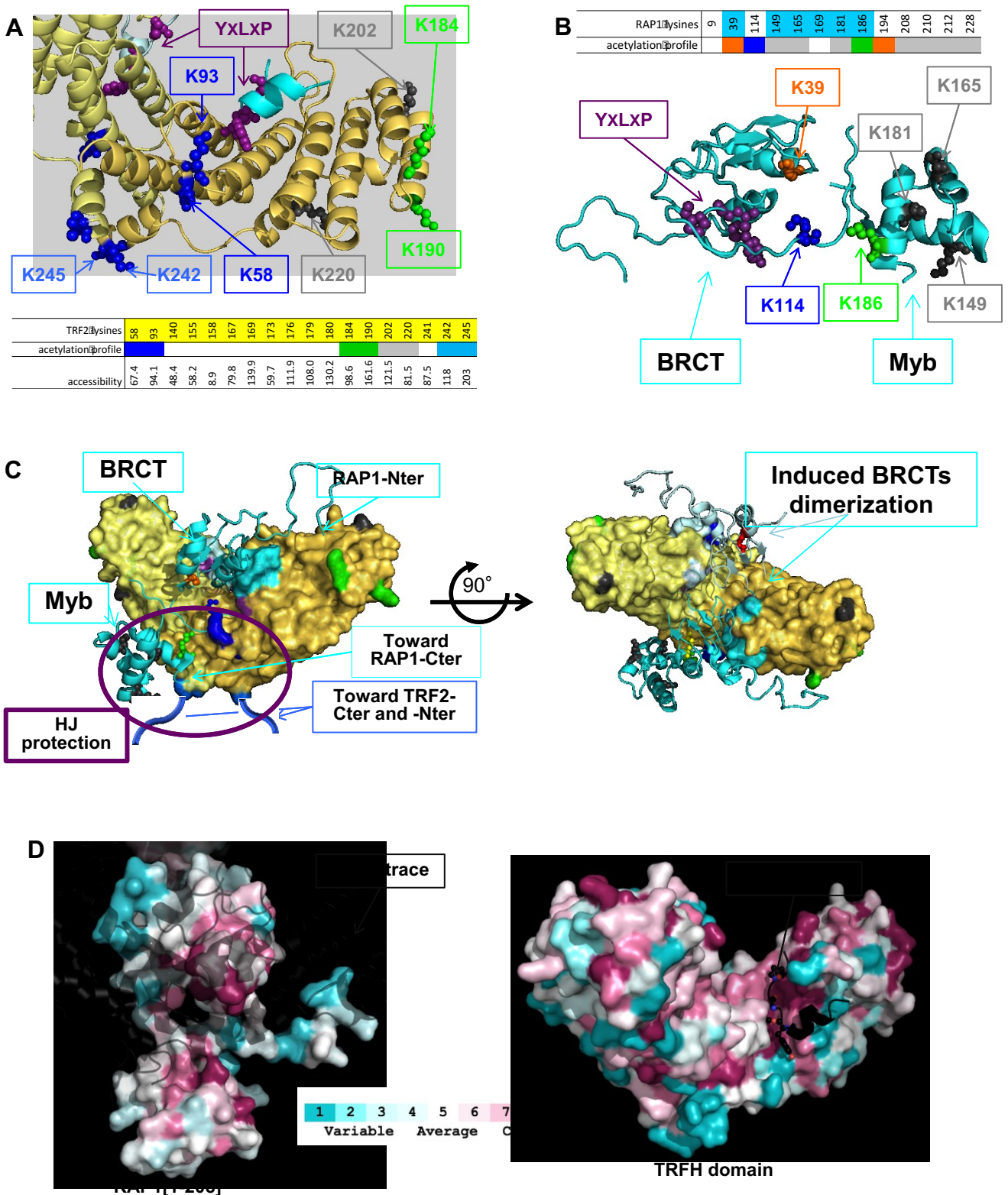
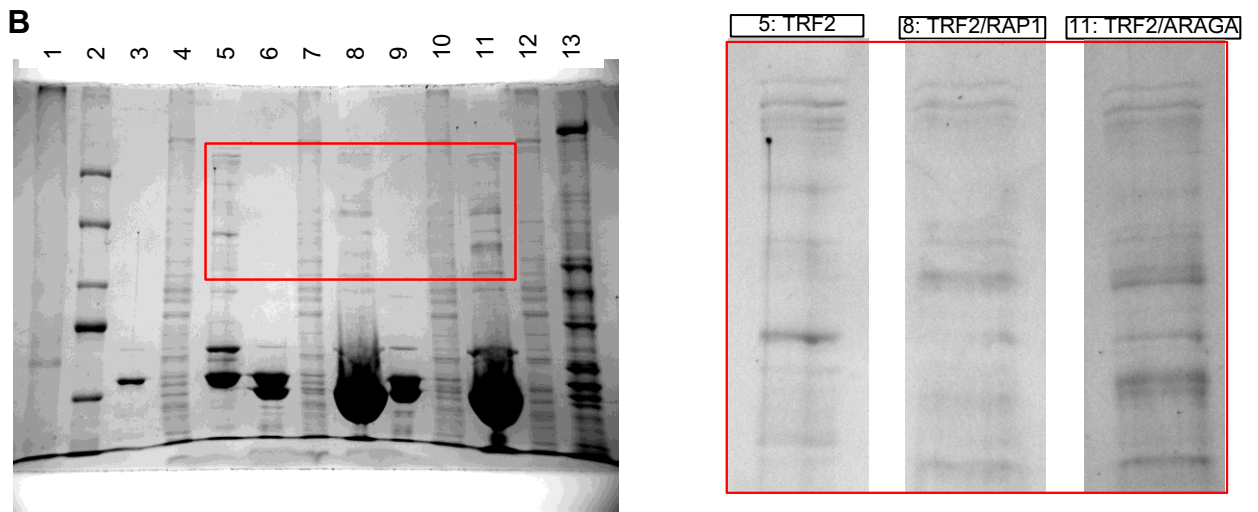
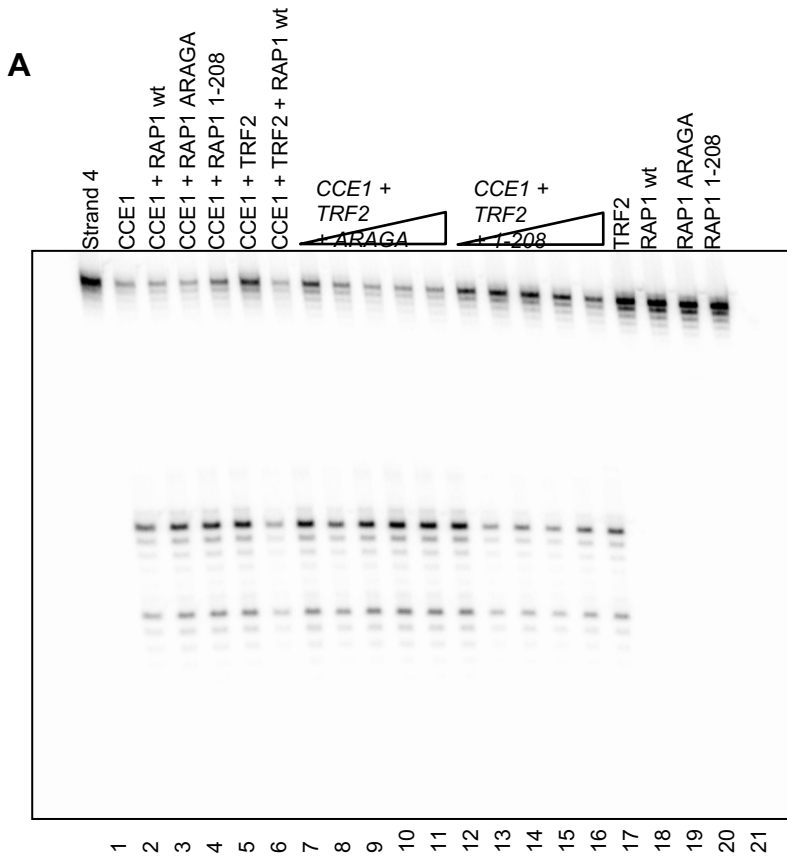


Figure 4: Complete interaction of RAP1 with TRF2 involves large area from RAP1 N-terminal moiety and TRF2 dimerization domain.



**Figure 5: Interaction of RAP1 with TRF2 is associated to different binding properties.**

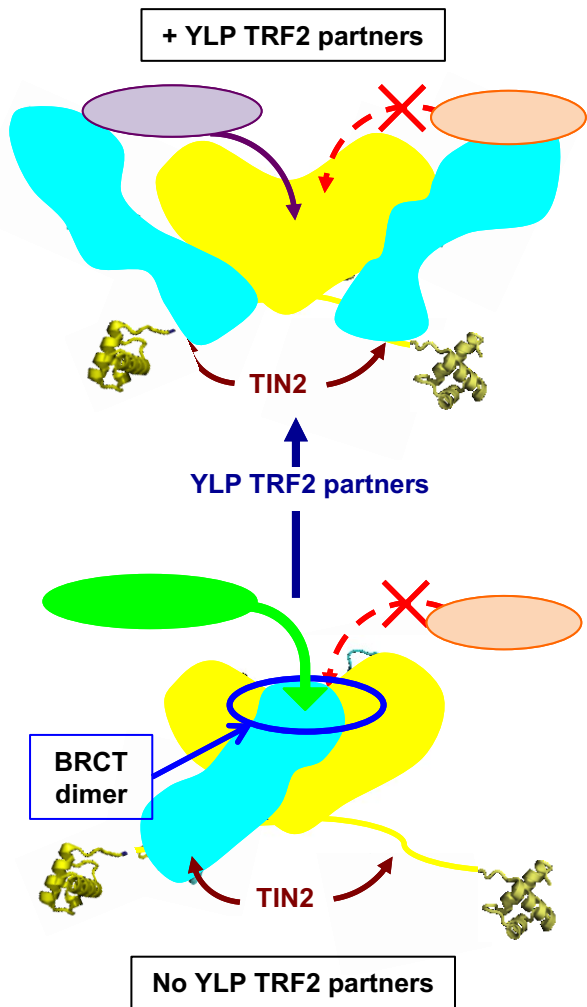


Figure 6: Model of binding properties of TRF2/RAP1 assembly.

## Supplementary information

### Protein expression

All proteins were expressed in *E. coli*, either in BL21DE3-star or in Rosetta strain (see table “Protein constructs used in this study” for details). After transformation by heat shock and selection of recombinant colonies with appropriate antibiotic (see table), bacteria were grown in LB medium + appropriate antibiotic in a shaking incubator at 37 °C until cultures reached an OD<sub>600</sub> ~0.6. Cultures were then cooled to 20 °C, expression was induced by addition of IPTG qsp 0.5 mM final concentration in medium when OD<sub>600</sub> reached ~0.8-1, and bacteria were further grown overnight (except for construct E, which was induced for 3 hours at 37°C).

Bacteria were harvested by centrifugation at 3,000 g at 4 °C for 30 min, LB medium was discarded and pellets were washed in cold PBS, centrifuged again and stored dry at -80 °C after flash-freezing in liquid nitrogen.

### Protein purification

For all proteins, cell pellets were thawed and suspended in 40 mL cold buffer A + Triton-X-100 0.5% v/v + PMSF 1 mM. Suspensions were treated by lysozyme (20 mg for one pellet) at room temperature for 15 min under gentle agitation, and then submitted to sonication at 4 °C (at least 4 cycles of 1 min with pulses of 1 second interspersed with 1 second pause, at 40, 50, 60 and 70% power of a Vibra-Cell 75042 sonicator). Lysates were then supplemented with MgCl<sub>2</sub> qsp 1.5 mM final concentration and treated by Benzonase nuclease (1 µL for one pellet) at room temperature for 15 min. Finally, lysates were centrifuged at 48,000 g at 4 °C for 30 min and resulting supernatants were passed through 0.22 µm filters before application on chromatography column.

The following purification strategy was followed for most protein constructs (specific details for some constructs are provided in next paragraphs): all steps were performed at 4 °C (cold room) or with buffers and sample on ice; filtered supernatant from last step of lysis was immediately applied on a HisTrap HP 5 mL column (GE Healthcare) previously regenerated with Ni<sup>2+</sup> ions and equilibrated in buffer A, loaded column was washed with several volumes of buffer A (until baseline equilibrated) and elution was performed with a linear gradient of 0 to 100% v/v buffer B over 20 column volumes. Fractions containing the protein of interest (as seen by SDS-PAGE) were pooled and set in buffer D (or in PBS for RAP1<sub>[270-399]</sub> construct E) through a HiPrep 26/10 Desalting column (GE Healthcare) to ensure full activity of protease for tag cleavage. Appropriate protease (either thrombin or TEV, see table “Protein constructs used in this study”) was added to the protein solution to a ratio of 2 to 4% w/w and the mix was incubated overnight at 4 °C. Next day, protein solution was adjusted to 300 mM NaCl (400 mM for RAP1 constructs B and C because pI < 5 gives stronger binding to Ni<sup>2+</sup>) and 20 mM imidazole and run through the HisTrap column previously stripped, Ni<sup>2+</sup>-regenerated and equilibrated in buffer C; flow through containing the tag-free protein was collected. Protein solution was finally set in buffer D through the Desalting column, glycerol was added to a final concentration of 5% v/v, and protein was stored as 1 mL aliquots at -80 °C after flash-freezing in liquid nitrogen.

For TRF2, an additional step of affinity chromatography on Heparin HP 5 mL column (GE Healthcare) was performed after the second HisTrap to ensure no DNA from the bacteria remained bound to the protein. For this step, the protein solution was adjusted to 100 mM NaCl before application on the column previously equilibrated in buffer E, and elution was performed with a linear gradient of 0 to 100% v/v buffer F over 20 column volumes. The migration profile of purified TRF2 during gel filtration reveals a first high peak and a

secondary peak as previously observed (Bae & Bauman, 2007, Amiard et al., 2007). SEC-MALS coupled to SDS-PAGE analysis and ITC measurements show that the first peak corresponds to highly pure TRF2 full-length homodimer with an experimental molecular weight of 114.5 kDa (theoretical expected molecular weight = 113.2 kDa, Sup-Figure 1A and 1B), compatible with SAXS analysis.

For RAP1 constructs A and D, an additional step of size exclusion chromatography was performed after the first HisTrap, on Superdex 200 and Superdex 75 columns (GE Healthcare), respectively, equilibrated in buffer G (high salt buffer, to separate potentially interacting contaminants). For ARAGA mutant Circular dichroism spectra in far-UV and temperature denaturation show that these three mutations do not affect the secondary and global structure of RAP1 (Supplementary Figure 7A-7C).

Prior to any experiment that required thawing of stored aliquots, all proteins were subjected to size exclusion chromatography on Superdex 200 column in order to remove eventual aggregates and to set the appropriate buffer for the experiment. It was of particular importance to avoid the presence of degradation fragments in experiments (TRF2 and RAP1 are prone to degradation due to long unfolded regions).

#### References:

Amiard S, Doudeau M, Pinte S, Poulet A, Lenain C, Faivre-Moskalenko C, Angelov D, Hug N, Vindigni A, Bouvet P, Paoletti J, Gilson E, Giraud-Panis MJ (2007) A topological mechanism for TRF2-enhanced strand invasion. *Nat Struct Mol Biol* **14**: 147-154.

Bae NS, Baumann P (2007) A RAP1/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends. *Mol Cell* **26**: 323-334.

#### Kinetics of RAP1 limited trypsin proteolysis

RAP1 (construct A) was dialyzed against the following buffer: 20 mM HEPES pH = 7.5, 150 mM NaCl. 45 µg of RAP1 were mixed with trypsin at a ratio of 1/10,000 w/w and incubated at 20 °C. Reaction content at 0, 30 min, 1h, 1h30, 2h, 3h, 4h, 5h, 6h, 7h, 8h, 23h, 24h and 25h was analyzed by SDS-PAGE.

#### Circular dichroism spectroscopy

RAP1 WT (construct A), RAP1 ARAGA (construct C) and TRF2 (construct E) were dialyzed against the following buffer: 9 mM Tris pH = 7.5, 300 mM NaF. For each protein, a concentration series of 3, 6 and 15 µM was used for spectra acquisition. Spectra were measured with a Jasco J-815 CD Spectrometer instrument. Each spectrum results from an accumulation of 20 acquisitions, scanning from 250 to 190 nm by steps of 0.1 nm and a bandwidth of 2 nm. Protein spectra were subtracted with buffer spectrum. To be able to compare spectra between the different proteins, they were expressed in molar ellipticity per residue according to the following formula:

$$[\theta] = \frac{\theta}{10 \times c \times l \times n}$$

where  $[\theta]$  is the molar ellipticity per residue ( $\text{deg.cm}^2.\text{dmol}^{-1}.\text{residue}^{-1}$ ),  $\theta$  is the raw CD data from the spectrometer (mdeg),  $c$  is the molar concentration of the protein sample ( $\text{mol.L}^{-1}$ ),  $l$  is the optical path (cm) and  $n$  is the number of residues in the protein. Spectra were also scaled so that their value at 209.5 nm (inflexion point) coincided. This allowed avoiding discrepancies in spectra amplitudes due to uncertainties in protein concentration determination (measured by absorbance at 280 nm with a Nanodrop instrument, which is less sensitive than the CD spectrometer).

Thermal denaturation was followed by measuring the CD signal at 222 nm (alpha helix main signal) while heating the samples from 10 to 95 °C by 1 °C steps. Thermal denaturation data were fitted using SciDAVis software (<http://scidavis.sourceforge.net/>) to determine  $T_m$  values.

### SAXS analysis

We performed Small Angle X-ray Scattering analysis and calculated *ab initio* envelopes that represent the statistical distribution of the three-dimensional conformation of the molecule in solution, including flexible regions (Koch et al., 2003). The data were analyzed using FOXTROT (from SWING beamline) and PRIMUS (Primary Analysis & Manipulations with Small Angle Scattering DATA) from ATSAS 2.1 (Konarev et al., 2006), from which Guinier and normalized Kratky plots (Durand et al., 2010) were generated (Figure 1, Supplementary Figure 5). From the corrected scattering curves, the pair-distribution functions were computed using GNOM (Svergun, 1992), leading to the radius of gyration  $R_g$ , and the maximal distance  $D_{max}$ . The molecular weights of the particles were derived from the extrapolated intensity at the origin  $I_0$  (Figure 4B). In order to calculate reliable *ab initio* averaged envelopes, we generated 100 low-resolution *ab initio* shape envelopes using GASBOR (Svergun, 1999). The best 50 envelopes were selected based on Chi2 factor. After superposition of the envelopes, files were selected based on the double criteria:

- $\chi^2 < \chi^2_{min} + 1 \text{ ecart.standard}(\chi^2)$
- $NSD < NSD_{min} + 1 \text{ average deviation}$  (NSD= Normalized Spatial Discrepancy);

The selected files were averaged with DAMAVER. From this approach, we do not necessary average more envelopes than from the classical approach (8 to 10 envelopes), but the tight selection criteria improved the accuracy of the final envelopes.

### References:

- Koch MH, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* **36**: 147–227
- Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006) ATSAS 2.1, a program package for Small-angle scattering data analysis. *J Appl Cryst* **39**: 277-286.
- Durand D, Vivès C, Cannella D, Perez J, Pebay-Peyroula E, Vachette P, Fieschi F (2010) NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J Struct Biol* **169**: 45-53.
- Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* **25**: 495-503.
- Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**: 2879-2886.

### Setup of protein footprinting protocol

Protocol was adapted from Delphine Benarroch-Popivker method (personal communication) as well as methods reported in the following references to our specific study of protein / protein interactions. Therefore, we had to take into account the number of lysines present in TRF2 and RAP1 and use an excess of one protein to ensure the other one is engaged in complexes at 100%. TRF2 (construct F), RAP1 (construct B) and RAP1-ARAGA (construct C) were set in buffer H by size exclusion chromatography (Superdex 200 10/300 GL, GE Healthcare).

The following mixtures were then prepared to a final volume of 20  $\mu$ L:

- free proteins: TRF2 at 20  $\mu$ M (monomers) +/- sulfo-NHS-acetate at 4.075 mM (same absolute concentration as in the condition where we look at TRF2 engaged in the complex) ; RAP1 or ARAGA at 20  $\mu$ M +/- sulfo-NHS-acetate at 7.85 mM (same absolute concentration as in the condition where we look at RAP1 or ARAGA engaged in the complex)

- complexes: TRF2 at 20  $\mu$ M (monomers) + RAP1 or ARAGA at 30  $\mu$ M (1.5-fold molar excess versus TRF2) + sulfo-NHS-acetate at 4.075 mM (2.5-fold molar excess versus total number of lysines in this condition) ; RAP1 or ARAGA at 20  $\mu$ M + TRF2 at 30  $\mu$ M (monomers, 1.5-fold molar excess versus RAP1 or ARAGA) + sulfo-NHS-acetate at 7.85 mM (2.5-fold molar excess versus total number of lysines in this condition).

Before addition of sulfo-NHS-acetate, all preparations were incubated 10 min at 20 °C to allow complex formation. Excess of one protein was necessary to ensure that the other one is engaged in complexes at 100%. Sulfo-NHS-acetate was dissolved in water shortly before use, as advised by the manufacturer (Thermo Scientific). Control experiments were also performed for free proteins by adding water instead of sulfo-NHS-acetate. Reactions were incubated 30 min at 20 °C, and finally quenched by addition of 2  $\mu$ L of 1M Tris buffer pH = 7.5 to consume the remaining excess of free sulfo-NHS-acetate. Samples were then analyzed by SDS-PAGE, and excised bands were analyzed by mass spectrometry PMF (peptide mass fingerprinting) after protease digestion.

#### *References:*

-Hassani O, Mansuelle P, Cestèle S, Bourdeaux M, Rochat H, Sampieri F (1999) Role of lysine and tryptophan residues in the biological activity of toxin VII (Ts  $\gamma$ ) from the scorpion *Tityus serrulatus*. *Eur J Biochem* **260**: 76–86

-Kvaratskhelia M, Miller JT, Budihis SR, Pannell LK, Grice SFJL (2002) Identification of specific HIV-1 reverse transcriptase contacts to the viral RNA:tRNA complex by mass spectrometry and a primary amine selective reagent. *PNAS* **99**: 15988–15993

-Liu Y, Kvaratskhelia M, Hess S, Qu Y, Zou Y (2005) Modulation of Replication Protein A Function by Its Hyperphosphorylation-induced Conformational Change Involving DNA Binding Domain B. *J Biol Chem* **280**: 32775–32783

-Mendoza VL, Vachet RW (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrometry Reviews* **28**: 785–815

-Nuss JE, Sweeney DJ, Alter GM (2006) Reactivity-Based Analysis of Domain Structures in Native Replication Protein A. *Biochemistry* **45**: 9804–9818

-Wang X, Kim S-H, Ablonczy Z, Crouch RK, Knapp DR (2004) Probing Rhodopsin–Transducin Interactions by Surface Modification and Mass Spectrometry. *Biochemistry* **43**: 11153–11162

#### **Mass spectrometry analysis of footprinting products**

Protein bands excised from SDS-PAGE gels were washed twice with 500  $\mu$ L ammonium bicarbonate 100 mM pH = 8, washed again twice with 500  $\mu$ L ammonium bicarbonate 50 mM pH = 8 / acetonitrile 50%, partially dried by lyophilization, and rehydrated by 20  $\mu$ L ammonium bicarbonate 100 mM pH = 8.

All bands were digested by trypsin at 10 ng/ $\mu$ L in ammonium bicarbonate 100 mM pH = 8, incubated 2h at 50 °C and under agitation. RAP1 and ARAGA replicate bands were also digested by V8 protease at 10 ng/ $\mu$ L in ammonium bicarbonate 100 mM pH = 8, incubated overnight at 30 °C and under agitation. Proteolysis reactions were stopped by addition of TFA at 1% v/v final concentration.

0.5  $\mu$ L of each sample were directly spotted on MALDI plates with 0.5  $\mu$ L 4-HCCA matrix at 10 mg/mL in acetonitrile 50% / water 49.9% / TFA 0.1%. For MS/MS acquisitions, samples were concentrated by MicroZipTip C18 before spotting.

MS data were acquired on a 4800 MALDI-TOF/TOF instrument (ABSciex, Foster City, USA). Spectra were analyzed with Mascot (<http://www.matrixscience.com>), against non-redundant NCBI protein database, to determine sequence coverage and identify TRF2 and RAP1 fragments containing lysine residues, which were subsequently confirmed by MS/MS sequencing. Peaks of interest were also identified by manual comparison of spectra acquired on free proteins treated or not treated by sulfo-NHS-acetate.

### Acetylation profiles analysis

Using mMass software, MS spectra were normalized relative to the highest peak (peak at  $m/z$  = 1064 for RAP1 and ARAGA digested by trypsin; peak at  $m/z$  = 1483 for RAP1 and ARAGA digested by V8 endopeptidase Glu-C; peak at  $m/z$  = 909 for TRF2 digested by trypsin). This relative scale allowed comparison of spectra obtained from different samples, which had otherwise different amplitudes on their absolute scales. For each fragment containing lysine residues and confirmed experimentally by MS/MS, we compared peak heights for the free protein versus the protein bound to its partner. Results are shown in Supplementary Figures 8 (TRF2) and 9 (RAP1). Identical results were obtained with RAP1 and ARAGA, except when a spectrum is displayed for ARAGA (Supplementary Figures 8 D and 9 G). Taken together, the modifications in acetylation profiles highlight regions in TRFH, TRF2-RBM, TRF2-Myb, RAP1-RCT, RAP1-Myb and RAP1-BRCT.

*Reference:* Strohal M, Kavan D, Novák P, Volný M, Havlíček V (2010) mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal. Chem.* **82**: 4648–4651

### TRF2

As expected, the known interaction between RAP1-RCT and TRF2-RBM is associated to protection of Lys293 from TRF2-RBM. In addition, Lys327, Lys329 and Lys333 from TRF2, and Lys383 and Lys384 from RAP1-RCT are affected by TRF2/RAP1 interactions, which extends the RBM region of TRF2 by about 20 residues beyond the one described in the crystal structure (Supplementary Figure 10, PDB entry 3K6G, 24). TRFH domain contains 17 accessible lysines located mainly on the external helices of the dimer or on the inner side. We observe acetylation for 6 of them and interestingly the acetylation profile is modified upon RAP1 interaction for 4 of them. Lys58, Lys93, located close to YRLGP interaction site, and the more external Lys183 and Lys190, are protected by RAP1. Interestingly, Lys58 and Lys93 are protected by RAP1-ARAGA but not the more external Lys183 and Lys190, suggesting that the interaction of the external region is driven by the YRLGP sequence. Finally, Lys202 and Lys220, located on the opposite face of the monomer, are always acetylated, which implies that this region does not belong to the interaction path.

Strong decrease in peak height for peaks corresponding to acetylated peptides 55-66 and 81-99 show that K58 and K93 are protected in TRF2/RAP1 complex (Supplementary Figure 8, B and C). Peak corresponding to di-acetylated peptide 181-192 decreases for TRF2 in complex with RAP1 (Supplementary Figure 8 D, left panel) but not in complex with ARAGA (Supplementary Figure 8 D, right panel), showing that K184 and K190 are only protected in TRF2/RAP1 complex and not in TRF2/ARAGA complex. No variation of peak height for peaks corresponding to peptides 201-220 and 203-220 show that these lysines are not affected by the presence of RAP1 nor ARAGA (Supplementary Figure 8 E). Strong decrease in peak height for peak corresponding to 4-acetylated peptide 242-274 shows that this region is



protected in TRF2/RAP1 or TRF2/ARAGA complex. MS/MS analysis revealed that this peak results from a mix of peptides 242-274 acetylated on different lysine residues, and the decrease in peak height indicates that one of the lysines is protected in the complex but we cannot say precisely which one. This is consistent with this region being the beginning of a flexible linker, immediately in C-ter with respect to TRFH domain. Variations of peak height for peaks corresponding to peptides 290-307 (acetylated), 294-307, 294-311 (acetylated), 312-327, 312-329 (acetylated) and 312-332 (di-acetylated) were all correlated due to the fact that trypsin is unable to cleave after an acetylated lysine, giving a complex cleavage pattern that depends on acetylation state. The data presented in Supplementary Figure 8 H can be explained by:

- A protection of K293, explaining the decrease seen in top left panel
- K307 becoming more accessible (thus more frequently acetylated), explaining the decrease seen in top middle and left panels and the increase seen in top right panel
- K311 (lysine marked by a star) becoming more accessible (thus more frequently acetylated), explaining the decrease seen in bottom left panel instead of the expected increase due to K327 protection
- A protection of K327, explaining the decrease seen in bottom middle and right panel
- A protection of K329, explaining the decrease seen in bottom right panel.

Supplementary Figure 8 I shows that one of the three lysines K333, K353 or K355 is protected in the complex, explaining the decrease seen for peaks corresponding to di- and tri-acetylated peptide 333-356 (middle and right panels, respectively) while peak corresponding to mono-acetylated peptide 333-356 does not vary between free and complexed TRF2 (left panel). It has been shown that TIN2 and RAP1 can interact together with TRF2. Since TIN2 interaction site includes K353 and K355, we concluded that the protected lysine is K333, which is also consistent with the proximal protected K327 and K329. Peaks corresponding to di-acetylated peptide 447-459 and mono-acetylated peptide 448-459 do not vary between free and complexed TRF2 (Supplementary Figure 8 J), showing that K447, K449 and K459 are identically exposed either in free TRF2 or in the complex. Decrease in peak height for peaks corresponding to di-acetylated peptide 460-482 and mono-acetylated peptide 465-482 (Supplementary Figure 8 K) show that K464 and K475 are protected in the complex.

#### RAP1

Slight increase in peak height for peak corresponding to acetylated peptide 24-40 (Supplementary Figure 9 B) shows that K39 is more accessible in the complex. This is observed only with RAP1-WT, not with ARAGA. Increase in peak height for peak corresponding to peptide 101-129, and no significant variation in peak height for peak corresponding to acetylated peptide 101-129 (Supplementary Figure 9 C) show that K114 is protected in the complex. K149, K165 and K181 remain equally accessible in the complex as in free RAP1, as shown by no significant variation of peak height for peaks corresponding to peptide 134-149 and acetylated peptide 134-153 (Supplementary Figure 9 D), acetylated peptide 154-169 (Supplementary Figure 9 E), and peptide 170-183 and its acetylated form (Supplementary Figure 9 F). K186 is protected in the complex with RAP1-WT only (not with ARAGA), as shown by a decrease in peak height for peak corresponding to acetylated peptide 183-192 (Supplementary Figure 9 G). Slight increase in peak height for peak corresponding to acetylated peptide 190-208 (Supplementary Figure 9 H) shows that K194 is more accessible in the complex, only with RAP1-WT. K208, K210, K212 and K228 remain equally accessible in the complex as in free RAP1, as shown by no significant variation of peak height for peaks corresponding to mono- and di-acetylated peptide 195-211 (Supplementary Figure 9 J) and

non-acetylated, mono-acetylated and di-acetylated forms of peptide 212-229 (Supplementary Figure 9 K). Strong increase in peak height for peak corresponding to acetylated peptide 324-344 (Supplementary Figure 9 L) shows that K324 is more accessible in the complex than in free RAP1. K383 is protected in the complex, as shown by an increase in peak height for peak corresponding to peptide 380-395 and a correlated decrease in peak height for peak corresponding to an acetylated form of this peptide (Supplementary Figure 9 M). K384 is also protected, as shown by the same variations for peaks corresponding to non-acetylated and acetylated peptide 384-392 (Supplementary Figure 9 N).

### **RAP1 / TRF2-Myb interaction assay**

Footprinting experiments showed that TRF2 K464 and K475 are protected in the complex (Supplementary Figure 10, A and B), which was not expected because these lysines are located in the Myb DNA-binding domain, a region not expected to interact with RAP1. We therefore tested the possibility of a contact between RAP1 and TRF2-Myb by size exclusion chromatography. The chromatogram displays two peaks at same elution volumes as the peaks obtained with both proteins run separately through the column (Supplementary Figure 11), and SDS-PAGE analysis of the contents of those peaks confirms that the two proteins do not co-elute. Protein concentration in the output from the column was estimated to be around 25-30  $\mu\text{M}$  for RAP1 and around 55-60  $\mu\text{M}$  for TRF2-Myb. We can therefore conclude that RAP1 does not interact with TRF2-Myb with a detectable affinity, and that the protection detected in footprinting experiments could be due to conformational adjustment of TRF2 or intra-TRF2 (potentially inter-monomer) contacts.

### **Preparation of nuclear extracts**

Nuclear extracts were prepared following the protocol described in (Abmayr *et al*, 2001), with the following minor modifications : all buffers contained 2-mercaptoethanol at a concentration of 5 mM instead of DTT; PMSF concentration in all buffers was 1 mM instead of 0.2 mM; conductivity of the extracts was not measured but final dialysis was performed against 100 volumes of dialysis buffer (molecular weight cut-off 6-8 kDa) during 1h30; final centrifugation was done at 16,100 g instead of 25,000 g. Quantification by Bradford assay showed that in our hands this protocol yielded nuclear extracts containing from 3 to 5 mg/mL total proteins.

## Supplementary data

**Supplementary Table 1: Protein constructs used in this study**

Protein construct		Expression vector	Tag (all at N-ter)	Cleavage site	<i>E. coli</i> strain	Antibiotic
A	RAP1 [1-399]	pET-16b	DsRed - 9-His	TEV	Rosetta	Amp 100 µg/mL Cm 35 µg/mL
B	RAP1 [1-399]	pET-16b	9-His	TEV	Rosetta	Amp 100 µg/mL Cm 35 µg/mL
C	RAP1 [1-399] ARAGA	pET-16b	9-His	TEV	Rosetta pLysS	Amp 100 µg/mL Cm 35 µg/mL
D	RAP1 [270-399] RCT	pET-14b	6-His	Thrombin	BL21DE3-star	Amp 50 µg/mL
E	RAP1[1-208]	pET-16b	9-His	TEV	Rosetta pLysS	Amp 100 µg/mL Cm 35 µg/mL
F	TRF2 [3-500]	pTrcHisB	6-His	TEV	BL21DE3-star	Amp 50 µg/mL
G	TRF2 [1-500]	pETM-13 <sup>a</sup>	8-His	TEV	BL21DE3-star	Kan 30 µg/mL
H	TRF2 [43-245] TRFH	pETM-13 <sup>a</sup>	8-His	TEV	BL21DE3-star	Kan 30 µg/mL
I	TRF2 [400-500] Myb	pETM-13 <sup>a</sup>	8-His	TEV	BL21DE3-star	Kan 30 µg/mL

Amp: ampicillin; Cm: chloramphenicol; Kan: kanamycin; a: from Genscript

**Supplementary Table 2: Oligonucleotides used in this study.**

Oligonucleotide	Number of TTAGGG repeats	Sequence	Ends	Used in
T2-S1-S2	2	5'-CTAGGGTTAGGGTTAG-3'	Blunt	Supp. Fig. 1
T2-S5-S8-biotin	2	biotin-5'-TACTAGGGTTAGGGTTAGCA-3'	1-base 5'-overhang	Supp. Fig. 12

### Buffers used for protein purification

All buffers are freshly prepared and filtered through 0.22 µm diameter pores.

Buffer A: HisTrap loading buffer

- Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH = 8, 20 mM
- NaCl 500 mM
- Imidazole 60 mM in case of 8- or 9-His tag, 20 mM in case of 6-His tag
- 2-mercaptoethanol 10 mM

Buffer B: HisTrap elution buffer

- Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH = 8, 20 mM
- NaCl 500 mM (1 M for RAP1 constructs B and C because pI < 5 gives stronger binding to Ni<sup>2+</sup>)
- Imidazole 1 M
- 2-mercaptoethanol 10 mM

Buffer C: “Reverse HisTrap” loading buffer

- Tris-HCl pH = 7.5, 20 mM
- NaCl 300 mM (400 for RAP1 constructs B and C because pI < 5 gives stronger binding to Ni<sup>2+</sup>)
- Imidazole 20 mM
- 2-mercaptoethanol 10 mM

Buffer D: Desalting buffer

- Tris-HCl pH = 7.5, 20 mM
- NaCl 150 mM
- 2-mercaptoethanol 5 mM

Buffer E: Heparin low salt buffer

- Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH = 7.5, 10 mM
- NaCl 50 mM
- 2-mercaptoethanol 5 mM

Buffer F: Heparin high salt buffer

- Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH = 7.5, 10 mM
- NaCl 2 M
- 2-mercaptoethanol 5 mM

Buffer G: Gel filtration high salt buffer

- Tris-HCl pH = 7.5, 20 mM
- NaCl 1 M
- 2-mercaptoethanol 5 mM

Buffer H: Protein footprinting buffer

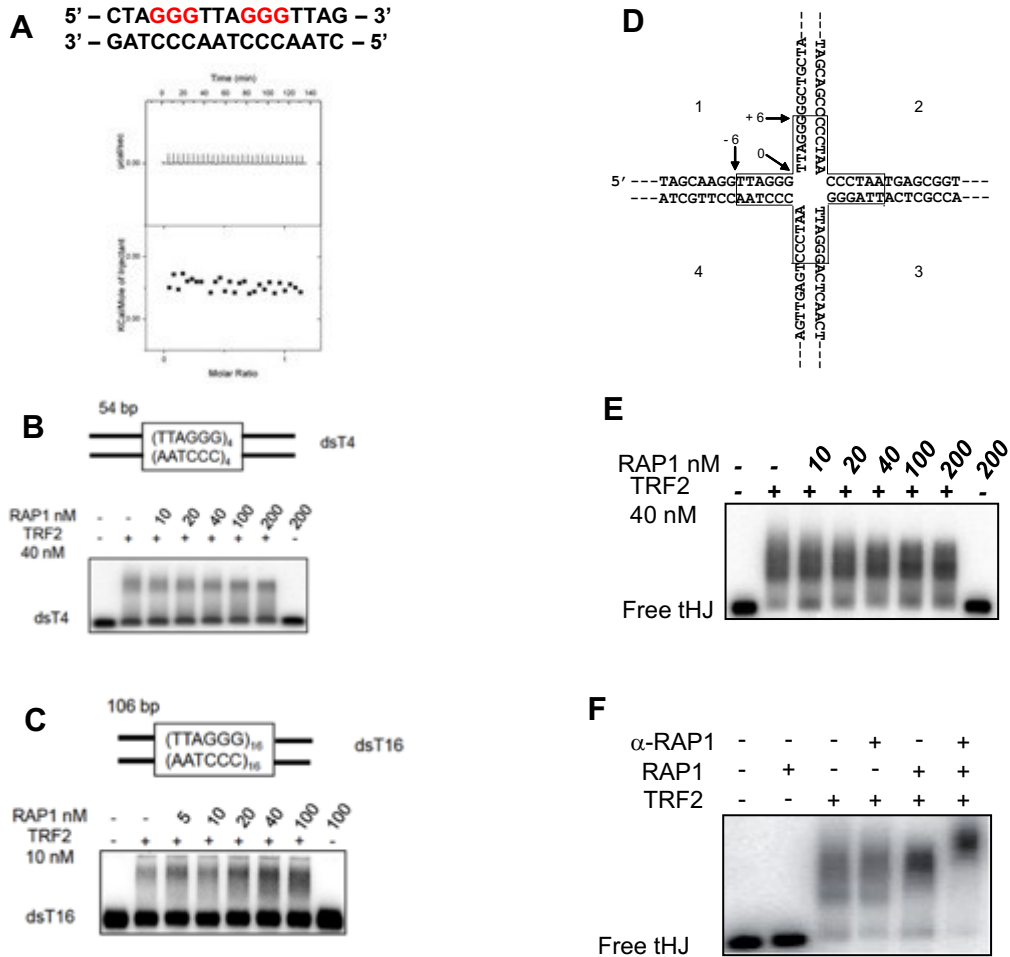
- HEPES-NaOH pH = 7.5,
- 20 mM NaCl 150 mM
- 2-mercaptoethanol 5 mM

**Supplementary Table 2: ITC experiments**

Cell	Syringe	Buffer	Temperature	Vol inj.
TRF2 (13 μM)	RAP1 (130 μM)	50 mM Tris pH 7.5, 150 mM NaCl, 10 mM βMe	283 K	10 μL
TRF2 (10 μM)	RAP1 <sub>[270-399]</sub> (91 μM)	50 mM Tris pH 7.5, 150 mM NaCl, 10 mM βMe	283 K	10 μL
TRF2 (13 μM)	RAP1-ARAGA (130 μM)	50 mM Tris pH 7.5, 150 mM NaCl, 10 mM βMe	283 K	10 μL
TRF2 (13 μM)	Apollo-YLP (130 μM)	50 mM Tris pH 7.5, 150 mM NaCl, 10 mM βMe	283 K	10 μL
TRFH (25 μM)	RAP1-YLP (250 μM)	25 mM Tris pH 7.5, 100 mM NaCl, 10 mM βMe	283K 293K 279 K	10 μL
TRF2/RAP1 (13 μM/13μM)	Apollo-YLP (130 μM)	50 mM Tris pH 7.5, 150 mM NaCl, 10 mM βMe	283 K	10 μL
Apollo-YLP (25 μM)	TRFH (250 μM)	25 mM Tris pH 7.5, 100 mM NaCl, 10 mM βMe	293 K	10 μL
TRF2 (13 μM)	Apollo-YLP (130 μM)	50 mM Tris pH 7.5, 150 mM NaCl, 10 mM βMe	283 K	10 μL
TRF2 (27 μM)	TIN2-YLP (270 μM)	50 mM Tris pH 7.5, 150 mM NaCl, 10 mM βMe	283 K	1.5 μL (ITC200)

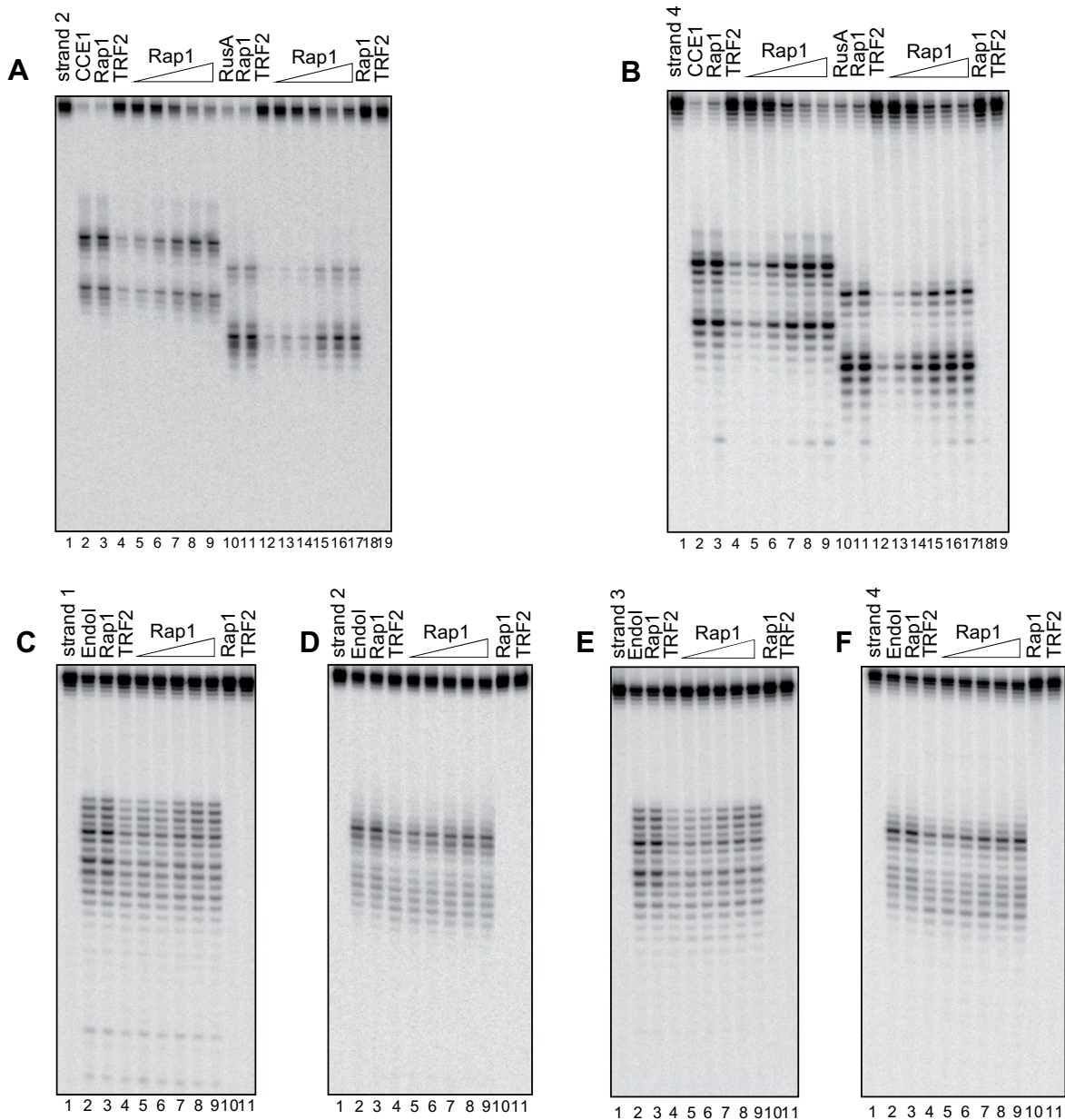
TRF2/RAP1 (27 $\mu$ M/27 $\mu$ M)	TIN2-YLP (270 $\mu$ M)	50 mM Tris pH 7.5, 150 mM NaCl, 10 mM $\beta$ Me	283 K	1.5 $\mu$ L (ITC200)
--------------------------------------	------------------------	--	-------	-------------------------

$\beta$ Me = 2-mercaptoethanol



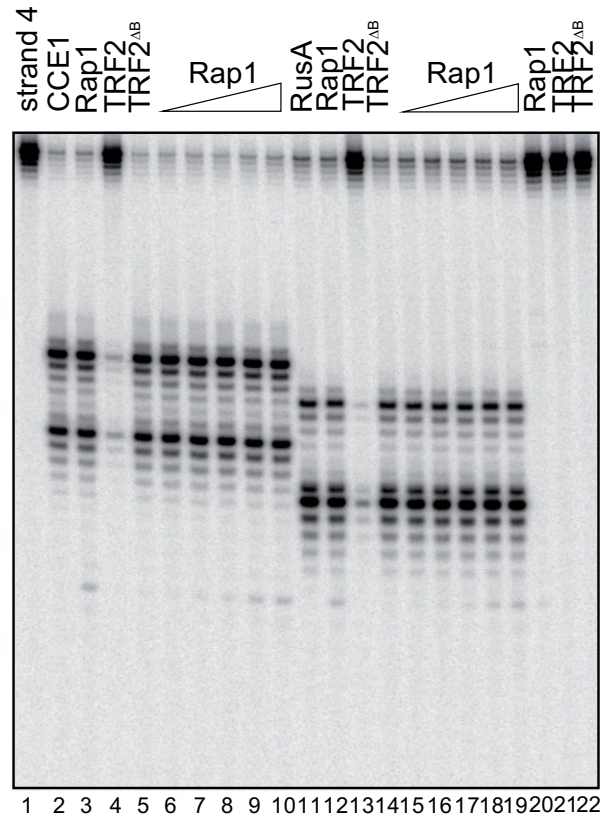
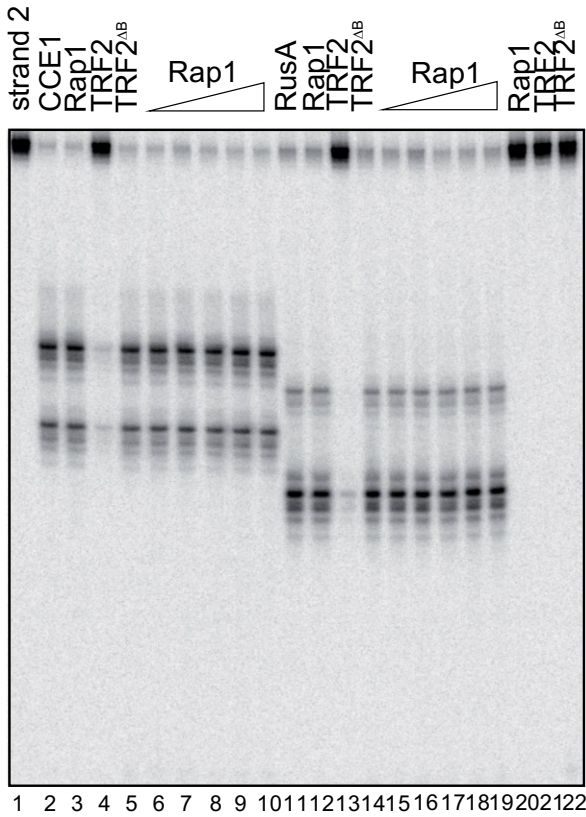
### Supplementary Figure 1: RAP1 does not bind DNA, does not affect TRF2 binding on linear DNA, and can form a ternary complex with TRF2 and tel H-J.

**A:** ITC experiments performed at 10° C with RAP1 and double-stranded DNA oligonucleotides in 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM β-ME. **B:** EMSA performed on 5 nM of a labelled 54 bp linear DNA that contains 4 telomeric repeats (dsT4) using 40 nM of TRF2 dimers and increasing amounts of RAP1. The addition of RAP1 causes a slight change in the aspect of TRF2-containing complexes but no visible changes in affinity. **C:** EMSA performed on 5 nM of a labelled 106 bp linear DNA containing 16 telomeric repeats (dsT16) using 10 nM of TRF2 dimers and increasing amounts of RAP1. The addition of RAP1 causes both a slight change in the aspect of the TRF2-containing complex and a mild increase in affinity (~10% increase in the amount of complex). **D:** Drawing of telomeric Holliday Junction (tHJ), a migratable Holliday Junction with 27 bp including 1 telomeric repeat per arm. **E:** EMSA performed on labelled tHJ (5 nM) using 20 nM of TRF2 dimers and increasing amounts of RAP1. The addition of RAP1 causes a change in the aspect of TRF2-containing complexes but no overt modification of affinity. **F:** Supershift experiment performed using 5 nM of tHJ incubated with 40 nM of TRF2 (dimer) with or without pre-incubation with 200 nM of RAP1 and post-incubation with 0.2 μg of anti-RPA1 antibody prior to agarose electrophoresis. The shifted DNA complexes obtained with TRF2 and RAP1 are supershifted by the anti-RAP1 antibody showing formation of a ternary TRF2/RAP1/junction complex.



**Supplementary Figure 2: RAP1 totally inhibits TRF2-dependent protection against CCE1 and RusA cleavage of tHJ, and largely inhibits TRF2-dependent protection against Endonuclease I cleavage of tHJ.**

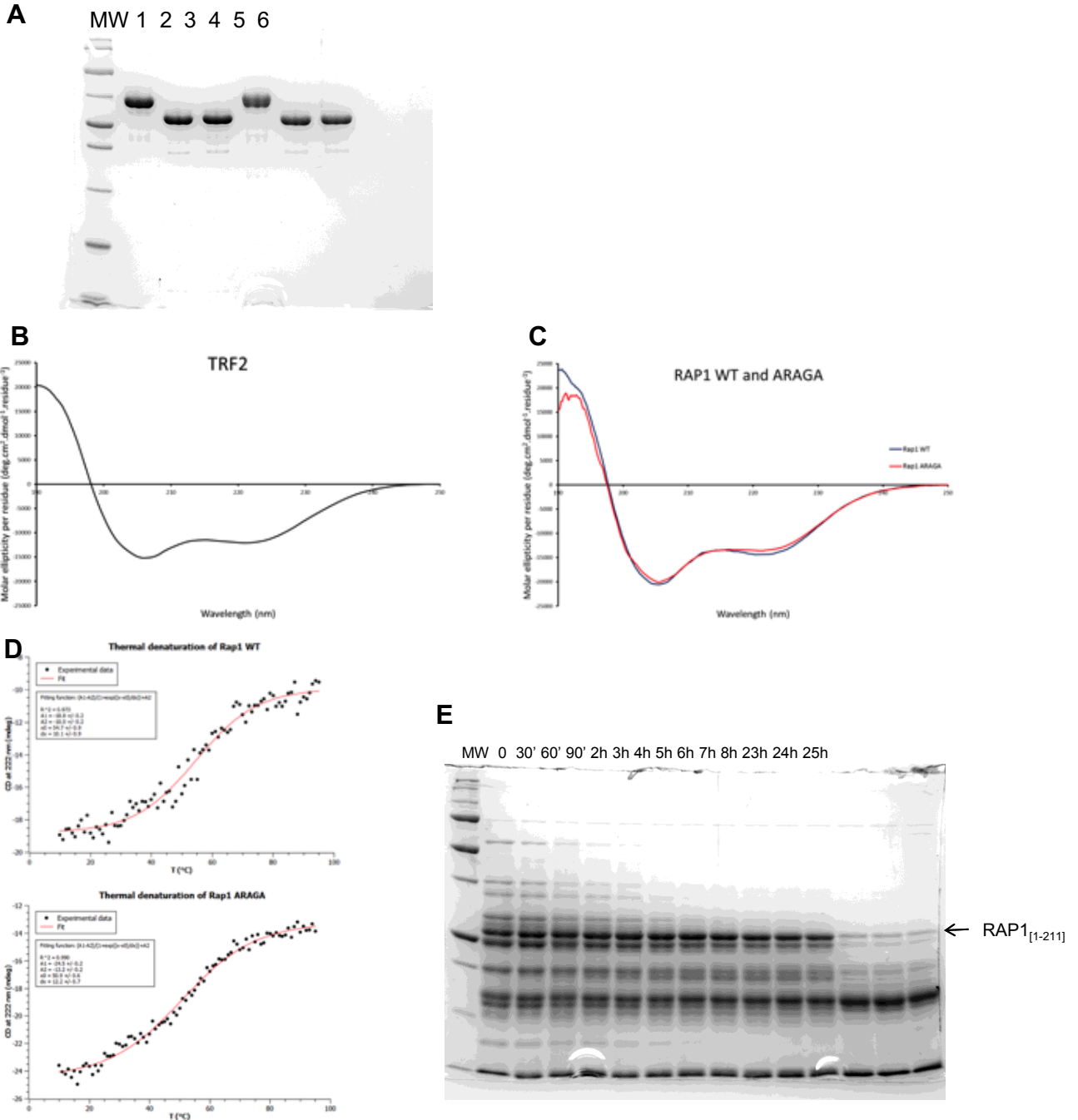
**A-B:** 5 nM of tHJ labelled on either strand 2 (A) or 4 (B) was incubated with 200 nM of TRF2 with and without increasing amount of RAP1 prior to cleavage with either CCE1 A or RusA . Concentrations of RAP1 used were 100, 200, 400, 1000, 1500 nM. In lanes 3, 11 and 18, the junction is incubated with 1500 nM of RAP1. **C-F:** 5 nM of tHJ labelled on either of the 4 strands was incubated with 200 nM of TRF2 with and without increasing amount of RAP1 prior to cleavage with Endonuclease I. Concentrations of RAP1 used were 100, 200, 400, 1000, 1500 nM. In lanes 3 and 10, the junction is incubated with 1500 nM of RAP1.



**Supplementary Figure 3: RAP1 does not modify tHJ cleavage in the presence of TRF2<sup>ΔB</sup>**

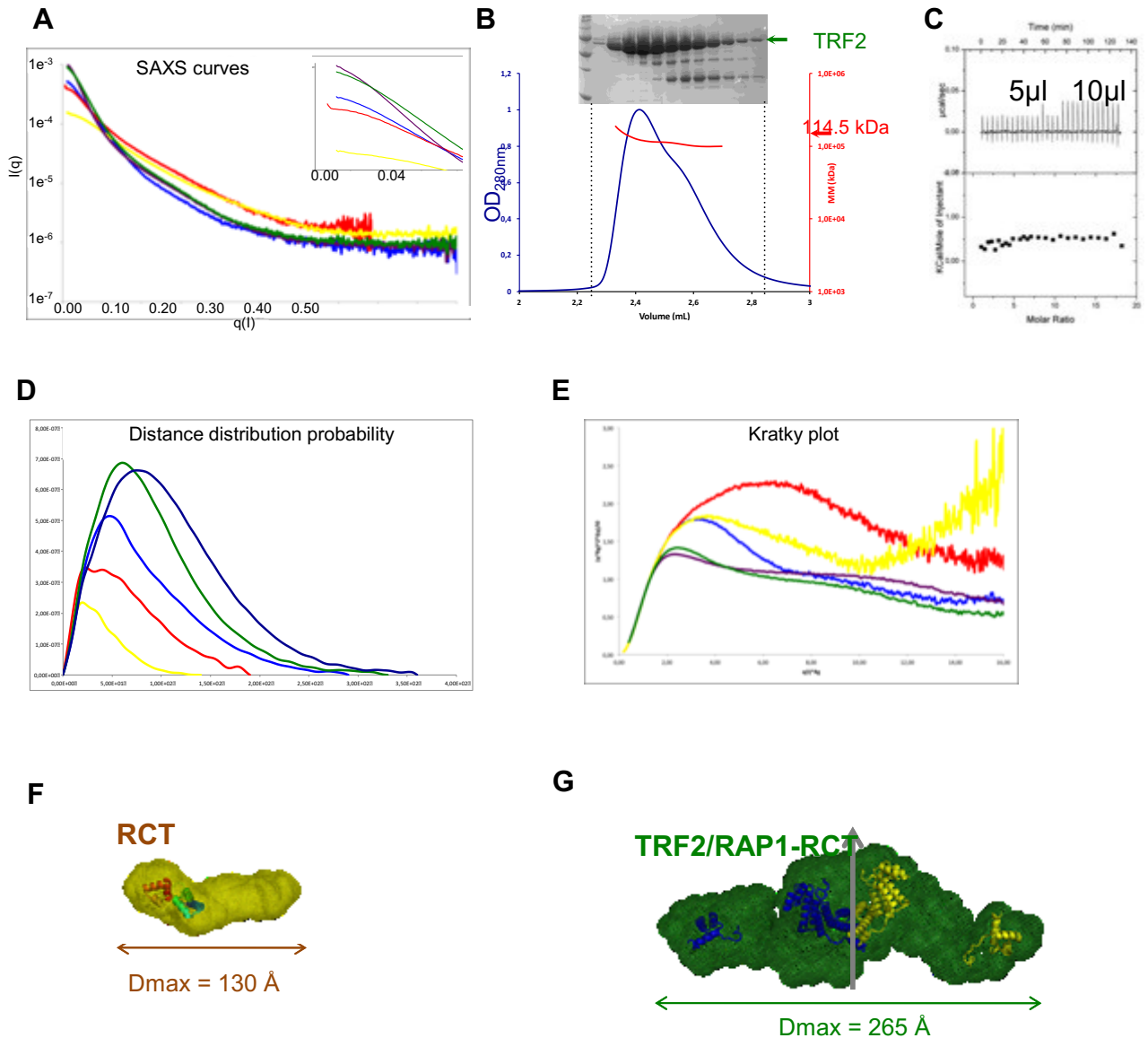
5 nM of tHJ labelled on either strand 2 or 4 was incubated with 200 nM of TRF2<sup>ΔB</sup> with and without increasing amount of RAP1 prior to cleavage with either CCE1 or Rusa. Concentrations of RAP1 used were 100, 200, 400, 1000, 1500 nM. In lanes 3, 12 and 20 the junction is incubated with 1500 nM of RAP1.





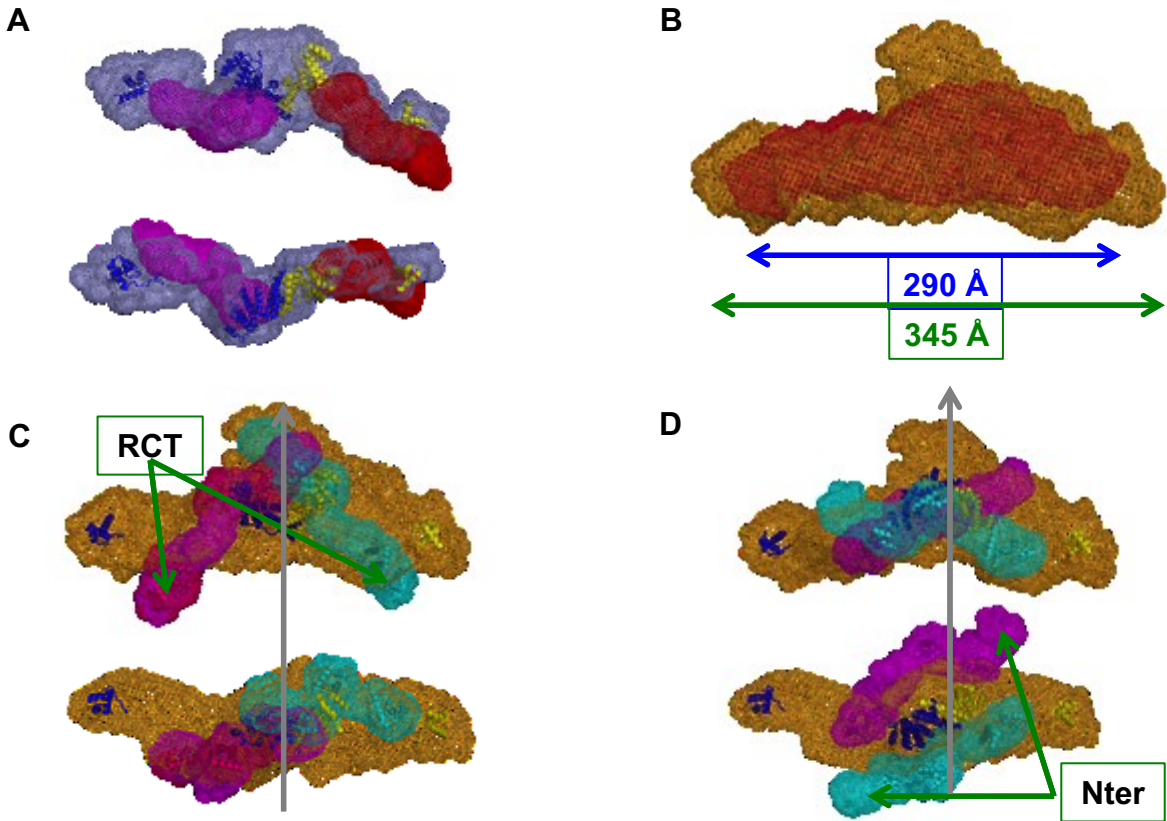
**Supplementary Figure 4: Qualitative analysis of TRF2, RAP1 and RAP1-ARAGA.**

**A:** SDS-PAGE of purified TRF2 (1), RAP1 (2), RAP1-ARAGA (3), and acetylated TRF2 (4), RAP1 (5), and RAP1-ARAGA (6) proteins (Coomassie blue stained); MW: Molecular Weights from top to bottom: 250, 150, 100, 70, 50, 40, 30, 20, 15, 10 and 5 kDa. **B-C:** Circular dichroism spectra of TRF2 (**B**) and of RAP1 and RAP1-ARAGA mutant (**C**). **D:** thermal denaturation curve of RAP1 WT (top), of RAP1-ARAGA mutant (bottom). **E:** Limited proteolysis of RAP. The band corresponding to RAP1[1-211] as identified by LC-MS is shown with an arrow. MW: Molecular Weights from top to bottom: 250, 150, 100, 75, 50, 37, 25, 20, 15 and 10 kDa.



**Supplementary Figure 5: SAXS analysis of Rap1<sub>[266-399]</sub> (yellow curves), RAP1 (red curves), TRF2 (blue curves), RAP1<sub>[266-399]</sub>/TRF2 (violet curves), and RAP1/TRF2 (green curves), and characterization of TRF2 dimer.**

**A:** SAXS curves with Intensity at the origin in the insert; **B:** SEC-MALS and SDS-PAGE analysis of TRF2; **C:** ITC analysis of TRF2 dilution; **D:** distance distribution function  $p(r)$ ; **E:** Normalized Kratky plots. **F-G:** *Ab initio* averaged envelopes RAP1-RCT (F) and TRF2/RAP1-RCT (G) with superimposed globular domains as references



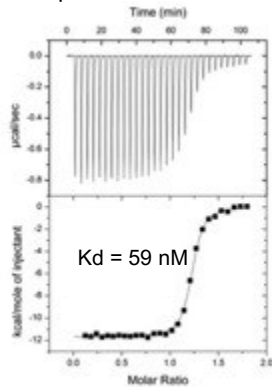
**Supplementary Figure 6: Superposition of isolated RAP1 or RAP1-RCT envelopes with envelopes of their complexes with TRF2.**

**A:** Superposition of two RAP1-RCT envelopes (red and magenta) with TRF2/RAP1-RCT envelope (grey) based on the known interaction between RAP1-RCT and TRF2-RBM; **B:** Superposition of TRF2 and TRF2/RAP1 *ab initio* envelopes with TRF2 Dmax in blue and TRF2/RAP1 Dmax in green; **C-D:** Superposition of two RAP1 envelopes (blue and magenta) with TRF2/RAP1 envelope (yellow), based on (B) the known interaction between RAP1-RCT and TRF2-RBM or (C) the optimal occupancy of the upper part of TRF2/RAP1 envelope by the N-terminal region of RAP1

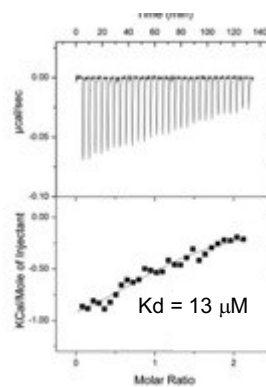
**A**

Condition	Janoušková et al (2015)	Our study
[TRF2-monomer]	5 $\mu$ M (cell)	13 $\mu$ M (cell)
[RAP1]	44 $\mu$ M (injector)	130 $\mu$ M (injector)
Injections	10 $\mu$ L	10 $\mu$ L
Temp.	25°C	10°C
Buffer	[Na <sub>2</sub> PO <sub>4</sub> ] = 20 mM pH = 7.0 [NaCl] = 50 mM	[Tris-HCl] = 50mM pH = 7.5 [NaCl] = 150 mM [2-mercaptoethanol] = 10 mM

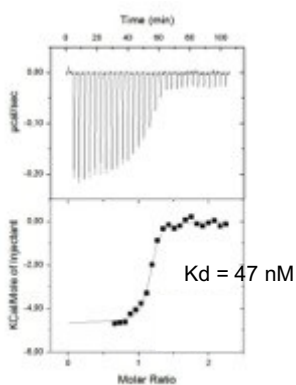
**B** Apollo-YLP / TRFH



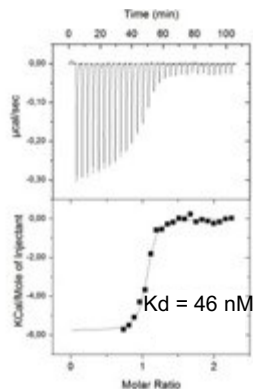
**C** TRFH / RAP1-YLP



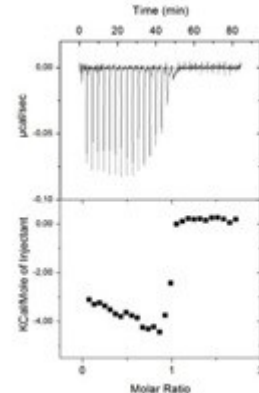
**D** RAP1-TRF2 / Apollo-YLP



**E** TRF2 / Apollo-YLP



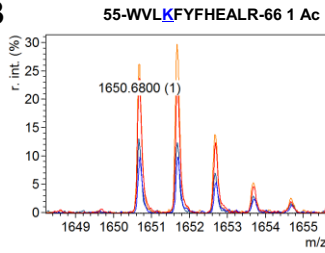
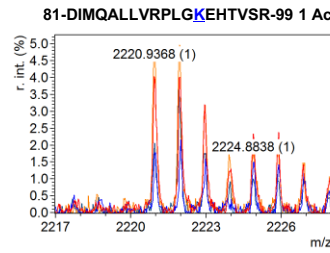
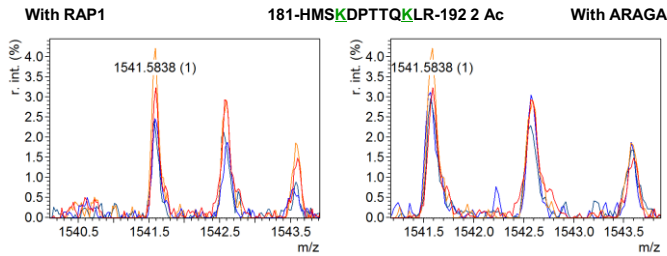
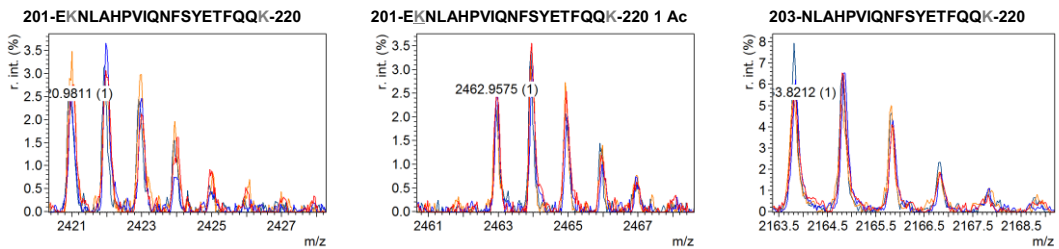
**F** Apollo-YLP+TRF2 / RAP1



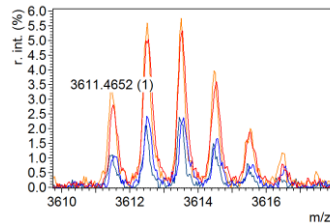
**Supplementary Figure 7: ITC experiments.** (A) Comparison of experimental condition of ITC measurement of the interaction between TRF2 and RAP1. (B-F) ITC characterisation of the interaction between TRFH and Apollo-YLP peptide <sup>494</sup>ATEFRGLALKYLLTPVNF<sup>511</sup> (B), TRFH and RAP1-YLP peptide <sup>89</sup>ERNERLELEAYRLGPASA<sup>106</sup> (C), the pre-formed TRF2/RAP1 complex with Apollo-YLP peptide (D), TRF2 and Apollo-YLP peptide (E), Apollo-YLP/TRF2 preformed complex and RAP1 (F).

**A**

Domains	TRFH																	Linker			
TRF2 lysines	58	93	140	155	158	167	169	173	176	179	180	184	190	202	220	241	242	245	255	258	267
acetylation profile	■											■	■	■	■	■	■	■	■	■	■

**B****C****D****E****F**

242-KALK?SESAASSTGK?EDK?QPAPGPVEK?PPREPAR-274 4 Ac

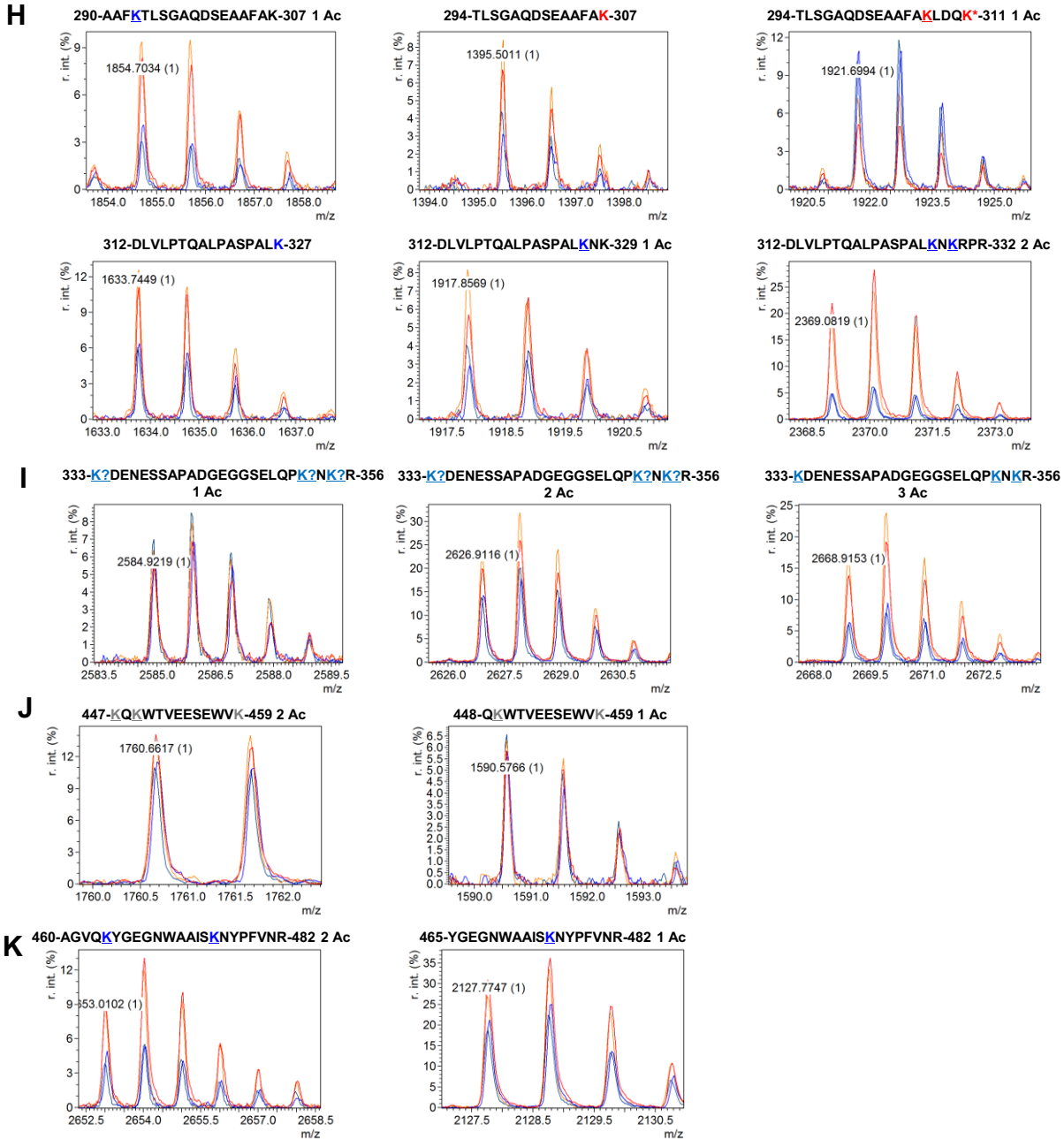


## Supplementary Figure 8: Comparison of MS spectra for TRF2 free vs in complex.

**A:** summary of footprinting results in TRF2 (K58 to K267). Lysine residues protected (blue) or not affected (grey) in TRF2/RAP1 and TRF2/ARAGA complex, or protected only in TRF2/RAP1 complex (green). Region protected in TRF2/RAP1 and TRF2/ARAGA complex, but precise lysine undetermined (light blue). Lysine residues not detected, or with signal too weak to be interpreted (white). **B-F:** MS peaks of interest, with corresponding fragment indicated above each peak. Lysine residues are color-coded according to **A**, acetylated lysines are underlined, and number of acetyl groups in each fragment is indicated. Red and orange curves: TRF2 alone; dark blue and blue curves: TRF2 in complex with RAP1 (with ARAGA in **D**, right panel).

**G**

Domains	RBM				Linker		TIN2		Linker				Myb											
TRF2 lysines	289	293	307	311	327	329	333	353	355	390	402	405	408	410	420	446	447	449	459	464	475	488	495	
acetylation profile																								

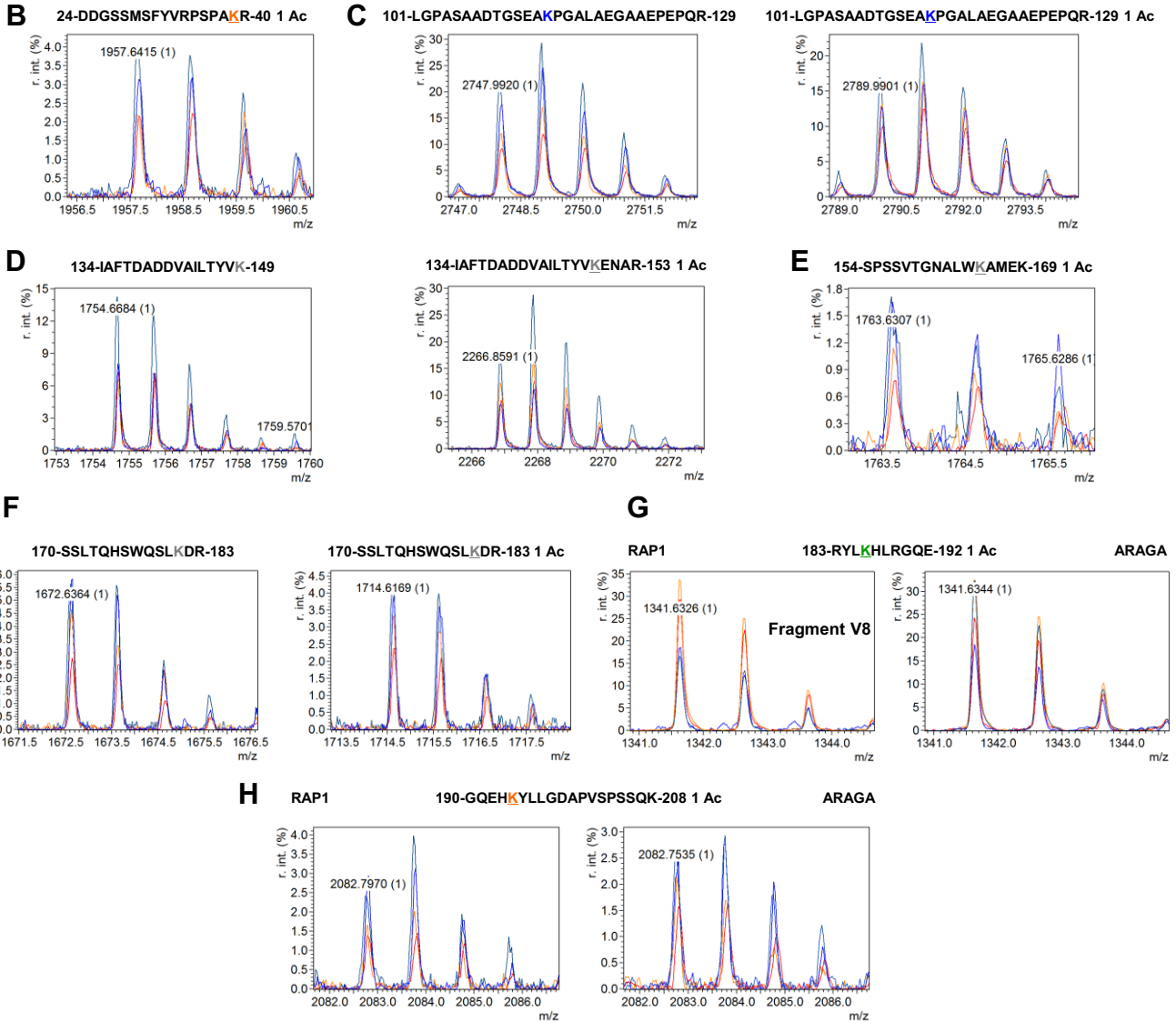


### Supplementary Figure 8 (continued)

**G:** summary of footprinting results in TRF2 (K289 to K495). Lysine residues protected (blue), more accessible (red) or not affected (grey) in TRF2/RAP1 and TRF2/ARAGA complex. Region protected in TRF2/RAP1 and TRF2/ARAGA complex, but precise lysine undetermined (light blue). Lysine residues not detected, or with signal too weak to be interpreted (white). **H-K:** MS peaks of interest, with corresponding fragment indicated above each peak. Lysine residues are color-coded according to **G**, acetylated lysines are underlined, and number of acetyl groups in each fragment is indicated. Red and orange curves: TRF2 alone; dark blue and blue curves: TRF2 in complex with RAP1.

**A**

Domains	N	BRCT	L1	Myb				L2	
RAP1 lysines	9	39	114	149	165	169	181	186	194
acetylation profile		orange	blue	grey	grey	grey	grey	green	white



### Supplementary Figure 9: Comparison of MS spectra for RAP1 free or in complex.

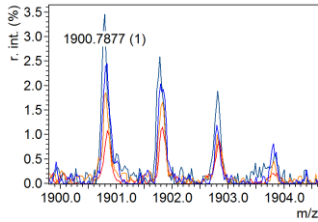
**A:** summary of footprinting results in RAP1 (K9 to K194). Lysine residues protected (blue) or not affected (grey) in TRF2/RAP1 and TRF2/ARAGA complex. Lysine residues protected (green) or more accessible (orange) only in TRF2/RAP1 complex. Lysine residues not detected, or with signal too weak to be interpreted (white).

**B-H:** MS peaks of interest, with corresponding fragment indicated above each peak. Lysine residues are color-coded according to **A**, acetylated lysines are underlined, and number of acetyl groups in each fragment is indicated. Red and orange curves: RAP1 alone; dark blue and blue curves: RAP1 in complex with TRF2 (ARAGA instead of RAP1 in **G**, right panel).

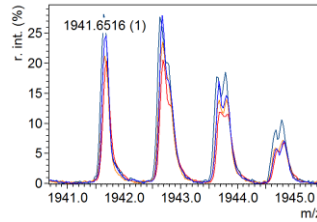
I

Domains	Linker 2							RCT								
RAP1 lysines	208	210	212	228	240	251	252	305	316	324	338	372	383	384	398	399
acetylation profile																

J 195-YLLGDAPVSPSSQK?L?R-211 1 Ac

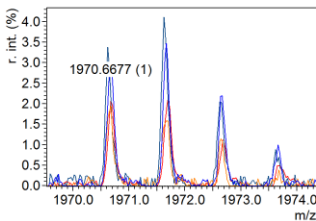


195-YLLGDAPVSPSSQKLKR-211 2 Ac

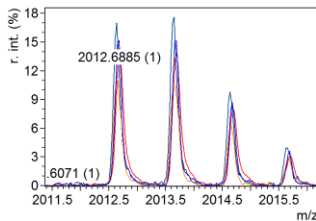


K

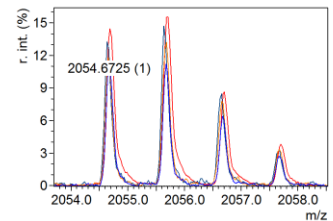
212-KAEEDPEAADSGEPQNKR-229



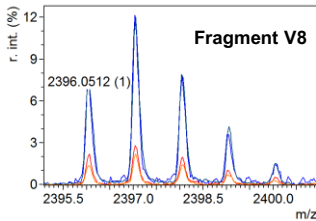
212-K?AEEDPEAADSGEPQNK?R-229 1 Ac



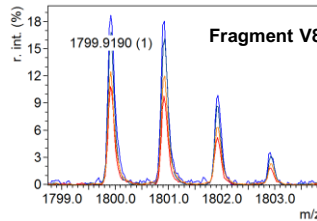
212-KAEEDPEAADSGEPQNKR-229 2 Ac



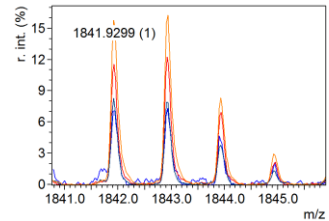
L 324-KFNLDLSTVTQAFKNSGELE-344 1 Ac



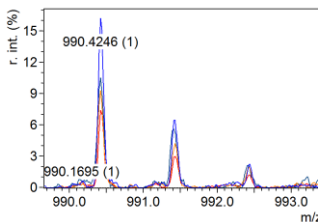
M 380-ALVKKFGAQNVARRIE-395



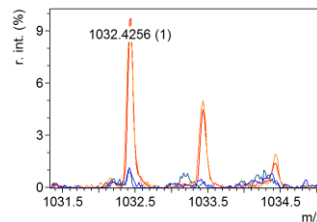
380-ALVKKFGAQNVARRIE-395 1 Ac



N 384-KFGAQNVAR-392



384-KFGAQNVAR-392 1 Ac

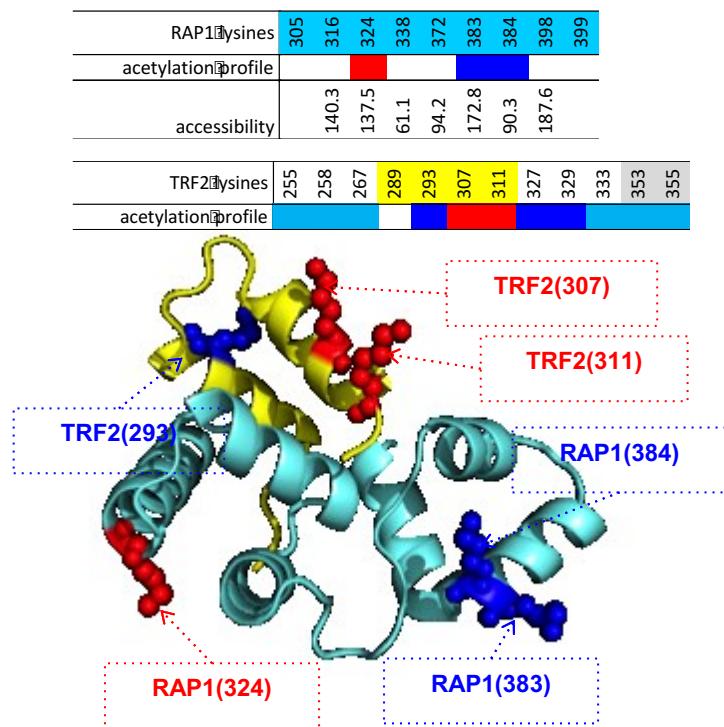


### Supplementary Figure 9 (continued)

I: summary of footprinting results in RAP1 (K208 to K399). Lysine residues protected (blue), more accessible (red) or not affected (grey) in TRF2/RAP1 and TRF2/ARAGA complex. Lysine residues not detected, or with signal too weak to be interpreted (white).

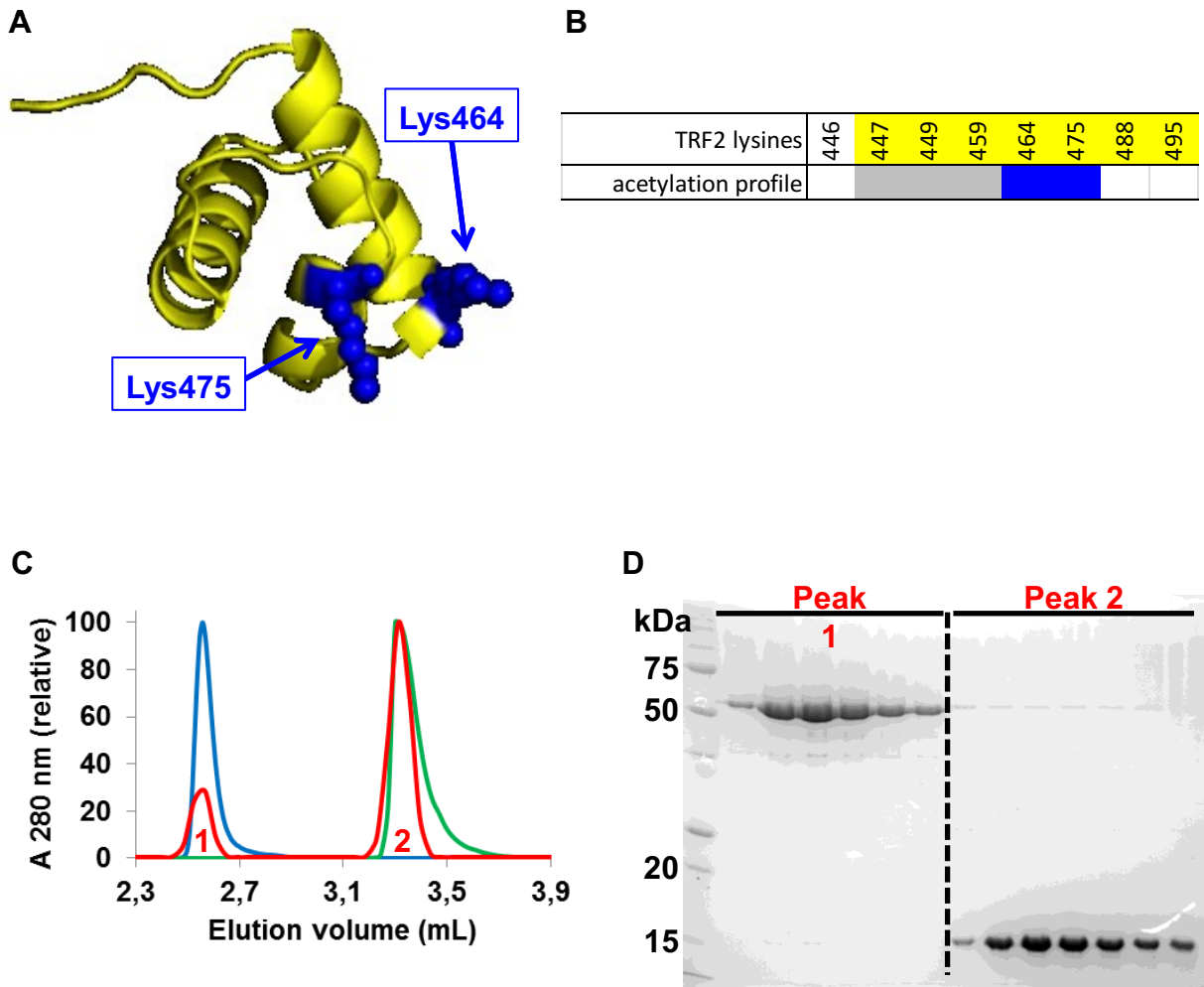
J-N: MS peaks of interest, with corresponding fragment indicated above each peak. Lysine residues are color-coded according to I, acetylated lysines are underlined, and number of acetyl groups in each fragment is indicated. Red and orange curves: RAP1 alone; dark blue and blue curves: RAP1 in complex with TRF2.





### Supplementary Figure 10 :

Acetylation profile of lysine residues in TRF2, RAP1 or RAP1-ARAGA alone, compared to TRF2/RAP1 or TRF2/RAP1-ARAGA shown on the RAP1-RCT / TRF2-RBM crystal structure (PDB entry code 3K6G). Lysine residues are represented in small sphere and coloured differently if they are protected (blue) or more accessible (red) in both TRF2/RAP1 and TRF2/RAP1-ARAGA complex. The inserted table shows the lysine number in the sequence, highlighted in blue in RAP1 globular domain, in yellow in TRF2 RBM, and in grey in TIN2 interacting region (Chen et al., 2008). The acetylation profile follows the same colour code as in the structure representation.



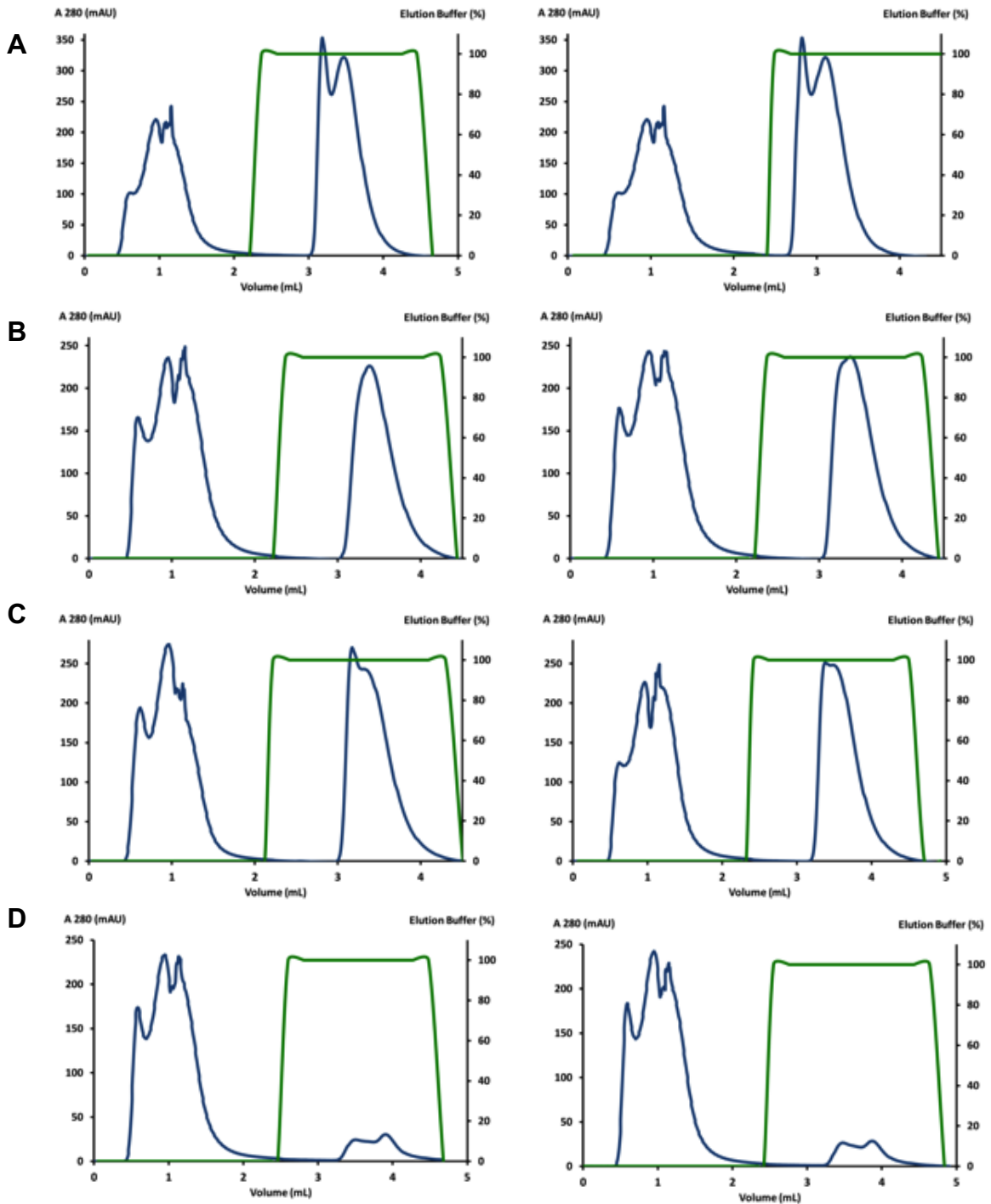
**Supplementary Figure 11: RAP1 does not interact with TRF2 Myb domain.**

**A:** TRF2 K464 and K475 (blue) on the structure of TRF2 Myb domain (yellow, drawn from coordinates in PDB 1W0U).

**B:** summary of footprinting results in TRF2 Myb domain. Lysines underlined in grey: partially acetylated or not acetylated, and not affected by RAP1 nor ARAGA binding. Lysines underlined in blue: protected when TRF2 is bound to RAP1 or ARAGA. Lysines underlined in white: not detected, or signal too weak to allow interpretation.

**C:** size exclusion chromatograms of RAP1 alone (blue curve), TRF2 Myb alone (green curve) and a mix of RAP1 and TRF2 Myb at a 1:2 molar ratio, 90 and 180  $\mu$ M respectively (red curve).

**D:** SDS-PAGE analysis (Coomassie blue stained) of the contents of the two peaks from the red chromatogram in C.



**Supplementary Figure 12: Streptavidin-DNA pulldown chromatography elution profile.**

HeLa nuclear extracts were loaded on streptavidin/(T2AG3)<sub>2</sub>-DNA column pre-equilibrated with TRF2 (A), RAP1/TRF2 (B), RAP1-ARAGA/TRF2 (C), or nothing (D) with two experiments for each case. The region corresponding to the elution appears in the green box.