Les structures et les mécanismes d'interactions des protéines de liaison à l'ADN simple brin aux télomères

Jean Chatain

▶ **To cite this version:**

Jean Chatain. Les structures et les mécanismes d'interactions des protéines de liaison à l'ADN simple brin aux télomères. Biophysique. Muséum National d'Histoire Naturelle PARIS, 2020. Français. NNT : . tel-03218825

HAL Id: tel-03218825
https://theses.hal.science/tel-03218825

Submitted on 5 May 2021

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.

MUSÉUM NATIONAL D'HISTOIRE NATURELLE

École Doctorale 227
Sciences de la Nature et de l'Homme : évolution et écologie

Année 2020

N ° attribué par la bibliothèque
|_|_|_|_|_|_|_|_|_|_|_|_|_|_|

THÈSE

pour obtenir le grade de

DOCTEUR DU MUSÉUM NATIONAL D'HISTOIRE NATURELLE

Spécialité : Biophysique

présentée et soutenue publiquement par

Jean CHATAIN

le 27 Novembre 2020

Les structures et les mécanismes d'interaction des protéines de liaison à l'ADN simple brin aux télomères

composition du jury :

Dr Hubert BECKER	Maître de conférences, Sorbonne Université	Rapporteur
Dr Sophie BOMBARD	Directrice de recherche, CNRS	Rapporteuse
Pr Claudine MAYER	Professeur, Université Diderot	Examinatrice
Dr Marie-Hélène LE DU	Chercheur, CEA	Examinatrice

sous la direction de :

Dr Carole SAINTOMÉ	Maître de conférences, Sorbonne Université	Directrice
Dr Patrizia ALBERTI	Maître de conférences, MNHN	Co-directrice

À Itzel Ontiveros,
À mon Père,
À ma mère.

Remerciements

Je remercie vivement le Dr Sophie BOMBARD et le Dr Hubert BECKER d'avoir accepté d'être rapporteurs de cette thèse. Je remercie également le Pr Claudine MAYER et le Dr Marie-Hélène LE DU d'avoir bien voulu être membres du jury et évaluer ce travail.

Je voudrais remercier le directeur de l'unité le Dr Jean-Baptiste BOULÉ pour m'avoir accueilli au sein du laboratoire : "Structure et Instabilité des Génomes".

Je remercie également le directeur de l'équipe SANTÉ (Structure des Acides Nucléiques, Télomères et Évolution), le Pr Jean-François RIOU, et tous ses membres (Pr Jian-Sheng SUN, Pr Anne-Laure GUIEYSSE-PEUGEOT, Dr Carole SAINTOMÉ, Dr Patrizia ALBERTI, Dr Patrick MAILLIET, Mr Gildas MOUTA CORDOSO, Mme Virginie HOSSARD) pour m'avoir accueilli pendant ces quatre années.

Je tiens à remercier mes directrices de thèse, le Dr Carole SAINTOMÉ et le Dr Patrizia ALBERTI pour la qualité de leur encadrement, leurs conseils, leur bienveillance et leur disponibilité.

Je remercie particulièrement le Dr Carole SAINTOMÉ pour la qualité des échanges et la confiance faite pendant les quatre années passées au laboratoire.

Je remercie les membres de mon comité de thèse : Pr Anne-Laure GUIEYSSE-PEUGEOT, le Dr Marie-Hélène LE DU et le Pr Stéphanie BONNEAU pour l'accompagnement, les conseils et les échanges.

Je remercie le Dr Emmanuelle DELAGOUTTE pour son expertise et son aide pour la purification des protéines de *RPA*,*POT1/TPP1* et *TPP1*.

Enfin je partage mes amitiés avec mes collègues de bureau : le Dr Gabriel LE BERRE, le Dr Sheldon DECOMBE, le (futur) Dr Marwan ANOUD et tout particulièrement le Dr Sylvain GÉNY.

Table des matières

Liste des publications **10**

Table des figures **13**

Liste des tableaux **17**

Liste des abréviations **19**

1 Introduction Générale **23**
1.1 Les rôles biologiques des télomères 23
1.2 L'ADN télomérique 26
1.2.1 Au sein des ciliés 26
1.2.2 Au sein des levures bourgeonnantes 27
1.2.3 Au sein des mammifères 28
1.2.4 Le motif télomérique 29
1.3 La télomérase 30
1.4 Le capuchon télomérique 34
1.4.1 Le complexe du *Shelterin* 34
1.4.2 La *T-loop* 38
1.4.3 Le complexe du *Shieldin* 40
1.5 Les structures et mécanismes de liaison des protéines de la famille *SSB-OB fold* humaine aux télomères 41
1.5.1 Le motif *OB fold* 42
1.5.2 Le complexe *POT1-TPP1* 44
1.5.3 La protéine *RPA* 51
1.5.4 Le complexe *CST* 58
1.6 Une structure non canonique de l'ADN présent aux télomères 61
1.6.1 La structure G-quadruplex 61
1.6.2 Polymorphisme structural des G-quadruplexes formés par le motif télomérique TTAGGG 63
1.6.3 Structures télomériques en G4 contigus 66

1.7 Mécanismes d'ouverture d'un seul G4 par la famille des *SSB-OB fold* aux télomères 68
1.7.1 Le complexe *POT1-TPP1* 69
1.7.2 La protéine *RPA* 71
1.7.3 Le complexe *CST* 72
1.8 Les problématiques de la thèse 73

2 Mécanismes de liaison des *SSB* à des séquences télomériques structurées en G4 contigus 75
2.1 Introduction 75
2.1.1 Étude de la liaison de *RPA* sur des séquences télomériques formant des structures G4s en tandem 76
2.1.2 Le complexe *POT1-TPP1* peut-il prendre en charge ces multimères de G4s en tandem ? 77
2.2 Présentation de l'étude 81
2.2.1 Multimères de *POT1-TPP1* : une coopérativité de liaison à l'ADN due à la structuration de l'ADN 81
2.2.2 Multimères de *POT1-TPP1* : une progression de liaison à l'ADN de 3' vers 5' 82
2.2.3 Multimères de *POT1-TPP1* : origine de la directionnalité de 3' vers 5' 82
2.3 Article 1 84
2.3.1 Résumé 84
2.4 Résultats complémentaires, discussions et perspectives 105
2.4.1 Étude de*POT1* : résultats et discussion 105
2.4.2 Étude de *RPA* 110
2.4.3 Étude de *CST* : en perspective 113
2.4.4 Compétition entre *RPA* et *POT1-TPP1* sur des multimères de G4s en tandem 113
2.5 Annexes méthodes 115
2.5.1 *FRET* 115

3 Structure du complexe ternaire *POT1-TPP1*-ADN 121
3.1 Introduction 121
3.1.1 La protéine *TPP1*, plus qu'un simple partenaire de *POT1* 121
3.1.2 Le modèle *TEBP*α - *TEBP*β chez *Oxytricha nova* . . . 122
3.1.3 Le modèle *POT1-TPP1* chez l'homme 126
3.2 Présentation de l'étude 127
3.2.1 *TPP1* améliore l'affinité de *POT1* en modulant les étapes d'association et de dissociation. 127
3.2.2 *TPP1* : Un contact avec l'ADN télomérique 128

3.2.3 Le complexe *POT1-TPP1* adopte une forme resserrée en présence d'ADN : une structure conservée au sein du vivant ? . . . 129
3.3 Article 2 . . . 130
3.3.1 Résumé . . . 130
3.4 Conclusion et perspectives . . . 143
3.4.1 Spectroscopie de masse . . . 143
3.4.2 Cryo-microscopie . . . 145
3.5 Annexes méthodes . . . 146
3.5.1 Résonance des plasmons de surface . . . 146
3.5.2 Technique de photopontage utilisant la 4-thiothymine . 149
3.5.3 Microscopie électronique . . . 151

4 Des répétitions non-canoniques au sein des télomères : structure et interaction avec *RPA* 155
4.1 Introduction . . . 155
4.1.1 Répétitions non-canoniques . . . 155
4.1.2 Un motif télomérique non-canonique instable . . . 155
4.1.3 Le motif CTAGGG forme un G4 non canonique. . . . 156
4.2 Présentation de l'étude . . . 156
4.2.1 Les répétitions du motif CTAGGG ne forment pas que des G4. . . . 156
4.2.2 Les répétitions CTAGGG structurées en tige-boucle posent un problème à la liaison de *RPA*. . . . 158
4.2.3 Et le brin C ? . . . 160
4.2.4 Le motif non-canonique stable TCAGGG . . . 161
4.3 Article 3 . . . 162
4.3.1 Résumé . . . 162
4.4 Conclusions, perspectives et résultats complémentaires . . . 210
4.4.1 Le motif instable TCAGGG : des structures en tiges-boucles qui affectent la liaison de *RPA* . . . 210
4.4.2 *POT1-TPP1* . . . 211
4.5 Annexes méthodes . . . 213
4.5.1 Dichroïsme circulaire . . . 213
4.5.2 Résonance magnétique nucléaire . . . 216

5 Conclusion générale 221
5.1 Les structures G4 aux télomères ne posent pas de problème à la liaison de *RPA* et de *POT1-/+TPP1* . . . 221

5.1.1 Les protéines *SSB-OB fold RPA* et *POT1-/+TPP1* : des modes de liaison différents aux séquences télomériques . . . 221
5.1.2 Une instabilité télomérique observée en présence d'autres structures que des G4s . . . 222
5.2 Conservation des mécanismes et des structures . . . 223
5.2.1 Un contact *TPP1*/ADN conservé au sein des protéines de type *POT1* . . . 223
5.2.2 Une structure conservée au sein des protéines de type *POT1* . . . 224
5.2.3 Un domaine de résolution de jonction de Holliday fonctionnelle au sein de *POT1* ? . . . 225

Bibliographie **228**

6 Annexes Articles **249**
6.1 The binding efficiency of RPA to telomeric G-strands folded into contiguous G-quadruplexes is independent of the number of G4 units . . . 249
6.2 Binding properties of mono- and dimeric pyridine dicarboxamide ligands to human telomeric higher-order G-quadruplex structures . . . 255
6.3 A multi-approach analysis highlights the relevance of RPA-1 as a telomere end-binding protein (TEBP) in *Leishmania amazonensis* . . . 260

7 Annexe technique : purification des protéines **275**
7.1 Vecteurs et cultures cellulaires . . . 275
7.1.1 Les vecteurs *POT1* et *TPP1* . . . 275
7.1.2 Cultures cellulaires . . . 276
7.2 Étapes de purification des protéines et contrôles qualités . . . 276
7.2.1 Le complexe *POT1-TPP1* . . . 276
7.2.2 La protéine *POT1* . . . 286

8 Annexes Figures **293**

Liste des publications

Articles publiés

(2018) Lancrey, A. ; Safa, L. ; **Chatain, J.** ; Delagoutte, E. ; Riou, J.-F. ; Alberti, P. & Saintomé, C. *The binding efficiency of RPA to telomeric G-strands folded into contiguous G-quadruplexes is independent of the number of G4 units* Biochimie, Elsevier BV, 2018, 146, 68-72

(2018) Saintomé, C. ; Alberti, P. ; Guinot, N. ; Lejault, P. ;**Chatain, J.** ; Mailliet, P. ; Riou, J.-F. & Bugaut, A. *Binding properties of mono- and dimeric pyridine dicarboxamide ligands to human telomeric higher-order G-quadruplex structures* Chemical Communications, Royal Society of Chemistry (RSC), 2018, 54, 1897-1900

(2020) Fernandes, C. A. ; Morea, E. G. O. ; dos Santos, G. A. ; da Silva, V. L. ; Vieira, M. R. ; Viviescas, M. A. ; **Chatain, J.** ; Vadel, A. ; Saintomé, C. ; Fontes, M. R. M. & Cano, M. I. N. *A multi-approach analysis highlights the relevance of RPA-1 as a telomere end-binding protein (TEBP) in Leishmania amazonensis* Biochimica et Biophysica Acta (BBA) - General Subjects, Elsevier BV, 2020, 1864, 129607

Articles soumis

(2020) **Chatain, J.** ; Hatem, G. ; Riou, J.-F. ; Alberti, P. & Saintomé, C. *Multiple hPOT1-TPP1 cooperatively coat G-overhang starting from 3'-end and proceeding toward 5', this directionality arises from a 3'-end binding preference and structuration of telomeric DNA into G4* Soumis le xx à Nucleic Acids Research (NAR).

Articles en cours

(2021) **Chatain, J.** ; Delagoutte, E. ; Bouceba, T. ; Vénien-Bryan, C. ; Alberti, P. & Saintomé, C. *Human TPP1 contacts DNA into a contracted horse shoe like conformation adopted by the ternary POT1-TPP1-DNA complex.*

(2021) **Chatain, J.** ; Blond, A. ; Phan, A.T. ; Alberti, P. & Saintomé, C. *(GGGCTA)>8 variant repeats in human telomeres fold into hairpin structures poorly unfolded by Replication Protein A : a possible origin of GGGCTA repeat length-dependent instability.*

Table des figures

1.1 Représentation schématique de l'horloge interne du vieillissement des télomères. 25
1.2 Schéma de l'ADN télomérique du macronoyau chez le cilié *Tetrahymena thermophila* 27
1.3 Schéma de l'ADN télomérique chez la levure *Saccharomyces cerevisiae* 28
1.4 Schéma de l'ADN télomérique chez le mammifère *Homo sapiense* 29
1.5 Les domaines structuraux et fonctionnels de la télomérase. . . 31
1.6 Structure Cryo-EM de la télomérase humaine liée à son substrat. 32
1.7 Cycle d'élongation par la télomérase. 33
1.8 Schéma du *Shelterin* au sein de différentes espèces 36
1.9 Visualisation de la *T-loop* chez les mammifères par microscopie électronique. 38
1.10 Modèle de structure de la *T-loop* chez l'homme. 39
1.11 Le dysfonctionnement des télomères du à la déplétion des protéines du *Shelterin* entraîne une fusion des télomères par *53BP1* 40
1.12 Modèle d'action du *Shieldin* avec *CST* dans le remplissage du brin C. 41
1.13 Schéma d'un motif canonique *OB fold*. 42
1.14 Les domaines structuraux et fonctionnels de *POT1*. 45
1.15 Les domaines structuraux et fonctionnels de *TPP1*. 46
1.16 Structure cristallographique de *POT1* N-term en présence de son ligand d'ADN. 47
1.17 Structure cristallographique de *POT1* C-term en présence de *TPP1*. 47
1.18 Structure cristallographique du domaine *OB fold* de *TPP1*. . . 48
1.19 Enveloppe déterminée par *SAXS* du complexe *POT1-TPP1* en présence d'ADN. 48
1.20 Détail des interactions de chaque nucléotide avec les différents résidus de *POT1*. 50
1.21 Schéma de liaison de *POT1* à son ligand H10 50

1.22 Les domaines structuraux et fonctionnels de *RPA*. 52
1.23 Structure cristallographique des domaines *DBD-A* et *DBD-B* de *RPA*1 en présence d'une polyCytosine (8 nt). 53
1.24 Structure cristallographique des domaines *DBD-C*, *DBD-D* et *DBD-E* de *RPA* assurant la trimérisation du complexe. 54
1.25 Structure cristallographique de *RPA* du champignon *Ustilago maydis* en présence d'un polyT. 55
1.26 Diagramme montrant les contacts de *RPA* avec un polyCytosine (C_8). 57
1.27 Les domaines structuraux et fonctionnels de *CST*. 59
1.28 Structure du complexe *CST* résolue par cryo-microscopie électronique en présence d'ADN, 5'-TAGG-3'. 60
1.29 Modèle de dimérisation du complexe *CST* en présence d'un ADN de 18 nt. 61
1.30 Modèle de la structure G-quadruplexe. 62
1.31 Localisation des structures G-quadruplexes au sein de la cellule. 63
1.32 Polymorphisme structural du G4 de la séquence canonique en présence de sodium. 64
1.33 Polymorphisme structural du G4 de la séquence canonique en présence de potassium. 65
1.34 L'extension simple brin des télomères peut former des structures G-quadruplex impliquées dans la protection terminale des télomères. 67
1.35 Prise en charge par les hélicases des structures G-quadruplexe et de la *T-loop* pour éviter le blocage de la fourche de réplication aux télomères 69
1.36 Modèle d'ouverture d'un G-quadruplexe par *POT1* et *POT1-TPP1*. 71
1.37 Modèle d'ouverture d'un G-quadruplexe par la protéine humaine *RPA*. 72

2.1 Schéma illustrant les sites potentiels de formation de multimères de G4 aux télomères et leur impact au cours de la phase S. 76
2.2 Microscopie électronique des complexes *POT1-TPP1* en présence d'un ADN télomériques (144 nt) 78
2.3 Liaison séquentielle de *POT1-TPP1*. 79
2.4 L'efficacité de liaison du mutant *POT1(H266L)-TPP1* est dépendante de la longueur du substrat de l'ADN simple brin. . . 80

2.5 Modèle de la coopérativité apparente de *POT1-TPP1* sur des longues séquences télomériques formant des multimères de G4 en tandem. 83
2.6 Expérience *EMSA* avec *POT1* en présence de longues séquences télomériques. 106
2.7 Mode de liaison partiel de *POT1* en fonction de la concentration chez l'homme et *S. pombe* 107
2.8 Spectre d'émission de *POT1* en fonction des G4s situés sur la séquence H69. 108
2.9 Cinétique d'ouverture par *POT1* des G4s situés sur la séquence H69. 108
2.10 Modèle de liaison de *POT1* sur des longues séquences capables de former des G4s contigus. 110
2.11 Modèle de la coopérativité de *RPA* sur des séquences formant des multimères de G4 en tandem. 112
2.12 Modèle du "switch" entre *RPA* et *POT1*. 114
2.13 Principe du FRET. 117
2.14 Couple Fluorescéine - TAMRA 118
2.15 Schéma des états ouvert et fermé d'un G4 marqué par fluorochromes. 119
2.16 Schéma de la séquence télomérique H69 marquée par des fluorochromes. 119

3.1 Les domaines structuraux et fonctionnels de *TEBP*α-*TEBP*β . 122
3.2 Structure du complexe *TEBP*α-*TEBP*β avec l'ADN télomérique 124
3.3 Représentation schématique des interactions *TEBP*α-*TEBP*β avec l'ADN simple brin (12 nt). 125
3.4 Comparaison des structures cristallines des motifs OB-fold de TPP1 humaine et de TEBPβ de O. nova 126
3.5 Expérience de photopontage du complexe *POT1-TPP1* à différentes positions le long que l'ADN. 129
3.6 Schéma du protocole MS/MS. 145
3.7 Principe de la résonance de plasmon de surface 147
3.8 Principe de fonctionnement du Biacore 148
3.9 Sensorgramme type issu du Biacore 148
3.10 Structure de la thymine et de la 4-thiothymine. 150
3.11 Schéma d'un microscope électronique à transmission (MET) . 152

4.1 Structure du G4 constitué de répétition du motif non-canonique CTAGGG en potassium 157

4.2 Efficacité de liaison de *RPA* à 50nM entre les séquences CTAGGG et les séquences TTAGGG. 160
4.3 Efficacité de liaison de *RPA* à 50nM entre la séquence H70TCAGGG et la séquence H70TTAGGG 161
4.4 Expérience *EMSA* avec *POT1-TPP1* en présence du motif TTAGGG et du motif CTAGGG. 211
4.5 Expérience d'*EMSA* de liaison de *POT1* et *POT1-TPP1* à des séquences mutées. 212
4.6 La substitution de T2 par U2 empêche *TPP1* d'améliorer l'affinité de *POT1* à son substrat. 213
4.7 Schéma du chemin d'un appareil de spectroscopie par dichroïsme circulaire. 214
4.8 Schéma d'une onde elliptique résultant de deux ondes polarisées circulairement gauche et droite. 215
4.9 Spectre *CD* des différents empilements de plateaux guanines dans le cas d'un G4 216
4.10 Schéma des différentes étapes du principe de la résonance magnétique nucléaire 218
4.11 Schéma des interactions des protons H_1 des guanines 219

5.1 Modèle de la réplication des télomères chez *Borrelia burgdorferi* 227

7.1 Représentation schématique des protéines recombinantes taguées *POT1* et *TPP1* 275
7.2 Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne His-Select. 277
7.3 Gel SDS-Page de *POT1-TPP1* après la colonne His-Select . . 278
7.4 Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne GSH. 279
7.5 Gel SDS-Page de *POT1-TPP1* après la colonne GST 280
7.6 Gel SDS-Page de *POT1-TPP1* après l'étape de digestions . . . 281
7.7 Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne Mono-Q. 282
7.8 Gel SDS-Page de *POT1-TPP1* après la colonne Mono-Q . . . 283
7.9 Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne d'exclusion. 284
7.10 Gel SDS-Page de *POT1-TPP1* après la colonne Superdex 200 285
7.11 Chromatogramme de purification du complexe *POT1* sur une colonne His-Select. 286
7.12 Gel SDS-Page de *POT1* après la colonne His-Select 287

7.13 Chromatogramme de purification du complexe *POT1* sur une colonne His-Select (2). 288
7.14 Gel SDS-Page de *POT1* après la colonne His-Select(2) 289
7.15 Chromatogramme de purification du complexe *POT1* sur une colonne Mono-Q. 290
7.16 Gel SDS-Page de *POT1* après la colonne Mono-Q 291

8.1 Comparaison des domaines *OB fold* en présence de leur ligand. 294
8.2 Comparaison des domaines *OB fold* en présence de leur ligand (suite). 295
8.3 Homologie de structure des domaines *OB fold* de *CTC1*. . . . 296
8.4 Comparaison d'homologie structurale des domaines *OB fold* de *CTC1* entre ceux de *RPA*70 et *Cdc13*. 297
8.6 Schéma des sens de propagation de la polarisation circulaire. . 298
8.5 Expérience de *single molecule* suivant la restructuration d'un G4 en présence de *POT1* 299
8.7 Spectre CD caractéristique d'une hélice α et d'un feuillet β . . 300

Liste des tableaux

1.1 Exemples de séquences télomériques 30
1.2 Tableau récapitulatif des rôles télomériques des protéines du complexe *Shelterin*. 37

Liste des abréviations

aa : acide aminé
***ACD* :** *Adrenocortical Dysplasia Protein*
ADN : Acide désoxyribonucléique (sous forme double brin)
ADP : Adénosine diphosphate
***ALT* :** *Alternative Lengthening of Telomeres*
***ATM* :** *Ataxia-Telangiectasia Mutated*
ATP : Adénosine triphosphate
***ATR* :** *Ataxia-Telangiectasia Related*
***BLM* :** *Bloom syndrom protein*
***CD* :** *Circular dichroism*
***Cdc13* :** *Cell division control protein 13*
***CST* :** *CTC1 - STN1 - TEN1*
***CTC1* :** *Conserved telomere maintenance component 1*
db : double brin
***DBD* :** *DNA-Binding Domain*
***D-loop* :** *Displacement-loop*
***EMSA* :** *Electrophoretic Mobility Shift Assay*
FAM : 6-Carboxyfluorescein
***FRET* :** *Fluorescence Resonance Energy Transfer*
G4 : G-quadruplexe
MET : Microscopie électronique en transmission
nt : nucléotide
***OB* :** *Oligosaccharide-Binding / Oligonucleotide-Binding*
***PIM* :** *POT1 Interaction Motif*
***POT1* :** *Protection of Telomeres 1*

p53 **:** *tumor protein 53*

RMN : Résonance Magnétique Nucléaire

RPA **:** *Replication Protein A*

RTEL1 **:** *Regulator of telomere elongation helicase 1*

RAP1 **:** *Repressor / Activator Protein 1*

RU **:** *Resonance units*

SAXS **:** *Small Angle X-rays Scattering*

sb : simple brin

SDS **:** *Sodium Dodecyl Sulfate*

SPR **:** *Surface Plasmon Resonance*

SSB **:** *Single-strand binding protein*

STN1 **:** *Suppressor of cdc13*

TAMRA : Tetramethylrhodamine

TBM **:** *TIN2 Binding motif*

TEBP **:** *Telomere end binding proteins*

TERC **:** *Telomerase RNA component*

TERT **:** *Telomerase reverse transcriptase*

TEN1 **:** *Telomere length regulation protein*

TIN2 **:** *TRF1 and TFR2 Interactiong Nuclear protein 2*

T-loop **:** *Telomeric loop*

TPP1 **:** *Telomere Protection Protein 1*

TRF1,2 **:** *Telomereic repeat binding faction 1,2*

WH **:** *Winged helix*

wHTH **:** *Winged helix-turn-helix*

WRN **:** *Werner syndrome protein*

53BP1 **:** *p53-binding protein 1*

---

* Les noms en italique correspondent à des mots étrangers.

Chapitre 1

Introduction Générale

1.1 Les rôles biologiques des télomères

Les télomères du grec $\tau\epsilon\lambda o\varsigma$, télos ("la fin") et $\mu\epsilon\rho o\varsigma$, méros ("la partie") désignent les complexes ADN :protéines situés aux extrémités des chromosomes linéaires chez les eucaryotes. Ils ont été nommés ainsi par H.J. Muller en 1938 suite à l'observation chez la drosophile de l'absence de délétions ou d'inversions induites par des irradiations aux rayon X au sein des régions terminales des chromosomes (Muller, 1938). Il suppose alors que les extrémités des chromosomes présentent une structure particulière capable d'assurer la stabilité de ces derniers. Cette hypothèse est reprise en 1939 par B. McClintock qui observe que les chromosomes de maïs présentant des cassures double brin peuvent fusionner, contrairement à leurs extrémités qui restent stables (McClintock, 1939). Ces auteurs présentent les télomères comme ayant pour rôle biologique dans la **protection des chromosomes**, en empêchant d'éventuelles fusions ou la dégradation de leurs extrémités par des exonucléases. Ce rôle est assuré par la présence de structures protectrices, qui seront décrites par la suite, jouant un rôle de "capuchon chromosomique".

En 1961, le biologiste L. Hayflick observe au sein de cultures cellulaires de fibroblastes humains que ces dernières ont un potentiel de réplication limitée à environ 50 divisions, appelé "limite de Hayflick" (Hayflick and Moorhead, 1961). De même il constate que les cellules ayant atteint cette limite entrent dans un état de "sénescence réplicative" qui décrit l'arrêt de la prolifération. James Dewey Watson en 1972 soulève le problème de la réplication terminale des chromosomes ; problème du à l'incapacité de l'ADN polymérase à synthétiser les extrémités des chromosomes linéaires (Watson, 1972). Cette incapacité entraîne une réplication incomplète et par conséquence un raccourcissement des chromosomes. A. Olovnikov propose en 1973 une théorie

qui relie la "sénescence réplicative" au problème de réplication des extrémités des chromosomes (Olovnikov, 1973). Il explique que l'érosion des télomères due aux divisions cellulaires successives mène à l'arrêt de la prolifération. Il met ainsi en relation le vieillissement avec la longueur des télomères. Cette relation décrit le raccourcissement des télomères comme une "horloge interne du vieillissement" (Figure 1.1) et confère ainsi un nouveau rôle aux télomères à savoir le **contrôle du nombre de divisions cellulaires**.

L'équipe d'E. Blackburn (Prix Nobel de Physiologie ou Médecine en 2009) découvre en 1985 chez le cilié *Tetrahymena* un complexe enzymatique capable de compenser le raccourcissement progressif des extrémités chromosomiques dû aux divisions cellulaires successives (Greider and Blackburn, 1985). Nommé télomérase, ce complexe enzymatique se compose d'une protéine de type transcriptase inverse et d'un fragment d'ARN matrice (voir section 1.3). De ce fait la télomérase permet l'élongation de l'extrémité simple brin des télomères, et ainsi compense l'érosion des chromosomes due à la prolifération cellulaire, évitant l'entrée en "sénescence réplicative" de la cellule. En 1989 G. B. Morin identifie une activité télomérase dans des extraits de cellules humaines (Morin, 1989). La mise au point d'un test pour la détection de l'activité télomérase, le TRAP (« Telomeric Repeat Amplification Protocol »), a permis de montrer que la télomérase est active dans la majorité (>85%) des cellules issues de biopsies tumorales, à la différence des cellules somatiques normales (Kim et al., 1994). La télomérase est active également au sein des cellules germinales (Bekaert et al., 2004). Quand les télomères atteignent une longueur critique, la cellule arrête de se diviser et entre dans un état de senescence. La télomérase, présentée par certains comme la "fontaine de jouvence", joue un rôle important au sein des cellules tumorales pour la sortie de "crise". L'absence de p53 (*tumor protein 53*, facteur de transcription régulant notamment l'apoptose (mort cellulaire)), autorise la prolifération cellulaire au-delà de la "limite de Hayflick". Dans ce cas les télomères de la cellule atteignent une longueur critique définissant un seuil appelé "crise" qui engendre une instabilité génomique (Figure 1.1). La crise est caractérisée par une mort cellulaire massive. Cependant, certaines cellules échappent à la crise en réactivant la télomérase, ce qui leur permet de maintenir une longueur minimale des télomères et de continuer à proliférer.

En 1995, l'équipe de R. R. Reddel observe que des cellules immortalisées de fibroblastes humains maintiennent la taille de leurs télomères sans activité détectée de la télomérase (Bryan et al., 1995). Ils mettent en évidence un nouveau mécanisme d'élongation des télomères, appelé *ALT* (*Alternative Lengthening of Telomeres*), qui repose sur le principe de la recombinaison homologue entre télomères. Ce mécanisme alternatif d'élongation indépendant de la télomérase représente une fraction importante (10 à 15 %) des cellules

cancéreuses (Cesare and Reddel, 2010).

Enfin, une étude réalisée en 2002 par l'équipe de T. de Lange observe que l'induction de la "sénescence réplicative" dépend plus de la présence de structures protectrices que de la longueur des télomères (Karlseder, 2002). Ces structures sont induites par la présence d'une série de protéines se fixant aux répétitions télomériques que nous présenterons par la suite. Cette étude révèle donc l'existence d'une relation entre **protection des télomères** et "sénescence réplicative".

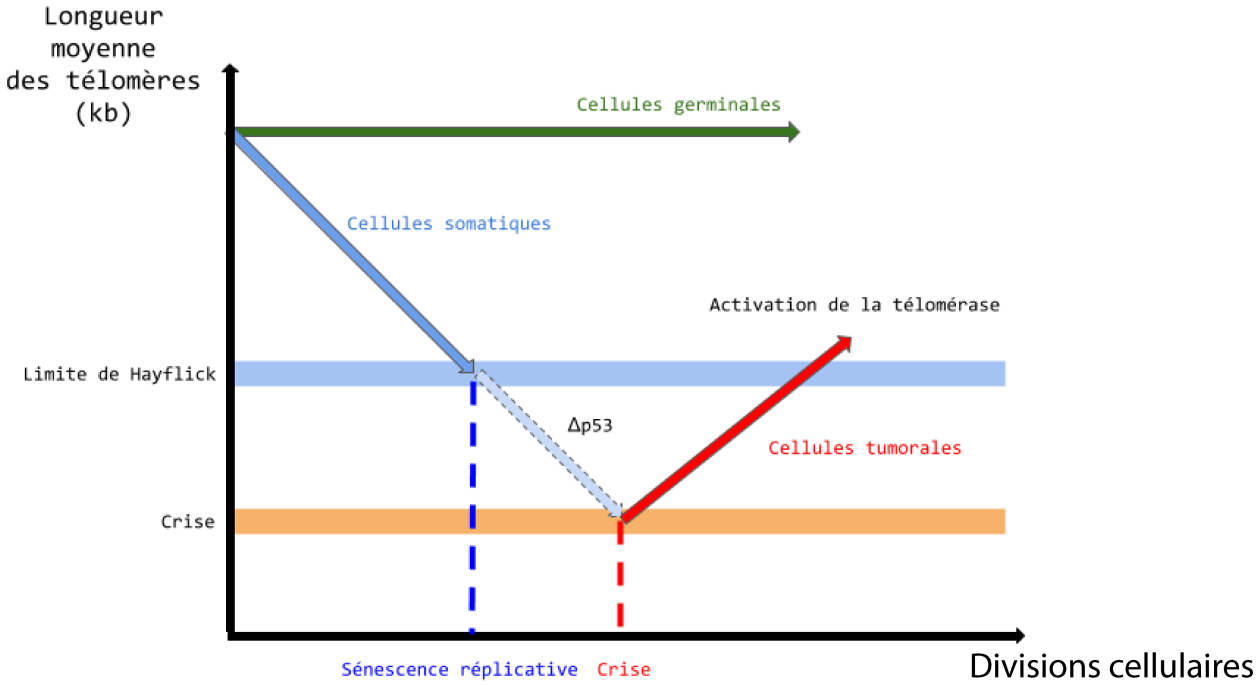


FIGURE 1.1 – **Représentation schématique de l'horloge interne du vieillissement des télomères.**

En résumé, les télomères assurent des fonctions essentielles au sein de la cellule de par leur longueur et leur structure. Parmi ces fonctions nous pouvons citer **le contrôle du nombre de divisions cellulaires** et **la protection contre la dégradation par des nucléases et contre la fusion entre télomères**. Les télomères permettent ainsi de résoudre le "problème de réplication des extrémités" (Wynford-Thomas and Kipling, 1997) et le "problème de protection des extrémités" (de Lange, 2009).

1.2 L'ADN télomérique

Chez la majorité des espèces, l'ADN télomérique se compose de courtes séquences répétées en tandem, appelées motif télomérique (Table 1.1) (Wellinger and Sen, 1997). Le nombre de répétitions varie selon les espèces, entre les individus, et même entre différents chromosomes dans la même cellule. Parmi tous les organismes étudiés, le brin d'ADN constitué de guanines successives (Brin G) est plus long que le brin complémentaire (Brin C), et forme ainsi une extension simple brin en 3' du chromosome appelée *G-overhang*. Si le concept de télomère a été posé à partir des observations faites chez la drosophile, cette espèce fait partie de celles qui dérogent à ces règles générales de constitution des télomères. En effet la drosophile présente des rétrotransposons spécifiques des télomères, c'est-à-dire des séquences d'ADN capables de se multiplier et de se déplacer au sein du génome.

1.2.1 Au sein des ciliés

Le premier motif télomérique identifié est celui du protozoaire cilié *Tetrahymena thermophila* (Blackburn and Gall, 1978). Les téloméres de cet organisme se composent de 50 à 70 répétitions du motif TTGGGG (T_2G_4), avec une extension du brin G en 3' de 12 à 16 nucléotides (Figure 1.2). Il est important de noter que chez les ciliés deux types de noyaux coexistent : le micronoyau et le macronoyau. Le micronoyau est diploïde et porte l'information génétique. Il est transcriptionnellement inactif. A l'inverse, le macronoyau est transcriptionellement actif et se compose de chromosomes fragmentés qui présentent également des extensions simple brin du motif télomérique (Klobutcher et al., 1981). Une étude réalisée par l'équipe de Carolyn M. Price montre que la majorité des extrémités des chromosomes du macronoyau présents chez le cilié *Tetrahymena thermophila* se terminent par la séquence 5'-TTGGGGT-3' (Jacob et al., 2001). Ils n'observent aucun changement significatif de la longueur de l'extrémité au cours du cycle cellulaire. Des changements ont été observés lorsque les cellules entrent dans un état de "crise", démontrant un équilibre modifié entre la synthèse de l'ADN et la dégradation ou la protection des extrémités. Ces résultats montrent que le maintien de la longueur des télomères dans le macronoyau est dû à la fois à la protection de l'extrémité des chromosomes, mais aussi à l'activité de la télomérase découverte par E. Blackburn et C. Greider (Greider and Blackburn, 1985).

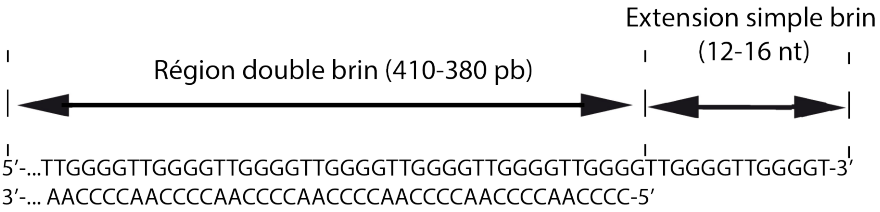


FIGURE 1.2 – **Schéma de l'ADN télomérique du macronoyau chez le cilié Tetrahymena thermophila**

1.2.2 Au sein des levures bourgeonnantes

Comme chez les ciliés, les télomères de la levure *S. cerevisiae* sont constitués d'une courte extension de 12 à 15 bases (Figure 1.3.A). Une étude réalisée par R. J. Wellinger et al, montre que la taille des extrémités chez *Saccharomyces cerevisiae* varie pendant le cycle cellulaire (Wellinger et al., 1993). Cette étude révèle par la technique d'"hybridation *Southern* non dénaturant", capable de détecter environ 60 à 300 bases d'ADN TG_{1-3}, que les télomères de levure acquièrent et perdent l'extrémité 3' simple brin de façon dépendante du cycle cellulaire. Notamment pendant la phase S, les levures présentent une augmentation de l'extension télomérique du brin G de plus de 30 nucléotides (Figure 1.3.B). Ces extensions ont été générées *in vivo* dans des cellules dépourvues de télomérase (Wellinger et al., 1996). Wellinger et al. en 1996 montrent que, en *S. cerevisiae*, des *G-overhang* peuvent être générées indépendamment de la télomerase et que leur longueur varie au cours du cycle cellulaire ; ils émettent alors l'hypothèse de l'existence d'une exonuclease spécifique du brin C, dont l'activité serait probablement régulée par *Cdc13* (Wellinger et al., 1996).

A l'inverse des ciliés, les télomères de *S. cerevisiae* se composent d'environ 300 paires de base de répétitions dégénérées qui comportent un nombre variable de G avec comme séquence consensus $G_{2-3}(TG)_{1-6}$ (Forstemann et al., 2000). D'autres levures bourgeonnantes ont des séquences télomériques régulières composées de la répétition de longs motifs d'une vingtaine de nucléotides. Toutefois les ciliés et les levures présentent une protection similaire de par la présence d'une structure semblable aux extrémités, à savoir une structure composée d'une extrémité simple brin (12-20 nucléotides chez les ciliés, de quelques dizaines de nucléotides chez *S. cerevisiae*) à laquelle est associée un ensemble de protéines spécifiques qui forment le "capuchon télomérique". Nous présenterons par la suite ces protéines spécifiques.

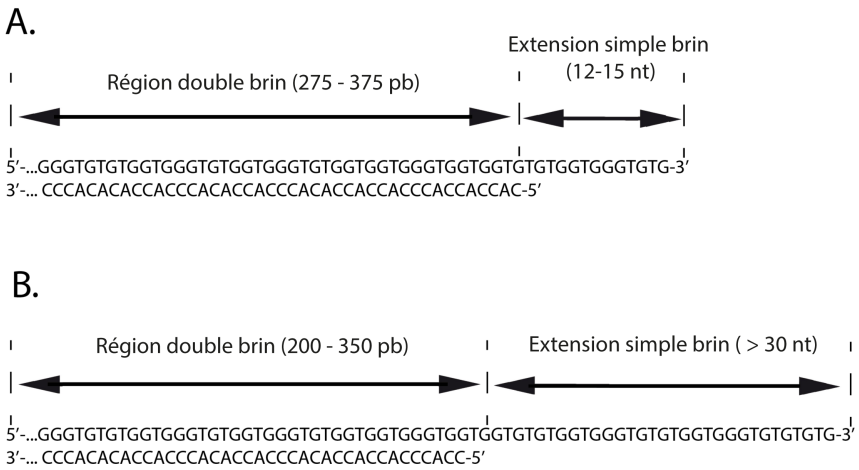


FIGURE 1.3 – **Schéma de l'ADN télomérique chez la levure Saccharomyces cerevisiae**
A. Schéma du télomère chez la levure *Saccharomyces cerevisiae*. **B.** Schéma du télomère chez la levure *Saccharomyces cerevisiae* pendant la phase S.

1.2.3 Au sein des mammifères

Contrairement aux organismes vus précédemment, les chromosomes des mammifères présentent un plus grand nombre de répétions télomériques et des extensions simple brin plus longues. Par exemple chez l'homme, la taille moyenne des télomères varie entre 2 à 15 kpb pour la région double brin et entre 50 à 300 nucléotides pour la région simple brin (Wright et al., 1997) (Figure 1.4). Cette extension en 3' du brin G est présente pendant tout le cycle cellulaire. Le maintien de la longueur de cette extrémité simple brin, comme pour le levure *Saccharomyces cerevisiae*, est assuré par l'activité d'exonucléases entrainant la dégradation du brin complémentaire (Brin C) (Makarov et al., 1997). (ICIICICI) En complément nous avons vu que la présence de l'activité de la télomérase au sein de cellules germinales, ou encore dans des lignées cellulaires tumorales, permet l'élongation du *G-overhang* (Morin, 1989). En revanche la télomérase est absente ou présente en très faible quantité dans les cellules somatiques normales. Suite à l'élongation par la télomérase du brin G, ou à la dégradation du brin C par des exonucléases, on retrouve une étape dite de remplissage du brin C (*C strand fill-in*) (Feng et al., 2017). La synthèse du brin C est médiée par un complexe *CST*, qui sera présenté par la suite, et le recrutement de l'ADN Polymérase *Pol*α. Cette étape permet de maintenir une longueur à-peu-près constante du *G-overhang*.

Contrairement aux ciliés et aux levures, la séquence télomérique est conservée chez tous les mammifères puisqu'on retrouve le même motif télomérique

à savoir TTAGGG (McElligott and Wellinger, 1997). De plus, à la différence des ciliés et des levures, l'extension simple brin des mammifères se compose de plusieurs centaines de nucléotides, ce qui conduit à une différence notable avec les ciliés et les levures. Cette différence s'exprime dans la formation et l'organisation d'une structure protectrice particulière. Cette structure particulière ainsi que les protéines participant à sa formation seront présentées dans la suite de la thèse.

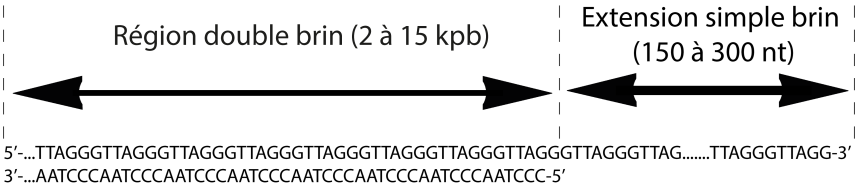


FIGURE 1.4 – **Schéma de l'ADN télomérique chez le mammifère Homo sapiens**

1.2.4 Le motif télomérique

En résumé, nous avons vu que l'ADN télomérique au sein de la majorité des espèces est constitué d'une région double brin avec une extension simple brin en 3'. La longueur de l'ADN télomérique varie selon les espèces (Wellinger and Sen, 1997). Les séquences télomériques sont composées de répétitions d'un motif en tandem, qui varie également selon les espèces (Table 1.1). Pour expliquer cette diversité dans les séquences télomériques on admet l'hypothèse de l'existence d'un motif télomérique ancestral duquel les autres motifs télomériques découleraient par la résultante de modifications de séquences de la matrice ARN de la télomérase. Parmi les arguments qui défendent cette hypothèse on peut citer l'existence de forte similarité entre les différents motifs. En effet on retrouve dans la très grande majorité des motifs une sur-représentation des thymines et guanines, ainsi qu'une absence quasi-totale de cytosine. Les exceptions présentant des cytosines dans le motif télomérique sont les télomères des levures ou encore ceux des nématodes qui présentent un dimère de guanines suivi d'une cytosine (TTAGGC). Une autre caractéristique conservée dans la majorité des motifs est la présence de guanines consécutives au sein de la séquence. On retrouve un nombre de guanines consécutives allant de G_2 pour les nématodes jusqu'à G_6 chez *Saccharomyces cerevisiae*. Il est admis de retenir comme motif télomérique ancestral la séquence TTAGGG en raison de sa sur-représentation au sein des espèces.

Organismes	Séquences répétées	Abréviations
Protozoaires		
Tetrahymena thermophila	TTGGGG	T_2G_4
Oxytricha nova	TTTTGGGG	T_4G_4
Euplotes crassus	TTTTGGGG	T_4G_4
Trypanosoma brucei	TTAGGG	T_2AG_3
Plantes		
Arabidopsis thaliana	TTTAGGG	T_3AG_3
Chalamydomonas	TTTTAGGG	T_4AG_3
Nématodes		
Ascaris lumbricoides	TTAGGC	T_2AG_2C
Caenorhabitis elegans	TTAGGC	T_2AG_2C
Vertébrés		
Homo sapiens	TTAGGG	T_2AG_3
Mus musculus	TTAGGG	T_2AG_3
Gallus gallus	TTAGGG	T_2AG_3
Ovis aries	TTAGGG	T_2AG_3
Insectes		
Bombyx mori	TTAGGG	T_2AG_3
Champignons		
Neurospora crassa	TTAGGG	T_2AG_3
Candida albicans	ACGGATGTCTAACTTCTTGGTGT	
Kluyveromyces lactis	ACGGCTTTGATTAGGTATGTGGTGT	
Saccharomyces cerevisiae	$T(G)_{2-3}(TG)_{1-6}$	
Schizosaccharomyces pombe	$TTAC(A)G_{2-3}$	

TABLE 1.1 – **Exemples de séquences télomériques**

1.3 La télomérase

La télomérase est une enzyme impliquée dans le maintien de la longueur de l'extrémité simple brin des chromosomes. La télomérase est un complexe ribo-nucléoprotéique constitué d'une sous-unité catalytique, *TERT* (*Telomerase reverse transcriptase*) et d'un ARN qui porte la matrice pour la synthèse des répétitions télomériques, *TERC* (*Telomerase RNA component*) (Figure 1.5), et de diverses protéines nécessaires à sa biogenèse et régulation. La sous-unité *TERT* possède un domaine catalytique *reverse transcriptase* (*RT*), assurant la synthèse du brin G, un domaine N-terminal (*TEN*) liant l'ADN télomérique et un domaine de liaison à *TERC* (*TRBD*). La sous-unité *TERC*

humaine (ou *hTR*), composée de 451 nucléotides, hautement structurée (présence de nombreuses régions en tige-boucle et en pseudonœud), possède du coté 5', près du pseudoneud(*PK*, pour *pseudo-knot*), la matrice (t, pour *template*), qui permet la synthèse de la répétition télomérique. Chez les mammifères la séquence de la matrice est 3'-CAAUCCCAAUC-5'.

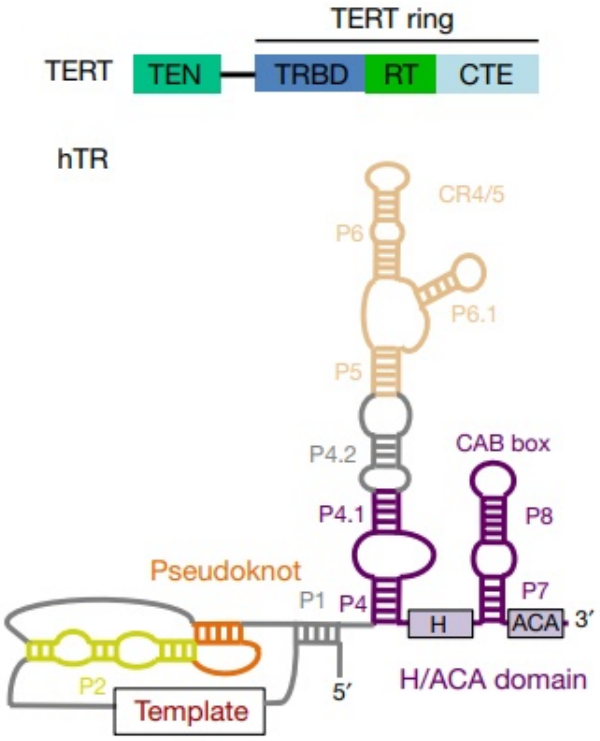


FIGURE 1.5 – **Les domaines structuraux et fonctionnels de la télomérase.** (En haut) Les domaines de la sous-unité *TERT*, (En bas) La structure secondaire de la sous-unité *hTR* (ou *TERC*) (Nguyen et al., 2018)

La région t/PK est essentielle à l'activité *in vitro* de la télomérase. Les régions tige-boucle localisées du côté 3' de l'ARN vont permettre d'ancrer un certain nombre de protéines essentielles pour l'assemblage et l'activité *in vivo* de la télomérase. Une étude réalisée par de Nguyen et al. en 2018 a résolu par cryo-microscopie électronique la structure de la télomérase humaine liée à son substrat (Figure 1.6) (Nguyen et al., 2018) Le complexe est composé de deux lobes dans lesquels l'ARN *hTER* fait office d'échafaudage pour différentes protéines. L'un des lobes constitue le cœur catalytique et contient le domaine t/PK de *TERC* et les domaines *TEN* et *RT* de *TERT*. L'autre lobe contient la région 5' de *TERC* et les différentes protéines intervenant dans la biogenèse

et la régulation de la télomérase.

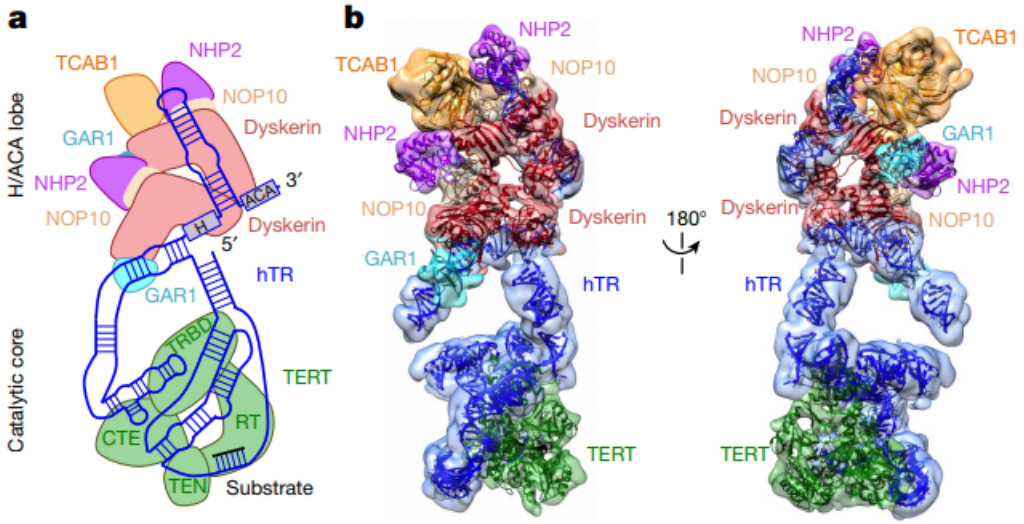


FIGURE 1.6 – **Structure Cryo-EM de la télomérase humaine liée à son substrat. a.** Schéma de la disposition des sous-unités. **b** Reconstructions par cryo-microscopie électronique pour la région H/ACA à 8,2 Å et le noyau catalytique à 7,7 Å. Les sous-unités utilisent un code couleur comme indiqué (Nguyen et al., 2018).

C. Greider et E. Blackburn ont été les premières à proposer un modèle de maintien de la taille des télomères par la télomérase (Blackburn, 1992). Dans ce modèle, la télomérase se sert de son ARN *TERC* comme matrice pour ajouter les nucléotides à l'extrémité des chromosomes. Le mécanisme consiste en trois étapes (Autexier and Lue, 2006) :

I- Étape d'association télomères-télomérase et liaison de la matrice ARN à l'extrémité 3' simple brin de l'ADN télomérique par complémentarité de séquence (Figure 1.7.a).

II- Étape d'élongation par transcription inverse de la matrice ARN et addition de nucléotides à l'extrémité 3' de l'ADN télomérique(Figure 1.7.b).

III- Étape de translocation : une fois la transcription inverse de la séquence matrice ARN achevée, une translocation du complexe de la télomérase a lieu pour permettre un autre cycle de synthèse (Etape II) (Figure 1.7.c-d).

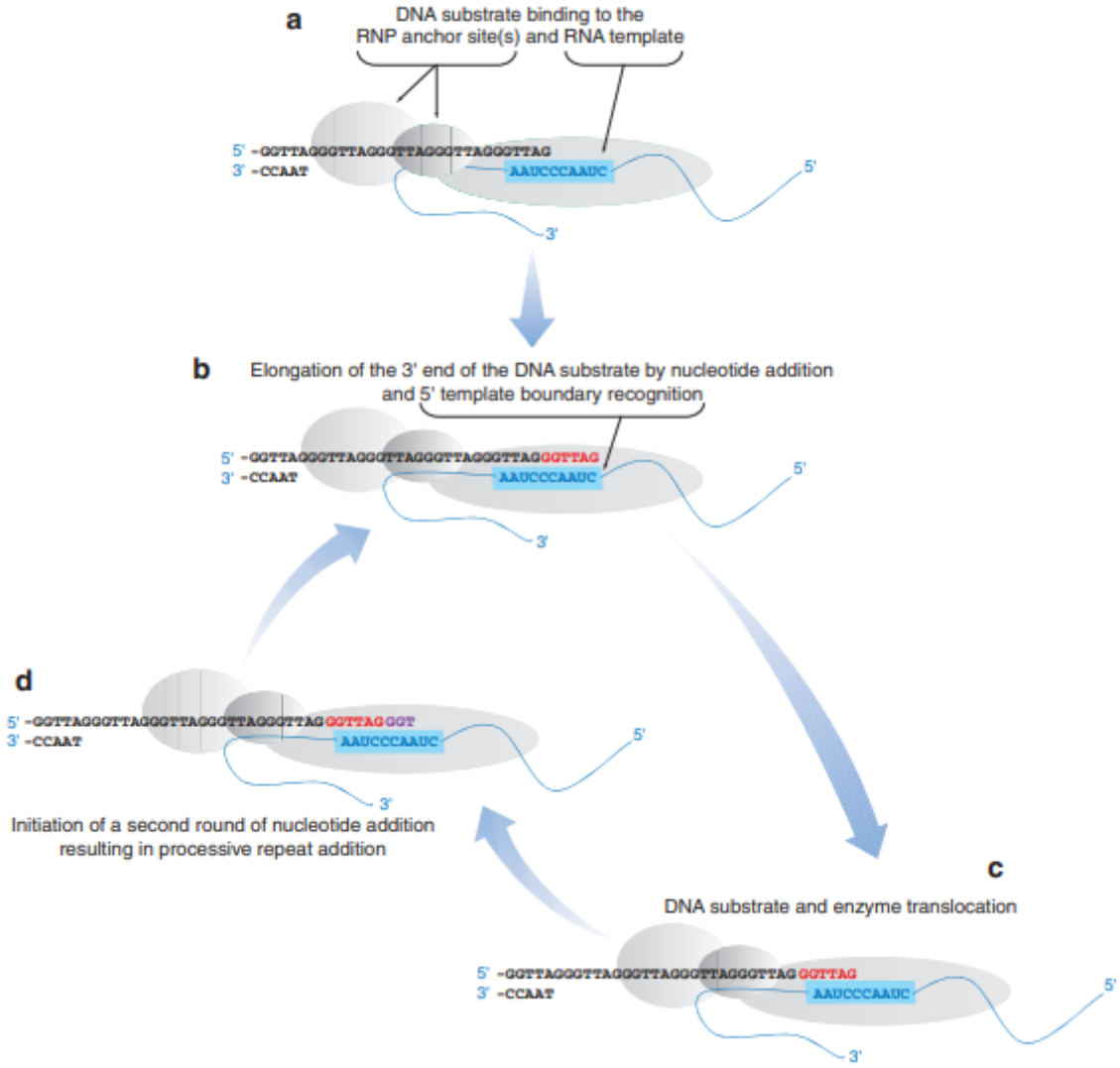


FIGURE 1.7 – **Cycle d'élongation par la télomérase.** (Autexier and Lue, 2006)

Ce mécanisme permet à la télomérase d'ajouter plusieurs répétitions télomériques, en faisant une protéine processive. Toutefois, si la télomérase présente une processivité importante *in vitro*, il n'en est pas de même *in vivo*. Il a été montré que des protéines telles que *TPP1* (Wang et al., 2007; Rajavel et al., 2014; Xu et al., 2019) ou encore *TIN2* (ho Kim et al., 1999; Chiang et al., 2004; Frank et al., 2015) augmentent la processivité de la télomérase, suggérant que ces protéines puissent jouer le rôle de facteur de

processivité pour la télomérase lors d'élongation du télomère. Il est important de noter que l'activité de la télomérase peut être inhibée notamment si cette dernière n'a pas accès à son substrat. La présence de la protéine de liaison à l'ADN simple brin télomérique *POT1* (section 1.5.2), capable de séquestrer l'extrémité 3' du *G-overhang*, peut par conséquent participer à l'inhibition de la télomérase (Zaug et al., 2005; Xu et al., 2019). De même l'extension simple brin télomérique constituée de guanines successives, en raison de la répétition du motif télomérique, est capable d'adopter une structure appelée G4, dont la présence empêche la télomérase d'accéder à son substrat (Zhou et al., 2016; Tan and Lan, 2020).

1.4 Le capuchon télomérique

Nous avons vu précédemment que le terme télomère désigne des complexes ADN-protéines situés aux extrémités des chromosomes. Ces complexes participent à la protection des extrémités des chromosomes en empêchant notamment l'activité de dégradation par des nucléases ou les événements de fusion entre télomères. En effet, pendant le cycle cellulaire, des protéines sont en charge de s'assurer du bon déroulement de l'interphas,(phases G1, S and G2). Parmi ces protéines nous pouvons citer la kinase *ATM* (*Ataxia-Telangiectasia Mutated*) et la kinase *ATR* (*Ataxia-Telongiesctasia Related*) qui participent aux voies de détection de dommages à l'ADN (points de contrôle). Ces dernières ne doivent pas signaler les extrémités des chromosomes comme des cassures de l'ADN et entraîner l'arrêt du cycle cellulaire. Ces phénomènes participent à ce qu'on appelle le "problème de protection des extrémités" (de Lange, 2009). Une réponse partielle apportée par la cellule au problème de la protection des extrémités consiste à "masquer" l'extrémité du chromosome à l'aide de structures particulières dues à la présence de protéines spécifiques de liaison à l'ADN télomérique.

1.4.1 Le complexe du *Shelterin*

Le complexe du *Shelterin* désigne l'ensemble des protéines spécifiques permettant de former une structure nucléoprotéique protectrice dite capuchon télomérique. Dans un modèle "linéaire" des extrémités des chromosomes, le complexe du *Shelterin* est capable d'assurer la protection de ces extrémités. Cet ensemble de protéines est propre à chaque espèce (Figure 1.8). On remarque que le complexe du *Shelterin* varie entre les espèces tant par le nombre de protéines le constituant que par son organisation. Derrière cette diversité apparente des complexes du *Shelterin* on peut remarquer des simi-

litudes. On retrouve au sein des différentes espèces étudiées la présence de protéine(s) spécifique(s) de liaison à l'ADN simple brin assurant la protection de l'extension 3' du chromosome (Figure 1.8). Par exemple : *TEBP*α (*Telomere End Binding Proteins*) chez *Oxytricha nova* :, *Cdc13* chez *Saccharomyces cerevisiae*, *Pot1* chez *Schizosaccharomyces pombe* ou encore *POT1* chez *Homo sapiens*. Certaines de ces protéines de liaison sont l'objet de notre étude et seront présentées plus en détail. De même on peut observer la présence de protéine(s) de liaison spécifique de la région double brin dans la majorité des espèces (Figure 1.8). On retrouve respectivement chez *Saccharomyces cerevisiae* et *Schizosaccharomyces pombe* les protéines *Rap1* et *Taz1* ; ainsi que chez les mammifères, notamment chez l'homme, deux homodimères capable de lier de manière spécifique l'ADN télomérique double brin à savoir *TRF1* et *TRF2* (*Telomeric Repeat-binding Factors*). Ces protéines protègent entre-autre la région sensible de la jonction télomérique (double brin / simple brin) en raison du fait qu'elles peuvent être détecté comme une cassure de l'ADN. En plus des protéines de liaison à l'ADN télomérique on retrouve un ensemble de protéines qui assurent la connexion entre les protéines de liaison à l'ADN télomérique simple brin (*Single-Strand Binding*(*SSB*)) et celles qui lient l'ADN télomérique double brin (*Double-Strand Binding*(*SSB*)) . Chez *Schizosaccharomyces pombe* la jonction entre les *SSB* et *DSB* est établi par *Rap1*,*Poz1* et *Tpz1* ; tandis que chez *Homo sapiens* la connexion est assurée par *TIN2* et *TPP1*. Ces protéines qui relient ces deux régions de l'ADN télomérique jouent un rôle crucial dans la formation de structures protectrices spécifiques des télomères qui seront présentées par la suite.

L'absence de protéines de liaison double brin chez les ciliés peut être liée à la non découverte de celles-ci ou s'expliquer par la nature de l'ADN télomérique : présence d'une courte extension (12 à 16 nt) pendant tout le cycle cellulaire. Il est admis que la protection des extrémités des chromosomes présents dans le macronoyau est assurée par le complexe *TEBP*α-*TEBP*β. Ce complexe permet la formation d'une structure non-canonique inter chromosomique de l'ADN télomérique, à savoir un G-quadruplex (la structure G-quadruplex sera présentée par la suite). On peut se demander si cette structure non canonique inter chromosomique n'est pas suffisante pour assurer la protection des télomères chez les ciliés.

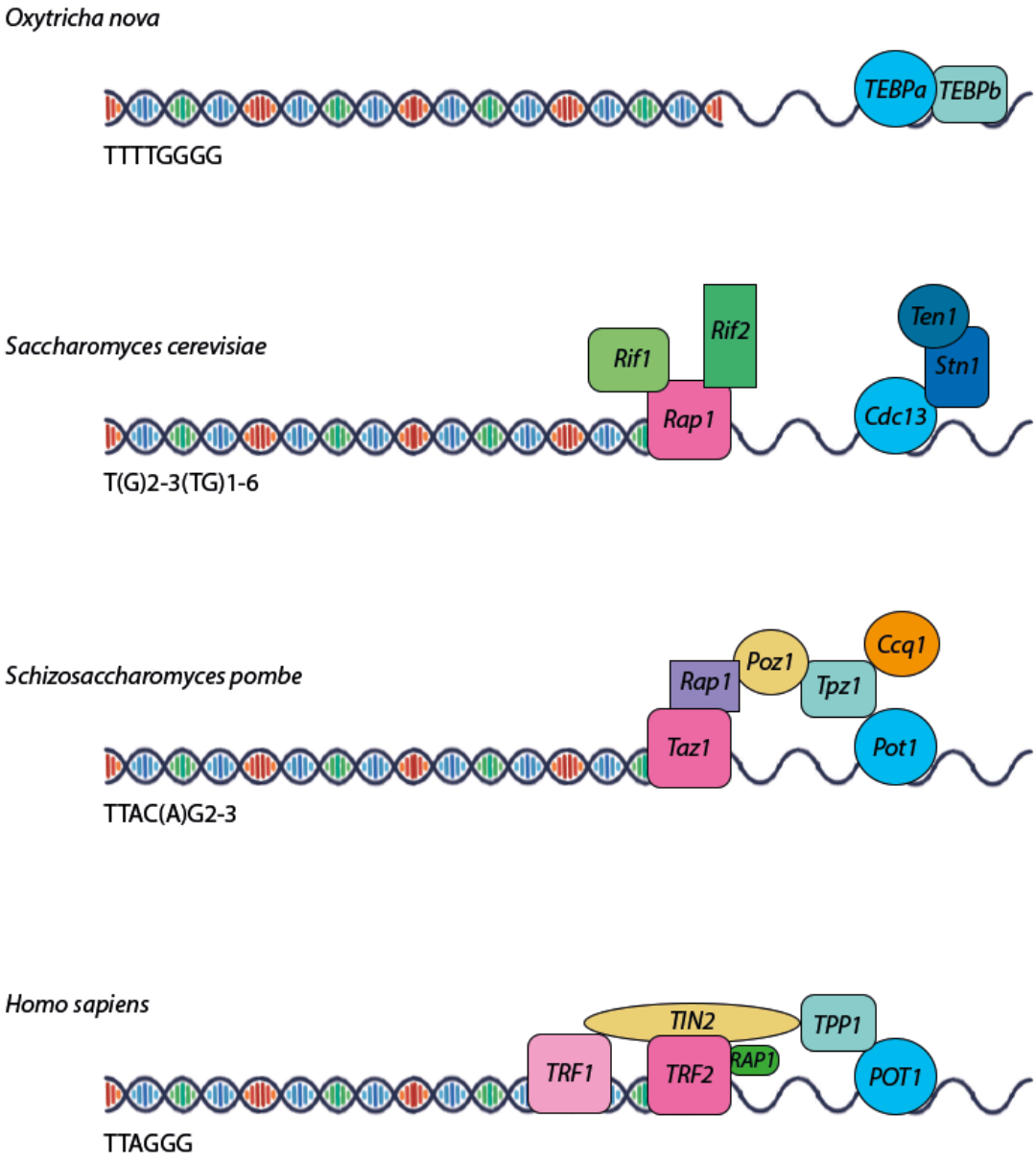


FIGURE 1.8 – **Schéma du Shelterin au sein de différentes espèces.**

Le shelterin humain

Mon travail de thèse porte principalement sur l'étude des télomères chez l'homme. C'est pourquoi je me propose de faire une description rapide du *Shelterin* humain. Le *Shelterin* chez l'homme se compose de six protéines : *TRF1*, *TRF2*, *Rap1*, *TIN2*, *TPP1* et *POT1*, assemblées comme dans la figure 1.8 (de Lange, 2005). Le complexe du *Shelterin* interagit avec l'ADN télomérique, par l'intermédiaire de *TRF1* et *TRF2* dans la région double brin et *POT1* pour la région simple brin. Les rôles spécifiques de chaque protéine sont présentés ci-dessous dans la Table 1.2. Le complexe du *Shelterin* chez l'homme participe à la protection des télomères notamment en inhibant les voies de réponses aux dommages de l'ADN ; mais également à la régulation de leur taille. Pour conclure, il est intéressant de noter que le *Shelterin* forme un complexe stable en absence de l'ADN télomérique (Liu et al., 2004).

Protéine	Fonctions
TRF1	Protection des télomères et recombinaison Régulation de la taille des télomères Inhibition de la réponse aux dommages de l'ADN Réplication des télomères
TRF2	Protection des télomères et recombinaison Régulation de la taille des télomères Inhibition de la réponse aux dommages de l'ADN
POT1	Protection du brin G Inhibition de la réponse aux dommages de l'ADN
TIN2	Protection des télomères et recombinaison Régulation de la taille des télomères
Rap1	Protection des télomères et recombinaison Régulation de la taille des télomères Inhibition de la réponse aux dommages de l'ADN Répression des gènes subtélomériques Régulation transcriptionnelle
TPP1	Protection des télomères et recombinaison Régulation de la taille des télomères Recrutement de la télomérase

TABLE 1.2 – **Tableau récapitulatif des rôles télomériques des protéines du complexe Shelterin.**

1.4.2 La *T-loop*

L'équipe de Titia de Lange a observé en 1999 par microscopie électronique en présence d'extraits purifiés d'ADN de mammifère la formation de larges boucles aux extrémités (Figure 1.9) (Griffith et al., 1999). En effet ils remarquent que les extrémités des chromosomes sont repliées sur elles-mêmes formant des boucles qu'ils appellent *T-loop* (*Telomere loop*). On observe la formation de ces boucles en présence des protéines du *Shelterin* et de l'ADN télomérique humain mais également de la souris, suggérant la conservation de la structure de la *T-loop* chez les mammifères. La taille de ces larges boucles est cohérente avec la longueur connue des télomères.

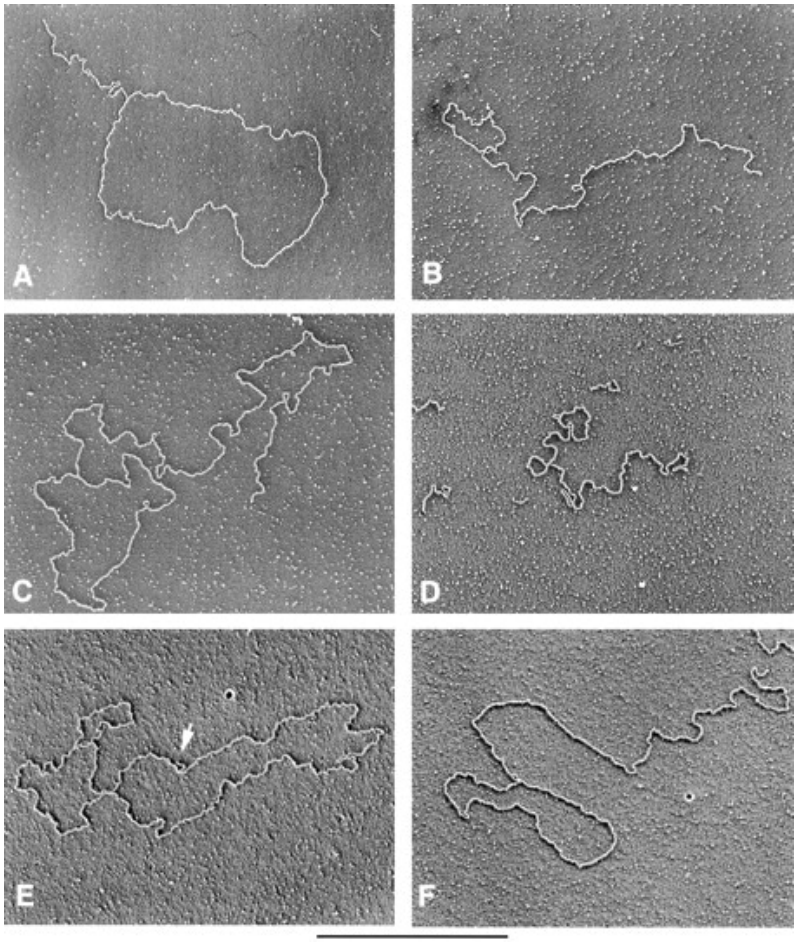


FIGURE 1.9 – **Visualisation de la T-loop chez les mammifères par microscopie électronique.** En présence d'ADN télomérique :**A-C.** de cellules HeLa ; **D.** Human PbLs ; **E-F.** de cellules de souris. (Griffith et al., 1999)

Il est connu que l'inhibition de *TRF2* entraîne une déprotection immé-

diate des extrémités chromosomiques qui se manifeste par : la perte de l'extrémité télomérique 3', l'activation de p53 et la présence d'événements de fusions chromosomiques dans la région des télomères (Biroccio et al., 2006). Ici les expériences de microscopie électronique ont montré que *TRF2* remodèle l'ADN télomérique linéaire en formant ces larges boucles (Figure 1.9.A-C). En effet, en absence de *TRF2* on assiste à une dégradation de l'ADN télomérique qui témoigne d'une déprotection de la région télomérique ce qui est cohérent avec d'autres études biologiques réalisées. L'équipe de Titia de Lange propose un modèle pour expliquer la formation de la *T-loop* basé sur l'invasion de l'extension télomérique dans la région double brin (Figure 1.10). Bien que le site exact du point d'invasion de l'extrémité 3' n'a pas été établi, dans la plupart des molécules la taille de la boucle correspond à plusieurs kpb et dans certains cas englobe clairement le télomère entier. Dans le modèle cette invasion par l'extrémité simple brin est stabilisée par la présence des protéines du *Shelterin* qui assurent la cohésion de la *T-loop*. De même la protéine de liaisons à l'ADN simple brin, à savoir*POT1*, avec son partenaire *TPP1*, assure la protection du simple brin au sein de la structure appelée *D-loop* (*Displacement-loop*). La structure de la *T-loop* permet de "masquer" l'extension simple brin et par conséquent de participer à la protection des extrémités des chromosomes notamment contre les dégradations induites par des nucléases.

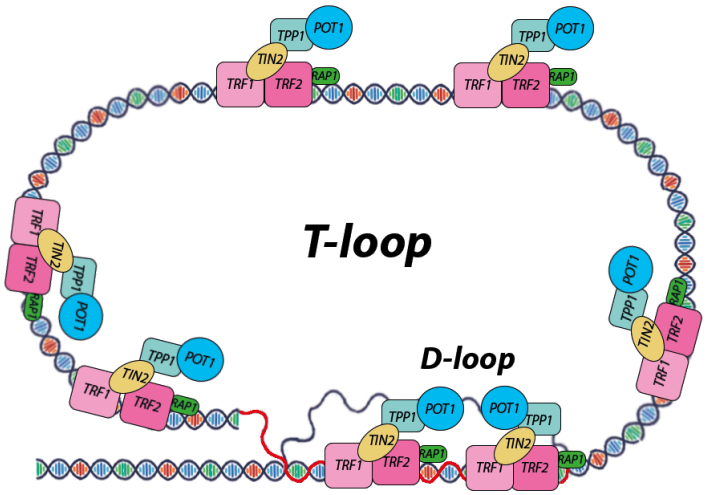


FIGURE 1.10 – **Modèle de structure de la T-loop chez l'homme.**

1.4.3 Le complexe du *Shieldin*

Une récente découverte a révélé l'existence d'un complexe protéique possédant une activité de liaison à l'ADN simple brin et impliqué dans la protection des télomères. Ce complexe désigné sous le nom de *Shieldin* se compose de quatre : *SHLD1*, qui assure la liaison à l'ADN simple brin par la présence de domaines *OB fold*, élément protéique de reconnaissance de l'ADN simple brin, *SHLD2*, *SHLD3* et la protéine *REV7* (Setiaputra and Durocher, 2019). Le complexe du *Shieldin* favorise de nombreuses activités liées à *53BP1* (*p53-binding protein 1*) qui joue un rôle crucial dans la signalisation et la réparation des cassures doubles brin par le *NHEJ Non-Homologous End-Joining* (Panier and Boulton, 2013). Par conséquent en absence du complexe *Shelterin*, le *Shieldin* grâce à la présence de *53BP1* permet la fusion des télomères dysfonctionnels offrant une voie de sortie de crise à la cellule (Figure 1.11.A).

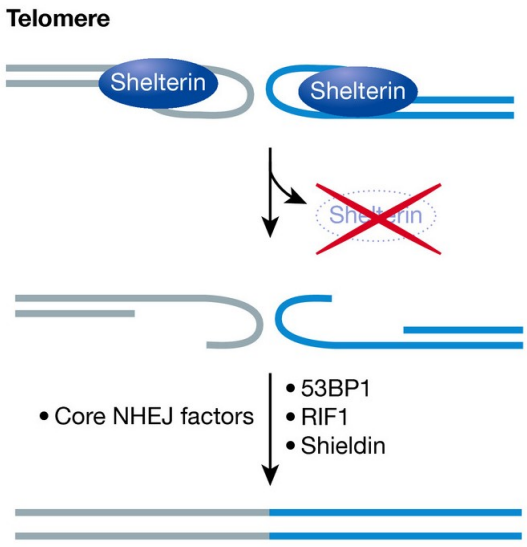


FIGURE 1.11 – **Le dysfonctionnement des télomères du à la déplétion des protéines du *Shelterin* entraîne une fusion des télomères par *53BP1*** (Setiaputra and Durocher, 2019)

Une étude en 2019 montre que la liaison du *Shieldin* empêche la résection des télomères (Setiaputra and Durocher, 2019). Ce résultat soulève la question du mécanisme de l'inhibition de la résection par la présence du *Shieldin*. La réponse se trouve peut-être dans le fait que le *Shieldin* a pour rôle de favoriser la synthèse du brin C télomérique lors de l'étape de remplissage (*C-strand fill-in*) (Figure 1.12.B) agissant de pair avec le complexe *CST* dans le recrutement de Polα et la synthèse du brin C télomérique (Mirman et al.,

2018).

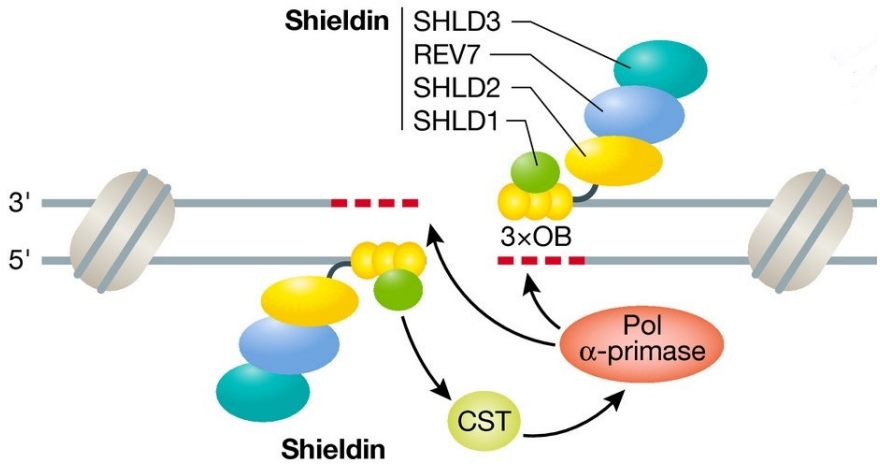


FIGURE 1.12 – **Modèle d'action du *Shieldin* avec *CST* dans le remplissage du brin C** (Setiaputra and Durocher, 2019)

1.5 Les structures et mécanismes de liaison des protéines de la famille *SSB-OB fold* humaine aux télomères

Dans le cadre de ma thèse je me suis intéressé aux protéines de liaison à l'ADN simple brin aux télomères chez l'homme. Ces protéines présentent une caractéristique commune à savoir de posséder au moins un domaine *OB fold*. Ces protéines forment ce qu'on appellera **la famille des SSB-OB fold aux télomères**. Chez l'homme cette famille est composée de plusieurs protéines dont fait partie le complexe *POT1-TPP1*, la protéine *RPA* et le complexe *CST*.

1.5.1 Le motif *OB fold*

Le motif *OB fold*, *Oligonucleotide/Oligosaccharide-Binding fold*, est un domaine structural relativement conservé parmi les êtres vivants, présent aussi bien chez les procaryotes que chez les eucaryotes. Comme son nom l'indique ce domaine protéique assure la liaison à l'ADN ou à l'ARN, bien qu'il ait été depuis observé également aux interfaces protéine-protéine. Les domaines *OB fold* ont une longueur comprise entre 70 et 150 acides aminés. Bien qu'il n'existe aucune relation forte de séquence entre les membres de la famille des *OB fold*, ce motif est reconnaissable par sa topologie caractéristique (Murzin, 1993). En effet, ce domaine est composé d'une hélice α, de cinq brins β et de trois boucles (*loop*, L) dont l'enchainement canonique est : (N_{term}) - β_1 - $\mathbf{L}_{12}$ - β_2 - β_3 - $\mathbf{L}_{34}$ - α - β_4 - $\mathbf{L}_{45}$ - β_5 - (C_{term}) (Figure 1.13). La variabilité de la longueur entre les domaines *OB fold* est principalement due à des différences dans la longueur des boucles, éléments variables, situés entre les éléments bien conservés de la structure secondaire. Parmi ces éléments secondaires, on retrouve les brins β qui s'enroulent orthogonalement pour former un tonneau β à cinq brins quelque peu aplati. Entre les brins β_3 et β_4, on trouve fréquemment une hélice α disposée au fond du tonneau β et orientée perpendiculairement le long de la section transversale de ce dernier. Il arrive que les brins β_3 et β_5 soient agencés de manière parallèle et ferment ainsi partiellement le tonneau β par liaison hydrogène de ces brins. Malgré des divergences de séquence, ces éléments structuraux ont été identifiés comme commun à tous les domaines *OB fold* (Bycroft et al., 1997).

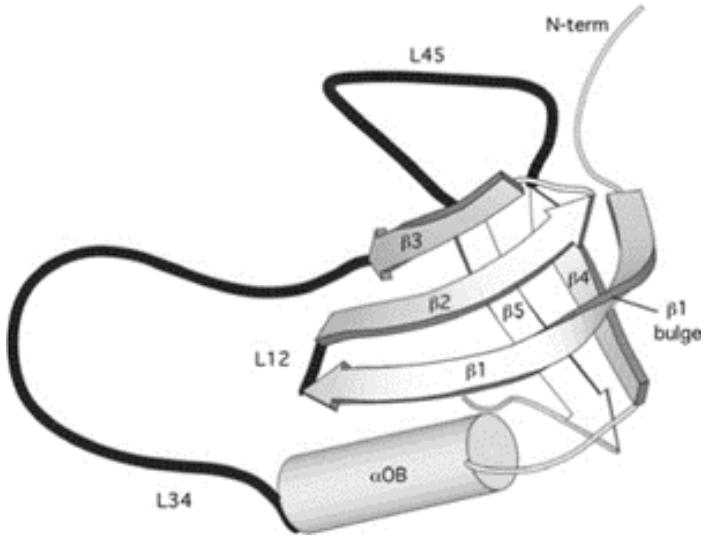


FIGURE 1.13 – **Schéma d'un motif canonique *OB fold*.**(Guardino et al., 2009)

Les données structurales semblent démontrer que les domaines *OB fold* possèdent une interface canonique de liaison avec leur ligand, centrée sur les brins β_2 et β_3. Cette interface est augmentée par la présence des boucles L_{12}, L_{34} et L_{45} comme en témoignent de nombreuses structures à haute résolution. Ces boucles semblent fournir une surface de reconnaissance idéale pour les acides nucléiques simple brin en permettant cette interaction par des liaisons hydrogène, des empilements aromatiques, des interactions hydrophobes et des interactions polaires (Annexe 8.1-8.2). Ces domaines *OB fold* sont divisés en trois catégories en fonction de la nature de leur substrat (Theobald et al., 2003) :

A. Les protéines qui se lient aux acides nucléiques sans spécificité de séquence (la protéine de réplication A *RPA*, la protéine de liaison à l'ADN simple brin d'*Escherichia coli SSB*,...).

B. Les protéines qui reconnaissent des régions monocaténaires spécifiques d'acides nucléiques (la protéine humaine *POT1*, la protéine $TEBP_\alpha$ de *O.nova*, le terminateur transcriptionnel *Rho* de *E. coli*, ...).

C. Les protéines qui interagissent avec des acides nucléiques ribonucléiques (les protéines ribosomales de *Thermus thermophilus S12* et *S17*, l'hélicase du *Thermatoga maritima RecG*, ...).

Un aspect intéressant de la reconnaissance des acides nucléiques par le domaine *OB fold* est la conservation de la polarité de liaison au ligand. En effet dans la grande majorité des domaines *OB fold* étudiés, la séquence d'acide nucléique se lie avec la polarité standard à savoir l'extrémité 5' dirigée vers les brins β_4 et β_5 et l'extrémité 3' orientée vers β_2. Les exceptions connues à cette règle sont la protéine *TEBP*β et la protéine de liaison à l'ADN simple brin *Ec. SBB*. Aucune étude à ce jour n'explique la raison physique de cette préférence de polarité au sein des motifs *OB fold*.

En résumé, nous avons vu que malgré une absence de similitude de séquence entre les différents domaines *OB fold*, ils présentent une structure commune caractéristique. Ce domaine protéique est crucial au sein de la cellule car il permet les interactions protéines-acides nucléiques. Parmi les rôles biologiques des protéines de la famille des *OB fold* nous pouvons citer leur implication dans la réplication, la recombinaison, la réparation et la transcription de l'ADN, ainsi que le maintien des télomères (Theobald et al., 2003). Au cours de ma thèse, je me suis intéressé à l'étude de certaines protéines de **la famille SSB-OB fold aux télomères** chez l'homme, à savoir le complexe *POT1-TPP1*, la protéine *RPA* et le complexe *CST*.

1.5.2 Le complexe *POT1-TPP1*

Son rôle biologique

La protéine *POT1* (*Protection of Telomere 1*), découverte par Baumann en 2001 (Baumann and Cech, 2001), joue un rôle de protection des télomères. Ce rôle de protection est assuré notamment par la participation de *POT1* comme seule protéine de liaison à l'ADN simple brin du complexe *Shelterin* (de Lange, 2005). Des études ont montré que la délétion du gène *POT1* dans des cellules humaines entraîne la dégradation des régions terminales et la fusion des chromosomes (Baumann and Cech, 2001). Ces instabilités chromosomiques induites par la délétion de *POT1* peuvent s'expliquer en partie par la non formation de la *T-loop* (Palm and de Lange, 2008). De même le modèle de liaison de *POT1* à l'extension 3' permet le coiffage de l'extrémité du chromosome, empêchant la liaison de *RPA* et par conséquent empêchant l'activation des voies de contrôle telle que la kinase *ATR* d'accéder à cette extrémité (Lei et al., 2004). Enfin la protéine *POT1* est également impliquée dans la régulation et le maintien de la longueur des télomères. Des expériences *in vitro* montrent que *POT1* inhibe l'activité de la télomérase à l'extrémité 3' de l'ADN en contrôlant l'accessibilité à son substrat (Lloyd et al., 2016).

La protéine *POT1* possède un partenaire au sein de la cellule, à savoir *TPP1*. Cette protéine a été découverte en 2004 par Liu et al (Liu et al., 2004), nommée à l'origine *PTOP*. Elle existe sous différentes appellations : *PTOP* (Liu et al., 2004), *TINT1* (Houghtaling et al., 2004), *PIP1* (Ye, 2004) ou encore *ACD* (Else et al., 2007). *TPP1* participe au complexe du *Shelterin* en assurant la liaison entre *POT1* et *TIN2* (de Lange, 2005). Une étude montre qu'en l'absence de *TIN2*, le complexe *POT1-TPP1* n'est plus présent aux télomères (Takai et al., 2011). De plus bien qu'aucune interaction de *TPP1* avec l'ADN télomérique n'ait été mis en évidence (aucun complexe entre *TPP1* et l'ADN n'a été détecté par *EMSA*), ce dernier améliore l'affinité de liaison de *POT1* avec son substrat d'un facteur d'environ 10 (Xin et al., 2007; Wang et al., 2007; Taylor et al., 2011). De ce fait *TPP1* est essentielle à la présence de *POT1* aux télomères. La protéine *TPP1* permet le recrutement et améliore la processivité de la télomérase, participant ainsi au maintien de la longueur des télomères (Zhong et al., 2012; Lloyd et al., 2016).

En résumé le complexe *POT1-TPP1* joue un rôle crucial dans la protection des télomères grâce à la liaison de *POT1* à l'extension 3', empêchant l'activation des voies de contrôle. De plus le complexe *POT1-TPP1* régule la longueur des télomères soit, en inhibant l'accès de la télomèrase par séquestration de son substrat en 3' par *POT1*, soit, en recrutant la télomérase et

en améliorant sa procéssivité via *TPP1*.

Sa structure

La protéine *POT1* est constituée d'une chaine de 634 acides aminés (71 kDa) et se compose de trois domaines *OB fold* (Figure 1.14). Les domaines *OB*1 et *OB*2 situés en N-term sont impliqués dans la liaison à l'ADN télomérique. En C-term on trouve un domaine *OB*3 qui permet l'interaction avec la protéine *TPP1*. Au centre du domaine *OB*3 est situé un domaine *HJRL* (*Holliday junction resolvase-like*). Des expériences *in vitro* en présence de *POT1* C-term (320 - 634 aa) n'ont pas mis en évidence d'interaction entre le motif *HJRL* et une séquence d'ADN mimant une jonction Hollidays (Chen et al., 2017) ; cependant ce domaine *HJRL* est essentiel pour la stabilité de l'interaction de *POT1* avec son partenaire *TPP1* comme cela a été révélé par cristallographie (Chen et al., 2017; Rice et al., 2017b).

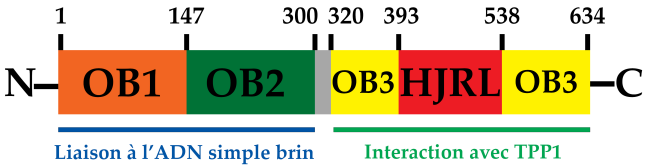


FIGURE 1.14 – **Les domaines structuraux et fonctionnels de *POT1***

La protéine *TPP1*, composée de 540 acides aminés, est présente sous deux isoformes au sein de la cellule. La présence de ces deux isoformes est induite par un épissage alternatif, dont la région située en N-term correspond à un "exon alternatif" encore peu étudié. La forme majoritaire est celle qui est dépourvue des 86 premiers acides aminés et qui nous intéressera dans la suite de la thèse. *TPP1* se compose d'un domaine *OB fold* en N-term, suivi d'un domaine d'interaction avec *POT1*, *POT1 interaction motif* (*PIM*), et d'un domaine de liaison avec *TIN2*, *TIN2 binding motif* (*TBM*) (Figure 1.15.A). Malgré la présence d'un domaine *OB fold*, aucune interaction entre l'ADN et *TPP1* n'a été observée (Xin et al., 2007; Wang et al., 2007). On remarque la présence d'un motif *TEL* présent dans le domaine *OB fold* de *TPP1*. Ce domaine *TEL* permet le recrutement de la télomérase par *TPP1*. Enfin il est important de noter que dans le cadre de mes expériences en présence de *TPP1*, j'ai travaillé avec une version tronquée de la forme majoritaire, à savoir sans le domaine d'interaction avec *TIN2* (Figure 1.15.B). Mon travail se limitant à l'étude des protéines de liaison à l'ADN simple brin aux télomères,

le système a été simplifié en retirant le domaine d'interaction avec *TIN2* (*TBM*) qui ne participe pas à la liaison du complexe *POT1-TPP1* à l'ADN.

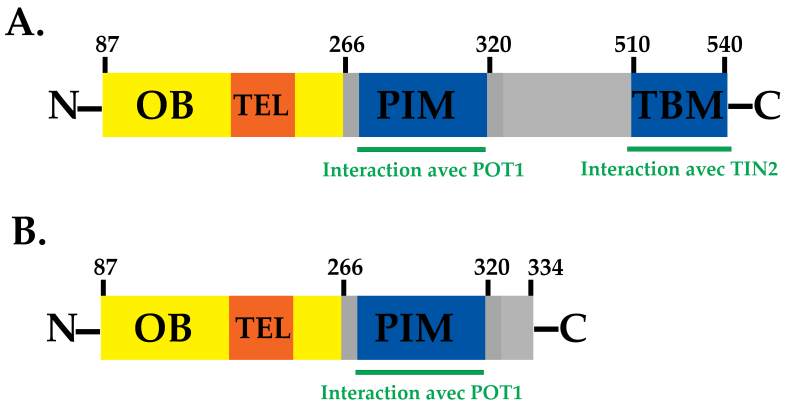


FIGURE 1.15 – **Les domaines structuraux et fonctionnels de *TPP1*.** **A.** Domaines de *TPP1*. **B.** Domaines de *TPP1* tronquée.

Il n'existe à ce jour aucune structure à haute résolution de la protéine *POT1*. En revanche, la structure de *POT1* N-term (1 - 300 aa contenant uniquement les domaines *OB*1 et *OB*2), en présence de son ligand, à savoir la séquence 5'-TTAGGGTTAG-3' (10 nt), a été résolue en 2004 par cristallographie (Figure 1.16)(Lei et al., 2004). En 2017 deux équipes ont déterminé simultanément par cristallographie la structure de *POT1* C-term en interaction avec *TPP1* (Figure 1.17)((Rice et al., 2017a; Chen et al., 2017)). Ces structures sont composées du domaine *OB*3/*HJRL* de *POT1* et du domaine d'interaction à *POT1* de *TPP1* (*PIM*).

De même que pour *POT1*, il n'existe aucune structure à haute résolution de *TPP1*. En revanche, l'équipe de M. Lei a résolu par cristallographie le domaine *OB fold* de *TPP1* (Figure 1.18)(Wang et al., 2007). On peut également citer que le domaine de *TPP1* permettant la liaison à *TIN2* (*TBM*) a été résolu par cristallographie en 2017 (Hu et al., 2017).

La structure du complexe *POT1-TPP1* n'est pas connue à ce jour. Cependant une expérience de *SAXS* (*Small Angle X-rays Scattering*) réalisée par Chen et al, en 2017 a permis de déterminer l'enveloppement du complexe *POT1-TPP1* en présence de la séquence minimale de liaison de *POT1* : 5'-TTAGGGTTAG-3' (Figure 1.19) (Chen et al., 2017). Ces résultats présentent le complexe *POT1-TPP1* sous la forme d'un V allongé, situant *TPP1* à l'opposé de *POT1* N-term et par conséquent de l'ADN également.

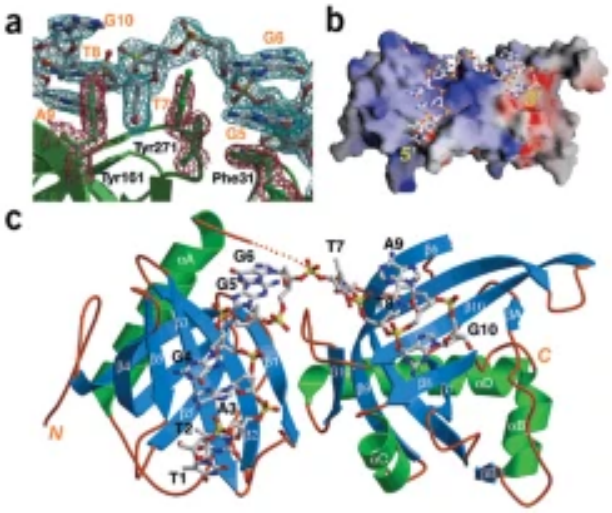


FIGURE 1.16 – **Structure cristallographique de *POT1* N-term en présence de son ligand d'ADN (PDB : 1XJV).**(**a.**) Carte expérimentale de densité d'électron à 1.73 Å. (**b.**) Représentation du potentiel électrostatique de surface : potentiel positif en bleu, potentiel négatif en rouge. (**c.**) Diagramme en ruban de *POT1* N-term avec son ligand 5'-TTAGGGTTAG-3'. (Lei et al., 2004)

A. **B.**

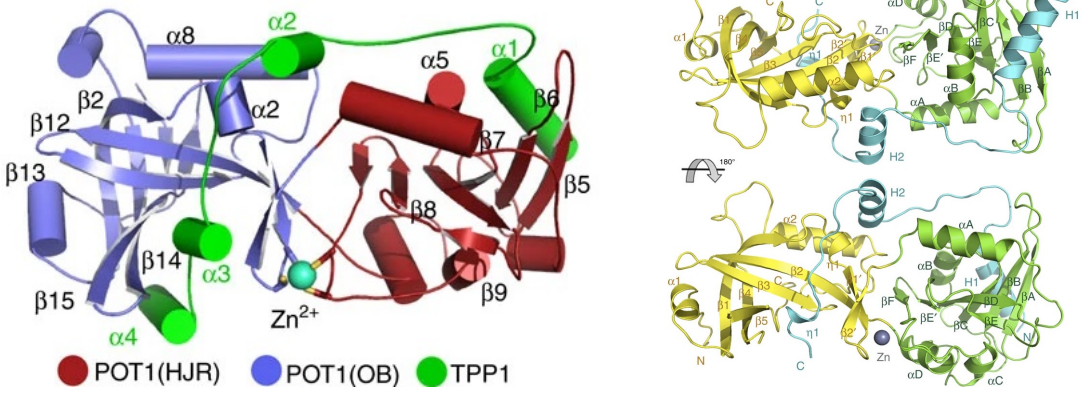


FIGURE 1.17 – **Structure cristallographique de *POT1* C-term en présence de *TPP1*.**Diagramme en ruban de *POT1* C-term en interaction avec *TPP1* (**A.**) (PDB : 5UN7) (Rice et al., 2017a).(**B.**) Les domaines *OB3* et *HJRL* de *POT1* respectivement en jaune et en vert, le domaine *PBM* de *TPP1* en cyan (PDB : 5H65) (Chen et al., 2017).

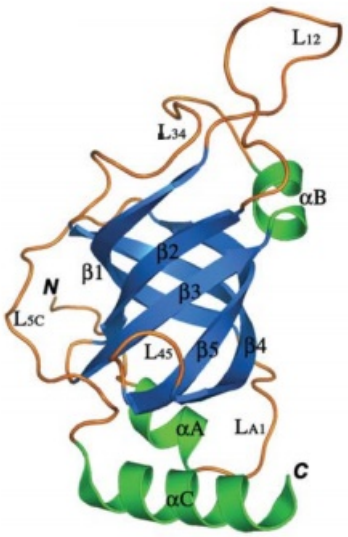


FIGURE 1.18 – **Structure cristallographique du domaine *OB fold* de *TPP1* (PDB : 2I46)**Diagramme en ruban du *OB fold* de *TPP1* (Wang et al., 2007)

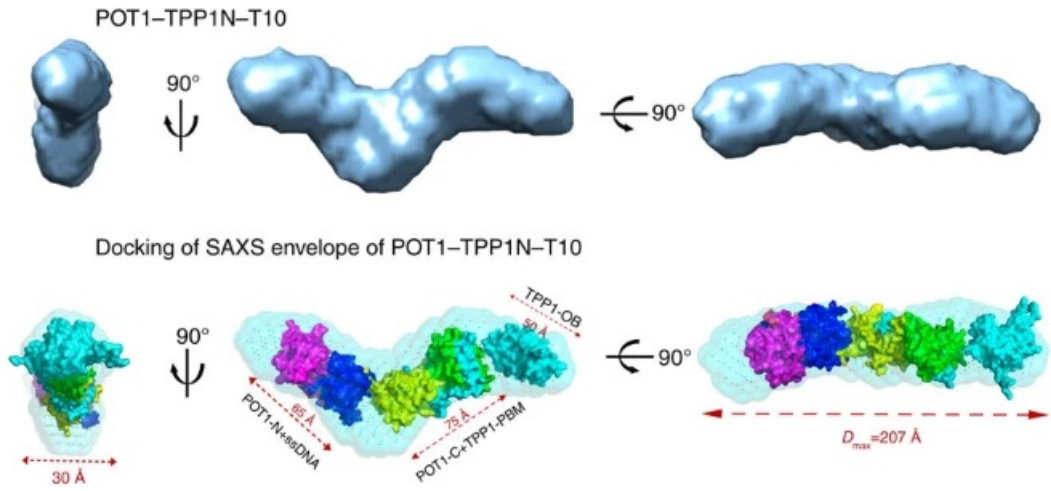


FIGURE 1.19 – **Enveloppe déterminée par SAXS du complexe *POT1-TPP1* en présence d'ADN.**Trois vues différentes de l'enveloppe en forme de V du complexe *POT1-TPP1* en présence d'ADN, T10 : 5'-TTAGGGTTAG-3'. Docking des structures cristallographique de *POT1* N-term (PDB : 1XJV), *POT1* C-term en interaction avec *TPP1* (PDB : 5H65) et *TPP1 OB fold* (PDB : 2I46) dans l'enveloppe de *SAXS*. (Chen et al., 2017)

Son mécanisme de liaison à l'ADN

La protéine *POT1* est une protéine *OB fold* de la catégorie des protéines qui reconnaissent des régions monocaténaires spécifiques d'acides nucléiques. En effet des expériences de SELEX, *Systematic Evolution of Ligands by EXponential enrichment* , ont montré que le domaine *OB*1 reconnaît le motif télomérique humain 5'-TTAGGG-3', alors que le domaine *OB*2 lie la séquence 5'-TTAG-3' (Choi et al., 2015). Cela implique que la protéine *POT1* a pour séquence minimale de liaison : 5'-TTAGGGTTAG-3' (10 nt), séquence que l'on nommera H10. Cette spécificité de séquence de *POT1* s'explique en partie par la nature des interactions mises en jeu et révélées par la structure cristallographique de Lei et al (Figure 1.16). Parmi les 31 liaisons hydrogène impliquées, 13 sont spécifiques des séquences et impliquent les groupes donneur-accepteur Watson-Crick des bases de l'ADN avec des résidus de la protéine (Figure 1.20) (Lei et al., 2004). Il est intéressant de remarquer que le domaine *OB1* comptabilise 22 liaisons hydrogène, c'est-à-dire plus des deux tiers de *POT1*, suggérant que le site principal de *POT1* soit 5'-TTAGGG-3' de *OB*1. Le domaine *OB*2 crée neuf liaisons hydrogène avec l'ADN télomérique, dont huit avec T7 et G10. A9 est le seul nucléotide du motif de liaison qui ne fait aucune interaction par liaison hydrogène, ce qui suggère que la position 9 ne participe pas à la spécificité de séquence. Le nucléotide à l'extrémité 3' du motif, G10, est pris dans une poche hydrophobe formée par deux brins β (β_3 et β_{47}. De plus la base G10 est en interaction par quatre liaisons hydrogène ce qui semble expliquer la reconnaissance hautement spécifique de ce nucléotide. Cette séquestration de la guanine présente en 3' pourrait expliquer le rôle protecteur de *POT1* de l'extension simple brin contre la dégradation par les exonucléases. L'interaction de POT1 à son substrat met en jeu un grand nombre d'empilements des bases avec des résidus d'acides aminés aromatiques mais aussi entre bases (Figure 1.20 et 1.21). De plus, les résultats de cristallographie montrent que la liaison des domaines *OB*1 et *OB*2 à l'ADN télomérique induit une courbure de l'ADN d'un angle d'environ 90° (Figure 1.16).

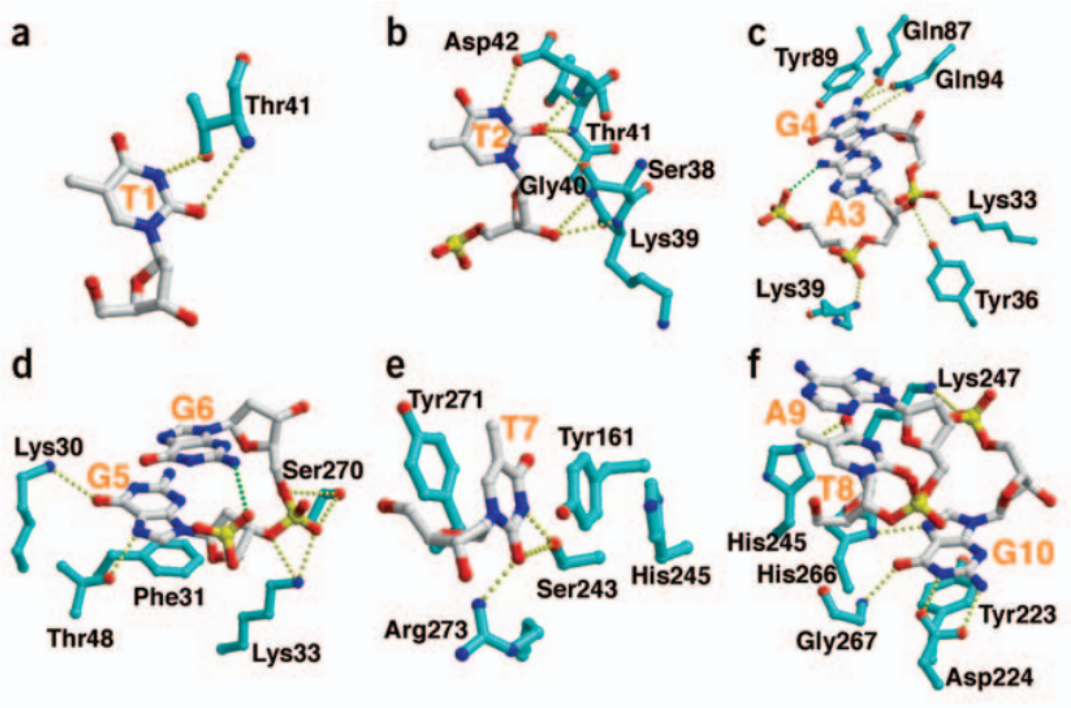


FIGURE 1.20 – **Détail des interactions de chaque nucléotide avec les différents résidus de POT1**(Lei et al., 2004)

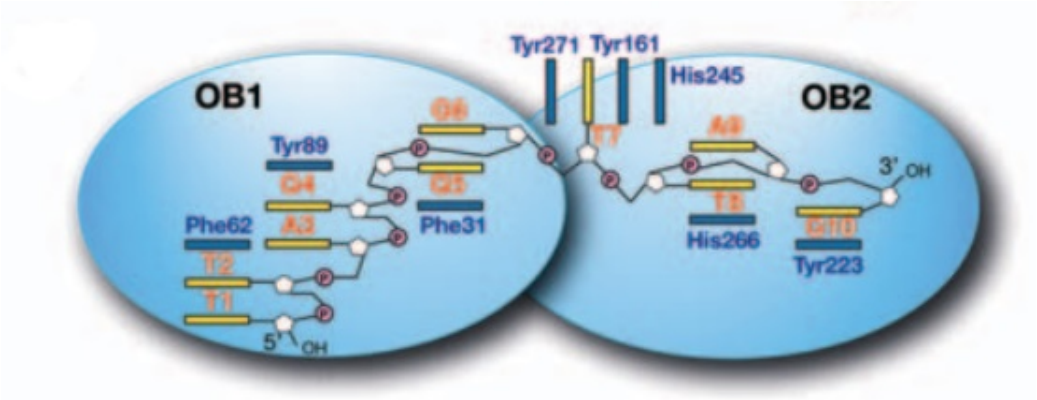


FIGURE 1.21 – **Schéma de liaison de POT1 à son ligand H10.** (Lei et al., 2004)

La présence de *TPP1* améliore l'affinité de *POT1* pour son substrat (Xin et al., 2007; Wang et al., 2007; Taylor et al., 2011), mais également augmente considérablement la discrimination de *POT1* pour l'ARN (Nanda-

kumar et al., 2009). En effet la récente découverte de l'ARN télomérique *TERRA* (Azzalin et al., 2007), dont la séquence est constituée de répétions $(UUAGGG)_n$, a soulevé la question suivante : comment *POT1* se lie fidèlement à l'ADN simple brin télomérique et évite la liaison avec l'ARN qui pourrait détourner *POT1* des télomères ? La réponse a été apportée par l'équipe de Thomas R. Cech en montrant *in vitro* qu'en présence de *TPP1* la protéine *POT1* ne reconnait plus son substrat si on remplace la deuxième thymine du motif de reconnaissance, 5'-TTAGGGTTAG-3', par une uracile, 5'-TUAGGGTTAG-3' (Nandakumar et al., 2009). Cette étude révèle que *TPP1* renforce la spécificité de *POT1* pour son substrat.

1.5.3 La protéine *RPA*

Son rôle biologique

La protéine *RPA*, identifiée par M. Wold en 1988 (Wold and Kelly, 1988), est un facteur essentiel pour la réplication de l'ADN, mais aussi pour la réparation de l'ADN et l'activation des points de contrôle (notamment l'activation des kinases *ATM* et *ATR*). Par conséquent la présence de *RPA* est régulée négativement aux télomères pendant l'interphase. Cependant l'implication de *RPA* dans le maintien des télomères a été mise en évidence au sein des levures *S. cerevisiae* et *S. pombe* (Smith et al., 2000; Ono, 2003), où elle est présente aux télomères au cours de la phase S. Notamment chez la levure *S. pombe*, il a été montré que *pRPA* est importante dans la synthèse correcte du *lagging strand* (Brin G) (Moser and Nakamura, 2009; Luciano et al., 2012; Audry et al., 2015). Chez l'homme, *RPA* est présente aux télomères humains lors de la réplication (Verdun and Karlseder, 2006; Déjardin and Kingston, 2009) et est impliquée dans le maintien des télomères (KOBAYASHI et al., 2010). Par ailleurs, Cohen et al. ont montré que la RPA humaine pouvait réguler l'activité *in vitro* de la télomérase probablement en déroulant les structures non usuelles présentes dans l'ADN télomérique (Cohen et al., 2004).

Sa structure

La protéine *RPA* se présente sous la forme d'un hétérotrimère composé de trois sous-unités : *RPA1, RPA2, RPA3*, de masses respectives 70, 32 et 14 kDa (Figure 1.22). Chaque sous-unité possède un ou plusieurs motifs *OB fold* qui permettent la liaison à l'ADN simple brin (*DBD = DNA binding Domain*) ou l'interaction avec d'autres protéines. La sous-unité *RPA*1 se compose d'un domaine d'interaction protéine-protéine situé en N-term (*DBD-F*), et de trois domaines *OB fold* de liaison à l'ADN simple brin : *DBD-A*, *DBD-*

B et *DBD-C*. La sous-unité *RPA*2 se compose de trois domaines. Un petit domaine non structuré situé en N-term qui contient de nombreux sites de phosphorylation (Gomes et al., 1996; Zernik-Kobak et al., 1997), un domaine *OB fold* central de liaison à l'ADN simple brin *DBD-D*, et enfin un domaine *Winged Helix*(*WH*) en C-term qui permet des interactions protéine-protéine, notamment avec celles impliquées dans la réparation comme *RAD52*, *XPA* (Mer et al., 2000). La sous-unité *RPA*3 est composée d'un unique domaine *OB fold* (*DBD-E*), dont un point de contact avec l'ADN simple brin a été mis en évidence en 2009 par photopontage (Salas et al., 2008). Ce domaine *OB-E* est essentiel pour la trimérisation de *RPA*. En effet cette trimérisation de la protéine *RPA* est assurée par les domaines : *DBD-C* (*RPA*1), *DBD-D* (*RPA*2) et *DBD-E* (*RPA*3).

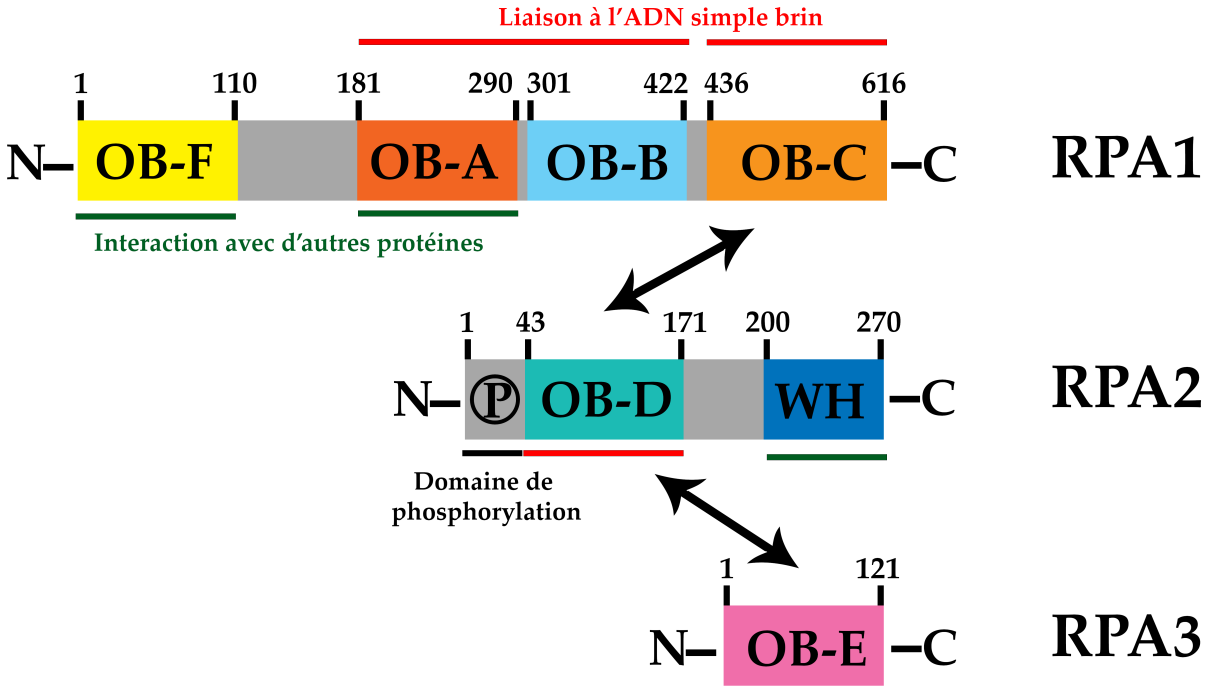


FIGURE 1.22 – **Les domaines structuraux et fonctionnels de *RPA*.**

Il existe plusieurs structures haute résolution des sous-unités de *RPA* ($\simeq$ 40 sur la *PDB*). Ne pouvant pas présenter chacune d'elle, j'ai pris le parti de choisir deux structures qui me semblent complémentaires et donnent une bonne vision structurale de *RPA*. La première structure est celle des domaines *DBD-A* et *DBD-B* de *RPA*1 résolue par cristallographie en présence d'une polyCytosine (8 nt) (Figure 1.23) (Bochkarev et al., 1997). On observe comme avec *POT1* deux domaines OB fold de liaison à l'ADN simple brin.

On notera que les domaines chez *RPA* sont alignés contrairement à ce qui avait été observé en présence de *POT1*. Nous reviendrons sur les différences entre *POT1* et *RPA* quand nous aborderons le mécanisme de liaison de *RPA*. La seconde structure que j'ai choisie de présenter est la structure cristallographique des domaines permettant la trimérisation des sous-unités de *RPA* : *DBD-C* (*RPA*1), *DBD-D* (*RPA*2) et *DBD-E* (*RPA*3) (Figure 1.24) (Bochkareva, 2002). On observe que la trimérisation est assurée par les trois hélices C-terminales des domaines *OB fold* disposées en parallèle.

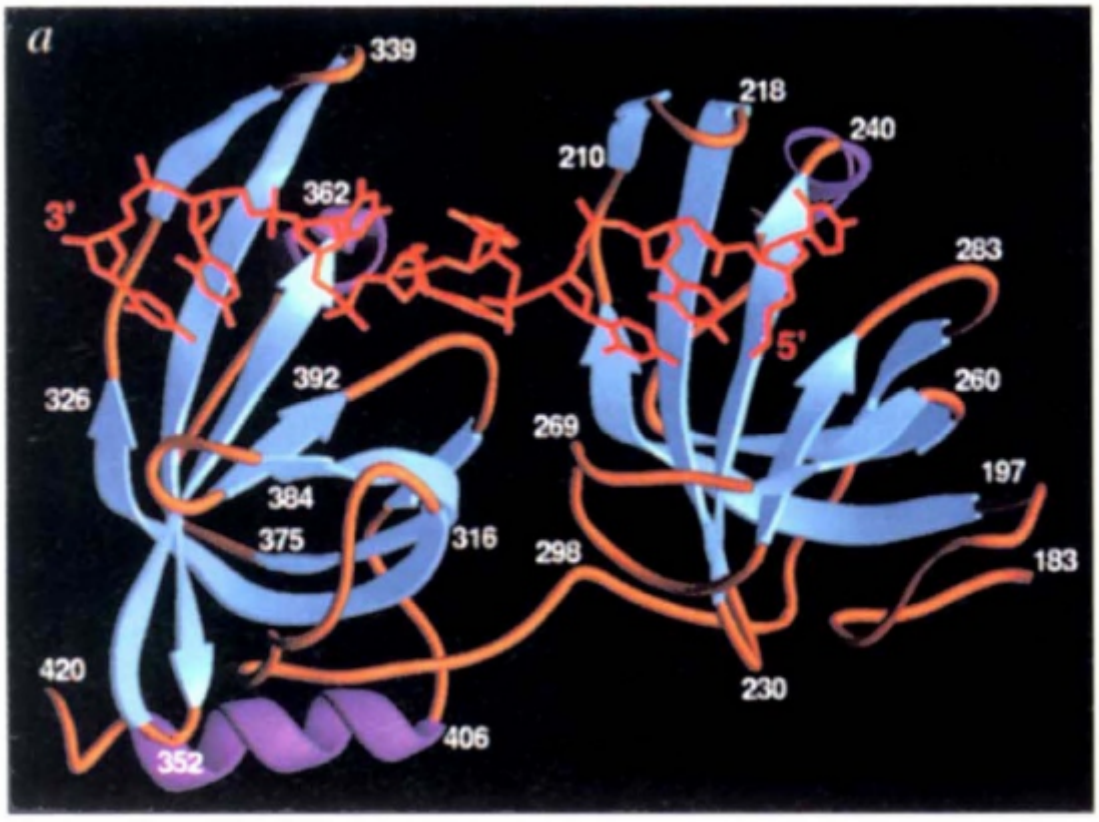


FIGURE 1.23 – **Structure cristallographique des domaines DBD-A et DBD-B de *RPA1* en présence d'une polyCytosine (8 nt) (PDB :1JMC) .** (Bochkarev et al., 1997)

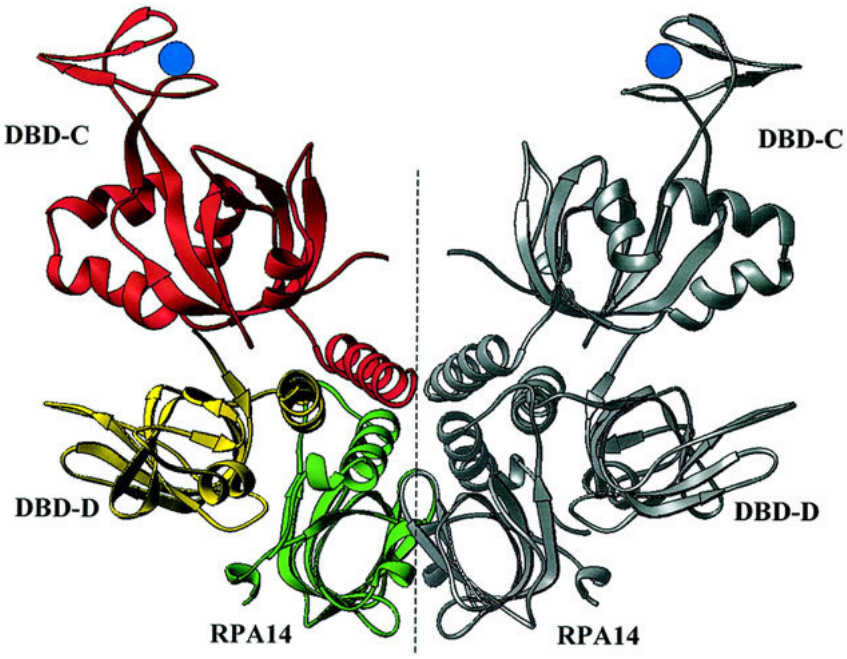


FIGURE 1.24 – **Structure cristallographique des domaines DBD-C, DBD-D et DBD-E assurant la trimérisation du complexe *RPA* (PDB :1L1O) .** (Bochkareva, 2002)

Même s'il n'existe aucune structure complète du complexe *RPA* humain, il est important de noter que l'hétérotrimère *RPA* est hautement conservé au sein des eucaryotes notamment de par sa séquence et de par sa structure. En 2012 Fan et al. ont résolu la structure quasi complète de la protéine *RPA* du champignon *Ustilago maydis* en présence d'un polyT avec une résolution de 3.1 Å (Figure 1.25) (Fan and Pavletich, 2012). Malgré l'absence du domaine *DBD-F* de la sous-unité *RPA*1 et du domaine *WH* de la sous-unité *RPA*2 cette structure cristallographique est la plus complète de la protéine *RPA* à ce jour. On retrouve la structure de la liaison à l'ADN médiée par les domaines *DBD-A* et *DBD-B* de *RPA*1 similaire à l'humaine vu précédemment ; ainsi que la structure de la trimérisation des trois sous-unités par l'intermédiaire des hélices C-terminal des domaines *DBD-C* (*RPA*1), *DBD-D* (*RPA*2) et *DBD-E* (*RPA*3). Cette structure quasi complète de *RPA* nous révèle l'agencement de ces différents domaines *OB fold* mais également décrit son mode de liaison par ces quatre domaines de liaison. On observe que *RPA* se lie à l'ADN simple brin à la fois par les domaines *DBD-A* et *DBD-B* d'une part, et d'autre par *DBD-C* et *DBD-D* induisant une conformation en forme de U à l'ADN.

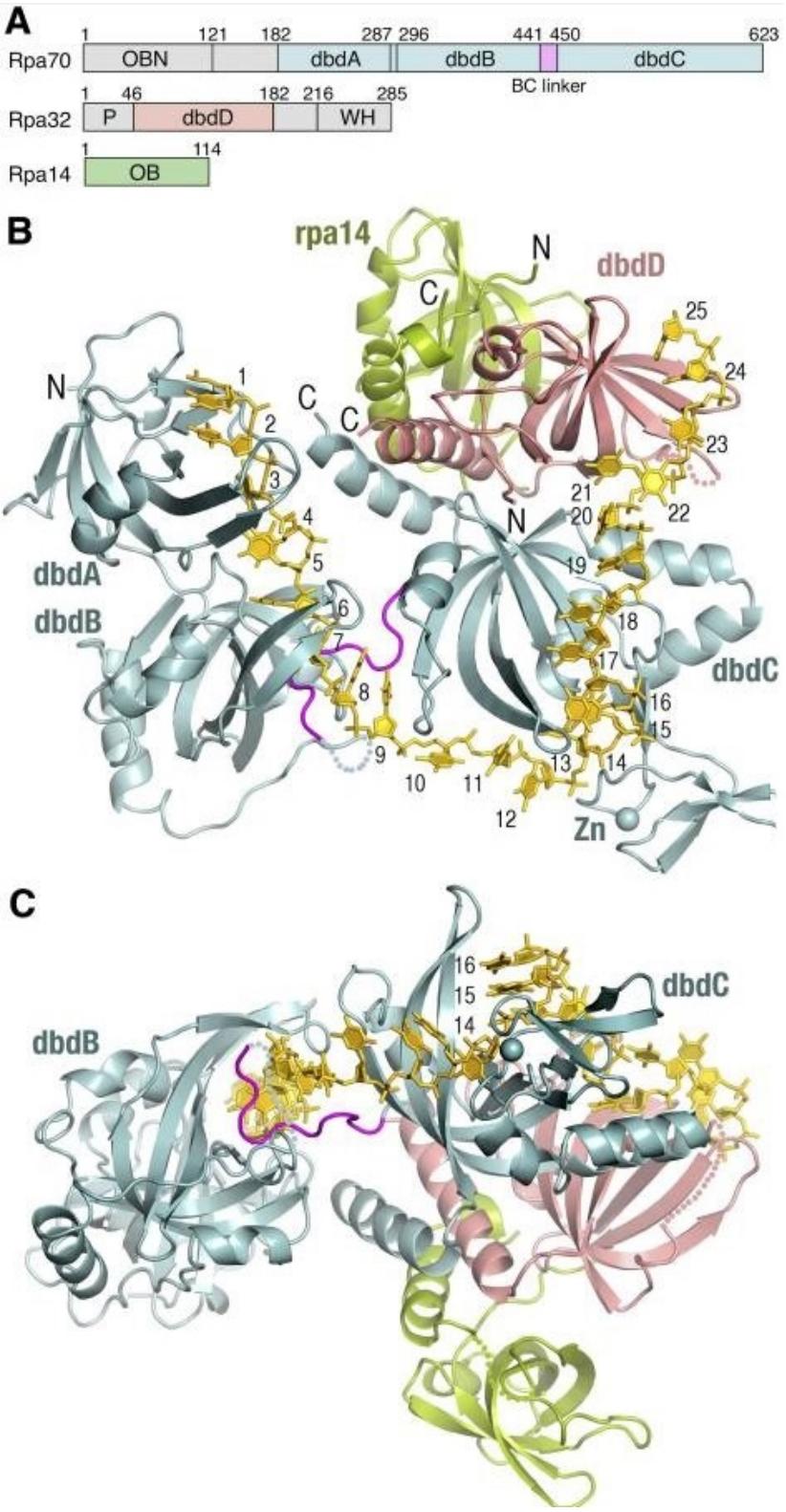


FIGURE 1.25 – **Structure cristallographique de *RPA* du champignon *Ustilago maydis* en présence d'un polyT (PDB :4GOP) A.** Les domaines structuraux de *RPA* du champignon *Ustilago maydis*, **B-C.** Structure cristallographique de *RPA* (Fan and Pavletich, 2012)

Son mécanisme de liaison à l'ADN

La protéine *RPA* est une protéine *OB fold* qui contrairement à la protéine *POT1* appartient à la catégorie des protéines qui se lient aux acides nucléiques sans spécificité de séquence. Les domaines *DBD-A* et *DBD-B* sont définis comme les domaines de haute affinité à l'ADN simple brin, alors que les domaines *DBD-C* et *DBD-D* montrent quant à eux une faible affinité (Bochkarev, 1999). Concernant les domaines *DBD-E* et *DBD-F*, aucune activité de liaison à l'ADN n'a jamais été observée. Toutefois des expériences de photopontage ont permis de mettre en évidence l'existence d'un contact de la sous-unité *RPA*3 avec l'ADN (Salas et al., 2008). Les domaines *DBD-A* et *DBD-B* lient l'ADN principalement à l'aide de liaison hydrogène. Les acides aminés impliqués dans ces liaisons hydrogène se situent majoritairement au sein des feuillets β_1 et β_3. On dénombre un total de 21 liaisons hydrogène entre les domaines de liaisons *DBD-A* et *DBD-B* et l'ADN simple brin (Figure 1.26) (Bochkarev et al., 1997), dont 12 liaisons hydrogène avec les bases et 9 liaisons hydrogène avec les groupements phosphate. Une partie des liaisons hydrogène avec les bases fait intervenir les couples donneur/accepteur Watson-Crick. En revanche, *RPA* présente un faible nombre, contrairement à *POT1*, de résidus aromatiques impliqués dans l'empilement des bases, à savoir 2 Phe pour *DBD-A* et 1 Phe et Trp pour *DBD-B*. L'absence de spécificité de séquence pourrait donc s'expliquer par la part non négligeable de liaison hydrogène avec les groupements phosphates, contrairement à ce qui est observé pour *POT1* ; mais également par un plus faible empilement des bases entre elles.

Le mécanisme de liaison de *RPA* à l'ADN simple brin se compose de plusieurs étapes et est associé à un changement conformationnel significatif démontré par des méthodes de biochimie, de microscopie électronique et de cristallographie (Iftode et al., 1999). Il existe trois modes décrits de liaison de la protéine *RPA* à l'ADN simple brin :

Une conformation compacte (recouvrant 8-10 nt). La liaison est assurée par les domaines *DBD-A* et *DBD-B* (Pfuetzner et al., 1997; Walther et al., 1999) qui recouvre une région de 8-10 nt. Dans ce mode de liaison, *RPA* lie de manière compacte et de façon instable (Blackwell and Borowiec, 1994).

Une conformation allongée contractée (recouvrant 12-23 nt). La liaison est entreprise par les domaines *OB fold* précédent (*DBD-A* et *DBD-B*) auquel s'ajoute le domaine *DBD-C* (Cai et al., 2007). Dans ce mode de liaison, *RPA* se lie à l'ADN simple brin avec une affinité plus élevée que celle de la conformation compacte (8-10 nt).

Une conformation allongée étendue (recouvrant 30 nt). La liaison à l'ADN

s'effectue par les domaines *DBD-A* et *DBD-B* auxquels s'ajoutent les domaines mineurs à savoir *DBD-C* et *DBD-D* (Kim et al., 1992; Bochkareva et al., 1998).

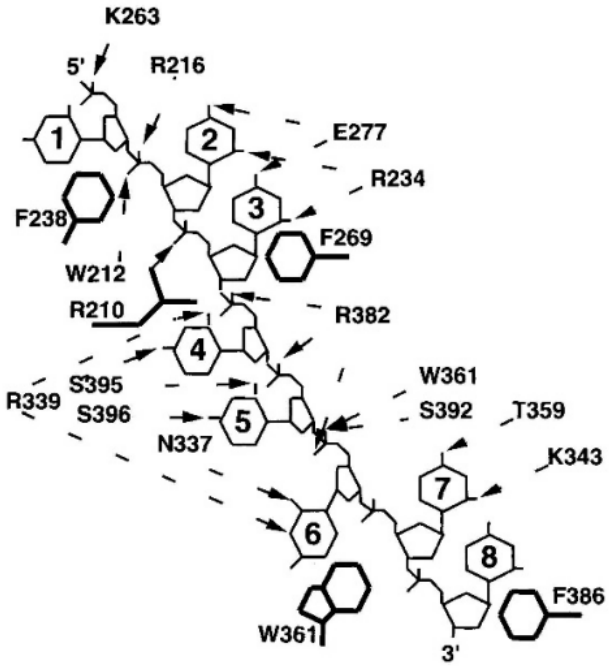


FIGURE 1.26 – **Diagramme montrant les contacts de *RPA* avec un polyCytosine (C_8).** Les bases sont numérotés de 1 à 8 à partir de l'extrémité 5 ', et les numéros ont été placés dans les bases correspondantes. Les liaisons hydrogène entre l'ADN et les résidus sont indiqués par des lignes pointillées. Les résidus aromatiques qui empilent les bases sont marquées en gras. (Bochkarev et al., 1997)

Une étude en présence d'un polyT montre que *RPA* lie l'ADN avec une faible coopérativité (Kim and Wold, 1995). Outre les modes de liaison cités ci-dessus, Pokherel et al. ont montré chez *Saccharomyces cerevisiae* que *RPA* possède une multitude de modes liaisons qui permettent de libérer soit le côté 3', soit le côté 5' afin de favoriser l'accès à différentes protéines, comme par exemple *RAD52* impliqué dans les voies de réparation de l'ADN et recruté par le domaine *DBD-D* de *RPA*2 (Pokhrel et al., 2019). En 2001 une étude a montré que la sous-unité *RPA*1 interagit principalement avec l'extrémité 5' de l'ADN, quelle que soit la structure de l'ADN, déterminant ainsi une polarité de liaison à l'ADN de 5' vers 3' de *RPA* (Kolpashchikov, 2001). *RPA* ne lie pas ou très peu l'ADN double brin (1000 fois moins affine pour le double brin et l'ARN), toutefois elle est capable de lier l'ADN double brin

contenant des régions simple brin internes (appelées boucles) ou présentes aux extrémités et de le dérouler (Iftode et al., 1999; Salas, 2005; Delagoutte et al., 2011).

1.5.4 Le complexe *CST*

Son rôle biologique

Dans les cellules humaines, le complexe *CST* joue un rôle majeur dans le maintien des télomères. En effet des études ont révélé que *CST* participe à la réplication des télomères (Miyake et al., 2009; Chen et al., 2012) ; notamment en permettant le redémarrage de la fourche de réplication dans la région télomérique (Zhang et al., 2019; Stewart et al., 2012). De même il a été montré que *CST* est nécessaire pour la synthèse du brin C télomérique (*C-strand fill-in*)(Feng et al., 2017). La longueur du brin C diminue continuellement en absence de *CST* indiquant une défaillance dans sa synthèse. En effet le complexe *CST* permet la synthèse du brin C en recrutant la polymérase Polα. *CST* joue également un rôle de terminateur de l'activité de la télomérase (Chen et al., 2012). Cette étude révèle une compétition entre *POT1-TPP1* et *CST* pour l'accessibilité à l'ADN télomérique. Or nous avons vu que le complexe *POT1-TPP1* recrute la télomérase permettant l'élongation de l'extension du brin G en 3'. Cette observation est renforcée par le fait que la liaison *CST* avec l'ADN télomérique augmente pendant la phase S / G2, coïncidant avec l'arrêt de l'élongation par la télomérase. Une déplétion de *CST* permet une activité excessive de la télomérase favorisant l'élongation des télomères. Enfin une étude réalisée en 2014 révèle que *CST* participe également à la réparation de l'ADN induit par des dommages (Wang et al., 2014).

Sa structure

Le complexe *CST* se présente chez les mammifères sous la forme de l'hétérotrimère suivant : *CTC1*, *STN1* et *TEN1* (Figure 1.27). La sous-unité *CTC1* se compose d'un enchaînement de sept domaines *OB fold*. A ce jour, seuls les domaines *OB-E* et *OB-F* ont été identifiés comme se liant à l'ADN. Le domaine *OB-G* présent en C-term permet la liaison avec la sous-unité *STN1*. L'organisation des domaines structuraux des sous-unités *STN1* et *TEN1* est très similaire à celle des sous-unités *RPA*32 et *RPA*14 du complexe *RPA* (Figure 1.22) : *STN1* et *TEN1* possèdent chacune un domaine *OB fold*, *OB-H* et *OB-I*, respectivement ; *STN1* possède en C-term deux domaines *wHTH* (*winged helix-turn-helix*) similaires au domaine *WH* présent dans la sous-unité

RPA2 et qui est impliqué, notamment, dans l'interaction avec des protéines de réparation de l'ADN. La présence de ces domaines *wHTH* pourrait expliquer le rôle de *CST* dans la réparation de l'ADN. Contrairement à *RPA2*, *STN1* ne possède pas de domaine de phosphorylation en N-term. On notera que l'interaction entre *CTC1* et *STN1* est renforcée par une interaction supplémentaire entre *OB-E* et $wHTH_1$.

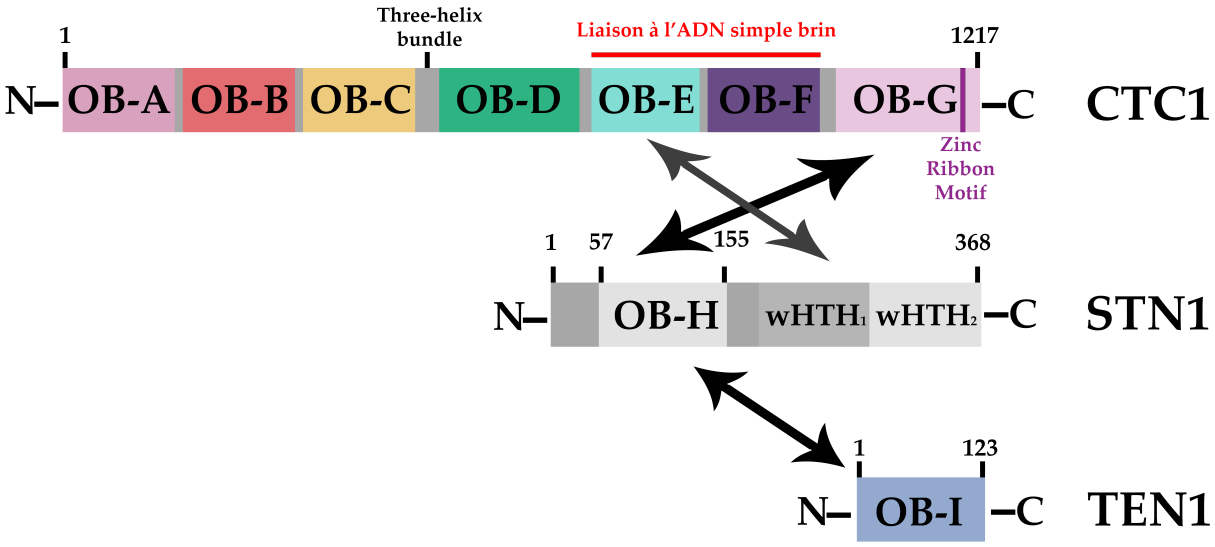


FIGURE 1.27 – **Les domaines structuraux et fonctionnels de *CST*.**

La structure du complexe *CST* humain en présence d'ADN a été résolue par l'équipe de Thomas R. Cech par cryo-microscopie électronique en 2020 (résolution 3.0 Å) (Lim et al., 2020). La structure globale montre une forme plutôt allongée du complexe ternaire avec les trois sous-unités en enfilade (Figure 1.28.A représentation en ruban). La trimérisation des sous-unités de *CST* est assurée par l'interaction des domaines *OB fold* : *OB-G* (*CTC1*), *OB-H* (*STN1*) et *OB-I* (*TEN1*) (Figure 1.28.B) *OB fold* (Figure 1.28.C). Ce mode d'interaction est différent à celui observé pour le complexe *RPA* (Figure 1.24.20). La liaison entre *CTC1* et *STN1* est renforcée par les domaines *STN1n* (*OB-H*) et *STN1c* (*wHTH1*), reliées avec un petit linker, interagissant avec la sous-unité *CTC1* en deux positions différentes : *OB-G* avec *OB-H* respectivement (Figure 1.28.A)

Une étude comparative des domaines *OB fold* des complexes *CST* et *RPA* a permis de mettre en évidence des similitudes structurales entre *OB-B* et *OB-C* de *CTC1* et *OB-A* et *OB-B* de *RPA*1, et *OB-G* de *CTC1* et *OB-C* *RPA*1 suggérant que *CTC1* et *RPA1* ont un ancêtre commun (Annexe 8.3)). Au regard de la similarité structurale existant entre les sous-unités *RPA*2 et

STN1, et *RPA*3 et *TEN1*, nous pouvons supposer que cette ancêtre commun était également un hétérotrimère.

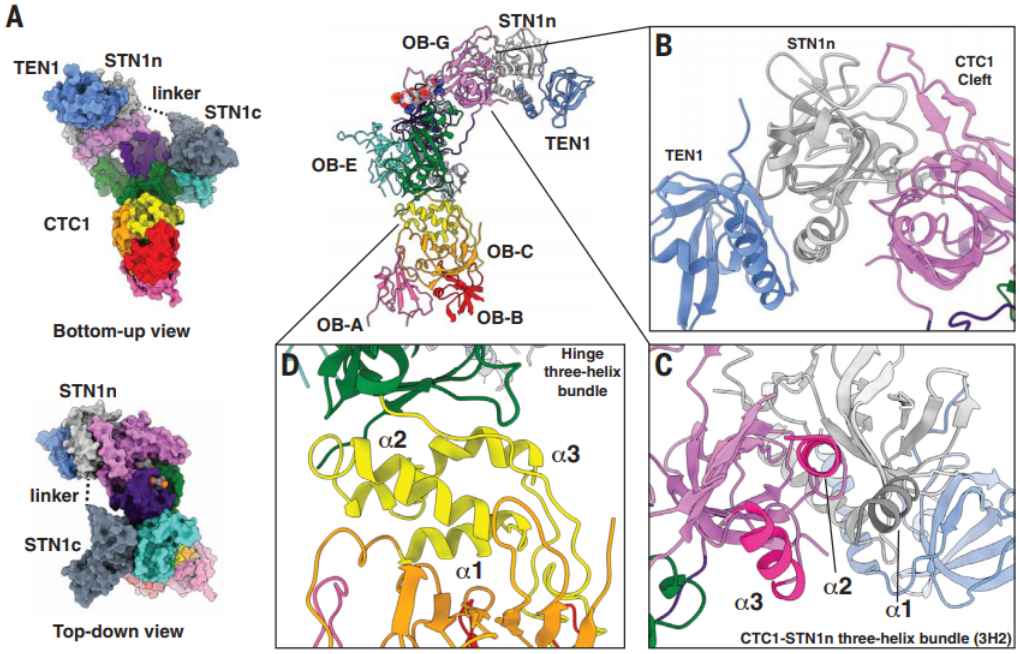


FIGURE 1.28 – **Structure du complexe CST résolue par cryo-microscopie électronique en présence d'ADN, 5'-TAGG-3' (PDB :6W6W).** (A) *STN1n* (N-term) et *STN1c* (C-term) interagissent indépendamment avec *CTC1* via un petit linker flexible ; (B) *STN1* et *TEN1* n'interagit pas avec *CTC1* en utilisant un faisceau d'hélice trimérique comme *RPA*, au lieu de cela *STN1* interagit directement avec un patch hautement conservé présent de *CTC1* ; (C) faisceau d'hélice *CTC1-STN1* qui est impliqué dans l'assemblage *CTC1* (a2 et a3 en rose vif provient de *OB-G*) et *STN1* (a1 en gris provient de *OB-H*) ; (D) le faisceau d'hélice (a1,a2 et a3) relie *OB-A*, *OB-B* et *OB-C* au reste des domaines *OB* C-term de *CTC1*. (Lim et al., 2020)

Son mécanisme de liaison à l'ADN

CTC1 est une protéine *OB fold* de la même catégorie que *POT1* à savoir une protéine qui reconnait des régions monocaténaires spécifiques d'acides nucléiques. En effet le complexe *CST* présente une préférence pour les séquences riches en guanines mais pas pour la séquence télomérique (Hom and Wuttke, 2017). En revanche le complexe *CST* peut se lier de manière moins spécifique sur de longs ADN simple brin (Miyake et al., 2009; Chen et al., 2012). Les mécanismes régissant les interactions entre le complexe *CST* hu-

main et l'ADN simple brin sont encore mal connus. La structure du complexe *CST* en présence de son ligand d'ADN a permis de montrer que le complexe *CST* a tendance à dimériser en présence d'un ADN de 18 nucléotides (Figure 1.29). Cette caractéristique a aussi été observée avec son homologue *Cdc13* chez la levure *S. cerevisiae* (Yu et al., 2011). Cependant les données structurales des domaines *OB fold* de *CTC1* montrent une plus forte ressemblance avec la sous-unité *RPA*1 humaine que pour la protéine *Cdc13* (Annexe 8.4). Ce résultat est basé sur l'utilisation d'un *Z-score* calculé avec *DALI structural homology* (Lim et al., 2020).

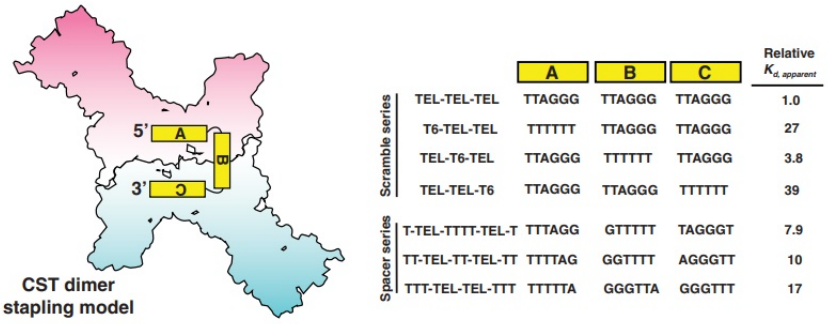


FIGURE 1.29 – **Modèle de dimérisation du complexe CST en présence d'un ADN de 18 nt** (Lim et al., 2020)

1.6 Une structure non canonique de l'ADN présent aux télomères

1.6.1 La structure G-quadruplex

Une étude réalisée en 1962 a montré par diffraction de rayon X que les guanines présentent une propension inhérente à s'auto-associer formant des structures hélicoïdales à quatre brins (Gellert et al., 1962). Depuis lors de nombreuses analyses biochimiques et structurales ont établi que les séquences d'acide nucléique simple brin, ADN et ARN, contenant des séries de guanines successives séparées par d'autres bases se replient spontanément en structures G-quadruplexes (G4) *in vitro* dans les conditions physiologiques (Bates et al., 2007). Il a été montré que les séquences télomériques, de par la présence au sein de leur motif d'une distribution en guanines successives (cf section 1.2.4), sont pour la plupart capables de se structurer en G4 comme c'est le cas par exemple chez le cilié *Oxytricha nova* (Lipps et al., 1982; Tran et al., 2010). Les éléments constitutifs des G4s sont les quartets de guanines qui désignent

un plateau de 4 guanines appariées entre elles par des liaisons hydrogène de type Watson-Crick, entre N(1) et O(6), et par des liaisons de type Hoogsteen, entre N(7) et NH_2(2), (Figure 1.30.A). C'est cet empilement de quartets les uns sur les autres qui forme des structures hélicoïdales à quatre brins appelées G-quadruplexes (Figure 1.30.C). L'empilement de quartets de guanines forme une cavité qui permet la présence de certains cations (comme le sodium (Na^+) et le potassium (K^+)) essentiels pour la stabilité des G- quartets (Figure 1.30.B). Depuis lors, des expériences dans des cellules ont permis de mettre en évidence la présence de structures G4s au sein des cellules. L'utilisation d'anticorps dirigés vers les G4 et couplés à des fluorophores a permis de mettre en évidence d'abord la présence de structures G4 aux télomères des mini-chromosomes du macronoyau d'un cilié (Schaffitzel et al., 2001; Paeschke et al., 2005), ensuite la présence et l'abondance de structures G4 au sein des cellules de mammifères (Biffi et al., 2013; Henderson et al., 2013). Cependant un de ces anticorps anti-G4 (nommé 1H6) est aussi capable de reconnaître les séquences simple brin riche en T et l'ADN dénaturé, limitant ainsi l'utilisation de ce dernier.

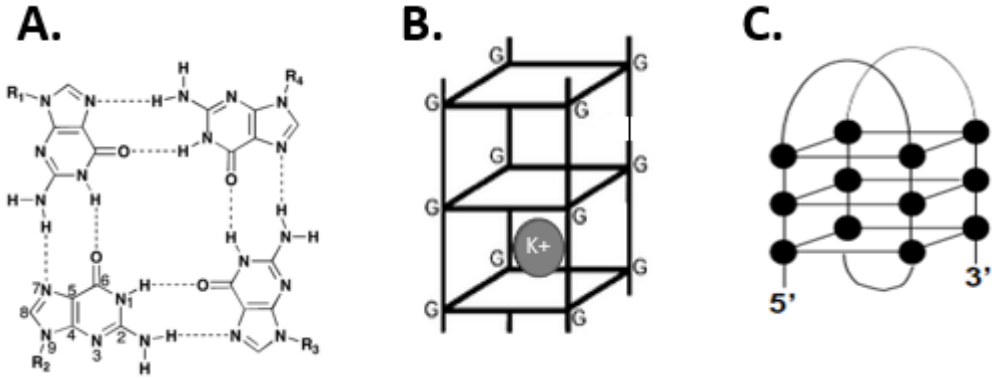


FIGURE 1.30 – **Modèle de la structure G-quadruplexe. A.** Un quartet de guanines, **B.** Cation présent entre deux quartets, **C.** Schéma d'un G4. Les • représentent les guanines.

Des analyses bioinformatiques, utilisant des algorithmes, ont révélé à partir de la séquence consensus (G_{3+}, N_{1-7}, G_{3+}, N_{1-7}, G_{3+}, N_{1-7}, G_{3+}) que le génome humain possède plus de 300 000 séquences susceptibles de former des G4 (Huppert, 2005). Plusieurs algorithmes ont été proposés pour la prédiction de structures G4 au sein des génomes (pour revue, (Lombardi and Londoño-Vallejo, 2019)). De manière significative il a été constaté que la localisation de ces séquences G4s n'est pas aléatoire. En effet ces séquences sont hautement conservées entre les espèces (König et al., 2010). On retrouve la

présence de ces structures G4s aussi bien chez les mammifères (Frees et al., 2014), que chez les bactéries (Beaume et al., 2012) ou encore au sein des virus ARN et ADN (Sundquist and Heaphy, 1993; Tuesuwan et al., 2008; Métifiot et al., 2014). Des séquences potentiellement G4 sont présentes dans différentes régions fonctionnelles du génome, par exemple les origines de réplication de l'ADN humain contiennent des motifs G4s. Les origines les plus fréquemment utilisées présentent une densité plus élevée en séquences G4 que les autres (Cayrou et al., 2012, 2011). Les structures G4s sont impliquées dans les divers mécanismes de la cellule à savoir la réplication de l'ADN, l'expression des gènes ou encore le maintien des télomères (Figure 1.31) (Rhodes and Lipps, 2015). En effet la région des télomères qui nous intéresse dans le cadre de la thèse est la région qui présente une des plus fortes abondances en séquences G4s au sein du génome humain en raison de la présence de répétition TTAGGG en tandem (Chariker et al., 2016).

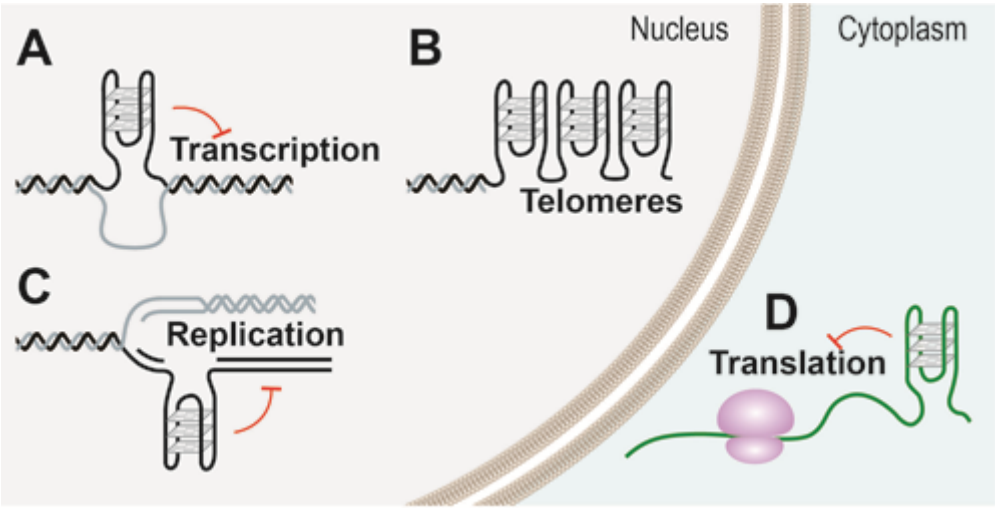


FIGURE 1.31 – **Localisation des structures G-quadruplexes au sein de la cellule.** (Rhodes and Lipps, 2015).

1.6.2 Polymorphisme structural des G-quadruplexes formés par le motif télomérique TTAGGG

Les structures G-quadruplexes sont très polymorphes : ils peuvent résulter d'un repliement intramoléculaire ou intermoléculaire de(s) séquence(s) d'ADN constituées de guanines successives et ils peuvent adopter une variété de conformations. Une même séquence peut se replier en plusieurs conformations, comme c'est le cas de la séquence télomérique humaine. Pour former un G4 intramoléculaire, une séquence doit comporter au moins quatre blocs

de guanines consécutives ; par conséquence, la séquence télomérique humaine minimale pour former un G4 est composée de trois répétitions et demi du motif télomérique, c'est à dire : 5'-GGG(TTAGGG)$_3$-3' (nous appelons cette séquence "H21").

La première structure G4 obtenue à partir la séquence télomérique humaine a été résolue en 1993 par RMN (résonance magnétique nucléaire) en présence de Na^+ (Figure 1.32.A) (Wang and Patel, 1993). Les boucles TTA assurant la connexion sont dans l'ordre suivant : latérale, diagonale, et latérale, de sorte que deux brins consécutifs sont antiparallèles ("*down, up, down, up*"). Ce type de structure G4 est dite de type "antiparallèle en panier". Une deuxième structure G4 en présence de Na^+ a été résolue par RMN en 2013 par Lim et al, avec la séquence 5'-(TTAGGG)$_3$TTA-3' (Lim et al., 2013) (Figure 1.32.B). L'agencement des boucles de connexion suit cette fois l'ordre suivant : latérale, externe (en "N", reliant deux brins parallèles), latérale, de sorte que le premier et le quatrième brin ont une direction opposée à celle du deuxième et du troisième brin ("*up, down, down, up*"). Ce type de structure est dite de type "mixte".

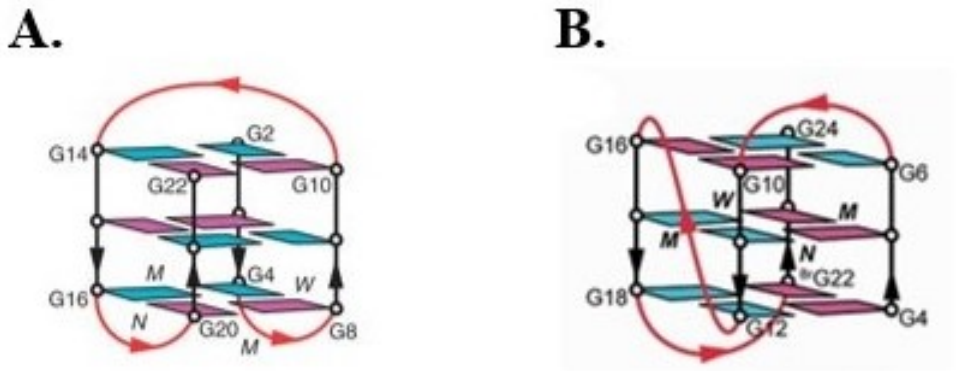


FIGURE 1.32 – **Polymorphisme structural du G4 de la séquence canonique en présence de sodium.** Les boucles sont colorées en rouge ; Les guanines anti et syn sont respectivement de couleur cyan et magenta. **A.** (Wang and Patel, 1993) , **B.** (Lim et al., 2013)

En 2002 l'équipe de S. Neidle a résolu par cristallographie une structure G-quadruplexe en présence de K^+ très différente des structures observées en sodium (Parkinson et al., 2002). Dans cette structure les quatre brins sont parallèles, les boucles TTA sont externes (en "N") et toutes les guanines sont dans une conformation anti (Figure 1.33.A). Cette structure, dite "parallèle" est aussi la structure formée en solution en présence de potassium et de PEG (Heddi and Phan, 2011). En 2006 l'équipe de D. J. Patel a résolu par RMN

deux structures G4 constituées de répétions TTAGGG en présence de K^+ (Luu et al., 2006; Phan et al., 2006). Désignées par *Form 1* (Figure 1.33.B) et *Form 2* (Figure 1.33.C), ces structures sont relativement similaires, elles contiennent deux boucles latérales et une boucle externe (en "N") générant une inversion des brins. Cependant elles diffèrent dans l'organisation de leurs boucles TTA : la boucle externe correspond à la troisième boucle dans la Forme 2 et à la première boucle dans la Forme 1. Il existe également une troisième forme en solution de potassium, dite *Form 3* : elle a une conformation du type antiparallèle en panier, elle repose sur deux plateaux de guanines et est aussi stable que la Forme 1 et la Forme 2 (Figure 1.33.D) (Lim et al., 2009).

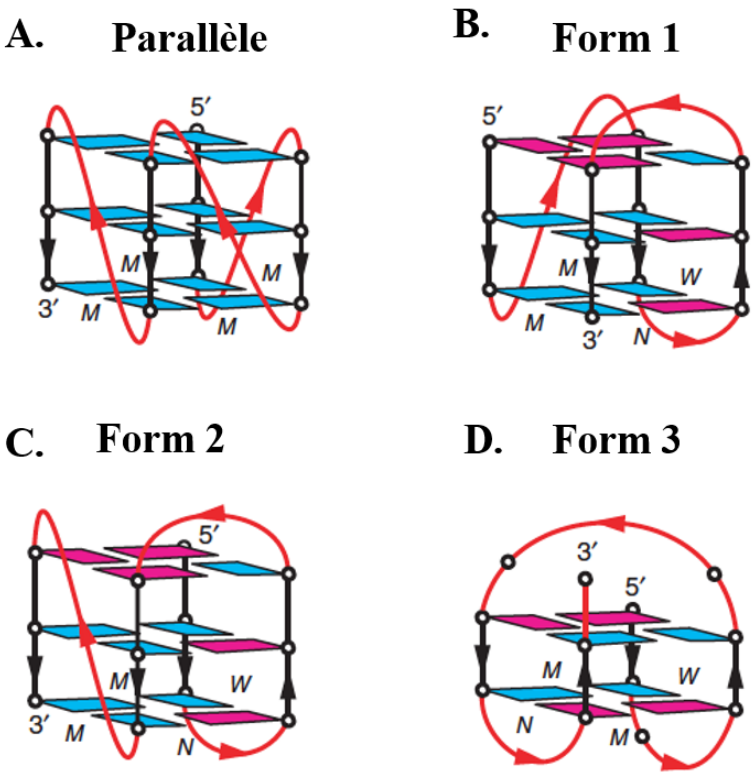


FIGURE 1.33 – **Polymorphisme structural du G4 de la séquence canonique en présence de potassium.** Les boucles sont colorées en rouge ; Les guanines anti et syn sont respectivement de couleur cyan et magenta. **A.** (Parkinson et al., 2002) , **B.** (Luu et al., 2006), **C.** (Phan et al., 2006), **C.** (Lim et al., 2009)

En résumé on dénombre six structures différentes résolues d'un G4 intramoléculaire formées à partir du motif télomérique TTAGGG : deux en présence de sodium et quatre en présence de potassium. On notera que la forme parallèle a été observée seulement en présence de PEG. Ces résultats illustrent bien le polymorphisme structural des G-quadruplexes qui s'exprime par l'orientation des brins (parallèle ou antiparallèle), la position des boucles de connexion et enfin la conformation des guanines au sein des quartets : syn ou anti.

1.6.3 Structures télomériques en G4 contigus

L'organisation en tandem des répétitions d'ADN télomérique constituées de guanines successives est presque universellement conservée chez les eucaryotes (section 1.2.4). *In vitro*, la majorité des séquences télomériques se structurent en G4 (Tran et al., 2010). Chez les ciliés, l'extension de l'extrémité 3' simple-brin est insuffisante pour former un G4 intramoléculaire. Des études ont montré qu'une structure G-quadruplex intermoléculaire antiparallèle est formée au niveau des télomères macronucléaires du cilié *Stylonychia lemnae* (Figure 1.34.A) (Schaffitzel et al., 2001; Paeschke et al., 2005). En revanche chez les mammifères la longueur de l'extrémité simple brin du Brin G est suffisamment importante pour former théoriquement plusieurs G4s. Des structures en G4 multiples peuvent également se former tout le long du Brin G télomérique lors de son ouverture, notamment au cours de la réplication ou de la transcription. Les études sur la formation, la nature et la stabilité des structures télomériques en G4 multiples ne sont pas nombreuses. Les premières études sur des longues séquences télomériques ont évoqué la formation de G4 contigus, sans véritablement démontrer leur formation (Yu et al., 2006; Vorlickova, 2005). Une étude en micro force atomique montrait la formation d'une structure en collier de perles compatible avec des structures en G4 multiples (Xu et al., 2009) Des études successives sur la stabilité de ces structures ont suggéré que la séquence télomérique humaine avait un potentiel limité pour se replier en G4 contigus (qu'elle pouvait se replier au plus dans 2 ou 3 G4 contigus) (Petraccone et al., 2011; Bauer et al., 2011). Toutefois, une étude menée au sein de notre équipe SANTÉ sur 8 12 et 16 répétitions télomériques a permis de montrer que les séquences télomériques se structurent en unités G4 contigues n'interagissant pas entre elles et de stabilité similaire, et que la stabilité de ces structures ne dépend pas du nombre d'unité G4 formées (Bugaut and Alberti, 2015). Dans le même temps, une étude en molécule unique sur une séquence de 24 répétitions du motif télomérique humain, confirmait la formation de structures G4 du type collier de perles et l'absence d'interactions entre unités G4 (Punnoose et al., 2014).

D'une part, il a été proposé que la formation de structures G4 à l'extrémité 3' simple-brin pourrait protéger cette extrémité contre les nucléases, moduler l'interaction avec des protéines de liaison à l'ADN simple-brin comme *POT1-TPP1* et *RPA* (Ray et al., 2014) et promouvoir la translocation de la télomérase (Jansson et al., 2019). D'autre part, les G4 aux télomères sont souvent considérés comme des structures problématiques pour la réplication des télomères, des structures qui nécessitent d'être déroulées (Figure 1.34.B).

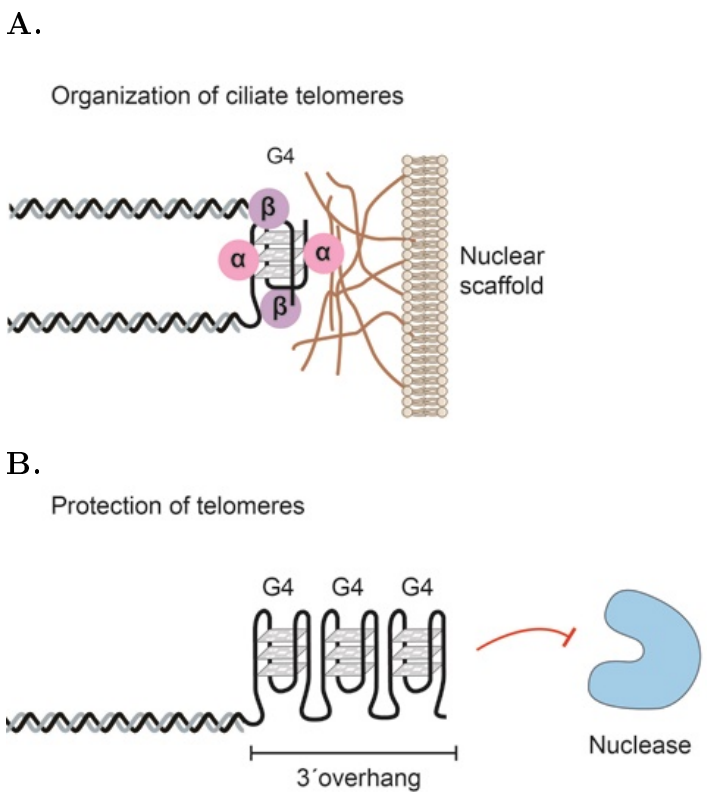


FIGURE 1.34 – **L'extension simple brin des télomères peut former des structures G-quadruplex impliquées dans la protection terminale des télomères. A.** Une structure G4 intermoléculaire chez les ciliées, **B.** La présence hypothétique de structures G4s en tandem au sein des mammifères (Rhodes and Lipps, 2015).

1.7 Mécanismes d'ouverture d'un seul G4 par la famille des *SSB-OB fold* aux télomères

Comme nous l'avons vu précédemment (section 1.6), des répétitions du motif télomérique TTAGGG se structurent en G4s. Ces structures G4s sont souvent considérées comme des structures problématiques pour la réplication des télomères. En effet ce point de vue repose sur l'observation qu'un défaut de certaines protéines de **la famille des hélicases**, aptes à dérouler *in vitro* ces structures, entraine une perturbation de la réplication des télomères. Ces études *in vivo* nous informent de l'importance des hélicases, notamment *WRN* (*Werner syndrome helicase*) (Crabbe, 2004), *BLM* (*Bloom syndrome protein*)(Zimmermann et al., 2014; Sfeir et al., 2009) ou encore *RTEL1* (*Regulator of telomere elongation helicase 1*)(Sfeir et al., 2009; Vannier et al., 2012), dans la réplication du brin G télomérique (Maestroni et al., 2017) (Figure 1.35). Il existe une autre famille de protéines susceptibles de prendre en charge cette activité de dérouler les G4s aux télomères : **la famille des SSB-OB fold**. Les protéines appartenant à cette famille ont souvent un rôle crucial au sein de la cellule rendant difficile des études *in vivo* présentant un défaut de ces protéines (Flynn and Zou, 2010). En revanche, des études *in vitro* ont permis de montrer qu'un certain nombre de protéines de liaison à l'ADN simple brin aux télomères, appartenant à **la famille des SSB-OB fold**, possèdent la capacité de dérouler les structures G4s. Chez l'homme il existe au moins trois complexes appartenant à cette famille possédant cette faculté à dérouler les structures G4s : *POT1-TPP1*, *RPA* et *CST*.

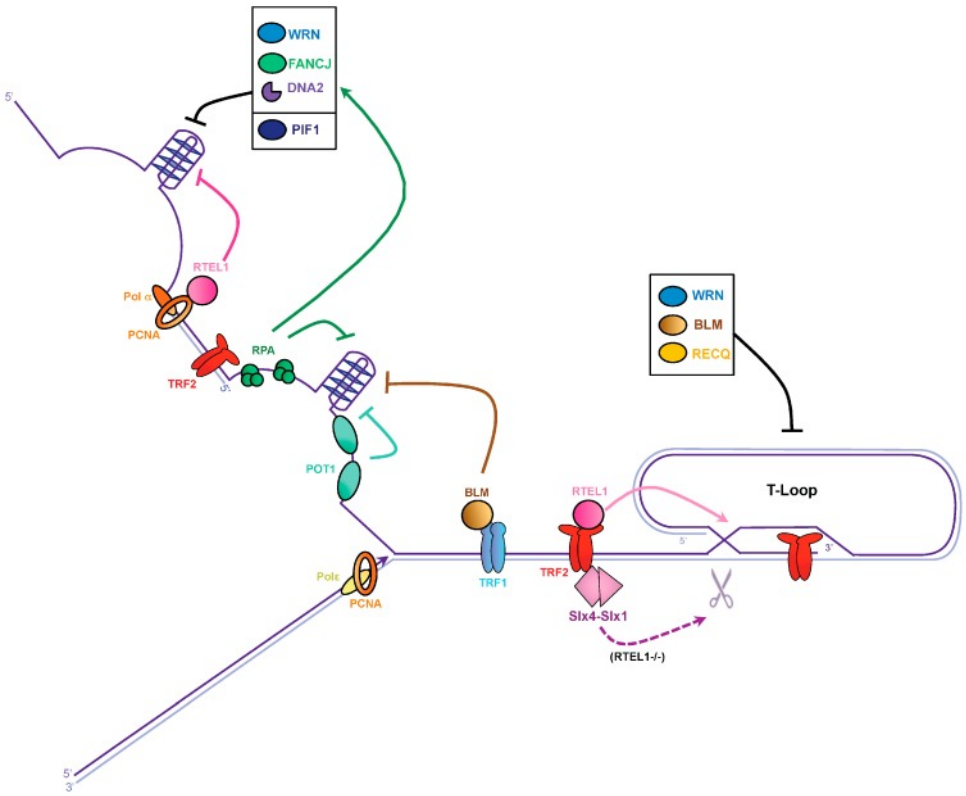


FIGURE 1.35 – **Prise en charge par les hélicases des structures G-quadruplexe et de la *T-loop* pour éviter le blocage de la fourche de réplication aux télomères.**

1.7.1 Le complexe *POT1-TPP1*

Des recherches *in vitro* réalisées en 2005 montrent que l'oligonucléotide $GGG(TTAGGG)_3$, replié en G4, est un mauvais substrat pour l'activité d'élongation de la télomérase, comparé à des substrats qui ne se structurent pas en G4. Cependant l'addition de POT1 restaure l'activité d'élongation de la télomérase. Ces résultats suggèrent que POT1 déroule le G4 facilitant ainsi l'activité d'élongation de la télomérase (Zaug et al., 2005). Des expériences en molécule unique entreprises par Hwang et al. en 2012 ont permis d'observer l'ouverture d'un G4 par la protéine *POT1* avec une directionnalité de 3' vers 5' (Hwang et al., 2012). Sur la base de la polarité de liaison des domaines *OB1* et *OB2* de *POT1* (section 1.5.2), les auteurs ont proposé un mécanisme d'ouverture du G4 dans lequel la liaison de *POT1* se fait dans un premier temps via le domaine *OB2*, du côté 3', suivi par le domaine *OB1*, du côté 5'(Figure 1.36). De plus, , l'observation de complexes Protéine :ADN de

stœchiométrie 2 :1 leur a permis de suggérer que deux protéines *POT1* sont nécessaires à l'ouverture d'un G4. Une étude publiée en mai 2020 par Chaires et al. montre que les courbes de liaison de la protéine *POT1* sur des séquences télomériques de 24 nt, composées de deux sites de fixation de *POT1* et structurées en G4, présentent une allure sigmoïdale suggérant une coopérativité de *POT1* dans l'ouverture du G4 (Chaires et al., 2020). A l'inverse, il semble qu'un seul complexe *POT1-TPP1* soit suffisant pour dérouler un G4 (Hwang et al., 2012; Ray et al., 2014). Afin d'expliquer cette différence, le modèle de Hwang et al. propose que le complexe *POT1-TPP1* glisse le long de l'ADN télomérique (Figure 1.36). Cette idée que le complexe *POT1-TPP1* glisse sur l'ADN télomérique est renforcée par le rôle du complexe dans la processivité de la télomérase. Cependant même s'il existe des modèles de protéines liant l'ADN en glissant le long de ce dernier, comme c'est le cas de *Cdc13* chez *S. cerevisiae*, il me semble difficile que le complexe *POT1-TPP1* ait un tel comportement au vu de la spécificité de séquence de *POT1*. Il faudrait que le complexe *POT1-TPP1* "saute" de site en site sur l'ADN télomérique, et non glisse comme proposé, ce qui ne semble pas cohérent.

Une étude réalisée en 2014 par Ray et al. montre à l'aide d'expérience en molécule unique que les structures G4 sont mieux déroulées par le complexe *POT1-TPP1* que par *POT1* seule (Ray et al., 2014). Mullins et al. ont étudié la liaison de *POT1-TPP1* à des séquences de 24 nt comportant deux sites de liaison et se structurant ou non en G4 (Mullins et al., 2016). Ils ont observé que la liaison du premier hétérodimère *POT1-TPP1* à la séquence qui se structure en G4 dépend fortement de la stabilité du G4, l'affinité en K^+ étant cinq fois inférieure à l'affinité en Na^+. En revanche, l'affinité de liaison du deuxième hétérodimère *POT1-TPP1* ne dépend pas de la structure secondaire et de sa stabilité, vraisemblablement en raison du dépliage résultant de la liaison du premier complexe *POT1-TPP1*.

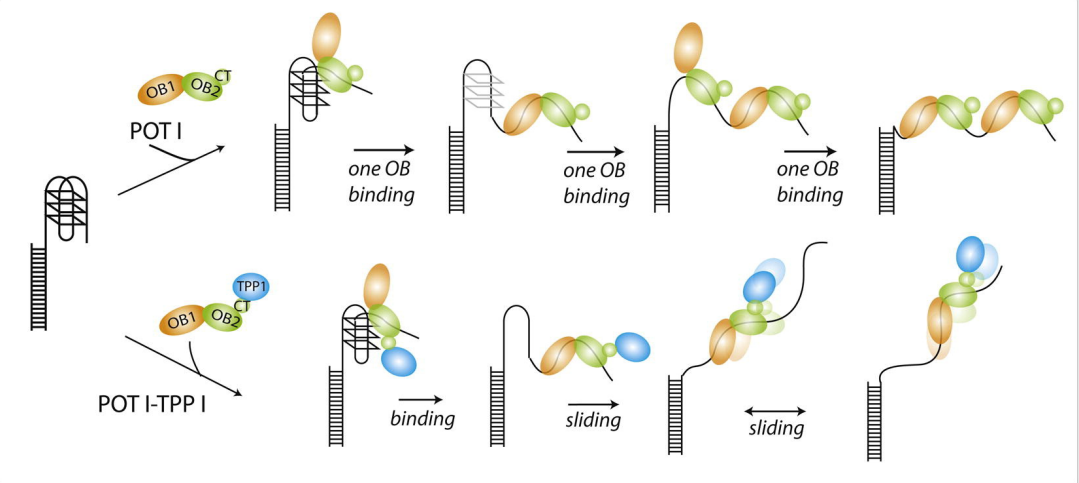


FIGURE 1.36 – **Modèle d'ouverture d'un G-quadruplexe par *POT1* et *POT1-TPP1*.** Proposé par Hwang et al. (Hwang et al., 2012).

1.7.2 La protéine *RPA*

De nombreuses données chez la levure ont suggéré que la protéine *RPA* pouvait intervenir aux télomères en maintenant le brin G sous une forme non structurée compatible pour la réplication des télomères et l'accès à la télomérase (Schramke et al., 2003; Cohen et al., 2004; Luciano et al., 2012; Audry et al., 2015). En effet, de nombreuses études *in vitro* ont montré que la protéine *RPA* se lie aux séquences structurées en G4 simple en déroulant la structure. Cela a été montré pour le G4 formé par la séquence télomérique humaine (Salas et al., 2006; Safa et al., 2014) et pour d'autres séquences du génome humain (Fan et al., 2009; Ray et al., 2013). Qureshi et al. observent que *RPA* ouvre les structures G-quadruplex avec différents degrés d'efficacité (Qureshi et al., 2012). Des études ont montré que *RPA* lie et ouvre préférentiellement les G4 les moins stables établissant ainsi une relation inverse entre affinité et stabilité des G4 (Ray et al., 2013; Safa et al., 2014).

Des recherches *in vitro* ont permis de mettre en évidence que les interactions des domaines *OBD* (*RPA2*) et *OBE* (*RPA3*) de *RPA* sont essentielles pour dérouler un G4 (Prakash et al., 2011). Enfin des études *in vitro*, menées au sein de l'équipe SANTÉ ont montré que la protéine *RPA* déroule les G4 de façon directionnelle, de 5' vers 3' (Safa et al., 2016) : une fois établi un premier contact de quelques nt avec l'ADN via la sous-unité *RPA1*, elle s'étend sur l'ADN vers 3' (Figure 1.37). Cette directionnalité de *RPA* dans le déroulement des G4 télomériques pourrait assurer une fixation rapide de *RPA* au brin G retardé lors de la progression de la fourche de réplication.

De récentes études ont montré que la capacité de *RPA* à dérouler les structures G4 se retrouve chez différentes espèces. Parmi ces études nous pouvons citer l'étude de Carlos A.H Fernandes et al., à laquelle j'ai eu l'opportunité de participer, portant sur la *RPA* du parasite protozoaire *Leishmania amazonensis* (Fernandes et al., 2020) (en Annexes Articles) ou encore l'étude de B. Aklilu et al. sur la *RPA* de la plante *Arabidopsis thaliana* (Aklilu et al., 2020).

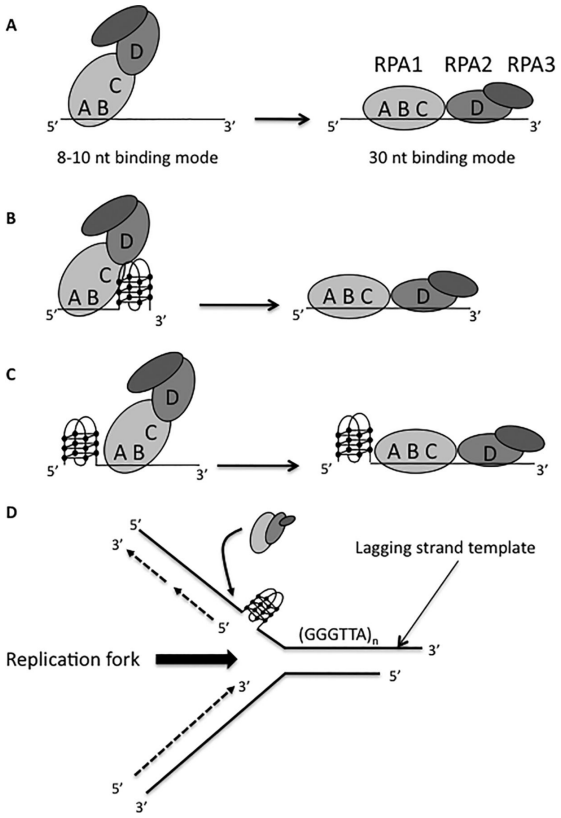


FIGURE 1.37 – **Modèle d'ouverture d'un G-quadruplexe par la protéine humaine *RPA*** (Safa et al., 2016).

1.7.3 Le complexe *CST*

Récemment, il a été observé que la présence de ligands de G4 exacerbe les problèmes de réplication du brin G observés en défaut de *CST* (Zhang et al., 2019). Ces résultats confortent les résultats obtenus *in vitro* par l'équipe de Carolyn M. Price en 2017 qui montre, outre la préférence de *CST* pour la jonction d'ADN double brin/simple brin, la capacité de ce complexe à dérouler le G4 formé par le motif télomérique humain (Bhattacharjee et al.,

2017). Cette étude montre que le complexe *CST* ne possède pas de préférence pour une extrémité 3' ou 5' de l'ADN simple brin. En effet l'efficacité de liaison de *CST* est similaire, qu'il y ait ou non présence d'ADN simple brin en 3' ou 5' du G4. Il est important de noter que le mécanisme qui régit l'ouverture des structures G4s par le complexe *CST* est aujourd'hui encore inconnu.

1.8 Les problématiques de la thèse

Des études, notamment au sein du laboratoire, ont montré que des longues séquences d'ADN télomérique sont capables de former *in vitro* des structures en G4 contigus (ou en tandem ou multiples). L'équipe SANTÉ s'intéresse au mécanisme de prise en charge de ces structures en G4 multiples par les protéines de la **famille des SSB-OB fold** présentes aux télomères. C'est pourquoi une étude récente, à laquelle j'ai pu participer, a été réalisée dans l'équipe avec l'objectif de caractériser l'interaction de la protéine *RPA* avec des longues séquences télomériques structurées en G4s en tandem (Lancrey et al., 2018). Cette étude a montré la capacité de *RPA* à se lier à ces séquences, en ouvrant les différentes unités G4, nous reviendrons sur ces résultats dans le chapitre 2 de la thèse. C'est dans ce contexte que j'ai commencé ma thèse en étudiant la liaison de *POT1* et *POT1-TPP1* à de longues séquences télomériques structurées en G4 multiples. La question posée était de déterminer les mécanismes de liaison de *POT1* et *POT1-TPP1* en présence d'oligonucléotides constitués de plusieurs répétitions télomériques pouvant mimer le *G-overhang in vitro*. Pour réaliser cette étude j'ai eu recours à des techniques de *EMSA* et de *FRET* qui seront présentées par la suite. Cette problématique est développée et détaillée dans le deuxième chapitre.

Le deuxième objectif de ma thèse a été de caractériser les mécanismes de liaison du complexe *POT1-TPP1* à l'ADN, notamment en s'interrogeant sur l'origine cinétique et structurale de la plus haute affinité du complexe *POT1-TPP1* par rapport à *POT1* seule. Pour répondre à cette question j'ai réalisé, par résonance plasmonique de surface (*SPR*), des expériences de cinétique d'association et dissociation de *POT1* avec son substrat d'ADN en présence et en absence de *TPP1*. J'ai également réalisé des expériences de photopontage du complexe *POT1-TPP1* avec l'ADN pour identifier les différentes sous-unités en contact avec l'ADN et, notamment, comprendre si *TPP1* est en contact avec l'ADN. Les résultats obtenus m'ont stimulé à poursuivre cet axe de recherche avec des expériences de microscopie électronique pour déterminer l'enveloppe du complexe *POT1-TPP1* en présence ou non d'ADN. Les résultats relatifs à cette partie sont développés dans le troisième chapitre.

Enfin j'ai participé à une étude portant sur un motif télomérique humain instable. En proximité des régions subtélomériques, les télomères humains présentent du polymorphisme : des répétitions de motifs non-canoniques sont présents au sein des répétitions canoniques TTAGGG. Pour comprendre l'origine de l'instabilité de ce motif, des études structurales ont été menées au sein de l'équipe. Ces études ont permis de montrer que'à partir d'une dizaine de répétitions, ce motif peut se structurer non seulement en G4 mais également en tige-boucle. Je me suis alors intéressé à la liaison de *RPA* et de *POT1-TPP1* avec les différentes structures formées par des répétitions de ce motif, pour comprendre si ces structures perturbent l'interaction de ces deux *SBB-OB fold* avec les télomères. Cette étude est abordée dans le quatrième chapitre de la thèse.

Chapitre 2

Mécanismes de liaison des *SSB* à des séquences télomériques structurées en G4 contigus

2.1 Introduction

Comme je l'ai présenté dans l'introduction générale (section 1.6.3), des études *in vitro* ont montré que le brin G télomérique peut se structurer dans le cas de longues séquences en G4s contigus (ou multimères de G4 en tandem) très stables, suggérant l'existence de telles structures au sein de la cellule. Comment ces multimères de G4s en tandem sont pris en charge et déroulées au sein de la cellule ? En effet, leur présence sur le brin G pourrait poser un problème au cours de la phase S, en bloquant la progression de la fourche de réplication et l'accès de la télomérase au *G-overhang*) (Figure 2.1). Nous nous sommes donc demandés si de telles structures pouvaient être prises en charge par les protéines de **la famille des SSB-OB fold** impliquées dans le maintien de télomères.

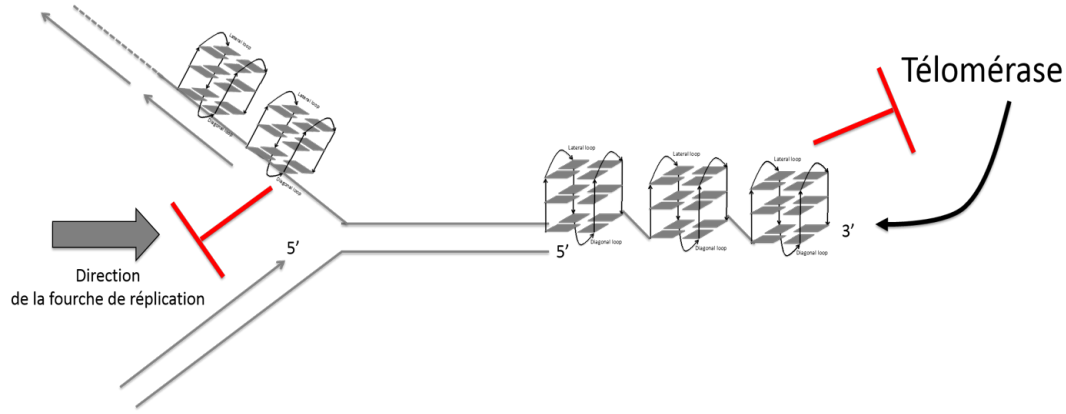


FIGURE 2.1 – **Schéma illustrant les sites potentiels de formation de multimères de G4 aux télomères et leur impact au cours de la phase S.**

2.1.1 Étude de la liaison de *RPA* sur des séquences télomériques formant des structures G4s en tandem

L'équipe SANTÉ a tout d'abord entrepris d'étudier la liaison de la protéine *RPA* à des longues séquences télomériques formant des multimères de G4s en tandem (étude à laquelle j'ai participé) (Lancrey et al., 2018) (en Annexes Articles). Le but de cette étude était de déterminer si *RPA* pouvait ouvrir ces G4 en tandem et si l'efficacité de la liaison de *RPA* aux brins G télomériques dépendait du nombre d'unités G4 en tandem. Pour répondre à cette question, nous avons eu recours à des expériences d'*EMSA* et de *FRET* en présence de différents oligonucléotides capables de se structurer respectivement en deux, trois ou quatre G4s contigus. Les résultats *in vitro* nous ont permis de conclure que le nombre n d'unités G4 contigus (n $\geq$2) n'affecte pas l'efficacité de *RPA* pour lier les brins G télomériques formant des G4s en tandem. Ce comportement s'explique par la nature des G4 télomériques, c'est-à-dire des unités structurales identiques et indépendantes (Bugaut and Alberti, 2015) et par la très faible coopérativité de *RPA*. Les expériences de *FRET* ont montré que *RPA* est capable d'ouvrir les différents G4s contigus présent sur ces longues séquences. Nous avons donc proposé que la prise en charge de ces structures par *RPA* sur le brin G pourrait jouer un rôle crucial pour la réplication des télomères et leur stabilité.

2.1.2 Le complexe *POT1-TPP1* peut-il prendre en charge ces multimères de G4s en tandem ?

De nombreuses équipes de recherche ont étudié les interactions de *POT1* ou *POT1-TPP1* sur de longues séquences télomériques contenant des répétitions importantes du motif TTAGGG. Ces études sont au nombre de quatre et sont présentés ci-dessous.

La première étude a été réalisée en 2010 par H. Wang et al., sur molécule unique, par microscopie à force atomique *atomic force microscope* (*AFM*)). Elle montre que la protéine *POT1* seule peut dérouler des structures G4 présentes sur de longues séquences télomériques (48-nt, soit 8 répétitions, et 96-nt, soit 16 répétitions), en présence de potassium (Wang et al., 2010). Toutefois, dans les conditions d'analyse, les auteurs ont observé que la majorité de l'ADN ne forme que la moitié des G4 théoriquement possibles, et que ces derniers ne sont pas disposés en tandem.

La seconde étude a été menée par l'équipe de Thomas R. Cech en 2011 avec de longues séquences télomériques de 72-(12 répétitions) à 144-nt (24 répétitions) (Taylor et al., 2011). Cette étude révèle que *POT1* et *POT1-TPP1* sont capables de recouvrir complètement ces longues séquences ; l'observation sur gel de complexes intermédiaires non saturés suggère que ces évènements de liaison sont peu ou pas coopératif. L'analyse par microscopie électronique (ME) montre que les complexes formés avec la séquence télomérique de 144-nt saturée par 12 hétérodimères *POT1-TPP1*, forment des structures globulaires compactes homogènes, reflétant un assemblage ordonné des protéines sur l'ADN (Figure 2.2 Toutefois, bien que les conditions expérimentales (présence de sodium) permettent à ces longs oligonucléotides de se structurer *in vitro* en G4s contigus, certes moins stables qu'en potassium, l'étude n'aborde pas cet aspect.

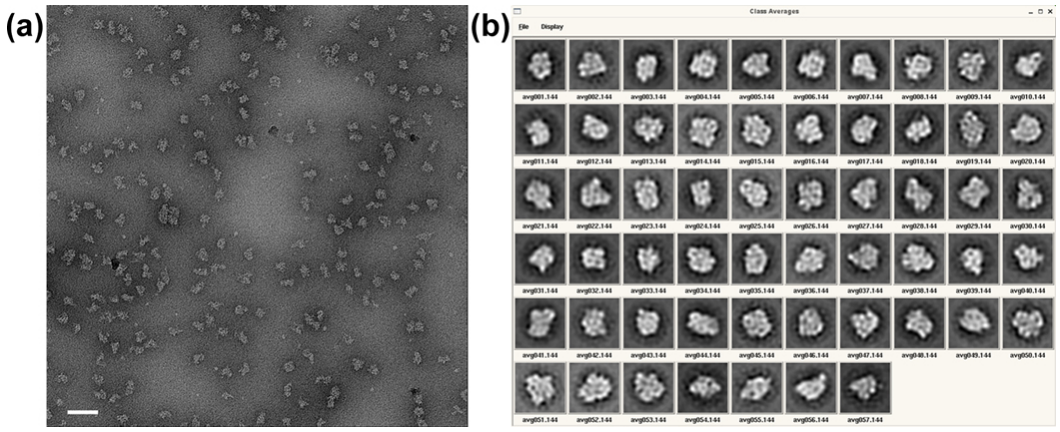


FIGURE 2.2 – **Microscopie électronique à coloration négative des complexes *POT1-TPP1* en présence d'un ADN télomériques (144 nt).(a)** Une micrographie typique de complexes télomériques assemblés à partir d'ADN télomérique de 144 nt saturé par douze hétérodimères *POT1-TPP1*. Les particules sont de forme globulaire et monodispersées sur la grille. **(b)** Classification 2D des complexes télomériques 144nt-*POT1-TPP1* (Taylor et al., 2011)

.

La troisième étude a été réalisée en 2013 par Corriveau et al. avec une séquence télomérique de 48-nt (8 répétitions), et a également montré que plusieurs complexes *POT1-TPP1* peuvent recouvrir complètement l'ADN (Corriveau et al., 2013). Toutefois l'oligonucleotide choisi pour leur étude, $(GGTTAG)_8$, ne présentant que sept blocs de trois guanines successives, ne permet pas de former deux G4s, mais uniquement un seul G4. La suite de l'étude a consisté à étudier l'interaction de *POT1-TPP1* avec des séquences mutées du 48-nt, ne pouvant plus former de structure en G-quadruplexe (remplacement des deux guanines situées entre les motifs de liaison de POT1, par des cytosines : 5'-(**CC**TTAGGGTTAG)$_4$-3') et possédant un motif de liaison de *POT1* muté (remplacement d'une des guanines du motif de liaison de POT1, par un cytosine : 5'-(TTAG**C**GTTAG)-3'). Dans ce contexte même si leurs séquences conservent le même nombre de sites de fixation de *POT1-TPP1*, elles forment maintenant des structures tige-boucle non représentatif du *G-overhang in vivo* (Mullins et al., 2016). L'objectif de cette étude était d'étudier l'effet de la structure G4 sur la liaison du complexe et de tester l'idée d'une coordination de la liaison de 3' vers 5' des protéines le long de l'ADN, idée basée sur l'observation que *POT1* possède une préférence pour l'extrémité 3' du brin G. Si les résultats montrent clairement que la présence d'un G-quadruplex peut influencer le mécanisme de liaison des protéines, les résultats obtenus avec les séquences mutées dans le motif de *POT1* ne permettent pas d'avancer un modèle de liaison coordonné et directionnel.

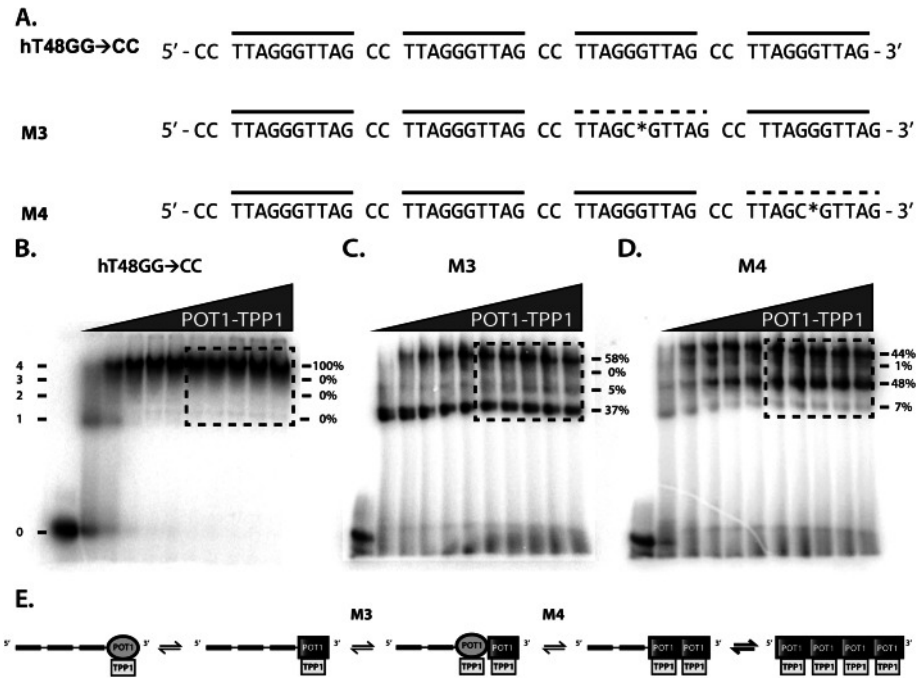


FIGURE 2.3 – **Liaison séquentielle de *POT1-TPP1*.A.** Diagramme des séquences d'ADN étudiées. Les barres pleines au-dessus de la séquence indiquent les sites de liaison *POT1* individuels. Les lignes en pointillées indiquent une mutation de G en C* dans le site de liaison de *POT1* de l'ADN. **B-C-D.** Les expériences de *EMSA* en présence des différentes séquences. Les complexes *POT1-TPP1*-ADN de différentes stœchiométries ont été quantifiés. (Corriveau et al., 2013)

.

Enfin, la dernière étude publiée en 2019 par l'équipe de Derek J. Taylor a été réalisée (Xu et al., 2019) avec une *POT1* possédant une mutation H266L, et, soit une séquence courte de 12-nt (H12 ; 5'-GGTTAGGGTTAG-3'), soit une séquence longue de 72-nt (H72), en présence de sodium. Les résultats montrent que l'acide aminé H266 est essentielle pour permettre la liaison de *POT1-TPP1* à la séquence (H12) (Figure 2.4.A-B), ce qui n'est pas le cas avec (H72) (Figure 2.4.C-D). Ce résultat montre que *POT1-TPP1* ne se comporte pas de la même façon sur de courtes séquences télomériques comparé à des séquences plus longues, sous-entendant un mode de liaison différents en fonction de la taille de l'oligonucléotide. Bien que la séquence télomérique H72 puisse se structurer *in vitro* en trois G4s contigus, l'article n'aborde pas cette dimension G4 et par conséquent ne traite pas du comportement de *POT1-TPP1* sur des séquences formant des G4s en tandem.

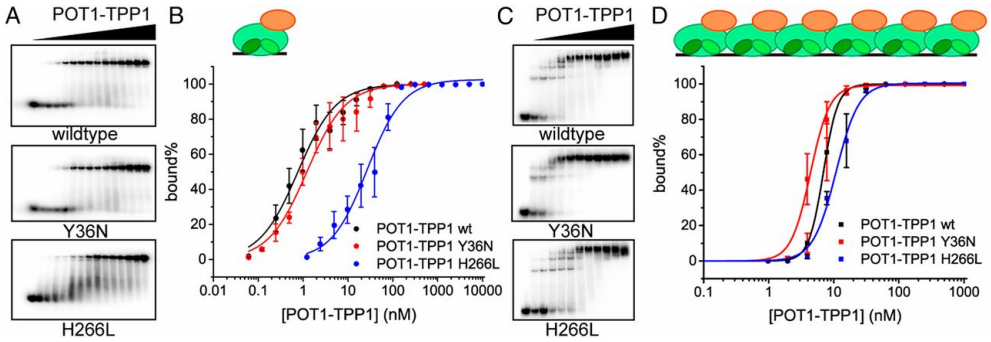


FIGURE 2.4 – **L'efficacité de liaison du mutant *POT1(H266L)-TPP1* est dépendante de la longueur du substrat de l'ADN simple brin.** **A.** *EMSA* pour comparer l'effet des mutations de *POT1* sur la liaison du complexe *POT1-TPP1* en présence d'un court ADN télomérique (12 nt). **B.** Quantification des données *EMSA* pour *POT1-TPP1* et les protéines mutants en présence d'un court ADN télomérique (12 nt). **C.** *EMSA* pour comparer l'effet des mutations de *POT1* sur la liaison du complexe *POT1-TPP1* en présence d'un long ADN télomérique (72 nt). **D.** Quantification des données *EMSA* pour *POT1-TPP1* et les protéines mutants en présence d'un long ADN télomérique (72 nt). (Xu et al., 2019)

En résumé, les différentes études des interactions de *POT1* et *POT1-TPP1* avec de longues séquences télomériques ont montré que ces protéines peuvent recouvrir entièrement l'ADN et ouvrir les G-quadruplexes. Toutefois, dans la grande majorité des cas, ces études ont été réalisées en présence de sodium, une condition saline qui n'est pas biologiquement représentative, puisque dans la cellule il y a une concentration importante de potassium qui joue un rôle stabilisateur des G4. Quant à la seule étude faite en potassium, elle a été réalisée sur des ADN qui ne forment que très rarement des multimères de G4 en tandem. A ce jour, il n'y a donc eu aucune étude sur la liaison et l'ouverture de de G4 télomériques en tandem en potassium par la protéine *POT1* et/ou le complexe *POT1-TPP1*.

2.2 Présentation de l'étude

Dans ce contexte, j'ai entrepris d'étudier *in vitro* les interactions de la protéine *POT1* et du complexe *POT1-TPP1* sur de longues séquences télomériques structurées en G4s contigus, en présence de potassium, cation biologiquement représentatif et qui est le plus favorable pour stabiliser les G4. Pour répondre à cette question, j'ai réalisé : (i) des expériences de *EMSA*, afin de visualiser les complexes nucléoprotéiques formés, et (ii) des expériences de *FRET* (titrages et cinétiques), afin de suivre l'ouverture des G4 (principe décrit en annexe de ce chapitre pages 115). Le travail réalisé avec *POT1-TPP1* est présenté ci-après sous forme d'un article dans la section 2.3 ; les résultats obtenus avec *POT1* sont présentés à la suite de l'article.

2.2.1 Multimères de *POT1-TPP1* : une coopérativité de liaison à l'ADN due à la structuration de l'ADN

Lors de cette étude, j'ai réalisé des expériences de *EMSA* permettant d'étudier la liaison du complexe *POT1-TPP1* avec de longues séquences d'ADN télomérique capables de se structurer en G4s contigus. Les résultats ont permis de montrer que le complexe *POT1-TPP1* peut prendre en charge de G4s en tandem en formant des complexes multimèriques dans lesquels tous les sites de liaison sont saturés. Nous avons remarqué que les courbes de liaison présentaient un aspect sigmoïdale, suggérant un phénomène de liaison coopérative de *POT1-TPP1* avec des multimères de G4s en tandem. Dans le but de comprendre cette coopérativité surprenante pour une protéine connue pour ne pas être coopérative (d'après les expériences en sodium), j'ai réalisé des expériences similaires en présence de lithium, condition saline ne permettant pas de stabiliser les G4s. Dans ce cas, les courbes de liaison ne présentaient plus cet aspect sigmoïdal, traduisant l'absence de coopérativité de liaison de *POT1-TPP1* sur de longues séquences télomériques non structurées. Á ce stade, nous pouvons donc conclure que le complexe *POT1-TPP1* ne présentent pas de coopérativité à proprement parler, c'est-à-dire d'interaction entre les protéines, mais plutôt une **coopérativité apparente** dû à la présence des multimères de G4s : la liaison d'un premier complexe POT1-TPP1 à l'ADN déstabilise un G4, facilitant ainsi la liaison d'un deuxième complexe et ainsi de suite.

2.2.2 Multimères de *POT1-TPP1* : une progression de liaison à l'ADN de 3' vers 5'

Dans le but de vérifier si la liaison de plusieurs complexes *POT1-TPP1* à l'ADN procède avec une directionalité ou non, j'ai réalisé des expériences de *FRET* avec un oligonucléotide capable de former trois G4s en tandem.

Les différents G4s ont été marqués par un couple de fluorochromes ; cela m'a permis de suivre indépendamment l'ouverture de chaque G4s au sein de la séquence. Les résultats de FRET (titrations et cinétiques) montrent que *POT1-TPP1* ouvrent plus rapidement le G4 en 3' que celui en 5'. Le fait que le premier G4 à être déroulé et celui du coté 3' de la séquence est cohérent avec la préférence de liaison de *POT1-TPP1* pour le site de liaison situé à l'extrémité 3', préférence déjà démontrée par des études précédentes (Lei et al., 2004; Hwang et al., 2012; Choi et al., 2015). La directionalité de 3' vers 5' dans la progression de la liaison de plusieurs complexes *POT1-TPP1* à l'ADN est un des résultats originaux de notre étude ; une telle directionalité avait été évoquée auparavant, mais pas véritablement démontrée (Corriveau et al., 2013).

2.2.3 Multimères de *POT1-TPP1* : origine de la directionnalité de 3' vers 5'

La directionalité que nous observons s'explique par deux facteurs : la préférence de liaison en 3' et la coopérativité de liaison en potassium (l'autre résultat original de notre étude). En compilant les résultats des expériences de *EMSA* et *FRET*, nous avons proposé un modèle qui permet de rendre compte du phénomène de la coopérativité apparente en KCl et de la directionalité de la liaison des complexes *POT1-TPP1*, observée en présence des G4 en tandem. La fixation préférentielle de *POT1-TPP1* à l'extrémité 3' induit l'ouverture du G4 en 3' La présence du complexe en 3' déstabiliserait le G4 suivant en favorisant ainsi la liaison d'un second complexe. C'est ce phénomène qu'on appelle **coopérativité apparente** (Figure 2.5) : **coopérativité** car la fixation d'une première protéine favorise la fixation d'une deuxième, mais **apparente** car elle découle de la présence des G4s contigus. Une coopérativité similaire a été observée lors d'une étude, portant sur la liaison de *POT1-TPP1* à une séquence H24, comportant deux sites de fixation distinct de *POT1-TPP1*, et capable de se structurer en G4 : la fixation d'un complexe *POT1-TPP1* déroule le G4 favorisant la fixation du deuxième complexe (Mullins et al., 2016).

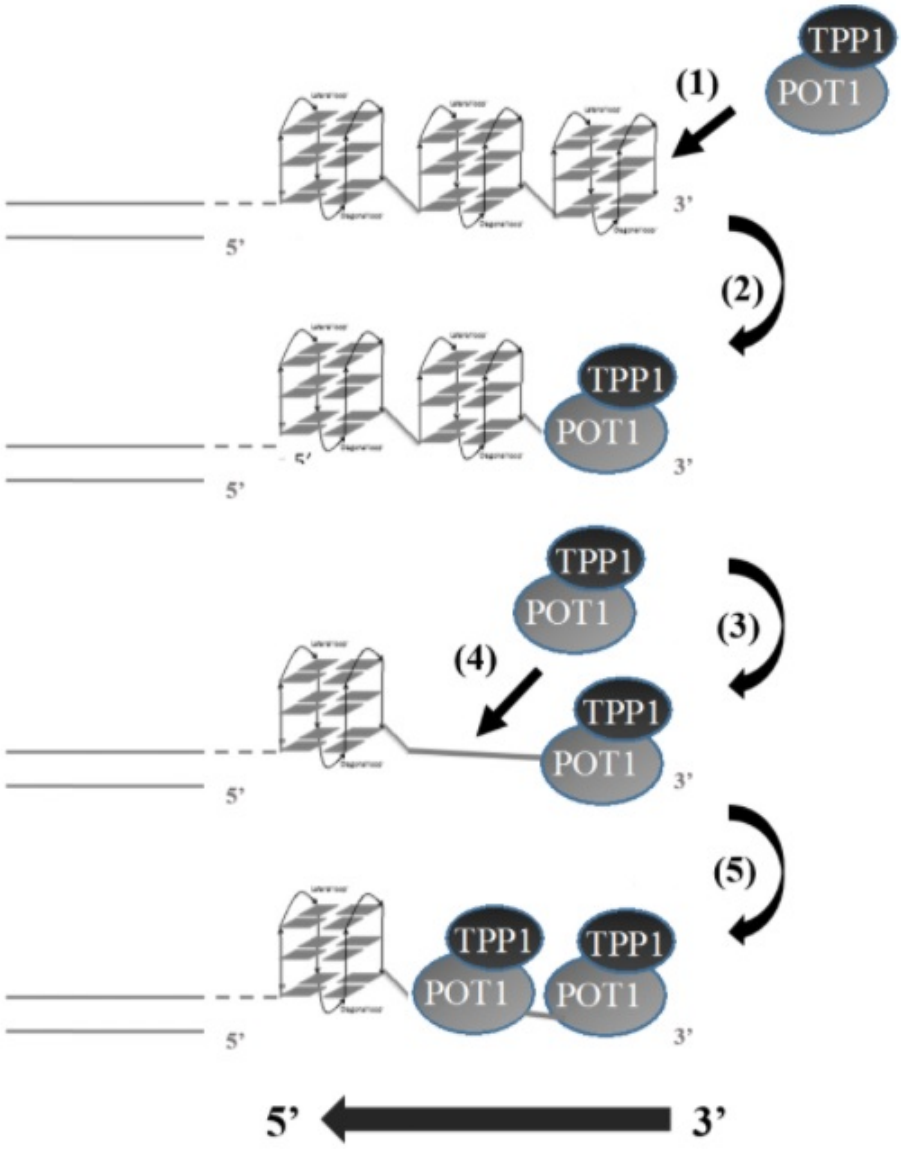


FIGURE 2.5 – **Modèle de la coopérativité apparente de *POT1-TPP1* sur des longues séquences télomériques formant des multimères de G4 en tandem.**

2.3 Article 1

2.3.1 Résumé

Les télomères sont des séquences d'ADN associées aux protéines du *Shelterin*, qui sont situées aux extrémités des chromosomes pour les protéger. Ils sont composés d'une région double brin et d'une région simple brin appelée *G-overhang*. Des études *in vitro* ont montré que le *G-overhang* humain forme des structures G-quadruplexe (G4) contiguës hautement stables. Ces structures en tandem doivent être résolues pendant la réplication des télomères et pour l'activité de la télomérase. La protéine humaine *hPOT1* (*Protection of Telomeres 1*), complexée avec son partenaire télomérique *TPP1*, recouvre de longues séquences télomériques formant un complexe multi-*hPOT1-TPP1* : ADN. Ici, nous montrons que de multiple complexe *hPOT1-TPP1* recouvrent plusieurs séquences télomériques structurées en G4 contigus, en dépliant ces structures G4 avec un mode de liaison séquentiel de 3' vers 5'. Nous proposons que cette directionnalité découle de deux caractéristiques : la coopérativité et la préférence de liaison à l'extrémité 3. En outre, nous montrons que la coopérativité de liaison *hPOT1-TPP1* n'est pas intrinsèque au mode de liaison *hPOT1* à l'ADN, mais découle de la structuration de l'ADN télomérique en unités G4 indépendante.

Multiple POT1-TPP1 cooperatively unfold contiguous telomeric G4 with a 3' to 5' sequential binding mode, arising from a 3'-end binding preference and structuration of telomeric DNA.

Jean Chatain[a], Georges Hatem[a], Jean-François Riou[a], Patrizia Alberti[a], and Carole Saintomé[*a,b]

[a]Structure et Instabilité des Génomes (StrInG), Muséum national d'histoire naturelle, CNRS, Inserm ; CP26, 57 rue Cuvier 75005 Paris, France.
[b]Sorbonne Université, UFR927, F-75005, Paris, France

September 24, 2020

Abstract

Telomeres are DNA sequences associated with shelterin proteins, that are located at the ends of the chromosomes to protect them. They are composed of a double-stranded region and a single-stranded region called the G-overhang. In vitro studies have shown that the human G-overhang forms stable contiguous G-quadruplex structures (G4) that need to be resolved during telomere replication and for telomerase activity. The human shelterin protein hPOT1 (Protection of Telomeres 1), in complex with its telomeric partner TPP1, coats long telomeric sequences forming a multi-hPOT1-TPP1:DNA complex. Here, we show that multiple hPOT1-TPP1 coat and unfold telomeric sequences structured in contiguous G4s, with a 3' to 5' sequential binding mode. We propose that this directionality arises from two features: cooperativity and 3'-end binding preference. Furthermore, we show that hPOT1-TPP1 binding cooperativity is not intrinsic to hPOT1 mode of binding to DNA, but arises from the structuration of telomeric DNA in G4 units.

1 Introduction

Human telomeres are composed of duplex TTAGGG repeats with a 3' single-stranded extension named G-overhang (50 -300 nt) and 6 telomeric proteins that constitute the shelterin complex. Telomeric DNA can be extended by telomerase, a ribonucleoprotein complex that adds GGTTAG repeats to the G-overhang[5]. It has been shown

*Corresponding author: carole.saintome@mnhn.fr

that 4 human telomeric repeats can fold into G-quadruplex structures (G4)[26] and the presence of such structures impedes the recruitment of telomerase[31]. Among shelterin proteins, the human telomeric single-stranded DNA binding protein hPOT1 (human Protection of Telomeres 1), involved in chromosome-end protection and in telomere-length regulation[3,4,14,8,1,13], is able to unfold these structures allowing telomerase extension in vitro[31]. hPOT1 consists of three OB fold (oligonucleotide/oligosaccharide binding) domains, OB1 and OB2 are involved in the DNA binding[13] and OB3 in the interaction with its shelterin partner TPP1 (Figure 1A). It has been shown that hPOT1 is sequence specific[15]; the OB1 domain recognizes the 5'-TTAGGG-3' sequence, while the OB2 domain binds the 5'-TTAG-3' sequence[13]. Thus, the 10 nt 5'-TTAGGGTTAG-3' sequence is the minimum binding site of hPOT1. hPOT1 interacts with TPP1 to form a heterodimer that protects chromosome ends and serves as a processivity factor for telomerase during telomere extension[24]. Although TPP1 also contains an N-terminal OB fold (Figure 1.A), no binding has ever been observed between DNA and TPP1 alone. Several biochemical and biophysical studies have been undertaken in order to understand the mechanism by which hPOT1 unfolds telomeric G4 structures[24,10,20,18,7]. These studies have focused on a single quadruplex formed by four repetitions of the human telomeric motif. However, long human single-stranded telomeric DNA sequences can form contiguous G4[30,29,19,6] that are likely to occur at the G-overhang made of up 50 tandem repeats. A few studies focused on the interactions of hPOT1 with long telomeric sequences and have shown that multiple hPOT1 fully coat the DNA[25,22,9,28]. In particular, it has been reported that in the presence of sodium cations, multiple hPOT1 or hPOT1-TPP1 binding presents little or no cooperativity[22,9]. Here, we investigated the interactions of hPOT1-TPP1 with long telomeric sequences containing contiguous G4 in the presence of the biological relevant potassium cations stabilizes contiguous G4 to a larger extent than sodium. Our data reveal that multiple hPOT1-TPP1 unfold telomeric sequences structured in contiguous G4s, with a cooperative 3' to 5' sequential binding mode.

2 Materials & Methods

2.1 Materials

Recombinant hPOT1-TPP1 complex were co-expressed in the baculovirus insect cell system (pFastBac HT-A vector) at the IGBMC-CERBM platform (Strasbourg). The full hPOT1 (residues 1-634) was expressed with 6 histidines in N-term and a truncated version of human TPP1 (87-334), with a GST-tag in N-term. Each protein contains a restriction site 3C (LGVLFQ/G) that allows removing tags. The cell pellet was resuspended in lysis buffer (50 mM Tris HCl pH 7.8, 300 mM NaCl, 2 mM DTT, 1 mM PMSF, 2 mM 2-Mercaptoethanol) supplemented with a cocktail of protease inhibitors (Roche), stirred on ice for 1h30. The cells were then lysed by sonication (60% Ampli, 1sec ON – 2sec OFF for 20 min) and the cell debris was removed by ultracentrifugation (17 000 rpm for 1h15) at 4°C. The supernatant was loaded on a 5 mL Ni-His Select column pre-equilibrated in lysis buffer. The column was next washed (flow

rate 0.15 mL/min) with 25 mL Ni-Wash buffer (50 mM Tris HCl pH 7.8, 300 mM NaCl, 2m DTT, 10 mM Imidazole). The hPOT1-TPP1 complex was eluted with 25 mL of Ni-Elution buffer (50 mM Tris HCl pH 7.8, 300 mM NaCl, 2m DTT, 300 mM Imidazole). The eluted fractions were loaded on a 5 mL GST column, pre-equilibrated with Ni-Elution Buffer. After a washing step with GST-Wash Buffer (50 mM Tris HCl pH 7.8, 75 mM NaCl, 1mM DTT), the hPOT1-TPP1 complex was eluted with GST-Elution Buffer (50 mM Tris HCl pH 7.8, 300 mM NaCl, 2m DTT, and 15 mM GSH). Tags were removed overnight with the PreScission (1.500 units) protease at 4°C in Low Salt Buffer (50 mM Tris HCl pH 7.8, 150 mM NaCl, 1mM DTT). Finally, the digested sample was loaded on a 1 mL MonoQ column, pre-equilibrated with Low Salt Buffer. The column was next washed with 5 mL of Low Salt Buffer and hPOT1-hTPP1 complex was eluted with a 25 mL gradient of NaCl (from 150 mM to 300 mM). Fractions containing proteins were pooled and stored at -80°C. **Recombinant hRPA** was expressed in the Escherichia coli BL21 (DE3) strain transformed with the plasmid pET11ahRPA that permits the co-expression of RPA1, RPA2 and RPA3. hRPA was purified as previously described[2]. Oligonucleotides were purchased from Eurogentec.

2.2 Electrophoretic mobility shift assays

Oligonucleotides were labelled with γ (^{32}P ATP using T4 polynucleotide kinase (NEB). Radiolabelled and doubly-labelled oligonucleotides were heated during 10 min at 85°C and slowly cooled at room temperature allowing the folding of oligonucleotide into G4 structures. For all EMSA experiments, proteins were diluted and pre-incubated (20 min at 4 °C) in a dilution buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl or LiCl, 1 mM DTT, 10% glycerol, 0.2 mg/mL BSA and 0.1 mM EDTA). In a standard reaction, oligonucleotide (2 nM for radiolabelled oligonucleotide and 100 nM for doubly-labelled oligonucleotide) was incubated with various amounts of protein in 10 μL reaction buffer (50 mM HEPES pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2% glycerol) for 20 min at 20 °C. Samples were then loaded on a native 5% Acrylamide/Bisacrylamide (29:1) in 0.5X TBE buffer at 4°C. After electrophoresis, in the case of radiolabelled oligonucleotide, the gel was dried and exposed on a phosphorimager screen, and scanned with a Typhoon instrument (Molecular Dynamics). Quantification was made with ImageQuant version 5.1. For each protein concentration, the fraction of radiolabelled oligonucleotide bound to protein was calculated as follow: $I_{DNA\,bound\,to\,protein}/I_{DNA\,bound\,to\,protein} + I_{free\,DNA}$, where I is the intensity of the band. In the case of doubly-labelled oligonucleotide, the gel was directly scanned with Typhoon instrument to detect FAM fluorescence.

2.3 Fluorescence Resonance Energy Transfert

2.3.1 Fluorescence titrations:

Fluorescence titrations of 100 mM H69-Q1,2,3 oligonucleotides (doubly-labelled with a FAM and a TAMRA) were carried out on a SPEX Fluorolog spectrofluorimeter

(Jobin Yvon) as previously described (Lancrey 2018). The Fluorescence spectra were recorded at 20 °C in a buffer containing 100 mM KCl, and 5 mM lithium cacodylate (pH 7.2) (excitation wavelength 470 nm). The protein (from 50 to 600 nM) was directly added to the solutions containing the oligonucleotides and incubated for 5 min before spectra recording. The percentage of opened labelled G4 units at 20 °C was calculated as previously described[11].

2.3.2 Fluorescence kinetics:

Each oligonucleotide H69-Q1,2,3 (100 nM) was incubated for 10 min at 20 °C, in a buffer containing 100 mM KCl, and 5 mM lithium cacodylate (pH 7.2), then mixed with variable molar excess of protein for 40 min and finally heated until 80°C. Fluorescence intensities at 518 nm (FAM) and 586 nm (TAMRA) were recorded at regular time intervals (10 secondes).

3 Results

The oligonucleotides we chose for our study are a 45mer (H45), a 69mer (H69) and a 93mer (H93) mimicking human telomeric sequences (Figure 1B). In the presence of potassium, these oligonucleotides fold into two, three and four contiguous G4 units, respectively,[6]. Each telomeric sequence contains multiple hPOT1-binding sites (TTAGGGTTAG underlined by black dashed lines), the number of which increases with the length of the DNA: 3 for H45, 5 for H69 and 7 for H93.

3.1 The structuration of telomeric sequences into contiguous G4 affects the binding curve of multiple hPOT1-TPP1

By electrophoretic mobility shift assays, we studied the binding efficiency of hPOT1-TPP1 to H45, H69 and H93 in potassium, where these oligonucleotides fold into 2, 3 and 4 contiguous G4, respectively (we define as "binding efficiency" the percentage of oligonucleotide bound to the protein as a function of protein concentration). In a standard experiment, 2 nM of radiolabelled DNA was incubated for 20 min at 20°C with different amounts of proteins. Figure 2A.1 illustrates the results of EMSA experiments.

The presence of retarded bands for H45, H69 and H93 indicated that hPOT-TPP1 binds telomeric DNA even in the presence of the physiological relevant cation K^+, where these sequences are structured into stable contiguous G4 (previous studies had been carried out in sodium[22]. The number of retarded bands increased with the length of the oligonucleotide, revealing hPOT1-TPP1:DNA complexes of increasing stoichiometry, where one, two, three, four, five (or more) hPOT1-TPP1 are bound (Figure 2A.1). For H45, we observed three retarded bands, indicating the formation of 1:1, 2:1 and 3:1 hPOT1-TPP2 complexes. For H69, we observed 5 bands, indicating the formation of up to 5:1 hPOT1-TPP1:DNA complexes. The number of retarded bands for H45 and H69, corresponds to the number of hPOT1 binding sites (Figure 1B). For H93, hPOT1-TPP1 complexes with a stoichiometry higher than 5:1 stayed in the well due to their high molecular weight. For the three oligonucleotides, the presence of complexes

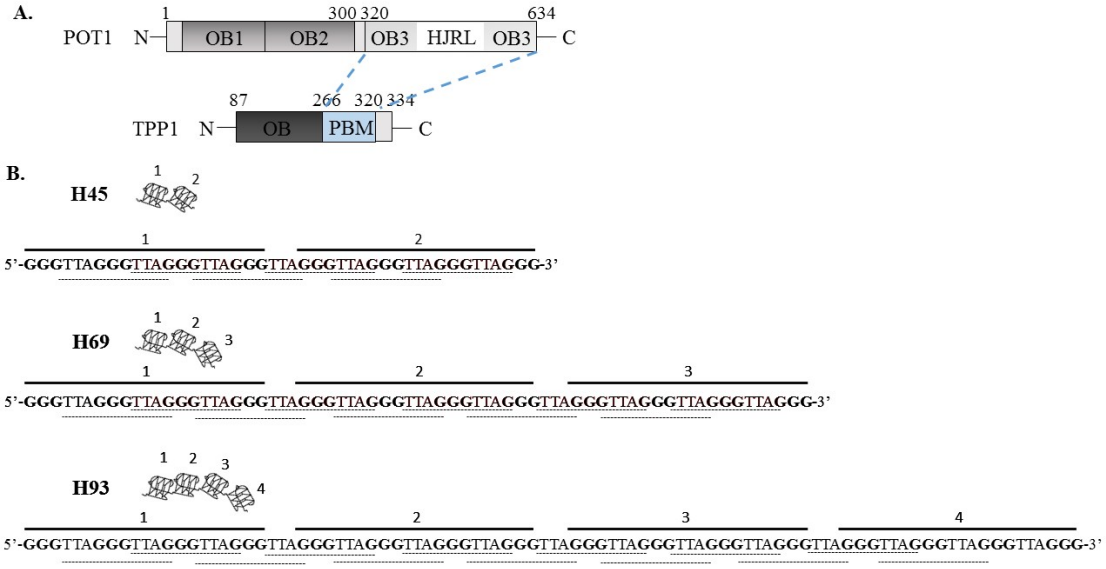


Figure 1: **Structure of the human POT1 and TPP1 proteins and telomeric repeat sequences investigated for our study.** (**A.**) Domain organization of the hPOT1-TPP1 complex. OB1 and OB2 of POT1 are colored in dark grey while OB3 and the embedded HJRL domain are colored in grey and white, respectively. TPP1 contains a OB domain (in black grey) and the PBM domain (in blue) for POT1 interaction (indicated by dashed blue lines); the TBM domain for TIN2 interaction is not present. (**B.**) H45, H69 and H93 are human telomeric sequences ; they contain respectively 3, 5 and 7 POT1 binding motifs (black dashed lines under each sequence indicate potential POT1 binding sites) and folds into 2, 3 and 4 contiguous G4 (black lines above each sequence indicate sequences implicated in G4 structure).

of intermediate stoichiometry (where not all the binding sites are occupied) indicated that hPOT1-TPP1 does not bind DNA in an "all-or-none" mode. Nevertheless, quantification of the percentage of DNA bound to hPOT1-TPP1 as a function of concentration resulted in sigmoidal curves that suggested a cooperative binding (Figure 2A.2).

We wondered if the sigmoidal shape of the binding curves could be due to the presence of contiguous G4. Thus, we performed EMSA experiments in the presence of 100mM LiCl. In these conditions, quadruplexe structures are not stabilized[?] . Here again, we observe protein:DNA complexes of increasing stoichiometry depending on the length of the DNA (Figure 2B.1). However, the shape of the binding curves of hPOT1-TPP1 toH45, H69 ang H93 oligonucleotides in lithium displayed a far less pronounced sigmoidal shape (Figure 2B.2) (see supplementary Figure S1 for comparison). It is worth noting that the hPOT1-TPP1 binding on a non-structured short telomeric sequence is not affected by the nature of cations[18]. Our results suggest that the sigmoidal shape of the binding curves of hPOT1-TPP1 to H45, H69 and H93 in potassium is not intrinsic to the mode of binding of multiple hPOT1-TPP1 to DNA, but comes from the structuration of the DNA substrate into G4: the binding of a first hPOT1-TPP1 unfolds a first G4 facilitating the binding of a second hPOT1-TPP1, and so on.

3.2 hPOT1-TPP1 first opens the G4 at the 3' end and proceeds towards the 5' end.

It has been shown that POT1 is able to open a single G4[24,10,20,18,7]. We wondered if the hPOT1-TPP1 binding was correlated with the opening of all contiguous G4s. To this purpose, classical fluorescence resonance energy transfer FRET experiments were carried out on three modified H69 oligonucleotides, doubly-labelled with a 6-carboxyfluorescein (FAM) and a tetramethylrhodamine (TAMRA) at different positions: in H69-Q1 the fluorophores are placed at both sides of the first G4 unit (the one at the 5'end, called 5'-G4 unit); in H69-Q2 they are placed at both sides of the second G4 unit (called inner-G4 unit); in H69-Q3 they are placed at both sides of the third G4 unit (the one at the 3' end, called 3'-G4 unit) (Figure 3)[6,11]. Temperatures of thermal transitions of these double-labelled oligonucleotides were previously determined by fluorescence[6] and showed that the two external G4 have identical stability (Tt=60°C), while the inner G4 has a lower stability (Tt = 51°C).

FRET data show for all H69-Q1,2,3 an increase of FAM fluorescence and a decrease in TAMRA fluorescence upon hPOT1-TPP1 complex addition, indicating that hPOT1-TPP1 complex opens each of the three G4 units (Figure 3A). However, quantification of the different G4 units opening shows that the opening efficiency is lower for the 5'-G4 unit than for the inner-G4 and the 3'-G4 units (Figure 3B). It was previously shown that stable secondary structures hinder the first binding event of hPOT1-TPP1[18]. Here we show that, in a multi-G4 context, the opening efficiency of the different G4 units in H69 by hPOT1-TPP1 does not correlate with their stability; notably, hPOT1-TPP1 does not preferentially open the less stable G4 (the inner-G4) but the one located at the 3' end, despite its higher stability.

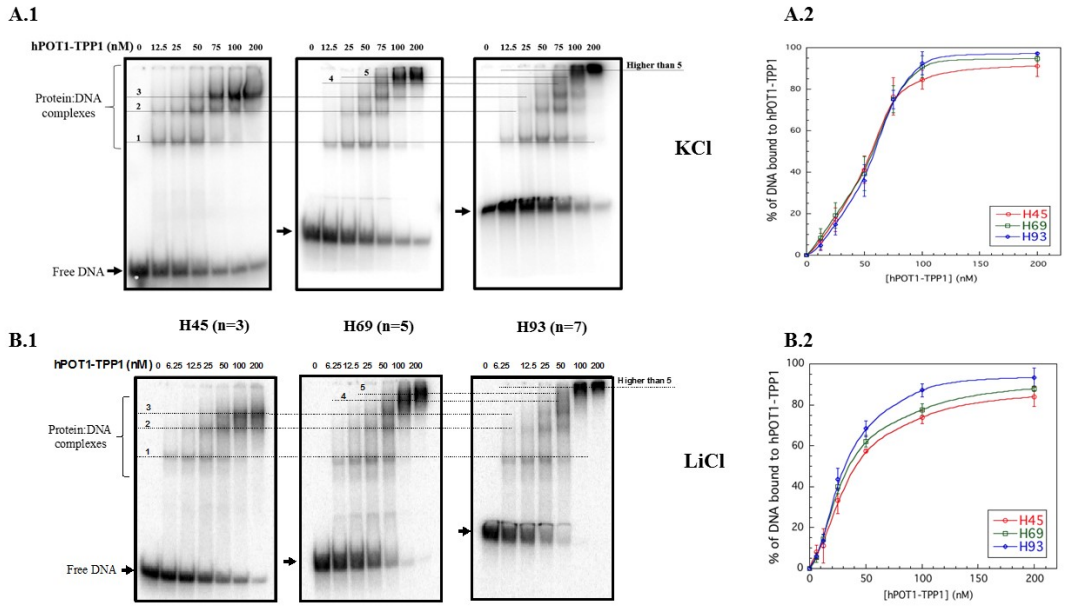


Figure 2: **Electrophoretic mobility shift assays.** P^{32}-oligonucleotides (2 nM) were incubated with increasing amounts of hPOT1-TPP1 in the presence of telomeric DNA sequences forming contiguous G4 in 100mM KCl (A) or in LiCl (B). Electrophoresis was run on a native 5% acrylamide:bisacrylamide 29:1 gel. (A1, B1) EMSA gels with hPOT1-TPP1. Arrows indicate free DNA ; dashed lines indicate protein:DNA complexes with different stoechiometry ; (A.2,B.2) Percentage of DNA bound to hPOT1-TPP1 : H45 (red circles), H69 (blue squares), and H93 (green diamonds) as a function of protein concentration. Error bars correspond to standard deviations calculated from three independent experiments.

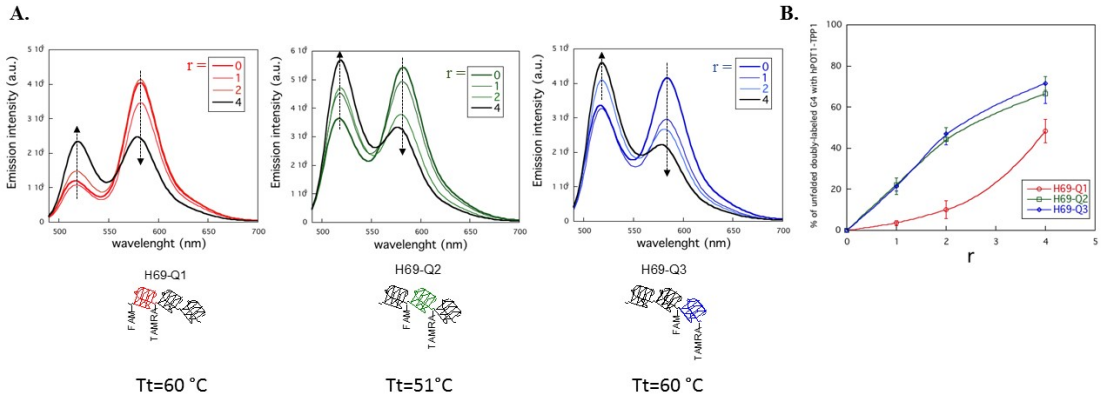


Figure 3: **Fluorescence titrations.** H69 was doubly-labelled at different position to follow the opening of individual G4 at 5' side (H69-Q1), at internal position (H69-Q2) and at the 3' side (H69-Q3). (A) Fluorescence emission spectra of H69-Q1, H69-Q2 and H69-Q3 (100 nM) with increasing amounts of or hPOT1-TPP1 in 100mM KCl solution. r is the protein/oligonucleotide concentration ratio. Spectra of doubly-labelled oligonucleotides in the absence of protein and 4 equivalents of hPOT1-TPP1, are in bold line. Arrows indicate the evolution of emission spectra with increasing protein concentrations. a.u., arbitrary units. Tt represent the thermal transition temperature for each doubly-labelled G4. (B) Percentage of unfolded labelled G4 units in H69-Q1 (red circles), H69-Q2 (blue squares) and H69-Q3 (green diamonds) as a function of r. The percentage of unfolded labelled G4 units was calculated as previously described [11]. Error bars correspond to standard deviations calculated from two independent experiments.

As a control, we verified that this difference was not due to a difference in protein binding efficiency to the three doubly-labelled H69 oligonucleotides by performing EMSA in the same conditions of FRET assays (Figure 4). EMSA experiments showed similar binding efficiency of hPOT1-TPP1 for the three doubly-labelled H69 (see quantification in Figure 4A) and similar migration pattern (Figure 4B). Altogether, these results show that hPOT1-TPP1 complex is able to unfold all the contiguous G4, but with a better efficiency for the 3'-G4 unit than the 5'-G4 unit, suggesting that hPOT1-TPP1 binds telomeric sequences structured into contiguous G4 preferentially from the 3' end.

To gain insight into the opening process of contiguous G4 by hPOT1-TPP1, we studied the unfolding kinetics of each doubly-labelled H69 in the presence of hPOT1-TPP1. H69 bears 5 hPOT1 binding sites (Figure 1B); quantification of EMSA migration pattern (Figure 4B) show that, whatever the nature of the doubly-labelled H69, 1 equivalent of hPOT1-TPP1 allows the formation of few hPOT1-TPP1:DNA complexes (13-15%) containing only 1 hPOT1-TPP1, while 4 equivalents of hPOT1-TPP1 allow the formation of hPOT1-TPP1:DNA complexes (90%) containing mainly 1 or 2 hPOT1-TPP1, and 10 equivalents allow the formation of hPOT1-TPP1:DNA complexes (100%) containing mainly 3 or 4 hPOT1-TPP1 (and a minor fraction of hPOT11-TPP1:DNA 5:1 complexes). These data indicate that, at 1 and 4 equivalents, hPOT1-TPP1 complexes cannot occupy all the binding site, while at 10 equivalents, most of the DNA binding sites are occupied. Thus, we performed fluorescence kinetics studies with 1, 4 and 10 equivalents of hPOT1-TPP1 (Supplementary Figure S2), and calculate the percentage of unfolded G4 as a function of time (Figure 5).

At 1 equivalent of hPOT1-TPP1 (where 1:1 complexes were the major fraction of bound DNA, fig. 4B), the 3'-G4 unit is preferentially unfolded (more than 40% of unfolded G4 for H69-Q3), while both the inner-G4 (less than 30% unfolded G4 for H69-Q2) and the 5'-G4 units (less than 25% of unfolded G4 for H69-Q1) were poorly unfolded, indicating that the 3'-G4 unit is the preferential binding site of hPOT1-TPP1. This preferential binding for the 3' end is not dependent on the presence of G4 as it is observed with a non-structured short telomeric sequence too (supplementary figure S3). At 4 equivalents (where 1:1 and 2:1 POT1-TPP1:DNA complexes were the major fraction of bound DNA, fig. 4B), a higher percentage of unfolded G4 is obtained at the steady-state, consistent with the formation of more complexes (Figure 4A), however similar percentages of unfolding were attained for the 3'-G4 and the inner-G4 unit (with around 60 % of unfolded G4 for H69-Q3 and Q2) while the 5'-G4 unit was less unfolded (less than 35 % of unfolded G4 for H69-Q1). At 10 equivalents (where 3:1 and 4:1 POT1-TPP1:DNA complexes were the major fractions of bound DNA, fig. 4B)), similar percentages of unfolded G4 were observed for the three H69-Q3, Q2 and Q1 (with around 80 +/- 5 % of unfolded G4). This result indicated that with a large excess of protein, the three G4 units are opened, including the 5'-G4 unit.

To verify that kinetics observed here are POT1-TPP1 specific and not an artifact from doubly-labelled oligonucleotides, we

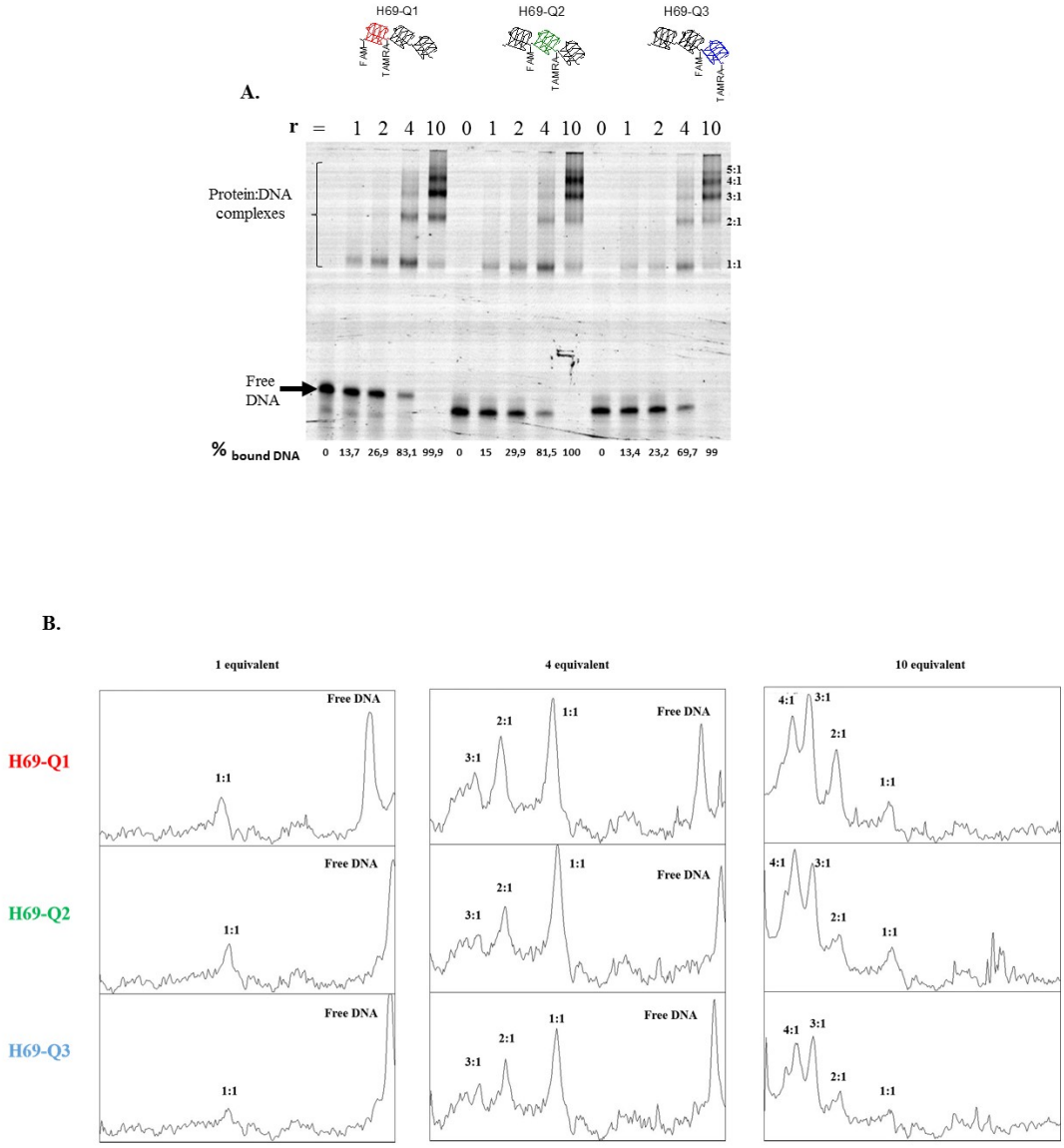


Figure 4: **Electrophoretic mobility shift assays in FRET conditions.** (A) Each doubly-labelled H69 (100 nM) were incubated with indicating amounts of hPOT1-TPP1 (r is the protein/oligonucleotide concentration ratio). Electrophoresis was run on a native 5% acrylamide:bisacrylamide 29:1 gel. Arrows indicate protein:DNA complexes with different stoechiometry ; % bound DNA corresponds to the percentage of DNA bound to hPOT1-TPP1. (B) Profils of EMSA pattern is presented for each line.; 1:1, 2:1, 3:1, 4:1 correspond to the complexes of different stoechiometry with 1, 2, 3 and 4 proteins per DNA, respectively.

carried out kinetics experiments with a different protein. We recently showed that RPA, the eukaryotic single-stranded DNA binding protein, is able to coat long telomeric DNA structured into contiguous G4[11] and to unfold all the G4. Here, we carried out opening kinetics with 6 equivalents of RPA (Figure 5D). Results show that the inner-G4 unit (which is the less stable) is better unfolded than the two external G4 units (which are more stable). This study shows that in a multi-G4 context, unlike hPOT1-TPP1, RPA binds and unfolds preferentially the less stable G4, corresponding here to the internal G4.

Thus, considering stoechiometry and kinetics data, we can conclude that in a multi-G4 context, hPOT1-TPP1 binding initiates from the 3' end and proceeds towards the 5' end, in contrast to RPA that binds and unfolds the inner and less stable G4 unit.

4 Discussion

In this work, we investigated how human POT1 with its partner TPP1, deals with telomeric contiguous G4 that may form at the G-overhang. To this purpose, we studied the interactions of hPOT1-TPP1 complex with oligonucleotides that fold into contiguous G4 stabilized by the presence of the biological relevant potassium cation.

We first showed that hPOT1-TPP1, binds long telomeric sequences structured into stable contiguous G4, and form multi-hPOT1-TPP1:DNA complexes where several hPOT1-TPP1 cover the DNA. Interestingly, we observed binding curves with sigmoidal shapes in the presence of potassium cations that are abrogated in the presence of lithium cations, suggesting a cooperative binding related to the presence of G4s structures (Figure 2 and S1). Although multi-hPOT1-TPP1:DNA have been already reported with long telomeric sequences structured into G4[22,25,28], no or little cooperativity has been observed[22] probably due to the fact that experiments were done in sodium, which stabilizes G4 to a lesser extend compared to potassium. Thus, the cooperative behavior observed in potassium is an apparent cooperativity due to the presence of contiguous stable G4 : the binding of a first hPOT1-TPP1 unfolds a first G4 facilitating the binding of a second hPOT1-TPP1, and so on. A similar G4-induced cooperativity has been observed on a monomeric G4 in potassium where two POT1 are necessary to open the G4[20]. In this case, the binding of one POT1, by unfolding the G4, could favor the binding of a second protein.

Then we investigated the unfolding mechanism of contiguous G4 with oligonucleotides containing three double-labelled G4. FRET kinetics show that hPOT1-TPP1 unfolds contiguous G4 differently according to their position : first the G4 at the 3' end, followed by the inner G4 and last the G4 at the 5' end (Figure 5). In addition, EMSA performed in FRET conditions with a telomeric DNA containing 5 binding sites shows that the 5'-G4 unit is opened only if three or four proteins are bound to the DNA. These data indicate a sequential G4-binding/opening mode from the 3' and to the 5' end.

We show that such directional and sequential opening is in accordance with (i) the preferential binding of hPOT1-TPP1 for the 3' end (supplementary Figure S3), as already previously observed for the human POT1-TPP1[24] but also for the fission yeast

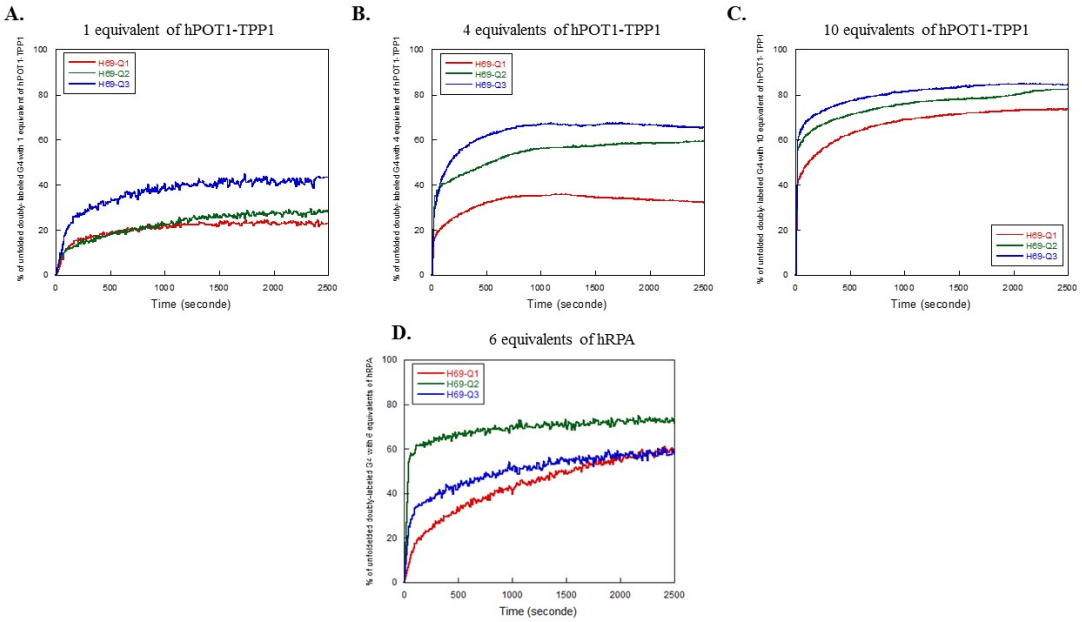


Figure 5: **Unfolding Kinetics.** Percentage of unfolded G4 upon time, for H69-Q1 (red), H69-Q2 (blue) and H69-Q3 (green) with 1 (A), 4 (B) and 10 equivalents (C) of hPOT1-TPP1, and 6 equivalents of hRPA (D), in 100 mM KCl. The oligonucleotide concentration was 100 nM. The percentage of unfolded G4 was calculated from fluorescence kinetics (supplementary Figure S2).

POT1[12], and with (ii) the 3' to 5' POT1 binding direction described by Hwang et al.[10]. Hwang et al. showed that 2 POT1 are necessary to open one G4 while only one POT1-TPP1 is needed[10].

Considering the whole set of data presented here, we propose a model (Figure 6) where the first hPOT1-TPP1 binds the G-overhang from the 3' end (1) and unfolds the G4 at the 3' end. When the first G4 is unfolded and bound by hPOT1-TPP1 (2), the second contiguous G4 is destabilized (3), favoring the binding of a second hPOT1-TPP1 (4), and so on towards the 5' end (5). Recently, we have observed a similar "apparent" cooperativity (arising from the structuration of the DNA substrate) with another human single-stranded binding protein, the replication protein A (hRPA)[11], which is known to have very little, if any, cooperativity, for unstructured ssDNA[27]. However, we show here that the binding modes of POT1-TPP1 and RPA are different, as hRPA first binds the less stable G4 (here the internal G4). This binding mode is in accordance with previous data showing that the affinity of hRPA for telomeric G4 is inversely correlated with the stability of the G4 structure[21]. Thus, the binding of RPA in a multi-G4 context is dictated by the stability of G4.

On one side, the ability of RPA to bind and unfold internal G4 in long telomeric sequence could be related to the role of RPA at telomeres where it participates to the G-strand during the replication phase[23,17,16,2]. On the other side, the preferential binding of POT1-TPP1 for the 3' end and its ability to unfold contiguous G4 with a 3' to 5' cooperative sequential binding mode, could contribute to telomere maintenance in suppressing G4 and in helping the access of telomerase[31,14,8,1,13]. The fact that telomeric contiguous G4s dictate differentially the binding mode of POT1-TPP1 and RPA, strongly argues that G4 at telomeres could be a steric factor that regulate the binding of ssDNA binding proteins and their interplay at telomeres[20].

Multiple POT1-TPP1 binding to long telomeric G-strand have already been reported, as well as the preferential binding to the site at the 3' end. Our work demonstrates that POT1-TPP1 unfold contiguous telomeric G4 in a 3' toward 5' directional mode and unravels the molecular origin of this directionality. We show that it arises from POT1-TPP1 binding preference for the 3' end of the G-strand and from its cooperative binding to the G-strand in potassium. Our data also show that this cooperative binding is not intrinsic to the mode of binding of POT1-TPP1 to DNA, but arises from the structuration of the DNA substrate into contiguous G4.

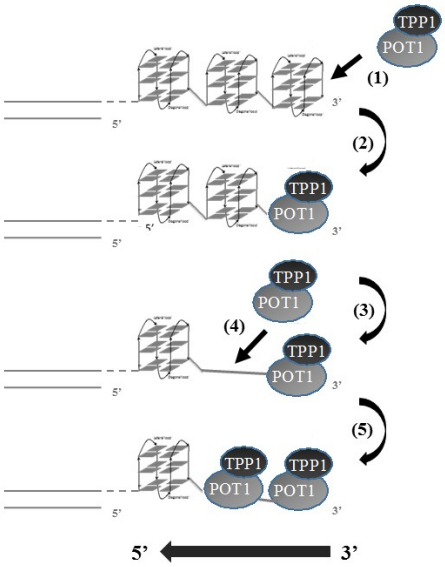


Figure 6: **Model of a 3' to 5' sequential binding mode of hPOT1-TPP1 on contiguous G4.** This first binding step of hPOT1-TPP1 occurs at the 3' end (1) and triggers unfolding of the G4 at 3' end (2). When the first G4 DNA is unfolded and covered by hPOT1-TPP1, it implies a destabilization of the contiguous G4 (3), favoring the binding of a second hPOT1-TPP1 (4), and so on towards the 5' end (5).

References

[1] Armbruster, B. N., Linardic, C. M., Veldman, T., Bansal, N. P., Downie, D. L., and Counter, C. M. (2004). Rescue of an hTERT mutant defective in telomere elongation by fusion with hPot1. *Molecular and Cellular Biology*, 24(8):3552–3561.

[2] Audry, J., Maestroni, L., Delagoutte, E., Gauthier, T., Nakamura, T. M., Gachet, Y., Saintomé, C., Géli, V., and Coulon, S. (2015). RPA prevents g-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *The EMBO Journal*, 34(14):1942–1958.

[3] Baumann, P. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*, 292(5519):1171–1175.

[4] Baumann, P., Podell, E., and Cech, T. R. (2002). Human pot1 (protection of telomeres) protein: Cytolocalization, gene structure, and alternative splicing. *Molecular and Cellular Biology*, 22(22):8079–8087.

[5] Blackburn, E. H. (2004). Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS Letters*, 579(4):859–862.

[6] Bugaut, A. and Alberti, P. (2015). Understanding the stability of DNA g-quadruplex units in long human telomeric strands. *Biochimie*, 113:125–133.

[7] Chaires, J. B., Gray, R. D., Dean, W. L., Monsen, R., DeLeeuw, L. W., Stribinskis, V., and Trent, J. O. (2020). Human POT1 unfolds g-quadruplexes by conformational selection. *Nucleic Acids Research*, 48(9):4976–4991.

[8] Colgin, L. M., Baran, K., Baumann, P., Cech, T. R., and Reddel, R. R. (2003). Human POT1 facilitates telomere elongation by telomerase. *Current Biology*, 13(11):942–946.

[9] Corriveau, M., Mullins, M. R., Baus, D., Harris, M. E., and Taylor, D. J. (2013). Coordinated interactions of multiple POT1-TPP1 proteins with telomere DNA. *Journal of Biological Chemistry*, 288(23):16361–16370.

[10] Hwang, H., Buncher, N., Opresko, P. L., and Myong, S. (2012). POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure*, 20(11):1872–1880.

[11] Lancrey, A., Safa, L., Chatain, J., Delagoutte, E., Riou, J.-F., Alberti, P., and Saintomé, C. (2018). The binding efficiency of RPA to telomeric g-strands folded into contiguous g-quadruplexes is independent of the number of g4 units. *Biochimie*, 146:68–72.

[12] Lei, M., Baumann, P., and Cech, T. R. (2002). Cooperative binding of single-stranded telomeric DNA by the pot1 protein ofSchizosaccharomyces pombe†. *Biochemistry*, 41(49):14560–14568.

[13] Lei, M., Podell, E. R., and Cech, T. R. (2004). Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature Structural & Molecular Biology*, 11(12):1223–1229.

[14] Loayza, D. and de Lange, T. (2003). POT1 as a terminal transducer of TRF1 telomere length control. *Nature*, 423(6943):1013–1018.

[15] Loayza, D. and de Lange, T. (2004). Telomerase regulation at the telomere. *Cell*, 117(3):279–280.

[16] Luciano, P., Coulon, S., Faure, V., Corda, Y., Bos, J., Brill, S. J., Gilson, E., Simon, M.-N., and Géli, V. (2012). RPA facilitates telomerase activity at chromosome ends in budding and fission yeasts. *The EMBO Journal*, 31(8):2034–2046.

[17] Moser, B. A. and Nakamura, T. M. (2009). Protection and replication of telomeres in fission yeastThis paper is one of a selection of papers published in this special issue, entitled 30th annual international asilomar chromatin and chromosomes conference, and has undergone the journal's usual peer review process. *Biochemistry and Cell Biology*, 87(5):747–758.

[18] Mullins, M. R., Rajavel, M., Hernandez-Sanchez, W., de la Fuente, M., Biendarra, S. M., Harris, M. E., and Taylor, D. J. (2016). POT1–TPP1 binding and unfolding of telomere DNA discriminates against structural polymorphism. *Journal of Molecular Biology*, 428(13):2695–2708.

[19] Petraccone, L., Spink, C., Trent, J. O., Garbett, N. C., Mekmaysy, C. S., Giancola, C., and Chaires, J. B. (2011). Structure and stability of higher-order human telomeric quadruplexes. *Journal of the American Chemical Society*, 133(51):20951–20961.

[20] Ray, S., Bandaria, J. N., Qureshi, M. H., Yildiz, A., and Balci, H. (2014). G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proceedings of the National Academy of Sciences*, 111(8):2990–2995.

[21] Safa, L., Delagoutte, E., Petruseva, I., Alberti, P., Lavrik, O., Riou, J.-F., and Saintomé, C. (2014). Binding polarity of RPA to telomeric sequences and influence of g-quadruplex stability. *Biochimie*, 103:80–88.

[22] Taylor, D. J., Podell, E. R., Taatjes, D. J., and Cech, T. R. (2011). Multiple POT1–TPP1 proteins coat and compact long telomeric single-stranded DNA. *Journal of Molecular Biology*, 410(1):10–17.

[23] Verdun, R. E. and Karlseder, J. (2006). The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres. *Cell*, 127(4):709–720.

[24] Wang, F., Podell, E. R., Zaug, A. J., Yang, Y., Baciu, P., Cech, T. R., and Lei, M. (2007). The POT1–TPP1 telomere complex is a telomerase processivity factor. *Nature*, 445(7127):506–510.

[25] Wang, H., Nora, G. J., Ghodke, H., and Opresko, P. L. (2010). Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting g-quadruplex unfolding. *Journal of Biological Chemistry*, 286(9):7479–7489.

[26] Williamson, J. R., Raghuraman, M., and Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: The g-quartet model. *Cell*, 59(5):871–880.

[27] Wold, M. S. and Kelly, T. (1988). Purification and characterization of replication protein a, a cellular protein required for in vitro replication of simian virus 40 DNA. *Proceedings of the National Academy of Sciences*, 85(8):2523–2527.

[28] Xu, M., Kiselar, J., Whited, T. L., Hernandez-Sanchez, W., and Taylor, D. J. (2019). POT1-TPP1 differentially regulates telomerase via POT1 his266 and as a function of single-stranded telomere DNA length. *Proceedings of the National Academy of Sciences*, 116(47):23527–23533.

[29] Yan Xu, Y. K., Loonnberg, T., and Komiyama, M. (2009). Human telomeric DNA sequence-specific cleaving by g-quadruplex formation. *Journal of the American Chemical Society*, 131(8):2871–2874.

[30] Yu, H.-Q., Miyoshi, D., and Sugimoto, N. (2006). Characterization of structure and stability of long telomeric DNA g-quadruplexes. *Journal of the American Chemical Society*, 128(48):15461–15468.

[31] Zaug, A. J., Podell, E. R., and Cech, T. R. (2005). Human POT1 disrupts

telomeric g-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences*, 102(31):10864–10869.

5 Supplementary figures

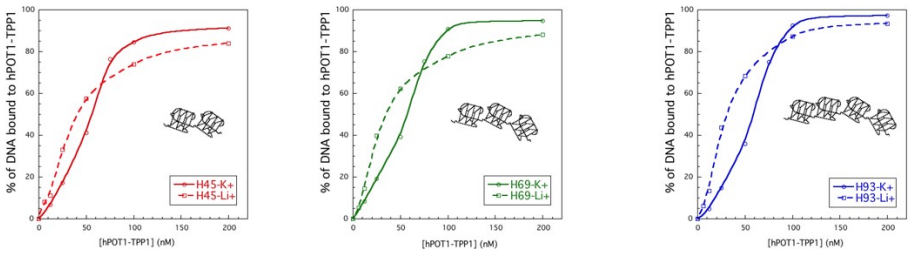


Figure S1: **Comparison of the hPOT1-TPP1 binding on telomeric sequences in potassium and lithium.** Overlays of binding curves of hOT1-TPP1 in potassium and lithium show for each sequence a significant shape difference indicating a different binding mode upon the nature of cation, and consequently to the presence of G-quadruplexes.

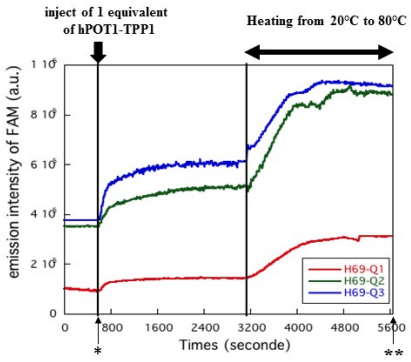


Figure S2: **Fluorescence kinetics.** Fluorescence intensities at 518 nm (FAM) (A) and 586 nm (TAMRA) (B) were recorded at regular time intervals (10s) for each oligonucleotide H69-Q1,2,3 (100 M) in a buffer containing 100 mM KCl, and 5 mM lithium cacodylate (pH 7.2). Fluorescence intensity was measured during three steps: (1) in the absence of protein at 20 °C for 10 min, where the G4 is entirely folded; (2) after the adding (black arrow) of one equivalent of hPOT1-TPP1, during this step the G4 is being unfolded by protein; (3) during the heating of the mixture (double-arrow) until 80°C, the G4 is being entirely unfolded. The percentage was calculated as follow: % of unfolded G4 = (I-FAM(t)– I-FAM(f)) x 100/ (I-FAM(u) – I-FAM(f)); where I-FAM(f) is emission intensity of FAM at 600s (+) before adding of protein ; I-FAM(t) is fluorescence intensity measured during 40 minutes after adding of one equivalent of hPOT1-TPP1 ; I-FAM(u) is emission intensity of FAM at 5640s after heating (++).

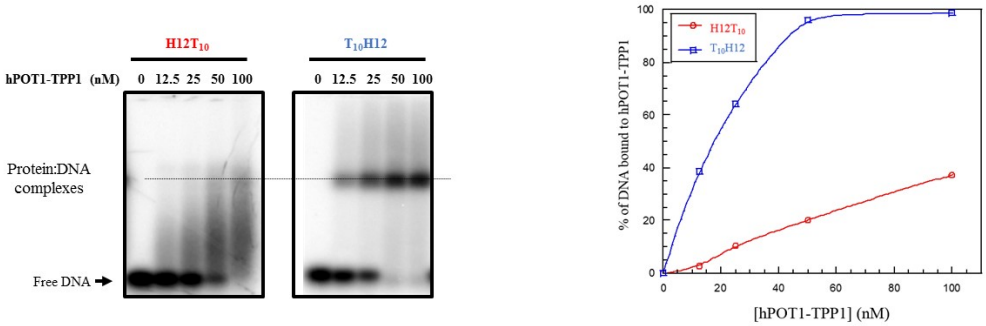


Figure S3: **Electrophoretic mobility shift assays with oligonucleotides containing POT1 binding site at the 3' end ($T_{10}H12$) and the 5' end ($H12T_{10}$).** (A) P^{32}-oligonucleotides (2 nM) were incubated with increasing amounts of hPOT1-TPP1 in 100mM KCl. Electrophoresis was run on a native 5% acrylamide:bisacrylamide 29:1 gel. H12: GGTTAGGGTTAG and T10=TTTTTTTTTT. Arrows indicate free DNA ; dashed lines indicate protein:DNA complexes. Presence of smear with H12T10 indicates that protein:DNA complexes are not stable. (B) Percentage of DNA bound to hPOT1-TPP1 : T10H12 (red circles) and H12T10 (blue circles) as a function of protein concentration.

2.4 Résultats complémentaires, discussions et perspectives

2.4.1 Étude de*POT1* : résultats et discussion

A la suite de l'étude réalisée sur *POT1-TPP1*, j'ai étudié les mécanismes de liaison de *POT1* seule sur ces longues séquences télomériques capables de se structurer en G4 contigus. Pour étudier ces mécanismes j'ai réalisé des expériences d'*EMSA* en présence des séquences télomériques décrites précédemment à savoir H45, H69 et H93 qui forment en potassium (K^+) respectivement deux, trois et quatre G4 contigus (Figure 2.6.A). Le résultat montre que *POT1* seule est capable de prendre en charge des multimères de G4 en tandem. Á partir des quantifications des complexes protéines :ADN, j'ai obtenu des courbes de liaisons (Figure 2.6.B) qui présentent aussi une allure sigmoïdale qui pourrait signifier une coopérativité apparente similaire à celle observée avec le complexe *POT1-TPP1*. Pour m'en assurer, j'ai réalisé des expériences *EMSA* en présence de lithium (Figure 2.6.C) Les courbes de quantification des complexes *POT1*-ADN ne présentent plus d'aspect sigmoïdal (Figure 2.6.D).

Cependant, dans le cas de *POT1* et contrairement à ce que nous avons observé avec *POT1-TPP1*, le nombre de bandes retardées ne correspond pas au nombre de sites de liaison. En effet regardons la liaison de *POT1* à la séquence H21. H21 possède deux sites de liaison possibles (TTAGGGTTAG), mais qui sont mutuellement exclusifs (si l'un est occupé, l'autre ne peut pas être occupé) :

* 5'-GGGTTAGGGTTAGGGTTAGGG-3'
* 5'-GGGTTAGGGTTAGGGTTAGGG-3'

Pourtant l'étude en *EMSA* montre la présence de complexe de stœchiométrie 1 :1, 2 :1 et même 3 :1 à haute concentration en protéine (Figure 2.7.A). Un résultat similaire a été observé chez *S. pombe*(Figure 2.7.B) (Dickey and Wuttke, 2014). Ceci suggère que *POT1* est capable de se lier de manière partielle à l'ADN télomérique lorsqu'elle est en large excès par rapport à son substrat, montrant que le mode de liaison de *POT1* dépend de sa concentration. La structure cristallographique du complexe *POT1*-ADN, caractérisée par Lei et al. (Lei et al., 2004) montrent que les domaines *OB1* et *OB2* se lient respectivement la séquence 5'-TTAGGG-3' et 5'-TTAG-3' (Figure 1.16) ; on peut imaginer que *POT1* soit capable de se lier par l'intermédiaire du seul domaine *OB1*, probablement le plus stable, réduisant sa séquence de liaison à 5'-TTAGGG-3'. Cette hypothèse peut expliquer la présence de complexes *POT1* :ADN de stœchiométrie 3 :1. L'hypothèse que *OB1* soit le domaine

de la *POT1* humaine de plus haute affinité pour l'ADN est confortée par les données relatives à la *Pot11* de *S. pombe* : à la différence de *POT1* humaine, les deux domaines *OB1* et *OB2* de *Pot1* (*S. pombe*) ont pu être produits et donc étudiés séparément ; cela a permis de montrer que, pour *Pot1*, le domaine *OB1* isolé a une plus forte affinité pour l'ADN que le domaine *OB2* isolé. Il est possible que, pour des raisons structurales à élucider, la formation du complexe *POT1-TPP1* rende impossible la liaison avec l'ADN via le seul domaine *OB1* de *POT1*.

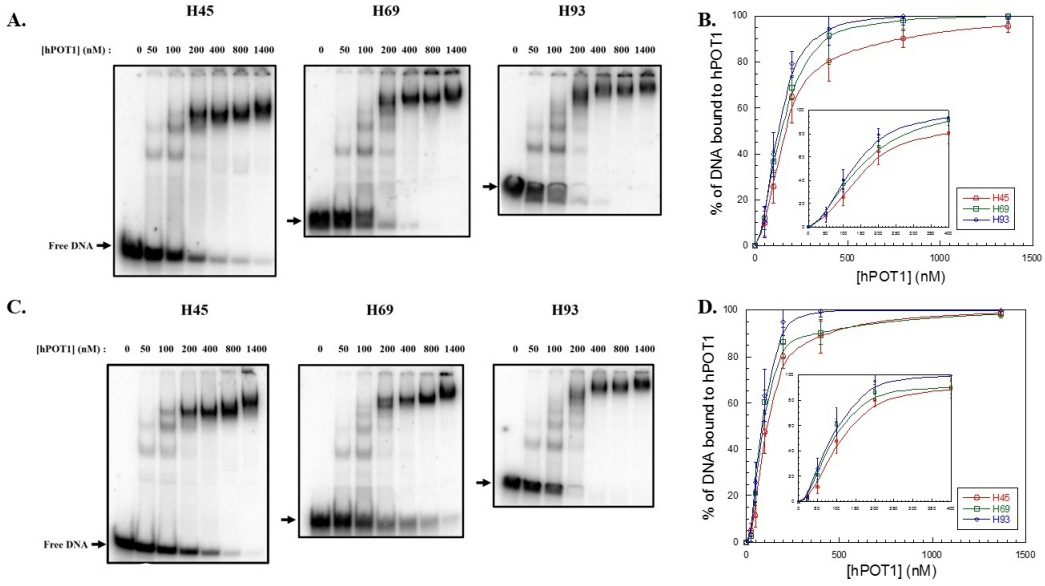


FIGURE 2.6 – **Expérience *EMSA* avec *POT1* en présence de longues séquences télomériques. A.** *EMSA* en présence de H45, H69 et H93 en potassium. [*POT1*] : 0 / 50 / 100 / 200 / 400 / 800 / 1400 nM. **B.** Quantification des expériences *EMSA* en potassium. **C.** *EMSA* en présence de H45, H69 et H93 en lithium. [*POT1*] : 0 / 50 / 100 / 200 / 400 / 800 / 1400 nM. **D.** Quantification des expériences *EMSA* en lithium.

Á ce stade ce mode de liaison de *POT1* via le seul domaine *OB1* reste hypothétique. Pour confirmer ou infirmer ce modèle des expériences plus approfondies sont nécessaires. Nous pouvons notamment imaginer d'utiliser une séquence courte d'ADN télomérique au sein de laquelle les différentes positions des thymines seraient remplacées par des 4-thioThymines. Par conséquent nous pourrions ainsi sonder l'environnement de notre ADN et mettre en évidence par photopontage la présence ou non de complexe de différente stœchiométrie en fonction de la concentration en *POT1*. Pour aller encore plus loin, une étude par spectroscopie de masse des complexes ADN :pro-

téine formés pourrait mettre en évidence les acides aminés impliqués dans les différents mode de liaison et ainsi démontrer la liaison partielle de *POT1* par l'intermédiaire d'un de ces domaines *OB fold*.

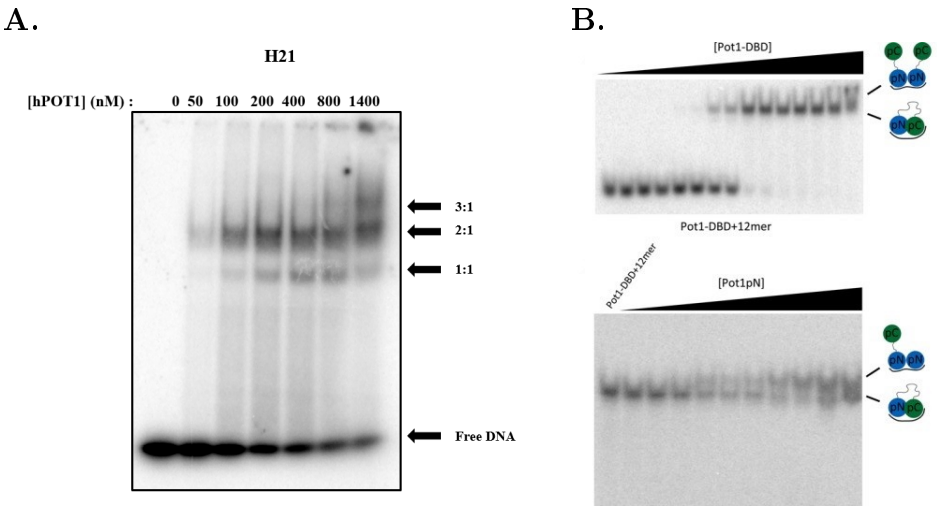


FIGURE 2.7 – .
Mode de liaison partiel de *POT1* en fonction de la concentration chez l'homme et *S. pombe*A. *EMSA* en présence d'un H21 en lithium. [*POT1*] : 0 / 50 / 100 / 200 / 400 / 800 / 1400 nM. **B.** *Pot1* de *S. pombe* forme un complexe de stœchiométrie plus élevé visible par *EMSA* à haute concentration en présence de 12mer. (Dickey and Wuttke, 2014)

J'ai également réalisé des expériences de *FRET* afin d'étudier la capacité de *POT1* à ouvrir un G4 en fonction de sa position au sein de la séquence. Pour ce faire j'ai travaillé avec le même système utilisé dans les expériences de *FRET* en présence du complexe *POT1-TPP1* à savoir un oligonucléotide capable de former trois G4s en tandem auxquelles un couple de fluorochrome donneur-accepteur a servi pour marquer chaque G4. Les résultats de FRET montrent que *POT1* ouvrent plus efficacement les G4 en position 3' et central que celui en 5' (Figure 2.8). L' unique expérience de cinétique, réalisée avec 6 équivalent de *POT1* (Figure 2.9) confirme que *POT1* ouvre préférentiellement les G4 situés à l'extrémité 3' et au centre, ce qui semble suggérer une directionalité de liaison 3' vers 5' de *POT1*. Une étude plus approfondie est nécessaire pour le confirmer.

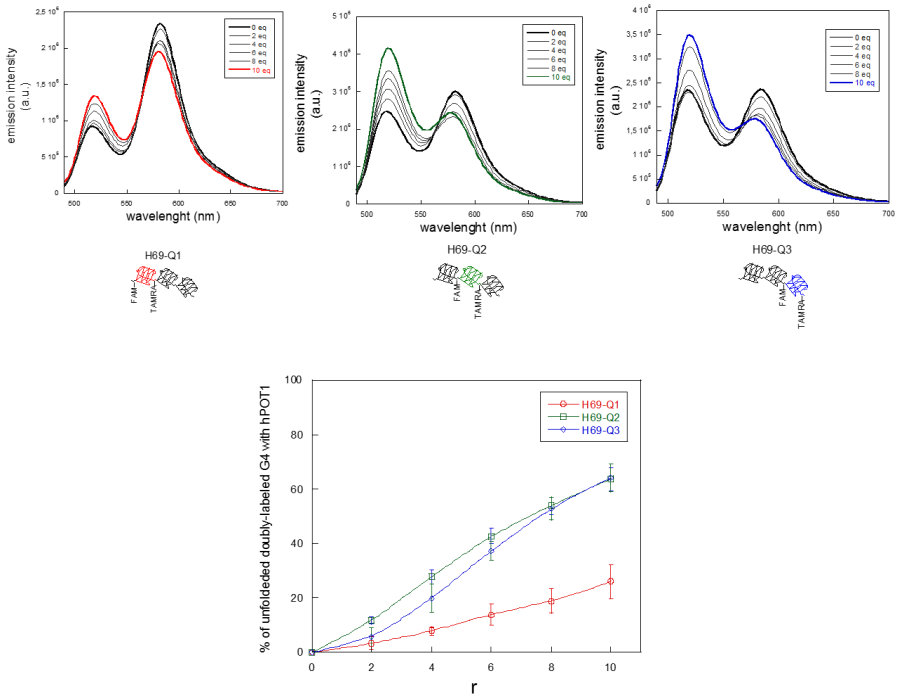


FIGURE 2.8 – **Spectre d'émission de *POT1* en fonction des G4s situés sur la séquence H69.** Les courbes correspondent aux quantifications du pourcentage d'ouverture en fonction de la concentration : la courbe bleu correspond au G4 à l'extrémité 3', en vert au G4 central et enfin en rouge au G4 en 5'.

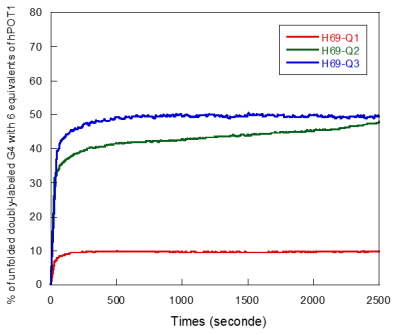


FIGURE 2.9 – **Cinétique d'ouverture par *POT1* des G4s situés sur la séquence H69.** La courbe bleu correspond au G4 à l'extrémité 3', en vert au G4 central et enfin en rouge au G4 en 5'.

En conclusion *POT1* seule est capable de se lier et d'ouvrir les structures G4 en tandem formées par de longues séquences télomériques. L'aspect sigmoïdal des courbes de liaisons de *POT1* à l'ADN en potassium laisse à imaginer un modèle de **coopérativité apparente** en présence de G4 contigus similaires à celui proposé pour *POT1-TPP1*. Cependant l'étude de la liaison de *POT1* à de longues séquences télomériques est plus complexe à appréhender et à interpréter. Les raisons à cette complexité sont multiples. Nous pouvons rappeler que le mode de liaison de *POT1* dépend de la concentration et qu'elle est capable d'adopter un mode de liaison partiel à haute concentration. De même les expériences de Hwang et al. en 2012 montrent que deux protéines *POT1* sont nécessaires à l'ouverture d'un G4 tandis qu'un seul complexe *POT1-TPP1* est suffisant (Hwang et al., 2012). Ce résultat indique qu'il faut donc deux événements de liaison de *POT1* pour ouvrir un G4. De plus, une étude réalisée en molécule unique par Ray et al. montre que la protéine *POT1* seule peut-être déplacée de son ADN, si ce dernier se restructure en G4 (Annexe 8.5) (Ray et al., 2014). Toutefois, leur étude ne donne pas suffisamment d'informations sur le mode de liaison de *POT1* ou s'il y avait une ou deux protéines *POT1* présente sur l'oligonucléotide, un H24 c'est-à-dire deux sites distincts de liaison de *POT1*. Cependant ce résultat nous montre que *POT1* seule peut être déplacé de l'oligonucléotide par la simple restructuration d'un G4. En compilant mes données préliminaires sur *POT1* et les données de la littérature, je propose le modèle d'interaction de *POT1* sur les G4 contigus suivant : une première protéine POT1 vient se lier en 3' de la séquence et ainsi dérouler le G4 à cette extrémité (Figure 2.10). Ensuite il existe deux possibilités :

1 : La vitesse de liaison d'une seconde protéine est plus rapide que la vitesse de reformation du G4. Dans ce cas la liaison de la seconde protéine assure ainsi le maintien de l'ouverture du G4. La présence de ces deux protéines sur l'ADN va déstabiliser le G4 contigu, ce qui va favoriser la fixation d'une troisième protéine et ainsi de suite. Dans ce cas de figure, on est dans un mode de liaison séquentiel **coopératif apparent** de 3' vers 5 décrite en présence de *POT1-TPP1*.

2 : La vitesse de liaison d'une seconde protéine est moins rapide que la vitesse de reformation du G4. Dans ce cas le G4 se restructure et déplace la protéine *POT1*.

Á mon avis il y a les deux phénomènes qui ont lieu en même temps dans nos expériences, ce qui pourrait expliquer en partie la complexité de notre système. Sans parler que ce modèle ne prend pas en compte les différents modes de liaison de *POT1* lié à sa concentration. La possibilité qu'un premier évènement de liaison échoue (cas de figure 2) pourrait expliquer en partie la

complexité de l'interaction de POT1 seule avec l'ADN.

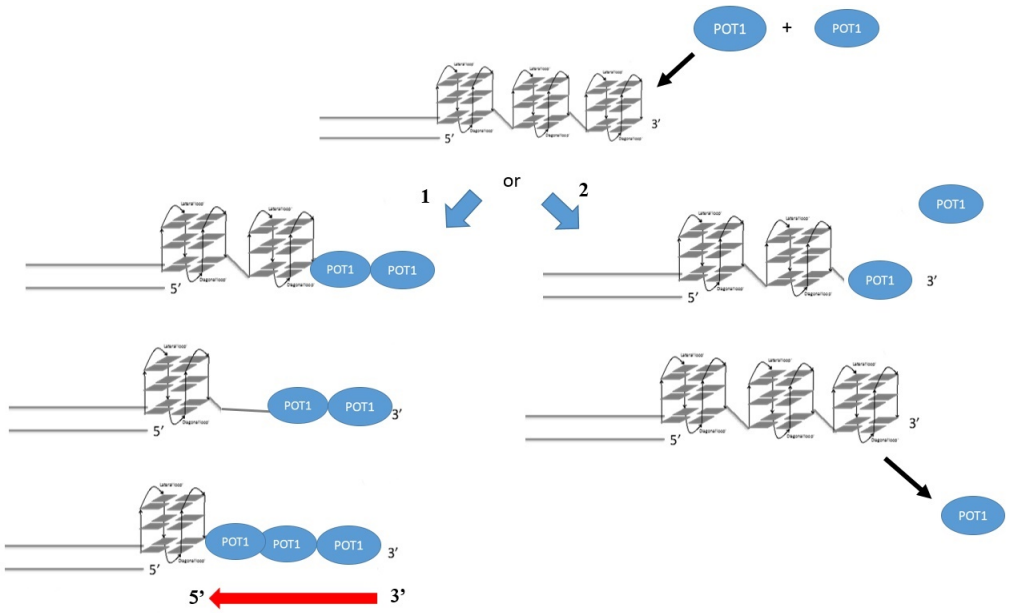


FIGURE 2.10 – **Modèle de liaison de *POT1* sur des longues séquences capables de former des G4s contigus.**

Il est important de ne pas voir les complexes *POT1*-ADN et *POT1-TPP1*-ADN comme figés mais plutôt dynamiques, ce qui se traduit ici par la "respiration" des G4s. C'est un aspect qui n'a pas été abordé dans l'article sur *POT1-TPP1*. Même si aucune expérience ne permet d'affirmer que le complexe *POT1-TPP1* résiste mieux à la formation d'un G4, sachant que *TPP1* améliore l'affinité de *POT1* à l'ADN télomérique et qu'un seul complexe *POT1-TPP1* suffit pour maintenir le G4 ouvert, on peut envisager que *TPP1* améliore la vitesse de liaison de *POT1* à l'ADN et ainsi la restructuration du G4.

2.4.2 Étude de *RPA*

L'étude réalisée en 2018 au sein de l'équipe SANTÉ a montré que *RPA* est capable de prendre en charge des G4s télomériques en tandem et, avec une mode de liaison légèrement coopératif, probablement du, à la lumière des résultats obtenus avec *POT1-TPP1*, à la présence des G4 contigus (Lancrey et al., 2018). Mes études de cinétique en *FRET* montrent que *RPA* ouvre plus facilement le G4 en position centrale que ceux situés aux extrémités. Ce résultat est en accord avec le fait que les G4s aux extrémités sont plus

stables que celui en position centrale (Bugaut and Alberti, 2015) et que *RPA* lie et ouvre préférentiellement les G4 les moins stables (Ray et al., 2013; Safa et al., 2014). A partir de ces résultats, je propose un mécanisme de liaison et d'ouverture des multimères de G4 contigus par RPA, rendant compte de sa coopérativité : RPA se lie et ouvre tout d'abord le G4 central, et cette liaison entrainerait la déstabilisation des G4s voisins facilitant la liaison d'autres protéines (Figure 2.11).

Il est à noter que la coopérativité observée pour *RPA* est faible. Cette faible coopérativité pourrait s'expliquer par le fait que *RPA* présente différents modes de liaison détaillés dans la section 1.5.3. On peut supposer que la protéine *RPA* se lierait au G4 central dans une "conformation allongée étendue" (30 nt), couvrant la totalité de la séquence impliquée dans la formation du G4. Dans cette situation, *RPA* ne génèrerait pas une région simple brin pouvant favoriser la fixation d'une autre protéine *RPA*. Du moins dans un premier temps, car au contact d'une autre protéine *RPA* adopterait une conformation plus compacte. Dans cette conformation la protéine *RPA* liée à l'ADN génère un simple brin et favorise dans cette situation la fixation d'une deuxième protéine traduisant une coopérativité.

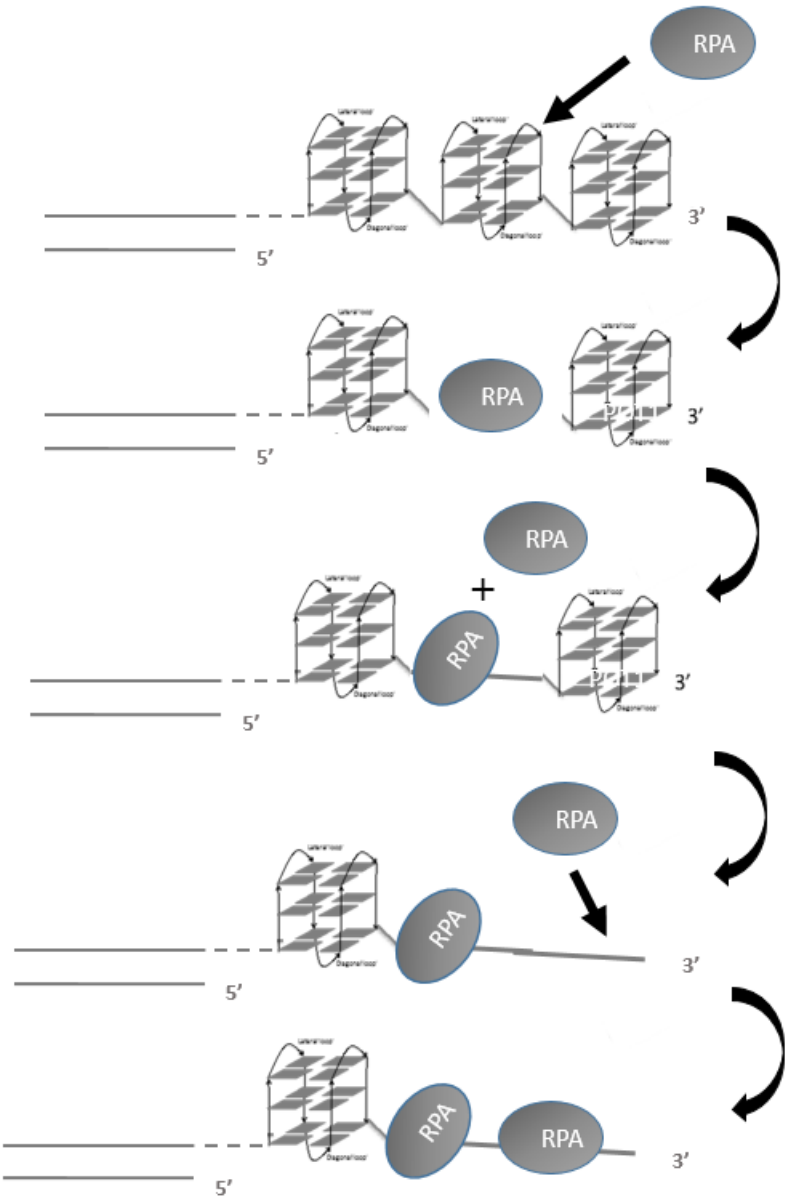


FIGURE 2.11 – **Modèle de la coopérativité de *RPA* sur des séquences formant des multimères de G4 en tandem.**

2.4.3 Étude de *CST* : en perspective

L'étude de l'équipe sur l'interaction de *RPA* avec des longues séquences télomériques et mes études sur *POT1+/-TPP1* montrent que *in vitro* les structures de G4s contigus ne semblent pas être problématiques pour la liaison de certaines de ces deux protéines *SSB-OB fold*. En effet, aussi bien la protéine *RPA* que le complexe *POT1-TPP1* sont capables de prendre en charge les structures télomériques en G4 contigus. Le troisième complexe de la famille *SSB-OB fold* impliqué dans le maintien des télomères est le complexe *CST*. En perspective, il sera intéressant de caractériser son interaction avec les séquences télomériques structurées comme nous l'avons fait pour *RPA* et *POT1*.

2.4.4 Compétition entre *RPA* et *POT1-TPP1* sur des multimères de G4s en tandem

Nous avons vu que la présence de *RPA* aux téloméres est régulée au cours du cycle cellulaire, elle est présente uniquement aux télomères lors de la réplication de ces derniers. Toutefois, il a été observé que la suppression de *POT1* entraine une accumulation de *RPA* aux télomères, même en dehors de la phase S (Gong and de Lange, 2010), indiquant que, dans des conditions normales, la liaison de *RPA* au *G-overhang* est inhibée par *POT1*. Cette inhibition s'explique en raison du rôle de *RPA* comme activateur des points de contrôle, voie de réponse aux dommage de l'ADN dépendant de la kinase *ATR* (Shiotani and Zou, 2009; Flynn and Zou, 2011). A l'inverse, la protéine *POT1* est présente en dehors de la phase S aux télomères pour protéger ces derniers et réprimer la voie *ATR* (Wu et al., 2006; Denchi and de Lange, 2007; Guo et al., 2007) ; elle est déplacée de l'ADN télomérique pendant la phase S de façon transitoire pour permettre la réplication du brin G. Ces résultats semblent présenter *RPA* et *POT1* comme deux antagonistes, ce qui a poussé Flynn et al. dans une revue de 2012 intitulée : "*RPA and POT1 : Friends or Foes at Telomeres ?* à s'interroger si *RPA* et *POT1* sont des "amies" ou des "ennemies" (Flynn et al., 2012). Pour expliquer la présence bien régulée de ces deux protéines aux télomères au cours du cycle cellulaire, Flynn et al. ont proposé un modèle dans lequel la protéine *hnRNP-A1* (*heterogeneous nuclear ribonucleoprotein A1*) et l'ARN non codant *TERRA* (*telomere repeat containing RNA*) sont impliqués.

Au cours de la phase S, la progression de la fourche de réplication déclencherait le déplacement du complexe *Shelterin* et ainsi de *POT1* du *G-overhang*. La séquestration de la protéine *hnRNP-A1* par *TERRA*, permettrait à *RPA*, présente aux fourches de réplication, de rentrer en compétition

avec *POT1* pour le *G-overhang*. En fin de phase S, *TERRA* déclinant, la protéine *hnRNP-A1* relarguée, déplacerait *RPA* (Sui et al., 2015). Au cours des phases G2/M, la surexpression de l'ARN *TERRA* permettrait de séquestrer à nouveau *hRNP-A1* ; la reformation du *Shelterin* favoriserait alors la liaison de *POT1* au *G-overhang*. Dans ce modèle, les auteurs n'ont pas pris en compte le rôle que pourraient avoir des structures de l'ADN télomériques dans la compétition *RPA-POT1*. Une étude plus récente menée par Ray et al. en 2014 montre que certaines structures G4s peuvent protéger *POT1-TPP1* du déplacement par *RPA* (Ray et al., 2014). On comprend donc qu'il est important de tenir compte des structures présentes sur le brin G pour mieux comprendre cette compétition existant entre *RPA* et *POT1-TPP1*. Il devient alors intéressant d'étudier cette compétition entre la protéine *RPA* et *POT1-TPP1* sur des longues séquences d'ADN simple brin capables de se structurer en G4s contigus. Cette étude est d'autant plus intéressante, que nous avons montré que ces protéines *SSB-OB fold* présentent une **coopérativité apparente** sur ces séquences d'ADN télomérique ; coopérativité qui pourrait intervenir dans les mécanismes arbitrant la compétition entre *RPA* et *POT1-TPP1*.

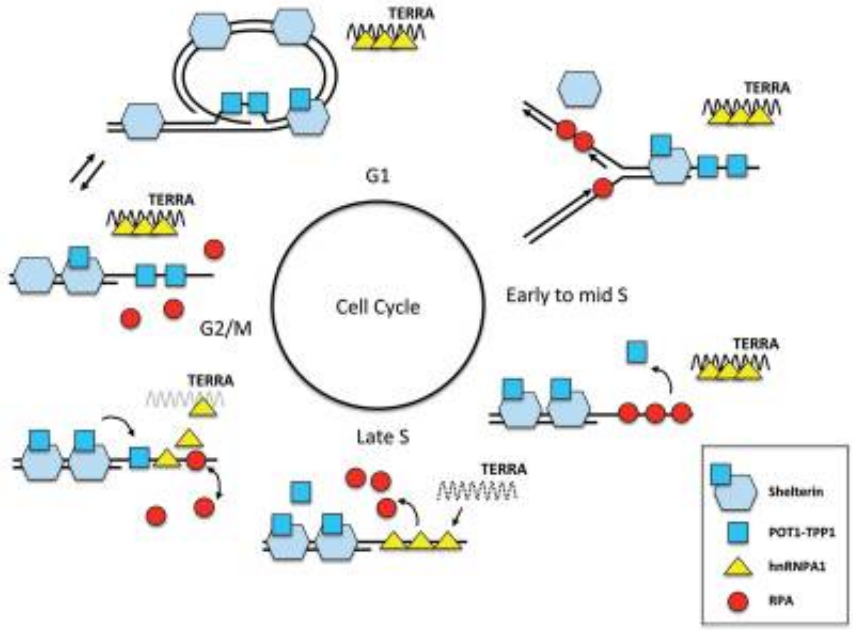


FIGURE 2.12 – **Modèle du "switch" entre *RPA* et *POT1*.** Proposé par Flynn et al. (Flynn et al., 2012).

2.5 Annexes méthodes

2.5.1 *FRET*

Le principe du FRET (*fluorescence resonance energy transfer*) repose sur le transfert d'énergie non radiatif, c'est à dire sans émission de lumière, entre deux fluorochromes : un accepteur et un donneur. Ce phénomène physique demande que le spectre d'émission du donneur recouvre, au moins partiellement, le spectre d'absorption de l'accepteur tel que cela est schématisé sur la figure 2.13.A. Un facteur de recouvrement spectral J_{AD}, calculé à partir de l'air du spectre d'émission du donneur recouvert par le spectre d'excitation de l'accepteur (eq. 2.1), permet de quantifier l'efficacité de transfert respective d'un couple accepteur - donneur.

$$J_{AD} = \int I_{Donneur}(\lambda)\epsilon_{Accepteur}(\lambda)\lambda^4 \mathrm{d}\lambda \tag{2.1}$$

avec :
J_{AD} : facteur de recouvrement spectral
$I_{Donneur}(\lambda)$: Intensité du donneur à la longueur d'onde λ
$\epsilon_{Accepteur}(\lambda)$: Coefficient d'extinction molaire de l'accepteur

L'efficacité E de transfert d'énergie dépend de la distance entre les deux fluorochromes (eq. 2.2). Le rayon de Föster R_0 (eq. 2.3) est défini comme la distance entre deux fluorochromes d'un couple accepteur - donneur pour laquelle l'efficacité de transfert équivaut à 50% (Figure : 2.13.B). Cette distance traduit le fait que la probabilité d'un fluorochrome donneur à se désexciter par transfert d'énergie non radiatif vers un fluorochrome accepteur correspond à 50% (Förster, 1965). De même, on estime être dans des conditions favorables au phénomène de FRET lorsque l'efficacité E de transfert d'énergie est supérieure à 5% de son efficacité maximale ce qui, traduit en distance effective du couple donneur - accepteur, correspond à 1,8 fois R_0. Il est intéressant de noter que le phénomène physique de FRET est le résultat d'une interaction dipôle - dipôle, expliquant sa très courte portée due à sa dépendance en R^{-6}. De manière générale, nous sommes dans des conditions favorables au FRET dans le cas où les deux fluorochromes sont séparés d'une distance inférieure à 10 nm.

$$E = \frac{R^{-6}}{R^{-6} + R_0^{-6}} \tag{2.2}$$

avec :
- R : Distance entre deux fluorochromes
- R_0 : Rayon de Föster

$$R_0 = (10^{-3}\kappa^2 n^{-4} Q_D J_{AD})^{\frac{1}{6}} \times 9730 \tag{2.3}$$

avec :
- κ : le facteur d'orientation
- n : l'indice de réfraction du milieu
- Q_D : le rendement quantique en l'absence d'accepteur
- J_{AD} : le facteur de recouvrement spectral

Cet aspect dipôle - dipôle entraîne également une contrainte sur l'orientation relative du dipôle donneur par rapport au dipôle accepteur. Il est important de savoir que le phénomène de FRET ne peut pas avoir lieu dans le cas où cette orientation relative forme un angle de 90° (Figure : 2.13.C). Dans notre étude, nous utilisons des fluorochromes attachés à des biomolécules en solution. Par conséquence nos fluorochromes ayant des degrés de liberté rotationnelle, nous pouvons nous affranchir des problèmes liés à l'orientation des dipôles.

L'intérêt d'utiliser la technique de FRET tient à la dépendance de ce phénomène physique à des distances très courtes, comme vu précédemment. Pour notre étude, nous avons choisi comme couple donneur - accepteur respectivement la Fluorescéine et le TAMRA (tetraméthyl-rhodamine), déjà utilisés en 1988 pour révéler l'hybridation de brins d'ADNs (Cardullo et al., 1988). En effet, ce couple de fluorochromes remplit les conditions du FRET, à savoir : la fluorescéine (donneur) possède une durée de vie de fluorescence suffisamment longue pour permettre un transfert d'énergie non radiatif ; le spectre d'émission du donneur (fluorescéine) recouvre partiellement le spectre d'excitation de l'accepteur (TAMRA) sans pour autant que ce dernier recouvre le spectre d'excitation de la fluorescéine (Figure : 2.14). En excitant la fluorescéine à sa longueur d'excitation λ_{ex} =470 nm, nous pouvons visualiser : soit dans les conditions de FRET le pic d'émission du TAMRA λ_{em}=582 nm, soit à l'inverse, en absence des conditions de FRET, le pic d'émission de la fluorescéine λ_{em}=518 nm.

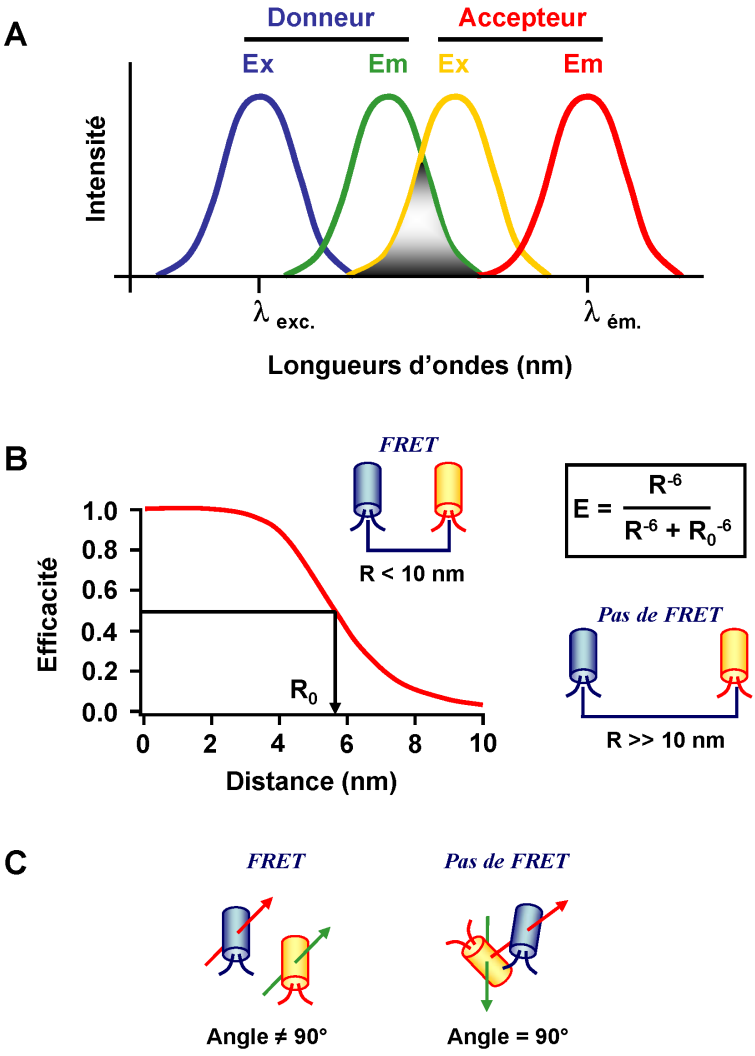


FIGURE 2.13 – **Principe du FRET.**
A. Schéma des spectres d'excitations et d'émissions respectifs du donneur et de l'accepteur. **B.** Courbe de l'efficacité de transfert d'énergie par rapport à la distance. **C.** Orientation relative entre le dipôle donneur par rapport au dipôle accepteur. ©Wikipédia

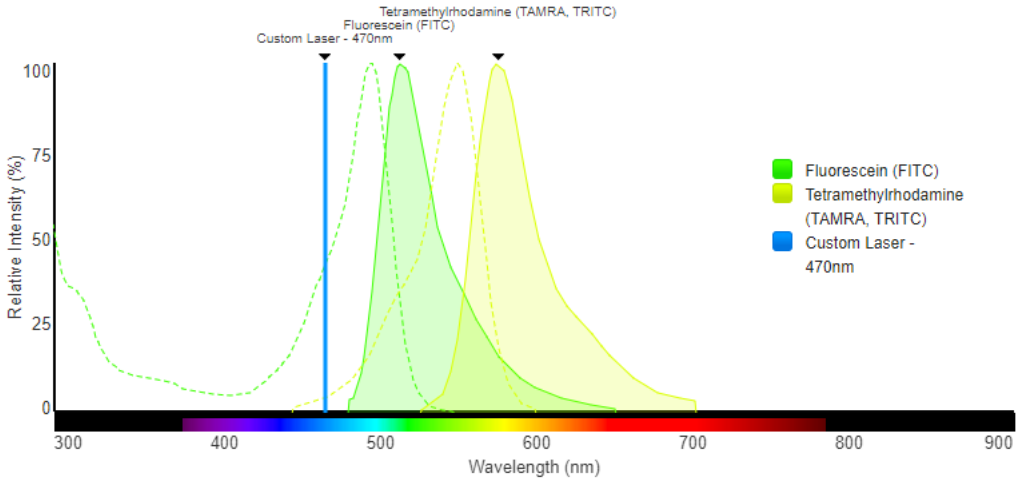


FIGURE 2.14 – **Couple Fluorescéine - TAMRA.**
Spectres d'excitation et d'émission respectifs de la Fluorescéine (donneur) et TAMRA (accepteur). ©ThermoFisher Fluorescence SprectraViewer

Pour suivre l'ouverture du G4, les deux fluorochromes sont fixés sur la première et dernière guanine constitutive d'un G4 (Figure : 2.15). L'efficacité de transfert d'énergie dépend de la distance entre les deux fluorochromes, c'est pourquoi la technique de FRET permet d'avoir une information spatiale du système étudié qui dans notre étude traduit l'état du G4 : fermé ou ouvert. En effet, en excitant la fluorescéine à sa longueur d'onde d'excitation, nous observons : soit l'émission du TAMRA, traduisant l'état fermé du G4 (Figure : 2.15.A), soit l'émission de la fluorescéine indiquant l'absence de FRET et l'état ouvert du G4 (Figure : 2.15.B). Ce système nous permet de marquer de façon indépendante chaque G4 présent dans un contexte de G4s en tandem (Bugaut and Alberti, 2015) (Figure 2.16). Nous avons vérifié par EMSA que la liaison de nos protéines à ces séquences d'ADN simple brin, marquées sur différentes positions, ne modifiaient pas la liaison des protéines à leur substrat. L'obtention de profil de migration très similaire, permet de conclure que, si notre système possède un biais expérimental, du à la présence des fluorophores, ce dernier est constant et influence de la même manière la liaison des protéines.

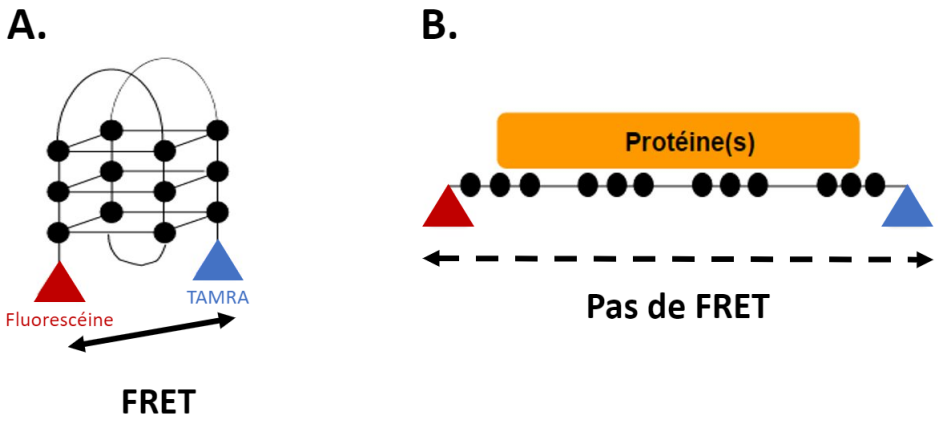


FIGURE 2.15 – **Schéma des états ouvert et fermé d'un G4 marqué par fluorochromes.**
Les • représentent les guanines.**A.** État fermé du G4 : FRET, **B.** État ouvert du G4 par la(les) protéine(s) : Pas de FRET.

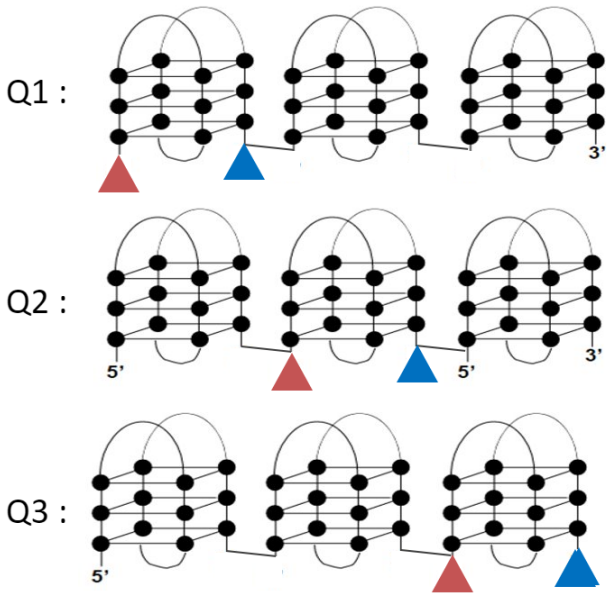


FIGURE 2.16 – **Schéma de la séquence télomérique H69 marquée par des fluorochromes** Les • représentent les guanines.

Chapitre 3

Structure du complexe ternaire *POT1-TPP1*-ADN

3.1 Introduction

3.1.1 La protéine *TPP1*, plus qu'un simple partenaire de *POT1*

Comme nous l'avons abordé dans l'introduction générale (section 1.5.2), la protéine *TPP1* découverte en 2004 par Liu et al. est essentielle à la présence de *POT1* aux télomères (Liu et al., 2004), et fait partie intégrante du complexe *Shelterin* qui, pour rappel, désigne le complexe multiprotéique constitué des six protéines : *TRF1*, *TRF2*, *RAP1*, *TIN2*, *TPP1* et *POT1* ; la protéine *TPP1* assure la liaison entre *POT1* et *TIN2* (section : 1.5.2). De même la suppresion de *TIN2* au sein de la cellule entraîne une délocalisation des protéines *POT1* et *TPP1* de la région télomérique (Takai et al., 2011), et provoque l'activation de la kinase *ATR* ainsi que des phénomènes de fusions inter chromosomiques caractéristiques d'une déplétion de *POT1* aux télomères. Par ailleurs, plusieurs études montrent que *TPP1* améliore l'affinité de POT1 pour son substrat (Xin et al., 2007; Wang et al., 2007; Taylor et al., 2011). Bien que *TPP1* possède un domaine *OB fold* en N-term aucune interaction entre l'ADN télomérique et *TPP1* n'a été observée jusqu'alors. Actuellement, il est donc couramment assumé que *TPP1* n'interagit pas avec l'ADN. En l'absence de toute structure du complexe ternaire *POT1-TPP1*-ADN, comment *TPP1* améliore l'affinité de *POT1* pour son substrat reste à ce jour une question ouverte, à laquelle je me suis intéressé. Pour répondre à cette question, je me suis intéressé à l'aspect cinétique de l'interaction de *POT1* et de *POT1-TPP1* avec leur site de fixation et à l'aspect structural,

en questionnant les données structurales disponibles pour *POT1* et *TPP1* et pour leurs protéines homologues chez les ciliés *TEBPα* et *TEBPβ*.

3.1.2 Le modèle *TEBPα* - *TEBPβ* chez *Oxytricha nova*

Le complexe *TEBPα*-*TEBPβ*, découvert en 1991 chez le cilié *Oxytricha nova* (*O. nova*), assure la protection des extrémités des chromosomes au sein du macronoyau (Gray et al., 1991). Gray et al, montre que chez le cilié *Oxytricha nova* *TEBPα* lie les répétitions du motif télomérique T_4G_4 de manière spécifique. *TEBPα* possède en N-term deux domaines *OB fold* de liaison à l'ADN télomérique (*OB*1 et *OB*2), et en C-term, un domaine OB-fold d'interaction avec son partenaire (*OB*3) (Figure 3.1.A) ; *TEBPβ* possède en N-term un domaine OB-fold, et en C-term, un domaine d'interaction avec son partenaire *TEBPα* (*TIM*) (Figure 3.1.B) (Fang and Cech, 1993). Cette organisation structurale des deux sous-unités n'est pas sans rappeler celle du complexe *POT1*-*TPP1* (Figures 1.14 et 1.15). De même, malgré la présence d'un domaine *OB fold* au sein de la sous-unité *TEBPβ* (Figure 3.1.B), *TEBPβ* seul ne lie pas l'ADN télomérique (expériences réalisées en utilisant la technique *EMSA*).

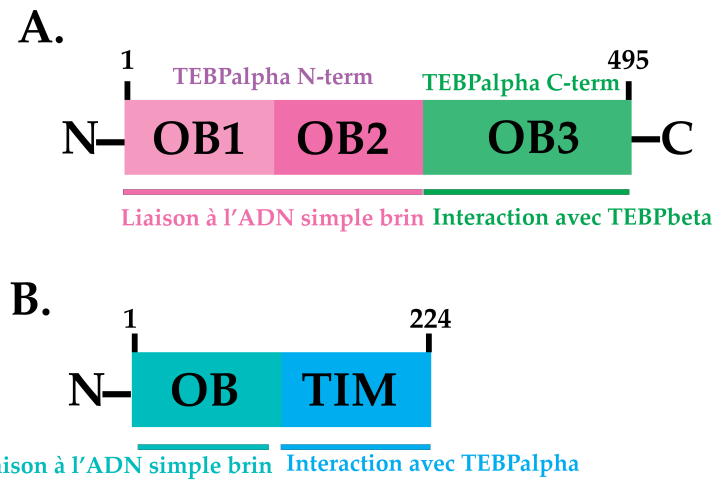


FIGURE 3.1 – **Les domaines structuraux et fonctionnels de *TEBPα*-*TEBPβ*. A.** de *TEBPα*, **B.** *TEBPβ*.

La structure du complexe ternaire *TEBPα-TEBPβ* (dont la séquence des 12 nucléotides est GGGGTTTTGGGG) a été déterminée par cristallographie en 1998 par Horvath et al, (Horvath et al., 1998) avec une résolution de 2.80 Å (Figure 3.2). Le complexe ternaire montre une structure repliée en forme de cœur dans laquelle l'ADN se place au centre (Figure 3.2.A, représentation atomique) et interagit avec les domaines *OB*1 et *OB*2 de *TEBPα* et le domaine *OB* de *TEBPβ* (Figure 3.2.B, diagramme en ruban). La figure 3.3 montre le détail des interactions ADN-domaines *OB* : *TEBPα* se lie à l'ADN, via les domaines *OB fold* situés en N-term, en établissant : (i) neuf liaisons hydrogène avec les bases G2, G3, G4, T8 et G11, (ii) trois interactions ioniques de *TEBPα* avec les groupements phosphate de l'ADN et (iii) 14 interactions hydrophobes dont 6 interactions hydrophobes avec la base G11. La présence de cette poche hydrophobe autour de cette guanine n'est pas sans rappeler le mode de liaison de *POT1* à l'ADN, où la guanine G9 est également prise dans une poche hydrophobe (section 1.5.2) (Lei et al., 2004). La structure cristallographique du complexe ternaire révèle un certain nombre d'interactions entre l'ADN et le domaine *OB* de *TEBPβ* : (i) trois liaisons hydrogène avec les bases G9 et G10, (ii) six interactions hydrophobes, dont quatre avec la base G9. On peut remarquer que l'ADN se replie sur lui-même et présente des empilements de bases (T6, T7 et G9) caractéristiques du mode de liaison des domaines *OB fold*.

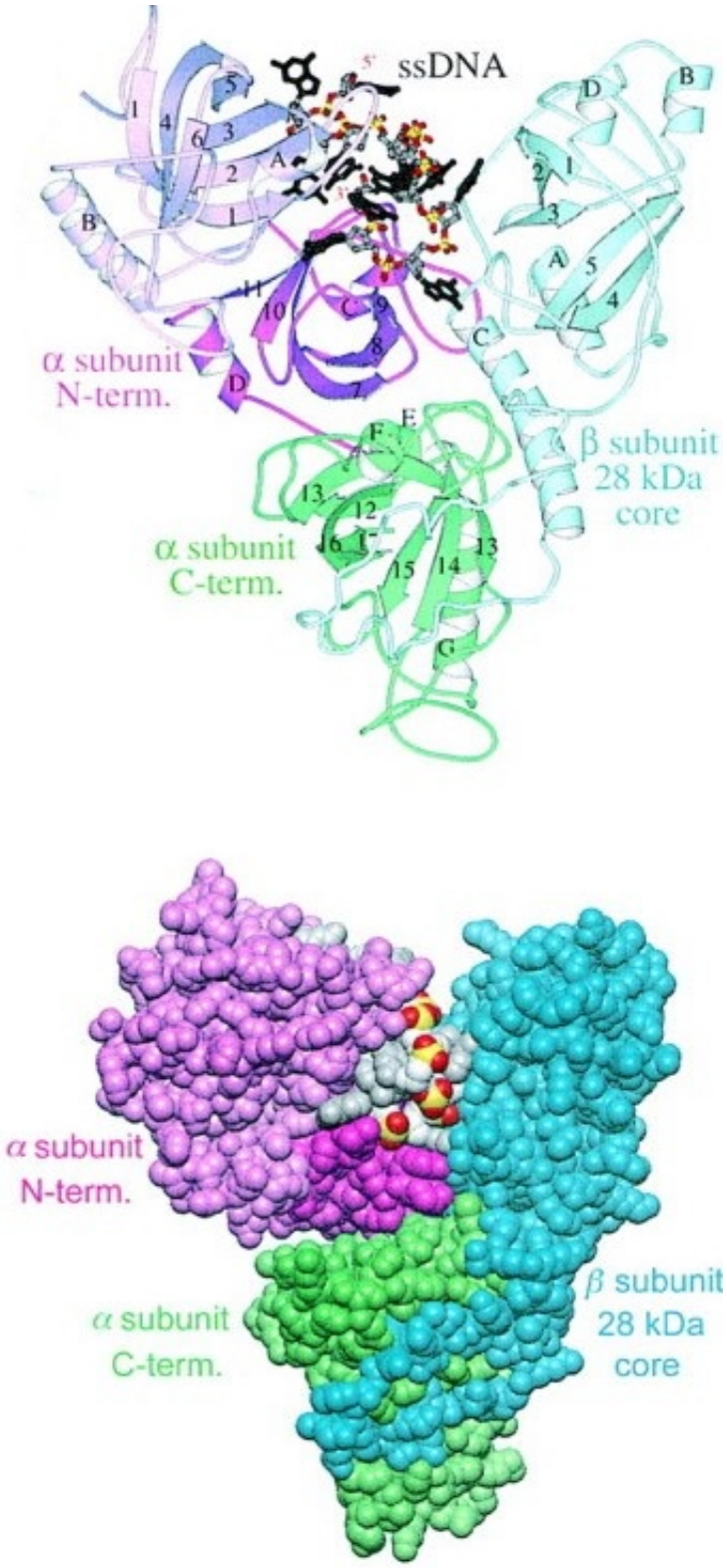


FIGURE 3.2 – **Structure du complexe TEBPα-TEBPβ avec l'ADN télomérique (12 nt) (PDB :1OTC).**(Horvath et al., 1998)

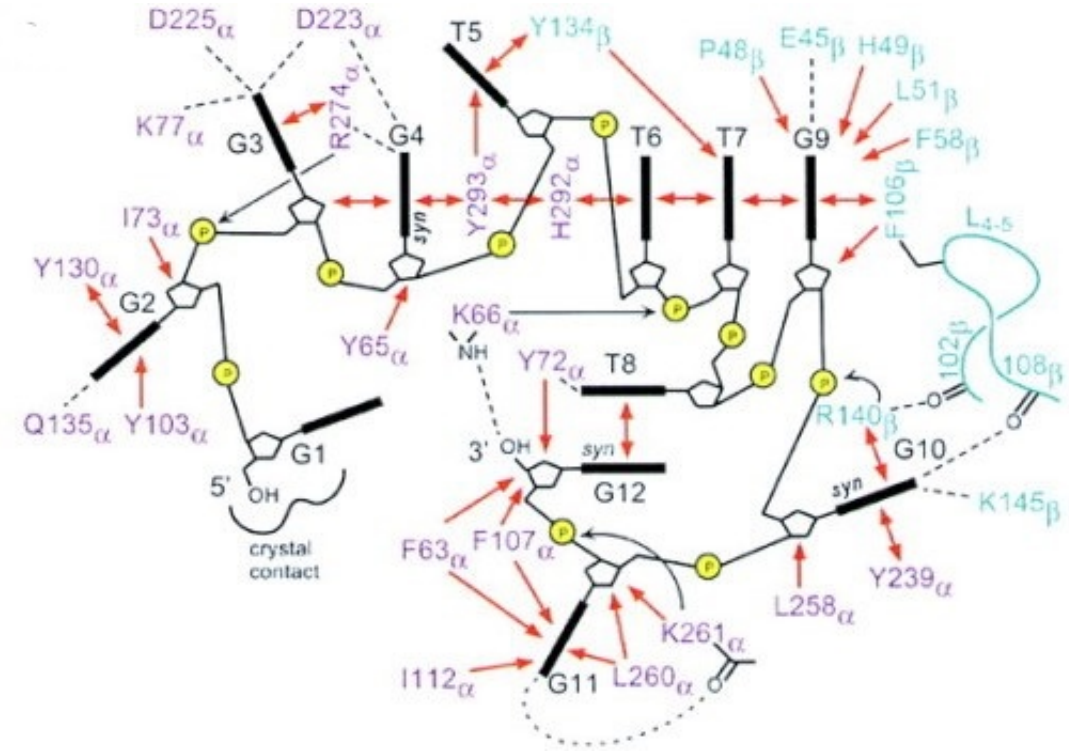


FIGURE 3.3 – **Représentation schématique des interactions *TEBP*α-*TEBP*β avec l'ADN simple brin (12 nt).** Les bases de l'ADN ss sont représentées par des rectangles noirs ; les groupes phosphate sont représentés par des cercles jaunes ; les résidus de *TEBP*α sont indiqués par des marqueurs violets les résidus de *TEBP*β sont indiqués par des étiquettes bleues. Les liaisons hydrogène sont représentées en pointillés ; les interactions ioniques sont représentées par des flèches noires pleines et les contacts étroits entre les groupes hydrophobes sont représentés par des flèches orange. Enfin si les groupes aromatiques sont empilés cela est indiqué par une flèche orange à deux pointes. (Horvath et al., 1998)

3.1.3 Le modèle *POT1-TPP1* chez l'homme

Pour rappel, des expériences de *SAXS* réalisées en présence du complexe *POT1-TPP1* et l'ADN de 10 nucléotides, ont été réalisées et une enveloppe a été déterminée (Figure **??**). Cette enveloppe se présente sous la forme d'un V allongé dans laquelle le domaine *OB fold* de *TPP1* est très éloigné de l'ADN. Cette structure en forme de "V" allongé va dans le sens d'une non conservation de la structure en forme de coeur, observée chez le cilié *O. nova*; et cela malgré une organisation et des domaines *OB fold* conservés (Figure 3.4) (Wang et al., 2007).

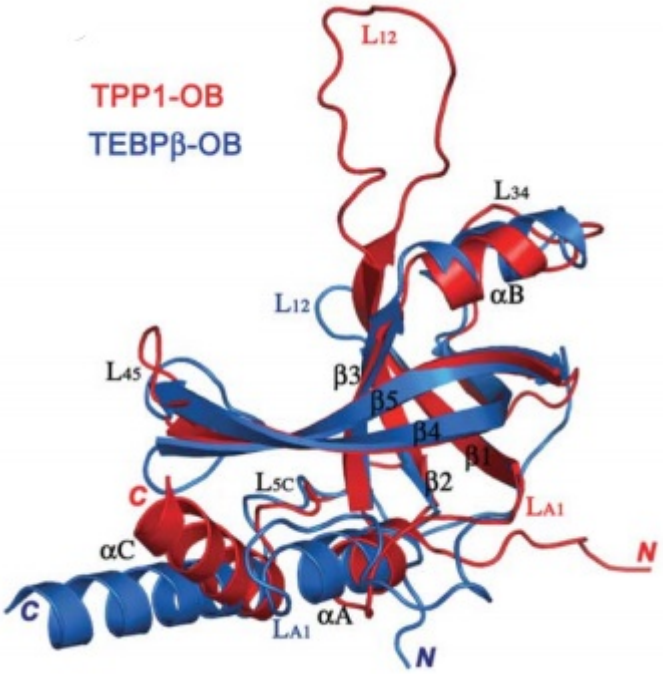


FIGURE 3.4 – **Comparaison des structures cristallines des motifs *OB-fold* de *TPP1* humaine et de *TEBP*β de O. nova**(Wang et al., 2007)

Toutefois, l'analyse *SAXS* restant une méthode de calcul de structure, il nous a semblé intéressant de réaliser une étude structurale en solution. Deux techniques complémentaires ont été utilisées : le photopontage utilisant la 4-thioThymine, qui permet de sonder les interactions protéine-ADN dans des complexes nucléoprotéiques (Annexe), et la microscopie électronique à transmission (MET), par coloration négative, qui permet d'observer l'enve-

loppe de ces mêmes complexes. Cette étude par MET a été réalisée par le Pr. Catherine Vénien-Bryan, à l'aide du microscope électronique à transmission (JEOL 2100 LaB6) de la plateforme IMPMC à Sorbonne Université, Les résultats de cette étude sont présentés dans l'article 2, ci-dessous.

3.2 Présentation de l'étude

Pour répondre à la question ouverte "comment *TPP1* améliore-t-elle l'affinité de *POT1* pour son substrat ?", j'ai réalisé un travail de recherche qui comprend trois volets.

1 : Une étude de la cinétique d'association et de dissociation des complexes nucléoprotéiques *POT1*-ADN et *POT1-TPP1*-ADN (par résonance plasmonique de surface), qui nous a permis de déterminer les origines cinétiques de la meilleure affinité du complexe *POT1-TPP1* pour le site de binding par rapport à *POT1* toute seule.

2 : Une étude de photo-pontage du complexe *POT1-TPP1*-ADN, qui nous a permis de mettre en évidence des contacts entre *TPP1* et ADN au sein du complexe.

3 : Une étude structurale de *POT1-TPP1* seule et en complexe avec l'ADN (par microscopie électronique), qui nous a permis de proposer un modèle de structure donnant une vision plus unitaire des structures des deux complexes homologues *POT1-TPP1* et *TEBP*α-*TEBP*β.

3.2.1 *TPP1* améliore l'affinité de *POT1* en modulant les étapes d'association et de dissociation.

Nos résultats de *SPR* montrent qu'en présence de la séquence H12, 5'-GGTTAGGGTTAG-3', c'est-à-dire dans les conditions favorables de fixation de *POT1*, à savoir la séquence TTAGGGTTAG à l'extrémité 3', le complexe *POT1-TPP1* a une meilleur affinité pour l'ADN que *POT1* toute seule. En comparant les constantes cinétiques k_a et k_d, on observe que la présence de *TPP1* augmente d'un facteur $\simeq$20 la constante d'association mais également, et de façon plus surprenant, d'un facteur $\simeq$4 la constante de dissociation. Ce résultat indique que l'augmentation de l'affinité de *POT1* pour son substrat est essentiellement due à une augmentation de la vitesse de liaison de *POT1* avec son substrat. Toutefois, le fait, qu'en présence de *TPP1*, le temps de résidence de *POT1* sur l'ADN diminue, indique que le complexe *POT1-TPP1* est plus dynamique que *POT1* seule. Ainsi, le temps de résidence plus long de *POT1* sur l'ADN peut protéger les télomères et inhiber l'accès à la télomérase,

tandis qu'en présence de *TPP1*, la dynamique de liaison du complexe peut participer à la processivité de la télomérase. En revanche le fait que *TPP1* impacte à la fois les étapes d'association et de dissociation peut questionner sur le fait que nous ne soyons pas dans le cas d'un effet allostérique.

3.2.2 *TPP1* : Un contact avec l'ADN télomérique

Le fait que l'étape d'association du complexe *POT1-TPP1* soit favorisée en présence de *TPP1*, suggère que *TPP1* pourrait faire un contact avec l'ADN favorable à la liaison du complexe. Cela signifierait donc que *TPP1* se situe à proximité de l'ADN télomérique. Pour répondre à cette question nous avons décidé de sonder l'environnement de notre ADN à l'aide d'expériences de photopontage. Pour cela, nous avons remplacé, sur une séquence télomérique 21 mer, en diverse positions (aux extrémités et position central) nos thymines par des 4-thioThymines (^{4s}T) (Figure 3.5). Les expériences réalisées en présence du complexe *POT1-TPP1* révèlent des bandes retardées (Figure 3.5.A) qui correspondent à des complexes covalents ADN-protéine, comme l'indique le résultat après traitement des échantillons irradiés à la protéinase K (Figure 3.5.B) ; la bande la plus retardée correspondant à un complexe covalent *POT1*-ADN et celle moins retardée à un complexe covalent *TPP1*-ADN. L'article montre qu'aucune bande n'est visualisé avec *TPP1* seul (avec l'ADN possédant la 4-thiothymine en position centrale) confortant le fait que *TPP1* seul n'interagit pas avec l'ADN télomérique, alors que *POT1* seule est bien pontée. En conclusion, nous montrons ici que *TPP1* est en contact avec l'ADN dans le complexe ternaire *POT1-TPP1*-ADN. Nous parlons de contact car la portée de notre sonde, la 4-thioThymine, est de quelques Å. Les résultats de photopontage ne sont pas sans rappeler les résultats obtenus chez le cilié *O. nova*, c'est-à-dire l'absence de contact avec l'ADN en présence de *TEBP*β seul mais une interaction de la sous-unité β avec l'ADN lorsqu'elle est en complexe avec *TEBP*α.

Dans le but de caractériser l'acide aminé ou les acides aminés de *TPP1* photoponté(s) avec l'ADN télomérique nous avons envisagé des expériences de MS/MS. Le faible rendement de photopontage (environ 10% d'ADN ponté) nous oblige à envisager un protocole d'enrichissement des complexes covalents, afin de s'affranchir de l'excès de protéine libre. Nous avons donc décidé de travailler avec des oligonucléotides 4-thioThymine biotinylés en 5'. Les grandes lignes du protocole expérimental sont décrites ci-dessous. Je n'ai malheureusement pas eu le temps pendant ma thèse d'optimiser cette expérience. Cette expérience fait partie des perspectives de mon travail de thèse.

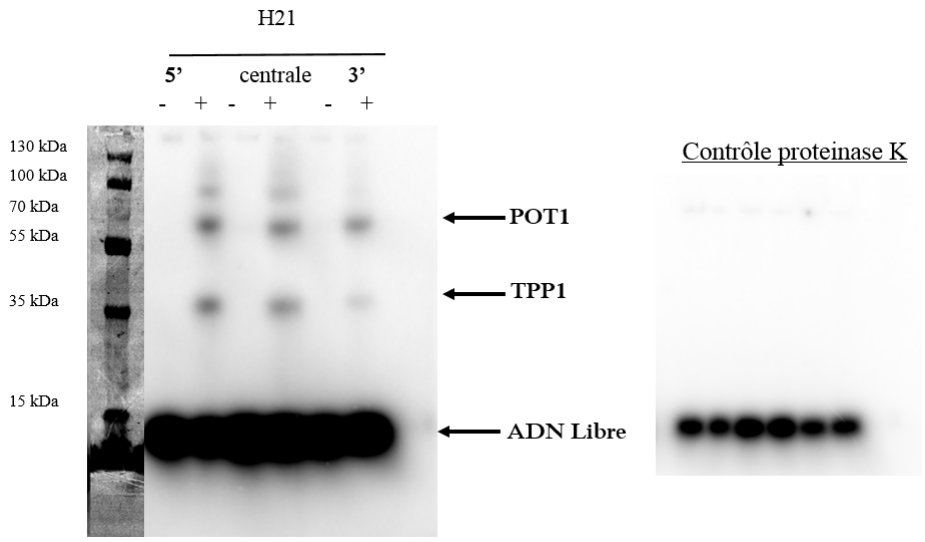


FIGURE 3.5 – **Expérience de photopontage du complexe *POT1-TPP1* à différentes positions le long que l'ADN.** *EMSA* dénaturant en présence de 300 nM de *POT1-TPP1* et 30 nM d'ADN. Les séquences utilisées correspond à différentes positions de la 4-thioT : la position 5' correspond à 5'-^{4s}**T**GGG(TTAGGG)$_3$-3', la position centrale correspond à 5'-GGGTTAGGG^{4s}**T**TAGGGTTAGGG-3' et enfin la position 3' correspond à 5'-GGG(TTAGGG)$_3$^{4s}**T**T-3'.

3.2.3 Le complexe *POT1-TPP1* adopte une forme resserrée en présence d'ADN : une structure conservée au sein du vivant ?

En parallèle des expériences de photopontage, nous avons entrepris des expériences de microscopie électronique afin de déterminer la structure du complexe *POT1-TPP1* en présence et absence d'ADN. Une classification des images 2D a été réalisée. L'analyse montre que le complexe *POT1-TPP1* adopte une forme allongée, semblable à une forme en v ou en fer à cheval, flexible en absence d'ADN. Cette forme est similaire à celle observée par la technique *SAXS* (Chen et al., 2017). En revanche, en présence d'ADN, la structure est moins flexible et plus compacte, avec un resserrage de la forme fer à cheval. Ce résultat pourrait suggérer un repliement de *TPP1* sur l'ADN, comme cela a été observé chez le cilié *Oxytricha nova* au sein de la structure en forme de cœur du complexe *TEBP*α-*TEBP*β. Cette étude de microscopie électronique ne permet pas d'avoir une résolution suffisante pour conclure à une similarité de structure entre *TEBP*α-*TEBP*β et *POT1-*

TPP1. Cependant elle est la première étape préliminaire pour la résolution d'une structure de *POT1-TPP1*. A ce jour aucune structure haute résolution du complexe *POT1-TPP1* n'existe.

3.3 Article 2

3.3.1 Résumé

Les télomères humains sont des complexes nucléoprotéiques aux extrémités des chromosomes eucaryotes qui protègent les chromosomes de la dégradation et de la fusion de leurs extrémités. Parmi les protéines télomériques, la protéine de liaison simple brin *POT1* (*(Protection of Telomeres*) se lie à l'ADN simple brin télomérique (ADNsb) et est associée à son partenaire *TPP1*. Plusieurs études ont montré que *TPP1* améliore l'affinité de *POT1* pour son substrat, mais aucune analyse approfondie n'a été entrepris pour comprendre le mécanisme sous-jacent de ce phénomène. Ici, nous avons réalisé une analyse biochimique et biophysique pour explorer le mode de liaison et la conformation du complexe ternaire *POT1-TPP1*-ADN. Nos données montrent que *TPP1* accélère l'étape de liaison *POT1* et réduit le temps de résidence de *POT1* sur l'ADN, indiquant que *TPP1* augmente la dynamique de liaison de *POT1*. De plus, *TPP1* contacte l'ADN dans un fer à cheval contracté comme la conformation adoptée par le complexe ternaire *POT1-TPP1*-DNA.

Human TPP1 contacts DNA into a contracted horse shoe like conformation adopted by the ternary POT1-TPP1-DNA complex

Jean Chatain[a], Emmanuelle Delagoutte[a], Tahar Bouceba[b], Catherine Vénien-Bryan[c], Patrizia Alberti[a], and Carole Saintomé*[a,d]

[a]Structure et Instabilité des Génomes (StrInG), Muséum national d'histoire naturelle, CNRS, Inserm ; CP26, 57 rue Cuvier 75005 Paris, France.
[b]Sorbonne Université - Campus Pierre et Marie Curie, Institut de Biologie Paris Seine (IBPS) Plateforme d'interactions moléculaires, CNRS-FR3631; 7, Quai Saint Bernard, F-75252, Paris Cedex 05, France.
[c]Institut de Minéralogie, Physique des Matériaux et Cosmochimie, UMR 7590 CNRS-Sorbonne Université-MNHN-IRD, 4 Place Jussieu, 75252, Paris, Cédex 05, France.
[d]Sorbonne Université, UFR927, F-75005, Paris, France.

October 3, 2020

Abstract

Human telomeres are nucleoprotein complexes at eukaryotic chromosomes ends that protect chromosomes from degradation and end-to-end fusion. Among telomeric proteins, the single-stranded binding protein POT1 (Protection of Telomeres) binds telomeric single-stranded DNA (ssDNA) and is associated with its partner TPP1. Several studies have shown that TPP1 enhances the affinity of POT1 for its substrate but no thoroughly analysis has been done to understand the mechanism that underlies this phenomenon. Here we performed biochemical and biophysical analysis to explore the binding mode and the conformation of the ternary POT1-TPP1-DNA complex. Our data show that TPP1 accelerates the POT1 binding step and reduces the residence time of POT1 on DNA, indicating that TPP1 increases the binding dynamics of POT1. In addition, TPP1 contacts the DNA into a contracted horse shoe-like conformation adopted by the ternary POT1-TPP1-DNA complex.

*Corresponding author:
carole.saintome@mnhn.fr

1 Introduction

Human telomeres are composed of double-stranded DNA region consisting of a repetitive hexameric sequence of TTAGGG that ends in a single-stranded G-rich DNA region called the G-overhang. This later region is specifically bound by the human telomeric POT1 protein (Protection of telomeres 1)[3,4]. Human POT1 (hPOT1) protects telomeres[10,7,2] and regulates telomeres elongation by telomerase, a ribonucleoprotein[6,13,11,26]. hPOT1 forms a heterodimer with the human telomeric protein TPP1 (hTPP1)[14,24], which recruits the telomerase and acts as a telomerase processivity factor[1,24,23,12,16,27]. hPOT1 binds specifically the human telomeric TTAGGG motif[15] thanks to the interaction of its two OB-fold domains, OB1 and OB2 (Figure 1). The crystal structure of the N-terminal half of hPOT1 bound to a telomeric decamer TTAGGGTTAG shows that OB1 binds the first six nts (TTAGGG) whereas the OB2 binds to the adjacent four nts (TTAG)[13]. hPOT1 binds DNA with a nanomolar affinity and a 3' end preference[15,23,24,17]. It has been shown that the affinity of POT1 for its substrate is increased in the presence of its protein partner TPP1[23,24,22,16,21], however the underlying mechanism remains unclear. Here, we studied interactions of hPOT1 alone and hPOT1-TPP1 complex with telomeric DNA in a kinetical and structural point of view.

2 Resultats & Discussion

We first performed Surface Plasmon resonance (SPR) assays to provide information both about the hPOT1 and hPOT1-TPP1 binding kinetics and affinities[18]. For this purpose, the $(T)_5$GGTTAGGGTTAG oligonucleotide T_5H12, containing the core telomeric sequence 5' TTAGGGTTAG 3', biotinylated at the 5' end was immobilized on a streptavidin-coated sensor chip.

The protein was injected at a range of concentrations and the resultant SPR response is showed on Figure 2A-B. Data show a specific concentration-dependent protein-DNA interaction. Curves were fitted with two different models (Langmuir binding 1:1 and steady state affinity) and we obtain the same value of KD. Fitting kinetic curves give an apparent KD value of 8,26 nM for hPOT1 and 1,7 nM for hPOT1-TPP1 complex (Figure 22C-D). These data correlates with the nanomolar affinity of POT1 and confirm that TPP1 enhances the affinity of POT1 for its substrate[23,24,22,16,21]. Determination of the kinetical parameters (k_{on} and k_{off}) (Figure 3), from fitting of binding curves obtained with the lowest protein concentrations (Figure 2 2E-F), indicates that POT1 is 20 times quicker to bind its substrate in the presence of TPP1 but its residence time is around 3 times lower. Here we show that TPP1, on one side improves the binding step of POT1 but, in the other side, it reduces the residence time of POT1 on DNA, indicating that TPP1 increases the binding dynamics of POT1. Thus, the longer residence time of POT1 on DNA may protect telomeres and inhibit telomerase access, while in the presence of TPP1, the binding dynamics of the complex may participate to the processivity of telomerase.

Although TPP1 also contains an N-terminal OB fold, no binding has never been observed between DNA and TPP1 alone[24] suggesting that TPP1 does not contribute to

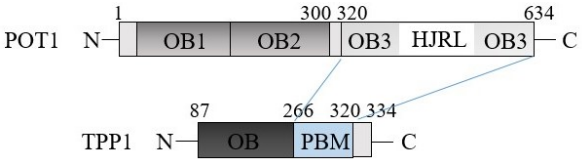


Figure 1: **Structure of the human POT1 and TPP1 proteins.**OB1 and OB2 of POT1 are colored in dark grey while OB3 and the embedded HJRL domain are colored in grey and white, respectively. TPP1 contains a OB domain (in black grey) and the PBM domain (in blue) for POT1 interaction ; the TBM domain for TIN2 interaction is not present.

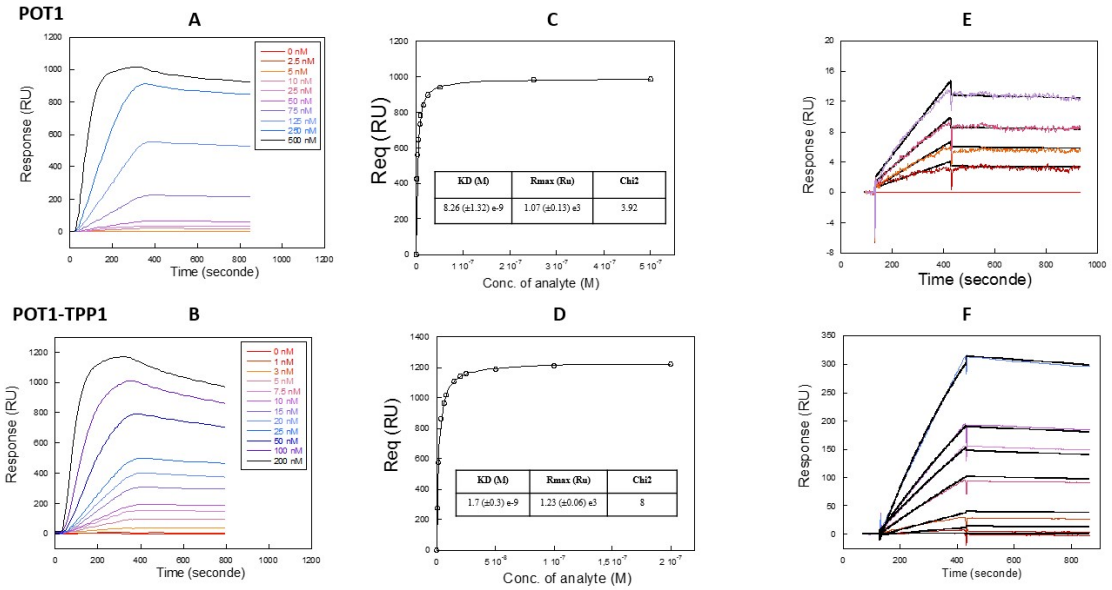


Figure 2: **Kinetic Measurement** kinetic curves of POT1 (A) or POT1-TPP1 (B) with T5-H12 were fitted with 1:1 langmuir binding model. Curves of responses at equilibrium as function of concentration of POT1 (C) and POT1-TPP1 (D) were fitted by a general fit using steady state affinity. Kinetic curve at low concentration of POT1 (E) or POT1-TPP1 (F)

Protein	k_a ($M^{-1}S^{-1}$)	k_d (s^{-1})	Rmax (Ru)	K_D(nM)	Chi_2
POT1	4.09 (±1.3) 10^3	3.97(±1.3) 10^{-5}	1030 (±15)	9.47(±0.4)	3.6
POT1-TPP1	8.05 (±2.01) 10^4	1.32 (±0.95) 10^{-4}	1260 (±23)	1.63 (± 0.23)	23.5

Figure 3: **kinetical parameters ($\mathbf{k}_{on}$ and $\mathbf{k}_{off}$)** from fitting of binding curves

increase affinity by additional contact with DNA. However, two recent studies, using hydrogen-deuterium exchange (H/DX) combined with mass spectrometry and hydroxyl radical footprinting assays, to probe DNA binding-induced conformational changes in POT1-TPP1 complex, have shown conformational changes in POT1 but also in TPP1[19,25]. We wondered if such change in TPP1 could result from a direct contact of TPP1 with DNA, that could contribute to the high improvement of the binding step of the complex, observed here. To reveal such contact, we used photolabeling with 4-thiothymine, which has been revealed as a very efficient photocrosslinkable agent to study nucleic acid-protein and nucleic acid-nucleic acid interactions[8,20]. In addition, its zero-length feature implies that only protein in direct interaction with the probe can be crosslinked. Thus, we used a 21 mer telomeric sequence analogue H21S containing a single photo-activatable 4-thiothymine residue ($^{S4}\mathbf{T}$) inside the hPOT1 binding motif (GGGTTAGGG^{S4}**T**TAGGGTTAGGG). Upon UV irradiation of the protein-DNA complex, the S4T residue can form a covalent bond linking the DNA with the protein in contact. Photocrosslinked complexes obtained in the presence of 300 nM of protein and 30 nM of H21-S, were resolved by SDS-PAGE(Figure 4). In the presence of POT1 alone (lane 1), we observed a retarded band, with an apparent molecular mass consistent with DNA cross-linking to POT1. This band is also present when the complex POT1-TPP1 was used (lane 4), but interestingly, a supplementary minor retarded band is observed with an apparent molecular mass consistent with DNA cross-linking to TPP1. This crosslinked species is only observed when TPP1 is associated with POT1 as no crosslink is obtained when TPP1 is added alone (lane 3), while it is present when we extemporaneously mixed both proteins before UV irradiation (lane 2). Thus, this data shows that TPP1 contacts DNA when it is associated with POT1 while it doesn't alone. Interestingly, this feature has been also observed with the β subunit of *Oxytricha* telomere-binding protein telomeric (TEBPβ) that form a heterodimer with the a subunit of Oxytricha telomere-binding protein telomeric (TEBPα)[9]. The formation of both crosslinks with POT1 or TPP1, in the POT1-TPP1-DNA complex, indicates that TPP1 is close to the DNA binding domain of POT1 (OB1 or OB2) suggesting a compact structure of the ternary POT1-TPP1-DNA complex.

Thus, we analyzed the shape of the ternary complex with H12 (GGTTAGGGTTAG) using electron microscopy. The observation of the complex POT1-TPP1 without DNA

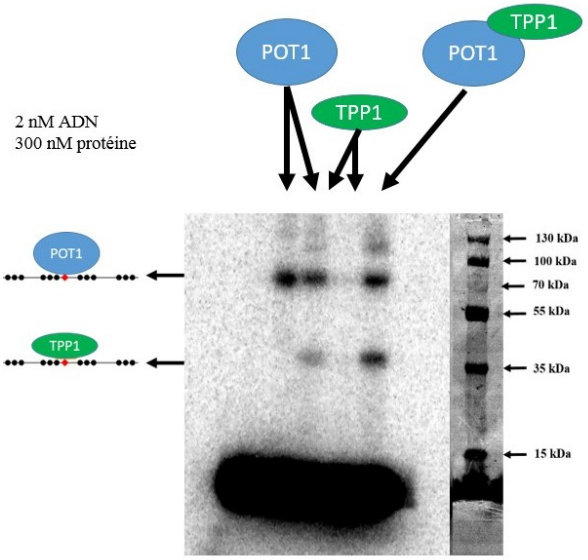


Figure 4: **Photocrosslinking experiment.** ^{32}P-labelled H21-S (30 nM) were mixed with 300 nM of protein then irradiated at 365 nm at 15°C for 45 min. SDS-PAGE (4-16%) were used to separate multiple crosslinked species (arrows) and free DNA. ^{32}P-labelled H21-S and covalent 32P-H21S-protein complexes were detected by phosphorimaging detection ; molecular weight size markers were detected by fluorescence detection.

under the electron microscope shows a V-shape conformation with a certain flexibility (Figure 5A). The characteristics structures are circled. POT1-TPP1-H12 presents some similarity with the complex without DNA, but with a tighter conformation (see circled characteristic structures on the figure 5B). In order to gain more structural information from the MET images, we performed a 2D classification. This step of classification allows to group similar images into classes, each class being an averaged 2D images of particles in the same orientation on the grid or with the same structural conformation.

2D classification of the POT1-TPP1 complex alone shows mostly the characteristic V-shape structure similar to the structure described using SAXS experiments[5]. Some classes show one arm longer than the other, particularly classes 1, 4 and 6 where the length is about 65Å for the shorter arm and 82 Å for the longer arm (Figure 5). This complex shows certain flexibility because some classes present a more linear structure with a V shape less visible (see classes 10 and 11 on Figure 5C). The narrow dimensions of the arms are from 32Å to 16Å which could correspond to the size of some protein domain and possible linker between these domains respectively. These dimensions are comparable with the data described[5].

2D classification of the POT-TPP1-H12 complex shows some similarity with the POT1-TPP1 alone with the characteristic V-shape or horse shoe-like conformation (Figure 5D, classe 3)[5]. Interestingly, we note a more compact structure where the two arms seem closer to each other(Figure 5D, classes 2, 5 and 6)as if the arms of the V-shape tighten around the DNA in a horse shoe-like conformation. We did not observe the linear shape seen with POT1-TPP1 alone(Figure 5C, classes 10 and 11)), indicating a less flexible structure. These structural data could explain the crosslinking obtained with TPP1. Thus, considering the whole set of structural data we suggest that POT1 and TPP1 associate with each other to form a horse-shoe like conformation, that looks like the heart-shape structure observed for the TEBPα-TEBPβ-ssDNA complex[?]

3 Materials & Methods

3.1 Productions and Purifications of hPOT1, hTPP1 and hPOT1/hTPP1 complex

The hPOT1, hTPP1, and hPOT1/TPP1 complex were co-expressed in the baculovirus insect cell system (pFastBac HT A) at the IGBMC. The full hPOT1 (residues 1-634) was expressed with 6 histidines in N-term and a truncated version of hTPP1 (87-334) with a GST-tag in N-term. Each protein contains a restriction site 3C that allows removing tags. The cell pellet was resuspended in lysis buffer (50 mM Tris HCl pH 7.8, 300 mM NaCl, 2 mM DTT, 1 mM PMSF, 2 mM 2-Mercaptoethanol) supplemented with a cocktail of protease inhibitors (Roche), stirred on ice for 1h30. The cells were then lysed by sonication (60% Ampli, 1sec ON – 2sec OFF for 20 min) and the cell debris was removed by ultracentrifugation (17 000 rpm for 1h15) at 4°C. The supernatant (for hPOT1 and hPOT1/hTPP1 complex) was loaded on a 5 mL Ni-His Select column pre-equilibrated in lysis buffer. The column was next washed with 25 mL Ni-Wash buffer (50 mM Tris HCl pH 7.8, 300

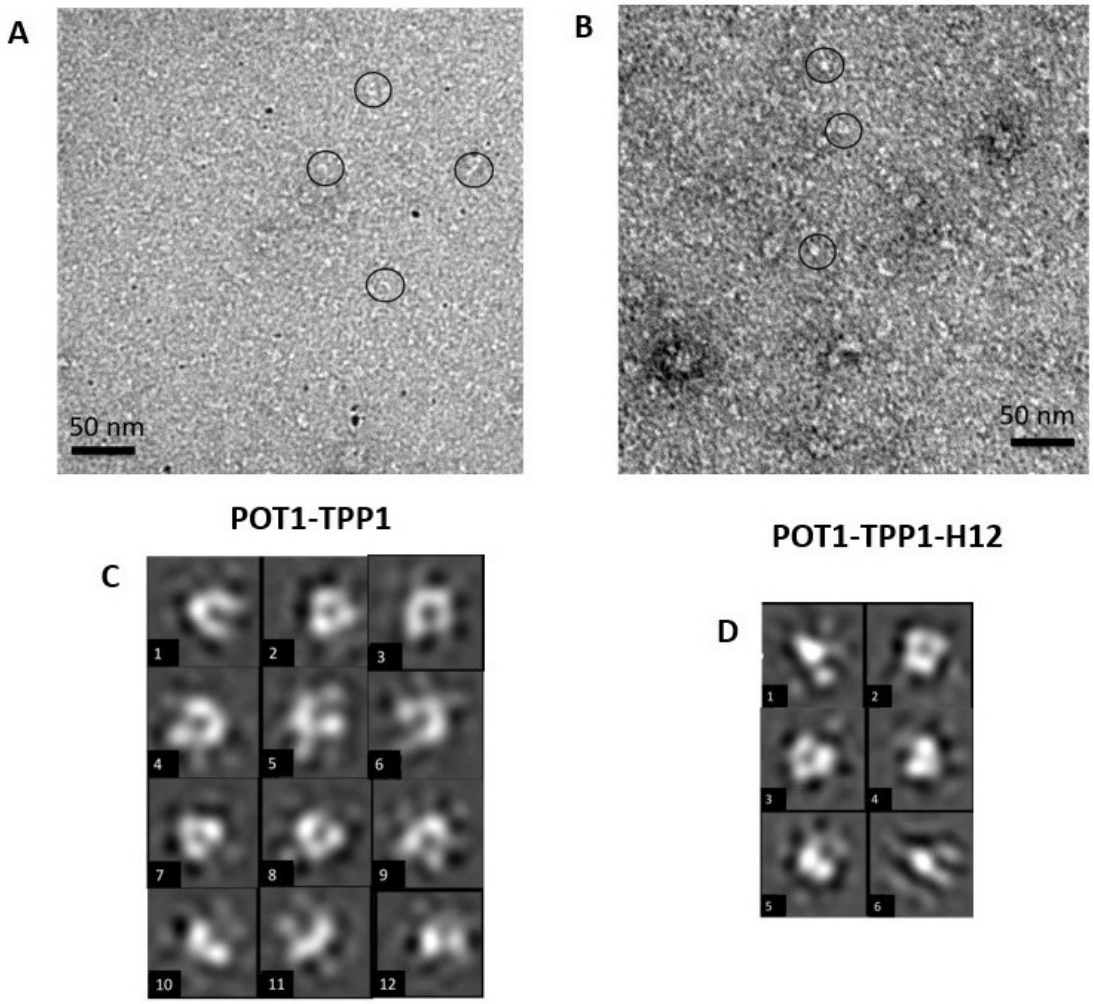


Figure 5: **Electron micrograph of negatively stained specimen.** The protein is bright against a darker background. (a): POT1-TPP1 Characteristics shapes (mostly horseshoe like and sometimes extended shapes) are circled. (b): POT1-TPP1 complexed with DNA (H12) shows some difference compared with POT1-TPP1 without DNA. The complex appears less extended. Characteristics shapes are circled Bar scale is 50 nm. 2D classification of negatively stained POT1-TPP1 (c) and POT1-TPP1-H12 (d). The box size is 46 pixels2 (198Å x 198Å)

mM NaCl, 2m DTT, 10 mM Imidazole). The protein was eluted with 25 mL of Ni-Elution buffer (50 mM Tris HCl pH 7.8, 300 mM NaCl, 2m DTT, 300 mM Imidazole). The eluted fractions (for hPOT1/hTPP1 complex) or supernatant for hTPP1 were loaded on a 5 mL GST column, pre-equilibrated with Ni-Elution Buffer. After a washing step with GST-Wash Buffer (50 mM Tris HCl pH 7.8, 75 mM NaCl, 1mM DTT), the hPOT1/TPP1 complex and hTPP1 were eluted with GST-Elution Buffer (50 mM Tris HCl pH 7.8, 300 mM NaCl, 2m DTT, and 15 mM GSH). Tags were removed overnight with the PreScission protease at 4°C in Low Salt Buffer (50 mM Tris HCl pH 7.8, 150 mM NaCl, 1mM DTT) (3 mL or 1500 Units). Finally, the digested sample was loaded on a 1 mL MonoQ column, pre-equilibrated with Low Salt Buffer. The column was next washed with 5 mL of Low Salt Buffer and the hPOT1, hTPP1 and hPOT1/TPP1 complex were eluted with a 25 mL gradient of NaCl (from 150 mM to 300 mM). Fractions containing proteins were pooled and stored at -80°C.

3.2 Sequences

Oligonucleotides were purchased from Eurogentec.

Names	Sequences
H12	$G_2T_2AG_3T_2AG$
Biotin-T_5-H_{12}	Biotin-$T_5G_2T_2AG_3T_2AG$
H21-S	$G_3T_2AG_3{}^{S4}\mathbf{T}TAG_3T_2AG_3$

3.3 Surface SPR

The surface plasmon resonance (SPR) biosensor experiments were performed on a Biacore 3000 instrument (Cityva Lifesciences) of the Platform of Molecular Interactions of the IBPS Institute (Sorbonne University). A carboxymethylated dextran surface called CM5 sensor chip was used. CM5 chip was first activated according to the recommendations of the Biacore amine coupling kit by injecting 70 μL of a mixture (1:1, v:v) of 100 mM NHS (N-hydroxysuccinimide) and 400 mM EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide). Then, streptavidin was covalently immobilized through primary amino groups to the carboxymethyl dextran matrix of CM5 sensor chip. Streptavidin (50μg/ml) diluted in immobilization buffer (10 mM sodium acetate, pH 4.5) was injected at a flow rate of 10 μl/min for 7 minutes in order to obtain 3900-4000 resonance units (RU). Streptavidin immobilization was followed by an injection of 70 μl of ethanolamine hydrochloride (1 M) pH 8.5, at a flow rate of 10 μl/min, to saturate the free activated sites of the matrix. Buffer HBSEP (10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% surfactant P20) was used as the running buffer. Streptavidin surface was preconditioned by three injections of 1M NaCl 50mM NaOH at a flow rate of 5μl/min, then, the biotinylated oligonucleotide T5H12 was immobilized on streptavidine, by injection of 3 μl at a flow rate of 5μl/min on the Flow Cell 2, folwed by a final wash with 1M Nacl 50 mM NaOH (5μl/min, contact time 1 min). A reference surface without oligonucleotide was prepared using the same procedure to appreciate blank surface signal.

The variations of SPR signal as a function of time (sensorgram) were obtained by passing various concentrations of the injected proteins over the oligonucleotide surface at a flow rate of 5 μl/min, with an association

phase of 5 min and a dissociation phase of 8 min. The surface of the sensor was regenerated with an injection of 0.05% SDS at a flow rate of 30 μl/min for 30 seconds. When necessary, the regeneration was carried out by a second injection of 0.05% SDS under the same conditions. Identical injections over blank surfaces executed in parallel (giving non significat signal) were subtracted from all experiments. The kinetics were evaluated using the BIAevaluation software, version 4.1 (Cityva Lifesciences). The data were processed by fitting the binding profiles to a 1:1 Langmuir interaction model and a steady state affinity model. The fitting of each data set yielded rates for association (k_a or k_{on}) and dissociation (k_d or k_{off}), from which the dissociation constant KD was calculated (KD = k_{off}/k_{on}).

3.4 Photocrosslinking experiment

Proteins (300 nM) was added to the labeled oligodeoxynucleotide (30 nM) in reaction buffer (50 mM HEPES pH 7.9, 0.1 mg/mL BSA, 100 mM KCl or NaCl and 2% glycerol). Samples were incubated 30 min at room temperature room and then irradiated for 45 min at 4°C with a 365 nm UV lamp (Vilber Lourmat, VL-6.L, 6W). Control samples were incubated at 4°C but not irradiated. Furthermore, an aliquot of irradiated sample was incubated with proteinase K at 37°C for 1h. All samples were heated at 90°C for 10 min, then loaded on a denaturing 4-16% acrylamide gel (iD Page Gel by Eurogentec) in 1X MOPs Buffer. Electrophoresis was performed for 60 min at room temperature to separate crosslinked species.

3.5 Electron Microscopy and image analysis

Negatively stained grids were prepared applying small aliquots (3.5 mL) of POT1-TPP1 and POT-TPP1-H12 diluted in a complexation Buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM DTT, 2% glycerol) (about 0.025 mg/mL) to glow-discharged, carbon-coated grids and allowed to absorb for 60 seconds. The sample was then stained with 2% uranyl acetate. For POT1-TPP1-H12 the complex was extemporaneously mixed with the H12 at a ratio POT1-TPP1:H12 of 0,1. Specimen were imaged at a nominal magnification of 25,000x with a 2 k x 2 k Gatan CCD camera (corresponding to a pixel size of 4.31 Å at the specimen level) in a JEOL 2100, LaB6 operating at 200 kV.

Image analysis: for POT-TPP1 specimen, a total number of 3,810 particles were selected manually from 66 micrographs with a box size of 46 pixels. Defocus was determined using CTFTILT. 2D classification was realized with EMAN2.21. For POT-TPP1-H12 specimen, a total number of 2,542 particles were manually picked from 17 images. 2D classification was realized following the same procedure as above.

References

[1] Abreu, E., Aritonovska, E., Reichenbach, P., Cristofari, G., Culp, B., Terns, R. M., Lingner, J., and Terns, M. P. (2010). TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. *Molecular and Cellular Biology*, 30(12):2971–2982.

[2] Barrientos, K. S., Kendellen, M. F., Freibaum, B. D., Armbruster, B. N., Etheridge, K. T., and Counter, C. M. (2008). Distinct functions of POT1 at telomeres. *Molecular and Cellular Biology*, 28(17):5251–5264.

[3] Baumann, P. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*, 292(5519):1171–1175.

[4] Baumann, P., Podell, E., and Cech, T. R. (2002). Human pot1 (protection of telomeres) protein: Cytolocalization, gene structure, and alternative splicing. *Molecular and Cellular Biology*, 22(22):8079–8087.

[5] Chen, C., Gu, P., Wu, J., Chen, X., Niu, S., Sun, H., Wu, L., Li, N., Peng, J., Shi, S., Fan, C., Huang, M., Wong, C. C. L., Gong, Q., Kumar-Sinha, C., Zhang, R., Pusztai, L., Rai, R., Chang, S., and Lei, M. (2017). Structural insights into POT1-TPP1 interaction and POT1 c-terminal mutations in human cancer. *Nature Communications*, 8(1).

[6] Colgin, L. M., Baran, K., Baumann, P., Cech, T. R., and Reddel, R. R. (2003). Human POT1 facilitates telomere elongation by telomerase. *Current Biology*, 13(11):942–946.

[7] Denchi, E. L. and de Lange, T. (2007). Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature*, 448(7157):1068–1071.

[8] Favre, A., Saintomé, C., Fourrey, J.-L., Clivio, P., and Laugâa, P. (1998). Thionucleobases as intrinsic photoaffinity probes of nucleic acid structure and nucleic acid-protein interactions. *Journal of Photochemistry and Photobiology B: Biology*, 42(2):109–124.

[9] Gray, J. T., Celander, D. W., Price, C. M., and Cech, T. R. (1991). Cloning and expression of genes for the oxytricha telomere-binding protein: Specific subunit interactions in the telomeric complex. *Cell*, 67(4):807–814.

[10] Hockemeyer, D., Sfeir, A. J., Shay, J. W., Wright, W. E., and de Lange, T. (2005). POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end. *The EMBO Journal*, 24(14):2667–2678.

[11] Kelleher, C., Kurth, I., and Lingner, J. (2005). Human protection of telomeres 1 (POT1) is a negative regulator of telomerase activity in vitro. *Molecular and Cellular Biology*, 25(2):808–818.

[12] Latrick, C. M. and Cech, T. R. (2010). POT1–TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *The EMBO Journal*, 29(5):924–933.

[13] Lei, M., Podell, E. R., and Cech, T. R. (2004). Structure of human POT1 bound

to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature Structural & Molecular Biology*, 11(12):1223–1229.

[14] Liu, J., Allspach, J. R., Feigenbaum, M., Oh, H.-J., and Burton, N. (2004). A STUDY OF FATIGUE EFFECTS FROM THE NEW SAT®. *ETS Research Report Series*, 2004(2):i–13.

[15] Loayza, D. and de Lange, T. (2004). Telomerase regulation at the telomere. *Cell*, 117(3):279–280.

[16] Nandakumar, J., Bell, C. F., Weidenfeld, I., Zaug, A. J., Leinwand, L. A., and Cech, T. R. (2012). The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature*, 492(7428):285–289.

[17] Nandakumar, J., Podell, E. R., and Cech, T. R. (2009). How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA. *Proceedings of the National Academy of Sciences*, 107(2):651–656.

[18] Patching, S. G. (2014). Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838(1):43–55.

[19] Rajavel, M., Orban, T., Xu, M., Hernandez-Sanchez, W., de la Fuente, M., Palczewski, K., and Taylor, D. J. (2016). Dynamic peptides of human TPP1 fulfill diverse functions in telomere maintenance. *Nucleic Acids Research*, page gkw846.

[20] Rechkunova, N. I. and Lavrik, O. I. (2020). Photoreactive DNA as a tool to study replication protein a functioning in DNA replication and repair. *Photochemistry and Photobiology*, 96(2):440–449.

[21] Rice, C., Doukov, T., and Skordalakes, E. (2017). Structure of the human POT1-TPP1 telomeric complex.

[22] Taylor, D. J., Podell, E. R., Taatjes, D. J., and Cech, T. R. (2011). Multiple POT1–TPP1 proteins coat and compact long telomeric single-stranded DNA. *Journal of Molecular Biology*, 410(1):10–17.

[23] Wang, F., Podell, E. R., Zaug, A. J., Yang, Y., Baciu, P., Cech, T. R., and Lei, M. (2007). The POT1–TPP1 telomere complex is a telomerase processivity factor. *Nature*, 445(7127):506–510.

[24] Xin, H., Liu, D., Wan, M., Safari, A., Kim, H., Sun, W., O'Connor, M. S., and Songyang, Z. (2007). TPP1 is a homologue of ciliate TEBPbeta and interacts with POT1 to recruit telomerase. *Nature*, 445(7127):559–562.

[25] Xu, M., Kiselar, J., Whited, T. L., Hernandez-Sanchez, W., and Taylor, D. J. (2019). POT1-TPP1 differentially regulates telomerase via POT1 his266 and as a function of single-stranded telomere DNA length. *Proceedings of the National Academy of Sciences*, 116(47):23527–23533.

[26] Zaug, A. J., Podell, E. R., and Cech, T. R. (2005). Human POT1 disrupts telomeric g-quadruplexes allowing telomerase extension in vitro. *Proceedings*

of the National Academy of Sciences, 102(31):10864–10869.

[27] Zhong, F. L., Batista, L. F., Freund, A., Pech, M. F., Venteicher, A. S., and Artandi, S. E. (2012). TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell*, 150(3):481–494.

3.4 Conclusion et perspectives

Les résultats de cette étude m'ont permis d'élucider les origines cinétiques de la meilleure affinité du complexe *POT1-TPP1* pour le site de fixation par rapport à *POT1* seule. Les expériences de photopontage m'ont permis de remettre en question l'acquis selon lequel *TPP1* n'interagit pas avec l'ADN. Les données de microscopie électronique m'ont permis de mettre en évidence une forme du complexe en fer de cheval plus resserrée par rapport à la structure proposée par l'études *SAXS* (Figure 1.19) qui suggère une structure plus proche de celle résolue pour le complexe homologue *TEBP*α-*TEBP*β des ciliée et donc une vision plus unitaire de ces complexes de protection. En perspective, cette étude mérite d'être poursuivi de façon plus approfondie. Pour cela j'ai déjà tracé des pistes de travail.

3.4.1 Spectroscopie de masse

Dans le but de caractériser le domaine de *TPP1* ponté, voire l'acide aminé (ou les acides aminés) ponté(s), avec l'ADN télomérique nous avons envisagé d'analyser les complexes covalents par spectrométrie de masse MS ou MS/MS, respectivement. Le faible rendement de photopontage (environ 10% d'ADN ponté) nous à obliger à envisager un protocole d'enrichissement des complexes covalents, afin de s'affranchir de l'excès de protéine libre. Nous avons donc décidé de travailler avec des oligonucléotides 4-thioThymine biotinylés en 5'. Les grandes lignes du protocole expérimental sont décrites ci-dessous. Je n'ai malheureusement pas eu le temps pendant ma thèse d'optimiser cette expérience. Cette expérience fait partie des perspectives de mon travail de thèse.

Protocole MS/MS

La sonde correspond à la séquence contenant une double modification à savoir une biotine en position T_2 et la 4-thioThymine (^{4s}T) en position 9 du motif de liaison de *POT1* : 5'-T$\mathbf{T}_{biot}$TTTGGTTAGGG4$^{4s}\mathbf{T}$TAG-3'. Nous avons choisi de positionner la biotine en interne afin de laisser l'extrémité 5' libre, dans un éventuel marquage radioactif de la sonde. Au préalable les sondes sont incubées sur la nuit à 4°C nos oligonucléotides modifiés en présence de billes magnétiques recouvertes de streptavidine. Puis à l'aide d'aimants nous réalisons différents lavages de sorte à éliminer les sondes qui ne seraient pas accrochés sur les billes magnétiques. Cette préparation est ensuite incubée en présence du complexe *POT1-TPP1*, ou de la protéine *POT1*, pendant 20 min à température ambiante dans un tampon favorable

à la liaison des protéines (50 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM DTT, 10% glycerol, 0.2 mg/mL BSA and 0.1 mM EDTA). Des contrôles négatifs sans protéine sont réalisés en parallèle. L'irradiation à 330 nm des échantillons est réalisée à l'aide d'une lampe UV, pendant 45 min à 15°C, puis incubés pendant 2 heures en présence de trypsine de sorte à réaliser des clivages ciblés des chaines protéiques. L'étape de digestion enzymatique est suivie de différents lavages de sorte à éliminer, d'une part les fragments protéiques non liés à l'ADN, et d'autre part la trypsine. La cristallisation de nos complexes peptide-ADN est une étape nécessaire pour l'étude en spectroscopie de masse en *MALDI-TOF* (*Matrix Assisted Laser Desorption Ionization - Time of Flight*). Or la cristallisation de complexes peptide -ADN n'est pas facile à obtenir. Pour contourner cette difficulté une digestion de l'ADN sur les billes a été effectuée à l'aide de la benzonase. Cette étape conduit alors au décrochage des peptides (lié à la sonde) des billes. L'éluat ainsi récupéré est concentré avant d'être analysé en spectroscopie de masse. L'analyse des spectres de masse des expériences réalisées avec complexe *POT1-TPP1* et la protéine *POT1*, devrait permettre de déterminer la(les) région(s) d'interaction de *TPP1* avec l'ADN mais également, si elles existent, de mettre en évidence des modes de liaison de *POT1* différents, en absence ou présence de *TPP1*. Cette expérience pourrait aussi être adaptée pour caractériser les différents modes de liaison de *POT1* seul en fonction de sa concentration. Il est important de noter que le signal obtenu est un signal positif puisque seules les régions photopontées seront détectées. L'autre avantage de cette méthode est de pouvoir mettre en évidence l'existence de plusieurs points de pontage à l'aide d'une seule expérience.

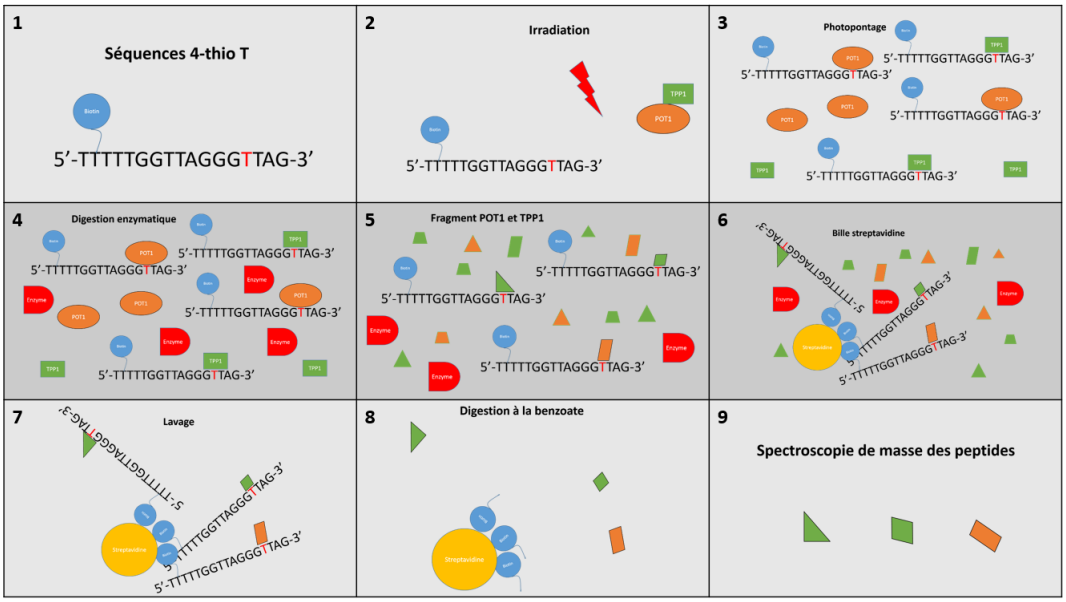


FIGURE 3.6 – **Schéma du protocole MS/MS.**

3.4.2 Cryo-microscopie

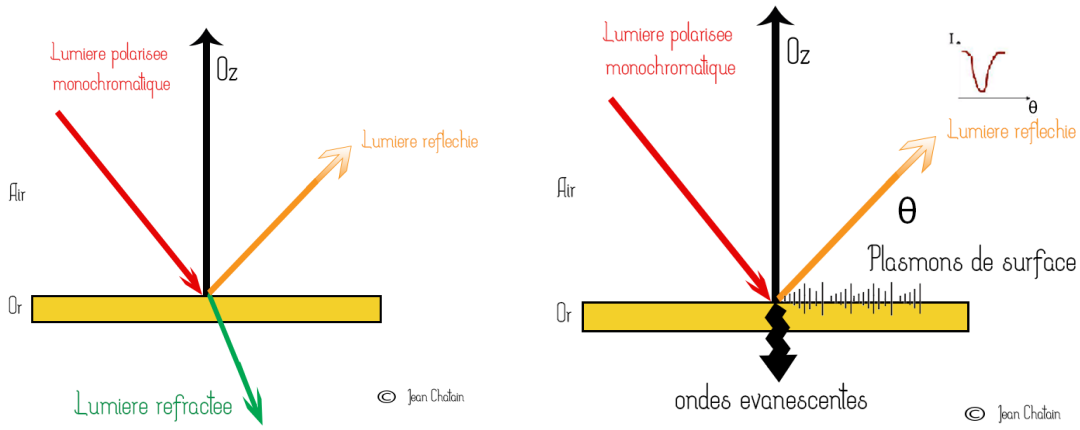
Nos résultats de microscopie éléctronique montrent que le complexe ternaire *POT1-TPP1*-H12 est peu flexible et semble homogène, ce qui nous permet d'envisager, en perspective de cette étude, de déterminer la structure à haute résolution du complexe *POT1-TPP1* en présence d'ADN par cryo-microscopie.

3.5 Annexes méthodes

3.5.1 Résonance des plasmons de surface

La résonance des plasmons de surface, en anglais Surface Plasmon Resonance (**SPR**), décrit un phénomène physique d'interaction lumière-matière. Lorsqu'une radiation polarisée illumine une interface constituée de deux milieux d'indices de réfraction différents, on observe qu'une partie de l'onde lumineuse est réfractée à travers la surface, tandis que l'autre partie est réfléchie sur la surface (Figure 3.7.A). Il est à noter qu'il existe un angle d'incidence pour lequel l'onde lumineuse est entièrement réfléchie. Dans cette situation, une des composantes électromagnétiques de la lumière génère une onde, qui se propage perpendiculairement à l'interface, appelée onde évanescente (Figure 3.7.B). Une onde évanescente a la particularité de décroître de façon exponentielle par rapport à la distance parcourue.

Dans le cas d'un métal, les électrons libres, présents dans la couche de valence, peuvent se déplacer d'environ 50 nm et constituent alors un "gaz" d'électrons, appelé plasma. De fait, une radiation lumineuse incidente, telle que sa longueur d'onde est très supérieure à l'épaisseur de la couche métallique, peut générer une résonance à la surface du métal en interagissant avec les électrons libres Polman and Atwater (2005). Cette résonance qui se traduit par l'oscillation cohérente des électrons délocalisés est appelée résonance des plasmons de surface (Figure 3.7.B). Il est important de noter que l'une des conséquences énergétiques de cette résonance plasmonique est visible dans la radiation réfléchie. En effet, il existe une valeur d'angle θ appelée angle de résonance pour lequel l'intensité de l'onde réfléchie est minimum. Cette valeur de l'angle θ dépend du milieu du champs évanescent, c'est pourquoi la technique de SPR permet de sonder la matière.



A. Lumière réfractée

B. Onde évanescente et résonance de plasmons de surface

FIGURE 3.7 – **Principe de la résonance de plasmon de surface**

La technique SPR est couramment utilisée pour étudier les interactions de biomolécules. En effet, nous avons vu que l'angle θ de résonance des plasmons de surface dépend du milieu du champs évanescent. Le biacore, appareil commercial reposant sur la résonance des plasmons de surface, permet de caractériser des interactions protéine/protéine ou protéine/ligand. Pour ce faire, le Biacore utilise des puces constituées d'une fine couche d'or, riche en électrons libre, recouvert de dextran (Figure 3.8). Le rôle du dextran est de permettre la fixation, d'un des partenaires, qui sera défini comme le ligand, l'autre partenaire (en solution), est appelée analyte et sera injecté dans la puce, afin de rentrer en interaction avec le ligand. Cette interaction entraine une modification du milieu et donc du champs de l'onde évanescence. La valeur de l'angle θ de résonance est par conséquent modifiée, ainsi que l'intensité de l'onde réfléchie en chaque point du détecteur. Ainsi, on peut suivre les phénomènes d'association de l'analyte avec le ligand au cours du temps, en réalisant des injections d'analyte à un flux constant. En modifiant la nature de l'injection (solution dépourvue de l'analyte), on peut alors suivre la dissociation de ce dernier au cours du temps. L'intérêt du flux pendant la dissociation est d'empêcher les analytes qui se sont décrochés du ligand d'interagir de nouveau avec ces derniers. Enfin, on effectue une étape de régénération de la puce en utilisant un détergeant, couramment du SDS, afin d'éliminer tous les analytes encore présents sur la puce. On obtient ainsi un sensorgramme exprimant la variation des signaux SPR, exprimés en unité de résonance RU au cours du temps (Figure 3.9).

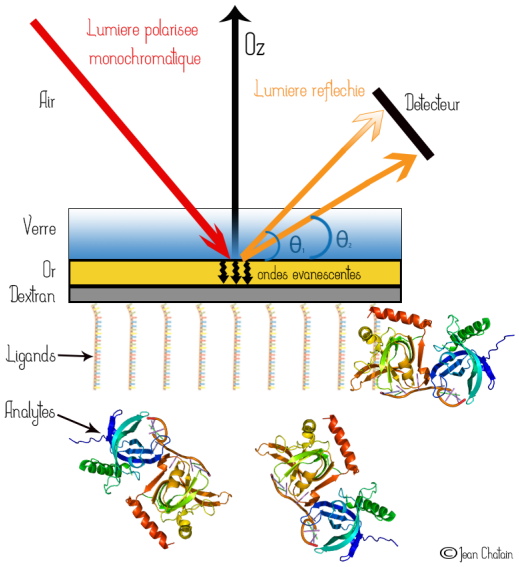


FIGURE 3.8 – **Principe de fonctionnement du Biacore**

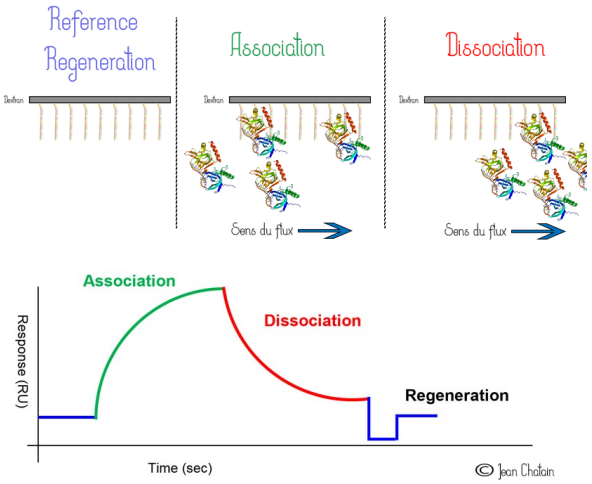


FIGURE 3.9 – **Sensorgramme type issu du Biacore**

Dans le cadre de notre étude, nous avons utilisé le Biacore 3000, de la plateforme "Intéractions moléculaires" de l'IBPS à Sorbonne Université, pour caractériser les interactions protéines/ADN ; plus précisément, les interactions entre la protéine *POT1*, ou le complexe h*POT1-TPP1*, avec l'ADN télomérique. Pour réaliser ces expériences, nous avons au préalable immobilisé sur la surface de dextran de nos puces des streptavidines à l'aide de couplage amino. En effet nos ligands, c'est-à-dire dans notre étude notre sé-

quence d'acide nucléique télomérique simple brin H12, ont été marqués en 5' par une biotine, afin de les immobiliser sur la puce à l'aide de l'interaction Streptavidine/Biotine. Nous avons réalisé des gammes de concentration croissantes d'analytes, c'est à dire dans notre cas de la protéine *POT1* ou du complexe *POT1-TPP1*. Á partir des courbes d'associations et de dissociations, obtenues à différentes concentrations d'analytes j'ai pu extraire les constances cinétiques d'association et dissociation, respectivement k_a et k_d, en utilisant le modèle de Langmuir 1 :1 (eq 3.1, c'est-à-dire une protéine interagissant avec un oligonucléotide) et calculer la constante de dissociation, K_D, selon le modèle et l'équation décrits en (eq 3.2). Afin de limiter le phénomène de transport de masse, nous avons appliqué un flux suffisamment élevé pour éviter que l'analyte ne se lie à nouveau au ligand lors de la phase de dissociation.

$$ADN + Proteine \underset{k_d}{\overset{k_a}{\rightleftharpoons}} ADN - Proteine \tag{3.1}$$

Ainsi on définit la constante d'équilibre de dissociation K_d (eq 3.2) et d'association K_a (eq 3.3) :

$$K_d = \frac{[A] + [B]}{[AB]} = \frac{k_d}{k_a} \tag{3.2}$$

$$K_a = \frac{[AB]}{[A] + [B]} = \frac{k_a}{k_d} \tag{3.3}$$

La technique de *SPR* permet donc de caractériser en déterminant les constantes cinétiques d'un système donné. Cependant, ces paramètres sont extraits de courbes. Pour ce faire, la *SPR* demande une connaissance préalable du système étudié afin de choisir au mieux le modèle mathématique. La *SPR* est une technique intéressante mais elle doit être utilisée avec d'autres techniques expérimentales pour corroborer les résultats. Dans mon cas, j'ai réalisé des expériences de *EMSA* en parallèle des expérience *SPR*.

3.5.2 Technique de photopontage utilisant la 4-thiothymine

Le photopontage est une technique d'investigation des interactions intra- et inter-moléculaires entre deux molécules qui ; dans notre cas il nous permet de visualiser des interactions protéine-ADN. La technique du photopontage, que nous avons utilisée, est basée sur l'utilisation d'une sonde photosensible, capable de lier de façon covalente une protéine à son ADN au sein d'un complexe nucléoprotéique. Dans notre étude, la sonde est l'ADN lui-même dans

lequel une thymine a été remplacée par une 4-thiothymine (^{4s}T). Cette dernière possède, à la place de l'atome d'oxygène du groupement carbonyle en position 4, un atome de soufre (Figure 3.10). Une étude a montré que le remplacement d'un atome d'oxygène par un atome de soufre entraine une faible variation de 0.45 Å du rayon de Van der Waals (Saenger, 1984), permettant à la 4-thioThymine de s'apparier avec une adénine par l'intermédiaire de liaison hydrogène classique de type Watson-Crick. L'avantage d'utiliser un analogue structural est que l'on ne perd pas la reconnaissance de la base modifiée ; cette propriété est très importante lorsqu'on étudie des protéines séquences spécifiques, ce qui est le cas de mon étude sur *POT1*.

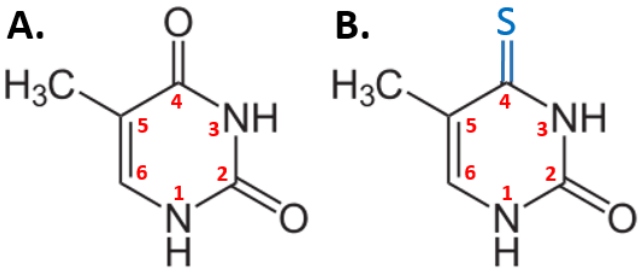


FIGURE 3.10 – **Structure de la thymine (A) et de la 4-thiothymine (B).**

Un autre avantage d'utiliser la 4-thiothymine provient de son spectre d'absorbance autour de 330 nm. Par conséquent en irradiant vers 330 nm, on excite spécifiquement la 4-thiothymine, sans affecter les autres bases classiques de l'ADN. L'irradiation se traduit par le passage de l'état fondamental singulet à un état excité triplet donnant lieu à un radical libre. C'est ce radical libre hautement réactif qui est capable de former une liaison covalente avec un acide aminé, proche dans l'espace (distance inférieure à 5 Angstrom), liant ainsi la protéine à la sonde. Pour visualiser les complexes ADN-protéines covalents formés, nous réalisons des gels *EMSA* dans des conditions dénaturantes en présence de nos sondes. Il est important de noter que nous avons recours au préalable à un marquage radioactif au phosphore 32 des sondes. La migration en parallèle de marqueurs de taille nous permet d'estimer la taille des complexes ADN-protéine covalents et ainsi d'identifier la nature des protéines liées covalentes. Ces points de contacts sont très utiles pour pouvoir proposer des modèles moléculaires.

3.5.3 Microscopie électronique

La microscopie électronique (ME) est une technique de microscopie basée sur l'utilisation d'un faisceau d'électrons. L'intérêt d'un microscope électronique est qu'il possède un pouvoir résolutif très supérieur à celui d'un microscope optique, en raison de la limite théorique de Abbe (eq 3.4). En effet, d'après de Broglie, la longueur d'onde λ d'un électron est beaucoup plus petite que la longueur d'onde d'un photon dans le milieu visible. C'est pourquoi un microscope électronique peut atteindre des grossissements allant jusqu'à $5\ 10^6$ fois, contre 2000 fois pour les meilleurs microscopes optiques.

$$d = \frac{\lambda}{2 \times O.N} \tag{3.4}$$

Avec

d	:	la résolution d'un microscope
λ	:	la longueur d'onde
O.N	:	l'ouverture numérique du microscope

Il est donc possible de faire une analogie entre le fonctionnement d'un microscope optique et d'un microscope électronique. En effet, c'est Louis de Broglie qui a généralisé en 1924 le concept de dualité onde-corpuscule aux objets microscopiques. On peut concevoir un faisceau d'électrons à l'aide d'un canon d'électrons, comme c'est le cas avec une lampe pour un faisceau de photon. Un canon d'électrons a pour principe d'extraire les électrons d'un conducteur, souvent un fil de tungstène, et de les accélérer par un champs électrique créant ainsi un faisceau d'électrons. Les électrons possèdent une charge négative, c'est pourquoi il est possible de concentrer un faisceau d'électrons à l'aide d'électro-aimants (Figure 3.11). Dans le cas de l'utilisation d'un microscope électronique à transmission (MET), on a souvent recours à la technique de coloration négative (negative stain en anglais) pour des échantillons biologiques. En effet, la matière biolofique est essentiellement composée d'atomes léger tel que le carbone, hydrogène, oyxgène, ... Ces atomes possèdent un faible pouvoir diffuseur du faisceau d'électron. La méthode de coloration négative consiste à préparer préalablement les échantillons en les recouvrant d'un film de carbone coloré avec 2% d'acétate d'uranyle. Ainsi l'acétate d'uranyle, atomes lourd, va se fixer préférentiellement à l'échantillon biologique. Un fort contraste est généré entre les régions riches en atomes lourd et les autres, générant une "empreinte" de l'échantillon. C'est pourquoi la technique de MET par coloration négative permet d'observer l'enveloppe d'un objet d'intérêt. Enfin dans le cas d'un microscope électronique à transmission, le faisceau d'électrons atteint un écran fluorescent qui émet un flash lumineux en chaque point de sa surface à l'emplacement où l'électron le

frappe. Un capteur CCD permet de convertir ce signal lumineux en un signal électrique analogique qui est traité et analysé afin de reconstituer l'image.

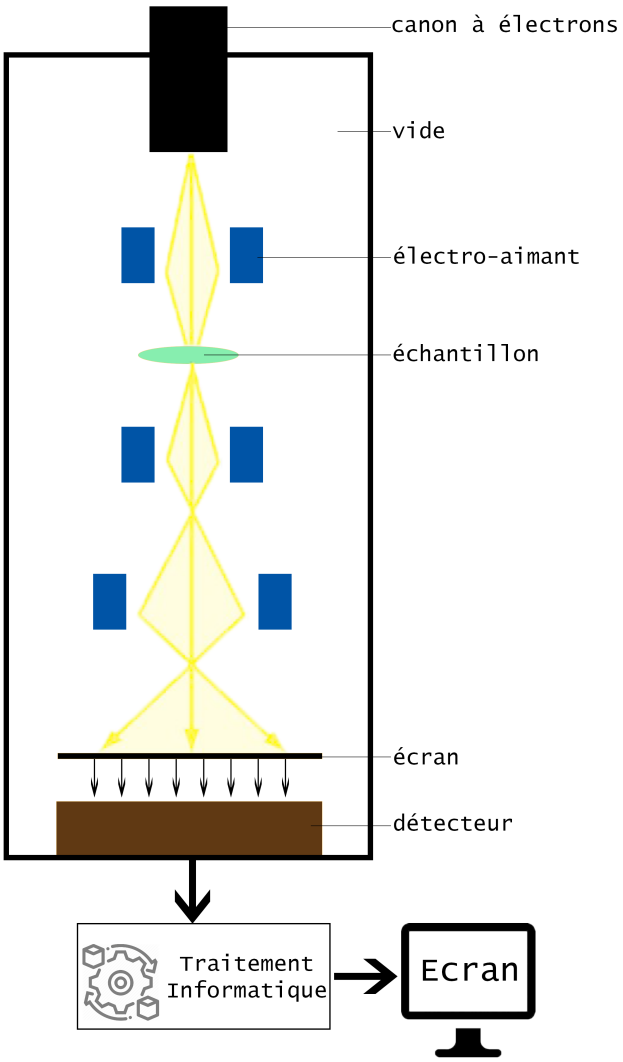


FIGURE 3.11 – **Schéma d'un microscope électronique à transmission (MET)**

La microscopie électronique est une méthode utilisée pour la biologie structurale. La résolution d'un ME classique, qui dépend de la longueur d'onde du faisceau d'électron (eq 3.4, est de l'ordre de la centaine de nm. C'est pourquoi la technique de ME donne une information sur l'enveloppe de la molécule étudiée. Ainsi il est possible de caractériser la forme globulaire ou fibrillaire de la protéine mais en aucun cas sa structure fine.

Nous avons entrepris dans un premier temps d'analyser l'enveloppe du complexe *POT1-TPP1* en présence et en absence d'ADN à l'aide de la technique de MET. La technique de MET a été choisie car elle présente comme avantages d' utiliser de faible quantité d' échantillon, de l'ordre de quelques μM/μL ; et de ne pas nécessiter des étapes d'optimisation, comme cela peut être le cas dans d'autres méthodes, exemple : cristallographie. De plus, la microscopie électronique permet une "visualisation" de l'enveloppe contrairement à des techniques utilisant des calculs, comme la diffusion des rayons X aux petits angles (*SAXS*). En revanche, la MET présente une limite au niveau de la taille de l'échantillon à analyser, à savoir 100kDa. C'est pourquoi dans notre étude, nous n'avons pas pu analyser la protéine *POT1* ($\sim$70 kDa) seule. Cette étude a été réalisée à l'aide du microscope électronique à transmission (JEOL 2100 LaB6) de la plateforme IMPMC à Sorbonne Université, par le Pr. Catherine Vénien-Bryan.

Chapitre 4

Des répétitions non-canoniques au sein des télomères : structure et interaction avec *RPA*

4.1 Introduction

4.1.1 Répétitions non-canoniques

Comme nous l'avons vu dans l'introduction générale (section 1.2.4), les télomères humains sont constitués de répétitions en tandem du motif TTAGGG (Moyzis et al., 1988). Cependant une étude menée en 1989 a montré la présence de trois répétitions non-canoniques (TTGGGG, TTAGGC, TTTAGGG) au sein des répétions canoniques humaines (Allshire et al., 1989). Dans les années suivantes d'autres motifs non-canoniques présents aux télomères humains ont été identifiés notamment TGAGGG, TCAGGG (Baird et al., 1995; Conomos et al., 2012). Les premières études sur l'identification et la distribution de motifs non-canoniques au sein des télomères humains ont porté sur les 1-2 Kb adjacents la région subtélomérique et ont mis en évidence un degré important de polymorphisme (Baird et al., 1995; Coleman et al., 1999; Baird et al., 2000). L'évolution des techniques de séquençage des génomes a ensuite permis de mettre en évidence la présence de répétitions non-canoniques tout au long des télomères (Conomos et al., 2012; Lee et al., 2013, 2018). Au-delà de quelques hypothèses spéculatives, l'origine de cette variabilité n'a pas été encore élucidée.

4.1.2 Un motif télomérique non-canonique instable

Aucune instabilité génomique liée à la présence des motifs non-canoniques cités ci-dessus n'a été à ce jour observée. Une étude de Royle et coll. a révélé, en proximité des régions subtélomériques (régions situées entre les télomères et les séquences d'ADN unique), la présence d'un motif télomérique non-canonique instable : le motif CTAGGG (Mendez-Bermudez et al., 2009). Ce motif représente 7% des télomères humains. L'instabilité de ce motif a été mise en évidence par une analyse des télomères de plusieurs familles d'individus (échantillons issus du CEPH - Centre d'Etude du Polymorphisme Humain). L'analyse a montré que, quand ce motif est présent chez le père, la progénitures porte un nombre de répétitions supérieur ou inférieur et variable d'un enfant à l'autre. Une instabilité de ce motif a été également observée dans des cellules somatiques d'un même individu (cellules du colon, normales et tumorales). Les données de cette étude montrent que, si un télomère contient environ une dizaine de répétitions CTAGGG flanquées de répétitions TTAGGG, il est susceptible d'être très instable, c'est-à-dire de présenter une perte ou un gain de répétitions. L'explication de l'origine de cette instabilité génomique en présence de répétition CTAGGG est encore inconnue.

4.1.3 Le motif CTAGGG forme un G4 non canonique.

En parallèle de cette étude génétique, une étude par RMN a montré que quatre répétions du motif CTAGGG forment en potassium un G4 "non canonique" (Figure 4.1.A). Cette structure G4 est dite "non canonique" car elle ne ressemble pas aux structures déterminées en présence de la séquence canonique TTAGGG (section 1.6). En effet le G4 formé à partir de répétitions CTAGGG se présente sous la forme de deux quartets de guanines et d'une tétrade GCGC où les G et les C forment des liaisons Watson-Crick (Figure 4.1.B) (Lim et al., 2009). Cette structure, pour son repliement en chaise et pour la présence d'une tétrade non-canonique est très différente des G4 formés par 4 répétitions du motif canonique TTAGGG en potassium. A l'époque, les auteurs de ces deux études avaient suggéré que cette structure G4 non-canonique pouvait être à l'origine de l'instabilité de ce motif.

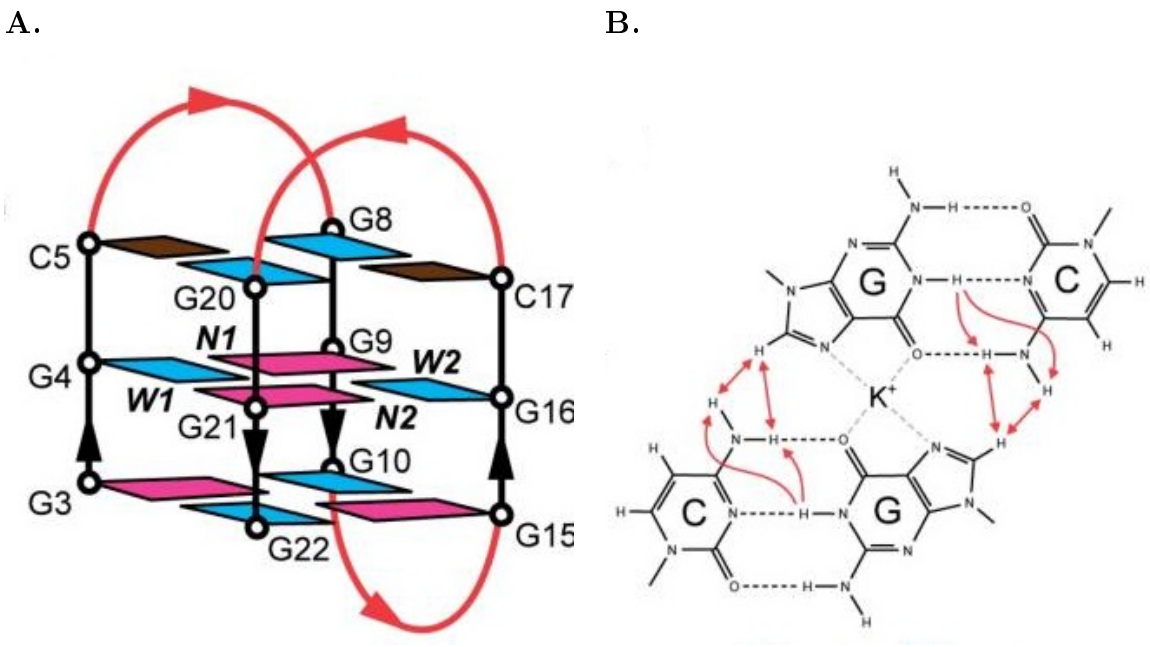


FIGURE 4.1 – **Structure du G4 constitué de répétition du motif non-canonique CTAGGG en potassium A.** Structure du G4 variant CTAGGG,**B.** Interaction Watson-Crick du tétrade GCGC, (Lim et al., 2009).

4.2 Présentation de l'étude

4.2.1 Les répétitions du motif CTAGGG ne forment pas que des G4.

L'étude de Royle et coll. (Mendez-Bermudez et al., 2009) mettait en évidence que l'instabilité du motif non-canonique CTAGGG ne se manifestait qu'à partir d'un nombre de répétitions supérieure à une dizaine. Néanmoins, aucune hypothèse n'avait été proposée pour expliquer pourquoi cette instabilité dépendait du nombre de répétitions. Dans l'équipe SANTÉ, après avoir étudié les structures en G4 en tandem formées par des répétitions du motif canonique TTAGGG (Bugaut and Alberti, 2015), Patrizia Alberti a caractérisé les structures formées par un nombre croissant de répétitions non-canoniques CTAGGG, à savoir 8, 12 et 16 répétitions, pouvant former en théorie 2, 3 et 4 G4 contigus. La question subjacente à cette étude était la suivante : et si l'instabilité du motif CTAGGG ne découlait pas du G4 non-canonique mais d'autres structures qui se forment à partir d'un certain nombre de répétions ? Des études structurales par *CD* et par RMN ont permis de mettre en évidence qu'à partir de 8 répétitions, le motif CTAGGG ne forme pas que des G4s non-canoniques en tandem, mais également des structures en tige-boucle, dont la stabilité augmente avec l'augmentation du nombre de répétitions. En présence des cations lithium (Li^+) et sodium (Na^+), l'équilibre est quasi-complètement déplacé vers la forme tige-boucle. En revanche en présence de potassium (K^+) (le cation qui stabilise le plus les G4), il existe un équilibre entre la forme G4 non-canonique et la forme en tige-boucle. La présence de ces structures tiges-boucles, se formant à partir de 8 répétitions CTAGGG, pourrait être à l'origine de l'instabilité génomique en présence de ce motif.

4.2.2 Les répétitions CTAGGG structurées en tige-boucle posent un problème à la liaison de *RPA*.

Dans le cadre de cette étude, en complément des données structurales obtenues par Patrizia Alberti, j'ai étudié la liaison de *RPA* aux répétitions du motif télomérique non-canonique CTAGGG. Pourquoi *RPA* ? Les données de Royle suggéraient que l'instabilité du motif CTAGGG était lié à la phase de réplication des télomères (Mendez-Bermudez et al., 2009), puisqu'elle ne se manifestait que lors de la transmission via la lignée germinale paternelle. *RPA* est présente aux télomères humains lors de leur réplication (Verdun and Karlseder, 2006). Elle participe chez la levure *S. pombe* à la réplication correcte du brin G (Moser and Nakamura, 2009; Audry et al., 2015). Ré-

cemment une étude au sein de l'équipe SANTÉ a montré que *RPA* se lie aux répétitions canoniques TTAGGG en déroulant les structures en G4 en tandem présentes (Lancrey et al., 2018).

Mes résultats montrent que *RPA*, en présence d'un oligonucléotide composé de 12 répétitions, en potassium, se lie moins efficacement à la séquence CTAGGG qu'à la séquence canonique TTAGGG. Étant donné qu'en potassium, il existe un équilibre entre forme G4 non-canonique et forme tige-boucle, nous nous sommes demandé laquelle parmi ces deux structures posait un problème à la liaison de *RPA* : la structure en G4 non-canonique ou la structure en tige-boucle ? Nous avons répondu à cette question en comparant :

1. l'efficacité de liaison de *RPA* à quatre répétions motif CTAGGG structurées en G4 non canonique (section 4.1.3) et à quatre répétitions canoniques TTAGGG structurées en G4 en potassium. On remarque la même efficacité de liaison en présence des deux structures G4s (Figure 4.2).

2. l'efficacité de liaison de *RPA*, en sodium, à 12 répétions du motif CTAGGG structurées en tige-boucle et à 12 répétions du motif TTAGGG structurées en G4 en tandem. On observe une moins bonne efficacité de liaison de *RPA* en présence des structures tige-boucle (CTAGGG) qu'en présence de G4 en tandem (TTAGGG) (Figure 4.2).

Les résultats montrent que les répétitions CTAGGG posent un problème à la liaison de RPA quand elles sont structurées en tige-boucle. Il est important de noter que dans la cellule les répétitions du motif CTAGGG sont entourées de part et d'autre par des répétitions canoniques. Nous nous sommes donc demandés, si la présence de ces répétitions canoniques ne pourrait pas aider la liaison de *RPA*, et par conséquent le déroulement des structures tige-boucle formées par le motif CTAGGG. Nous avons choisi de nous mettre dans les conditions les plus favorables pour la liaison de *RPA* , à savoir l'ajout d'une extension simple brin polyT (10 thymine) en 5'. En effet, 10 nt sont suffisants pour permettent la fixation partielle de *RPA* (via les domaines *DBD-A* et *DBD-B*) de la sous-unité *RPA*1 (Safa et al., 2016)) ; leur position en 5' peut faciliter l'ouverture de structures d'ADN telles que des G4 stabilisés par des ligands et des jonctions simple-brin/double-brin, cela en raison du mode de liaison orienté et séquentiel de *RPA* à l'ADN (d'abord *RPA*1, suivie par *RPA*2 du côte 3' de l'ADN) (Safa et al., 2016). L'ajout de cette courte extension aux répétitions CTAGGG a pour conséquence de déplacer l'équilibre vers les structures tige-boucle (données de Patrizia Alberti). J'ai pu montrer que *RPA* se lie peu efficacement aux répétitions CTAGGG structurées en tige-boucle, même en présence d'une extrémité simple-brin côté 5'. En conclusion,

même dans des conditions favorables de liaison *RPA* présente des difficultés à prendre en charge les structures tige-boucle formées par plusieurs répétitions CTAGGG.

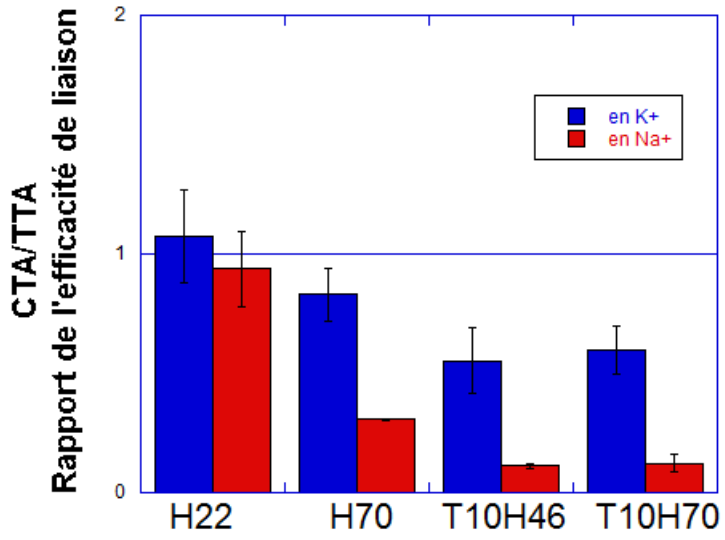


FIGURE 4.2 – **Efficacité de liaison de *RPA* à 50nM entre les séquences CTA et les séquences TTA.** Les barres en bleus indique en présence de K^+, les barres en rouges en présence Na^+.

4.2.3 Et le brin C ?

Des répétitions du motif CTAGGG sur le brin G sont associés à des répétitions TAGCCC sur le brin C. Ces répétitions variantes du brin C peuvent, en théorie, se structurer elles aussi en tige-boucle. Cependant des études réalisées par Patrizia Alberti par spectroscopie UV-visible (*TDS*,*CD*, *Tm*) indiquent que, contrairement aux répétitions CTAGGG, les répétitions TAGCCC ne forment pas de structures stables à pH et température physiologiques. Mes expériences EMSA indiquent que *RPA* se lie aussi bien aux répétitions TAGCCC qu'aux répétitions canoniques TAACCC. Ces résultats suggèrent que l'instabilité des répétitions non-canoniques CTAGGG / TAGCCC découle des répétitions CTAGGG sur le brin G et non des répétitions TAGCCC sur le brin C.

4.2.4 Le motif non-canonique stable TCAGGG

Nous avons également étudié un autre motif télomérique non-canonique : le motif TCAGGG, qui, contrairement à CTAGGG, est stable (Mendez-Bermudez et al., 2009). Ce motif serait particulièrement fréquent dans les cellules *ALT*, où il constituerait un site de liaison pour des récepteurs nucléaires, qui pourraient être impliqués dans la dynamique des télomères des cellules ALT (Conomos et al., 2012). Des études spectroscopiques indiquent que 4 répétitions de ce motif forment en potassium des G4, mais pas si réguliers que les G4 en tandem formés par les répétitions canoniques. Les spectres *TDS* et *CD* en présence de sodium (Na^+) et en présence de lithium (Li^+) révèlent la présence d'une deuxième structure non-G4, dont le spectre *CD* ne correspond à aucune structure connue. Cependant les expériences de *EMSA* réalisées montrent que cette deuxième structure ne pose pas de problème à la liaison de *RPA* (Figure 4.3). Nos résultats montrent également que *RPA* se lie aussi bien aux répétitions non-canoniques TCAGGG qu'aux répétitions canoniques TTAGGG structurées en potassium.

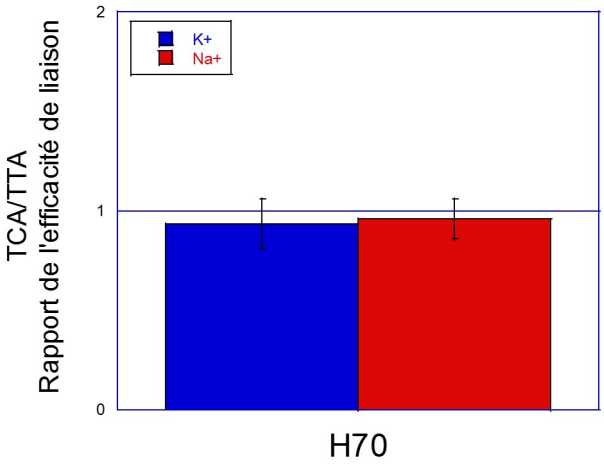


FIGURE 4.3 – **Efficacité de liaison de *RPA* à 50nM entre la séquence H70TCAGGG et la séquence H70TTAGGG**Les barres en bleus indique en présence de K^+, les barres en rouges en présence Na^+.

4.3 Article 3

4.3.1 Résumé

Dans les télomères humains, les répétitions canoniques GGGTTA sont entrecoupées de répétitions variantes. Des études antérieures ont montré que les répétitions de variantes de GGGCTA présentaient une instabilité dépendante de la longueur (les séquences de plus de dix répétitions de GGGCTA étaient instables dans les lignées germinales mâles) et se repliaient en un G-quadruplex non canonique (G4), reposant sur deux G-tétrades et un Watson -Crick GCGC tétrade. Le mécanisme sous-jacent à l'instabilité de GGGCTA est inconnu ; on a émis l'hypothèse que le G4 non canonique pourrait être un obstacle à la réplication télomérique du brin G. Dans la présente étude, nous démontrons que les séquences $(GGGCTA)_{n>7}$ se replient en G4 et tige-boucle. Nous démontrons également que la protéine de réplication humaine A (*RPA*) lie les répétitions de variantes de GGGCTA aussi efficacement que les répétitions canoniques de GGGTTA lorsque les répétitions de GGGCTA sont structurées en G4, mais moins efficacement lorsqu'elles sont structurées en tige-boucle. Nos résultats suggèrent comme origine possible de l'instabilité le nombre de répétions GGGCTA : lorsque les séquences du brin G télomérique sont exposés de manière transitoire en simple brin (par exemple lors de la réplication ou de la transcription), la structuration de répétitions $(GGGCTA)_{n>7}$ en tige-boucle et la mauvaise liaison du RPA à ces structures peut entraîner des structures tige-boucle non résolues, dont la réparation conduit à de petits gains ou pertes de répétitions de GGGCTA. Notre étude peut amener à réfléchir sur la présence des G4 télomériques et sur l'évolution du motif télomérique.

$(GGGCTA)_{n \geq 8}$ repeats can fold into hairpins poorly unfolded by Replication Protein A: a possible origin of the length-dependent instability of GGGCTA variant repeats in human telomeres.

Jean Chatain [1], Alain Blond [3], Anh Tuân Phan [4,5], Carole Saintomé [1,2], Patrizia Alberti[1*]

[1] Laboratoire Structure et Instabilité des Génomes (StrInG), Muséum national d'Histoire naturelle, CNRS, Inserm, Paris 75005, France

[2] Sorbonne Université, UFR927, F-75005, Paris, France

[3] Laboratoire Molécules de Communication et Adaptation des Microorganismes (MCAM), Muséum national d'Histoire naturelle, CNRS, 75005 Paris, France.

[4] School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore;

[5] NTU Institute of Structural Biology, Nanyang Technological University, Singapore 636921, Singapore

* To whom correspondence should be addressed: patrizia.alberti@mnhn.fr

ABSTRACT

In human telomeres, GGGTTA canonical repeats are interspersed with variant repeats. Previous studies showed that GGGCTA variant repeats displayed length-dependent instability (arrays of more than ten GGGCTA repeats were unstable) and folded into a non-canonical G-quadruplex (G4), relying on two G-tetrads and a Watson-Crick GCGC tetrad. The mechanism underlying GGGCTA instability is unknown; it was speculated that the non-canonical G4 might be a supplementary obstacle for replication. In the present study, we demonstrate that $(GGGCTA)_{n\geq 8}$ sequences fold into G4s and into hairpins. We demonstrate that human Replication Protein A (RPA) binds GGGCTA variant repeats as efficiently as GGGTTA canonical repeats when GGGCTA repeats are structured into G4s, but less efficiently when they are structured into hairpins. Our results suggest a possible origin of GGGCTA length-dependent instability: when tracts of the telomeric G-strand are transiently exposed as single-strands, the structuration of $(GGGCTA)_{n\geq 8}$ arrays into hairpins and the poor binding of RPA to these structures may result in unresolved hairpins, that can lead to GGGCTA instability via a variety of possible mechanisms. Our study inspires a working hypothesis on short tandem repeats instability and a few considerations about telomeric G4s and the evolution of telomeric motif (maybe, G4 are telomeres' best friend).

Keywords: telomere, instability, variant-repeats, G-quadruplex, G4, hairpin, RPA

INTRODUCTION

Telomeres are the nucleoprotein complexes that constitute the end of eukaryotic linear chromosomes; they safeguard genome integrity, by protecting chromosomes from erosion, degradation and inappropriate repair (1). In eukaryotes where telomeres are maintained by telomerase, the telomeric DNA strands running toward the 3' ends of chromosomes (the G-strands) are generally composed of repeats of a short motif (5-8 nt) bearing consecutive guanines (2, 3 or 4) (2). Because of the presence of consecutive guanines, telomeric G-strands from a variety of organisms are prone to fold into G-quadruplexes (G4s) (3), four-stranded structures relying on the stacking of tetrads of guanines (G-tetrads or G-quartets) and stabilised by several cations (notably K^+ and, to a lesser extent, Na^+) that fit into the central cavity of the structure and coordinate with the carbonyl oxygens of the guanines.

In humans, the telomeric motif is the hexamer GGGTTA (Moyzis et al. PNAS 1988); nevertheless, human telomeres display a certain degree of sequence polymorphism in the proximity of the subtelomeric regions (the proximal region): GGGTTA canonical repeats are interspersed with variant repeats (such as GGGGTT, GGGTGA, GGGTCA, ...) (4)(5)(6)(7). Advances in genome sequencing technologies have allowed detecting the presence of variant motifs all along the telomeres (8)(9)(10). About ten years ago, N. Royle and coll. identified, within the proximal region of telomeres, a variant motif that was highly unstable in the male germline: the GGGCTA motif, estimated to be present in 7% of human telomeres (11). When a homogenous array of at least a ten of GGGCTA repeats was present in the telomere of a father, the progeny displayed small gains or losses of these variant repeats; GGGCTA repeat instability (including complete loss of GGGCTA repeats) was also detected in colon samples of a single individual (11). At the same time, an NMR study revealed that about four repeats of this telomeric variant motif (the A(GGGCTA)$_3$GGG sequence) folded, in potassium, into a robust non-canonical G4: an antiparallel chair-type G4 formed by two G-tetrads sandwiched between a Watson-Crick GCGC tetrad and a Watson-Crick GC pair in a lateral loop (in sodium no stable Watson-Crick GC base-pairs were detected) (12). The presence of a non-canonical GCGC tetrad and the chair-type conformation made this structure very different from the G4 structures formed by the GGGTTA canonical motif (by the GGGTTA)$_3$GGG core sequence) (Figure S1). The mechanism underlying GGGCTA instability is unknown. It was shown that the GGGCTA variant motif did not challenge the binding of the shelterin proteins POT1 and TFR1,2; it was suggested that GGGCTA instability was replication-dependent, since it was

observed only in the male germline, and speculated that the non-canonical G4 structure could be an additional obstacle for G-strand replication (11).

After having investigated the structure and the stability of long telomeric sequences of GGGTTA canonical motif (13), we reconsidered the GGGCTA variant motif in the light of a question: why GGGCTA instability is observed only starting from a ten of repeats? If the non-canonical G4 conformation were at the origin of the instability of this variant motif, the instability should be observed starting from four GGGCTA repeats. We hence wondered what kind of structures are formed by more than four repeats of GGGCTA variant motif: do eight, twelve, sixteen, ... GGGCTA repeats fold into two, three, four, ... contiguous G4s, as GGGTTA canonical repeats do (13), or do they fold into other structures? With the aim to get some insights into the origin of the length-dependent instability of the GGGCTA telomeric variant motif, we undertook a structural study of sequences composed of an increasing number of GGGCTA repeats and studied the interaction of human Replication Protein A (RPA) with these structures.

Why RPA? RPA is considered *"the single-stranded DNA's first responder"* (14); it binds to transiently exposed single-stranded DNA and participates to the orchestration of DNA metabolic pathways. It is present at human telomeres during replication, suggesting a transient exposure of single-stranded DNA(15). Studies about telomere replication in *S. pombe* fission yeast indirectly highlight transient accumulation of RPA on lagging strand (G-strand) (16) and provide evidence that impairing RPA binding to DNA leads to problems in lagging strand replication (17). The study of Nakamura and coll. shows that, during telomere replication, lagging strand polymerases Polα and Polδ arrive at telomeres later than the leading strand polymerase Polε and that, during this delay time, RPA accumulates at telomeres; this suggest that lagging strand synthesis is delayed compared to leading strand synthesis and that long tracks of lagging strand are transiently exposed and covered by RPA (16). The study of Coulon and coll. shows that a mutation in RPA1 subunit impairing RPA binding to DNA results in a longer time of persistence of the lagging strand polymerase Polα at telomeres (but not of the leading strand polymerase Polε) and to defective telomere replication characterized by the accumulation of telomeric DNA high molecular weight structures; this suggests that efficient binding of RPA to lagging strand is essential for maintaining it unstructured and for its replication (17). Overall these studies support the implication of RPA in telomere replication and the importance of its efficient binding to lagging strand tracts transiently exposed as single-strands.

In a previous works we showed that human telomeric sequences folds into contiguous G4 units (13) that are efficiently bound and unfolded by human RPA, independently of their length (18). In the present work, we provide evidence that, starting from eight repeats, GGGCTA arrays can fold into G4 as well as into hairpin structures and that the hairpin structures (and not the G4 structures) impair the binding of human RPA. Our results suggest a working hypothesis on the origin of GGGCTA length-dependent instability in telomeres (and more widely of Short Tandem Repeats instability) and inspire a few considerations about telomeric G4s and the evolution of telomeric motifs.

MATERIALS AND METHODS

Oligonucleotides

Oligonucleotides were purchased from Eurogentec; GGGTTA, GGGCTA and GGGTCA repeat oligonucleotides were PAGE purified, all the other oligonucleotides (dxGG, dx22 and dx44 strand 1 and strand 2, GTTCTA, CCCTAA and CCCTAG repeats oligonucleotides) were Reverse-Phase Cartridge• Gold™ purified. Oligonucleotides were dissolved in bi-distilled water at a concentration of 200 μM and stocked at -20°C; concentrations were determined using molar extinctions coefficients provided by the manufacturer. Oligonucleotide sequences are listed in Table S1.

Proteins

Recombinant human RPA1, RPA2 and RPA3 were co-expressed in the *Escherichia coli* BL21-DE3 strain transformed with the pET11a-hRPA plasmid (a generous gift from Dr Klaus Weisshart, IMB, Jena, Germany), purified following published protocols (17) and stocked in 25 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol; aliquots were stored at -80°C.

UV-absorption measurements: melting curves, thermal difference spectra and circular dichroism spectra

Oligonucleotides were diluted in a cacodylic acid buffer (10 mM) at pH 7.2 (adjusted with LiOH), supplemented with LiCl or NaCl or KCl (oligonucleotide and salt concentrations are reported in figure legends).

Melting curves were obtained by recording the absorbance at fixed wavelength (335, 295, 273, 260 and 245 nm) as a function of temperature, first upon decreasing the temperature from 95°C to 5°C at 0.2°C min^{-1} and then increasing the temperature from 5°C to 95°C at 0.2°C min^{-1}; between the cooling and the heating steps, the temperature was kept at 5°C for 20 min. Melting curves at 295, 273, 260 and 245 nm were corrected by subtracting the absorbance at 335 nm as a function of temperature. For each sample, cooling and heating curves were superimposable (no hysteresis), this indicated that they corresponded to equilibrium states. Melting experiments were performed at least twice, melting profiles were reproducible. We defined "transition temperature *T*t" the maximum of the first derivative of absorbance as a function of temperature.

For each sample, the thermal difference spectrum (TDS) was obtained by subtracting to the absorbance spectrum at high temperature (95°C) the absorbance spectrum at low temperature (5°), recorded after cooling the sample from high to low temperature at 0.2°C/min; TDS provide information about nucleic acid structures (19).

Circular dichroism (CD) spectra were recorded at 5°C after annealing the samples from high temperature at 0.2°C min^{-1}; each CD spectrum was obtained by averaging three scans at a speed of 500 nm/min^{-1} and was corrected by subtracting the spectrum of a water filled quartz cell.

UV-absorption measurements were acquired on an Uvikon XS spectrophotometer (Secoman), coupled to a circulation thermocryostat (Julabo) for temperature control; temperature was measured with a sensor immersed into a water filled quartz cell. CD spectra were recorded on a J-810 spectropolarimeter (Jasco), equipped with a Peltier thermostatted cell holder. Measurements were acquired in 1 cm optical pathlength quartz cuvettes (Hellma).

Nuclear Magnetic Resonance

Oligonucleotides were diluted in a cacodylic acid buffer (10 mM) at pH 7.2 (adjusted with LiOH), supplemented with LiCl or NaCl or KCl (oligonucleotide and salt concentrations are reported in figure legend), heated at 90°C for a couple of minutes and slowly cooled to room temperature. D_2O (10% of total volume) was added to sample solutions before measurements. One dimensional 1H NMR spectra were acquired at 298K on an Avance III HD 600 MHz spectrometer (Bruker), equipped with a 5 mm triple resonance TCI cryoprobe (^{1}H-^{13}C-^{15}N) including shielded Z-gradients. For each sample, DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid), dissolved in the same solvent used for sample measurement, was used as external chemical shift standard. Excitation sculpting was used for peak solvent suppression (pulse program zgesgp). For each sample, 4096 scans were accumulated (with the exception of H46cta in KCl 50 mM, for which 10400 scans were accumulated).

Polyacrylamide gel electrophoresis (PAGE)

Oligonucleotides were diluted in a cacodylic acid buffer (10 mM) at pH 7.2 (adjusted with LiOH), supplemented with LiCl or NaCl or KCl 100 mM (oligonucleotide concentrations are reported in figure legends), heated at 90°C for a couple of minutes and slowly cooled to 4°C. After addition of sucrose (10%), oligonucleotide solutions were loaded (10 or 20 μL) in a 12% polyacrylamide gel (acrylamide:bisacrylamyde 19:1), prepared in a TBE buffer supplemented with LiCl or KCl or NaCl 20 mM. Electrophoresis were run in a TBE buffer supplemented with LiCl or KCl or NaCl 20 mM, in a cold room. Depending on their concentration,

oligonucleotides were detected by UV-shadow at 254 nm with a G:BOX (Syngene), or stained with methylene blue (0.02%) and detected with a ChemiDoc imager (Bio-Rad), or stained with ethidium bromide (0.5 μg/mL) and detected with a Typhoon FLA9500 (GE Healthcare Life Sciences).

Electrophoresis mobility shift assays (EMSA)

Oligonucleotides were labelled with [γ^{32}P]ATP using a T4 polynucleotide kinase (NEB). Radiolabelled oligonucleotides were diluted at 2.5 nM strand concentration in a buffer containing KCl or NaCl (HEPES 50 mM pH 7.9, BSA 0.1 mg mL^{-1}, KCl or NaCl 100 mM and glycerol 2%), heated for 10 min at 85°C and slowly cooled to room temperature. Serial dilutions of proteins were incubated (20 min at 4 °C) in a dilution buffer containing KCl or NaCl 100 mM (Tris-HCl 50 mM pH 7.5, KCl or NaCl 100 mM, DTT 1 mM, BSA 0.2 mg mL^{-1}, EDTA 0.1 mM and glycerol 10%,). 2 uL of serial dilution protein solutions were then added to 8 uL of radiolabeled oligonucleotide solutions and incubated for 20 min at 20 °C. Samples were then loaded on a native 5% polyacrylamide gel (acrylamide:bisacrylamide 29:1) in a 0.5X TBE buffer. Electrophoresis were run in a 0.5X TBE buffer, at room temperature. After electrophoresis, gels were dried, exposed on a Phosphorimager screen and scanned with a Typhoon FLA9500. The intensities *I* of the bands were quantified using ImageQuant software. For each protein concentration, the fraction of radiolabelled oligonucleotide bound to the protein was calculated as:

$$I_{\text{oligonucleotide bound to protein}} / (I_{\text{free oligonucleotide}} + I_{\text{oligonucleotide bound to protein}}).$$

RESULTS

GGGCTA unstable variant repeats fold into G4 and into hairpin structures.

In order to investigate the structure of GGGCTA repeats, we carried out an UV-absorption investigation of A(GGGCTA)*n*GGG oligonucleotides, with $n = 3, 7, 11, 15$. The $n = 3$ sequence (here named H22cta) is the sequence previously studied by Lim et al. (12): in potassium, it folds into a chair-type G4 composed of two G-tetrads and a Watson-Crick GCGC tetrad. The $n = 7$, 11, 15 sequences (named H46cta, H70cta and H94cta, respectively) bear eight, twelve and sixteen runs of GGG and may fold, in principle, into two, three and four contiguous G4 units, respectively. Hereafter, GGGCTA repeats bearing more than four runs of GGG (*i.e.* H46cta, H70cta and H94cta) will be globally referred to as "GGGCTA repeats".

We recorded thermal difference spectra (TDS) and circular dichroism (CD) spectra at the following strand concentrations 6 μM for H22cta, 3 μM for H46cta, 2 μM for H70cta and 1.5 μM for H94cta, so to have the same concentration of potential G4 units (6 μM) for all the sequences. At these strand concentrations, folding of H46cta, H70cta and H94cta into two, three and four contiguous G4 units, respectively, would result in identical TDS and identical CD spectra (identical in shapes and amplitudes), as for GGGTTA canonical repeats (13)(20).

The TDS and CD spectrum of H22cta in potassium (figures 1A and 1B, red spectra) were similar to those previously reported (12): TDS was characteristic of a G4 structure (19) and CD spectrum of G4 antiparallel folding (where stacked G-tetrads display alternating clockwise/anticlockwise hydrogen-bond orientation (21)); the *T*t was about 64°C.

With increasing the number of repeats, TDS and CD spectra of H46cta, H70cta and H94cta in potassium progressively shifted from those of the single G4 formed by H22cta oligonucleotide (figures 1A and 1B, blue, green and black spectra); in particular, the negative peak at 295 nm in TDS (characteristic of G4 folding) decreased and CD spectra shifted toward a signature that appeared to result from a H22cta-*like* CD component and from a second component. Overall, TDS and CD in potassium in figures 1A and AB indicated that H46cta, H70cta and H94cta did not simply fold into contiguous G4 units, differently from their canonical counterparts (for comparison, TDS and CD spectra of GGGTTA canonical repeats folding into contiguous G4 units, published in our previous studies (13)(20), are shown in figure S2).

To test the hypothesis of an equilibrium between a G4 and a duplex structure, we carried out measurements under conditions that allowed decreasing G4 stability without affecting duplex

stability: we studied the behaviour of GGGCTA repeats in the presence of lithium and sodium cations (Li^+ does not stabilize G4s, Na^+ stabilizes G4s to a lesser extent than K^+; the three cations stabilise duplexes to a similar extent). Replacing K^+ with Li^+ highlighted the formation of non-G4 structures with a high thermal stability: H46cta, H70cta and H94cta TDS and CD spectra lose the characteristic G4 signatures (figures 1C and 1D) and melting curves displayed high transition temperatures that increased with the number of repeats (from 54°C for H46cta to 57°C for H94cta) (figure 1D, inset). In Na^+, H22cta still folded into a G4 with a *T*t of 55°C (as already reported (12)) (figures 1E and 1F, red spectra), while H46cta, H70cta and H94cta TDS and CD spectra shifted toward non-G4 signatures similar to those displayed in Li^+ (figures 1E and 1F, blue, green and black spectra). Finally, upon decreasing KCl concentration from 100 mM to 10 mM, the CD spectrum of H46cta progressively shifted toward the non-G4 CD spectrum displayed in LiCl (figure 1G).
Overall, the above data suggested that GGGCTA repeats folded into G4 and into non-G4 structures and that both structures coexisted in potassium solutions.

In principle GGGCTA repeats may fold into duplex structures where four Watson-Crick base-pairs alternate with a couple of GG base-pairs (figure 1H); in order to ascertain the formation of this kind of local structure, we recorded the CD spectrum of the duplex formed by the autocomplementary sequence 5'CTAGGGCTAG (named dxGG) (figure 1H). The CD spectra of GGGCTA repeats in LiCl were similar to the CD spectrum of dxGG (figure 1J): a negative band around 255 nm, a major positive band around 275 nm and a minor positive band around 297 nm (in order to compare CD spectra, we fixed dxGG strand concentration to 10 μM: this corresponds to 5 μM dxGG duplex concentration, and 5 μM is the concentration of "dxGG duplex units" harboured in 1.5 μM H94cta, 2 μM H70cta and 3 μM H46cta). The deeper negative band around 255 nm in GGGCTA repeats compared to dxGG might arise from a higher proportion - about twice - of GG base-pair couples in GGGCTA putative duplexes than in dxGG duplex (*e.g.* 8 GG base-pair couples per 30 Watson-Crick base-pairs in H94cta putative hairpin *versus* 1 GG base-pair couple per 8 Watson-Crick base-pairs in dxGG (figure 1H)). In KCl and in NaCl, the CD spectra of GGGCTA repeats appeared to result from the contribution of a H22cta-*like* CD spectrum and of a dxGG-*like* CD spectrum (figures 1I and 1K).

In order to further ascertain the presence of an equilibrium between a G4 and a duplex structure, we carried out NMR measurements on H46cta; we recorded H22cta and dxGG spectra as reference spectra (figure 2).

The spectrum of H22cta in KCl (figure 2A) was similar to the one published in KCl 100 mM, with peaks in the 10-12 ppm Hoogsteen imino-proton region, characteristic of G-tetrads, and peaks in the 12-14 Watson-Crick imino-proton region that were previously ascribed to two G•C base-pairings in the GCGC tetrad and to a G•C base-pair in the central loop (12).

In NaCl and LiCl, H46cta displayed major peaks in the 12-14 ppm Watson-Crick region and only minor signals in the 10-12 ppm Hoogsteen region (figures 2D and 2E), supporting a strong shift toward a duplex structure, in agreement with TDS and CD spectra in NaCl and LiCl (shown in figures 1 C-F).

In KCl, H46cta displayed important peaks in both Watson-Crick and Hoogsteen regions (figure 2B and 2C). At least three features suggested that H46cta spectra in KCl resulted from the presence, in solution, of a G4 and of a duplex structure. **(1)** The ratio between the integrals of Watson-Crick peaks and Hoogsteen peaks (here defined as *WC/H* ratio) in H46cta spectra was higher than the *WC/H* ratio in H22cta G4 spectrum (60/40 for H46cta and 10/90 for H22cta, in KCl 10 mM) (figure 2B and 2A). **(2)** H46cta *WC/H* ratio increased with decreasing KCl concentration (40/60 in KCl 50 mM, 60/40 in KCl 10 mM) (figure 2C and 2B), suggesting a G4-to-duplex shift in agreement with CD spectra shown in figure 1G. **(3)** In the 10-12 ppm region, H46cta spectra displayed a 10.4 peak also present in dxGG spectrum in KCl (figure 2F); this peak did not appear in H22cta G4 spectrum and is characteristic of GG base-pairing within Watson-Crick duplexes (22).

By PAGE experiments, we assessed the molecularity of the duplex structures formed by GGGCTA repeats (in a 1-30 μM strand concentration range) under conditions where the equilibrium was completely shifted toward the duplex structure, *i.e.* in LiCl. H46cta migrated in two bands with the mobility of a 22 and a 44 base-pair duplexes (mimicking GGGCTA duplexes), indicating the formation of both intra- and intermolecular structures; the intramolecular structure (hairpin) was the major fraction. For longer repeats, only the hairpin structure was detected (figure S3).

In conclusion, UV-absorption, NMR and PAGE data demonstrated that GGGCTA repeats (H46cta, H70ctan H94cta) folded into G4 structures and into duplex (hairpin) structures. In

LiCl and NaCl, the equilibrium was strongly shifted toward the duplex conformation. In KCl, the equilibrium markedly depended on the number on GGGCTA repeats and on K^+ concentration: increasing the number of repeats or decreasing K^+ concentration shifted the equilibrium toward the duplex conformation.

Based on CD spectra shown in figure 1, we roughly estimated the fractions of oligonucleotides folded into the duplex structure, as detailed in figure S4: in NaCl 100 mM (CD in figure 1F) the duplex fraction was about 0.9; in KCl 100 mM (CD in figure 1B), the duplex fraction increased from about 0.35 for H46cta to about 0.60 for H70cta and H94cta; when decreasing the concentration of KCl from 100 mM to 10 mM (CD in figure 1G), the duplex fraction of H46cta increased from about 0.35 to about 0.75. These duplex fractions, roughly estimated from CD spectra, are reported in the histogram in figure 3.

Our data also provided some information about the conformations of the G4 and of the duplex structures adopted by GGGCTA repeats. With respect to the conformation of the duplex structure, CD, NMR and PAGE data supported the formation of hairpins as the one proposed in figure 1H for H94cta (long hairpins based on a dxGG-*like* local structure). With respect to the conformation of the G4 fraction in potassium, H46cta, H70cta and H94cta CD spectra displayed a H22cta-*like* antiparallel component (figures 1B and 1G). We argue that the G4 fraction in potassium may be structured into contiguous non-interacting G4 units (similarly to GGGTTA canonical repeats (13)), where each G4 unit is folded into a H22cta-*like* conformation.

RPA binds GGGCTA repeats structured into G4s as efficiently as GGGTTA canonical repeats, but poorly binds GGGCTA repeats structured into hairpins.

RPA binds GGGTTA canonical repeats structured into contiguous G4s, unfolding them, independently of the number of G4 units (18). We investigated, by EMSA, whether GGGCTA variant repeat structures (G4s or/and hairpins) challenged RPA binding compared to contiguous G4 structures formed by GGGTTA canonical repeats.

In potassium, RPA bound H70cta (A(GGGCTA)$_{11}$GGG sequence) less efficiently than its canonical counterpart H70tta (A(GGGTTA)$_{11}$GGG sequence) (figure 4A) (we define "binding efficiency" the percentage of DNA bound to RPA as a function of RPA concentration). Having demonstrated, for GGGCTA repeats in potassium, the existence of an equilibrium between H22cta-*like* G4 structures and duplexes, we wondered which of these structures impaired RPA efficient binding to GGGCTA repeats in potassium: the non-canonical H22cta-*like* G4 conformation, the duplex structure or both?

In KCl, we could not completely shift the equilibrium of H46cta, H70cta and H94cta toward the G4 or toward the duplex structure. Hence, in order to assess whether the non-canonical H22cta-*like* G4 conformation challenged RPA binding, we compared the binding of RPA to H22cta (A(GGGCTA)3GGG sequence) and to its canonical counterpart H22tta (A(GGGTTA)3GGG) in potassium; while, in order to assess whether the hairpin structure challenged RPA binding, we studied the binding of RPA to H70cta in sodium, where the equilibrium of H70cta was strongly shifted toward the hairpin structure. EMSA results showed that RPA bound the non-canonical G4 formed by the GGGCTA variant motif in potassium as efficiently as the G4 formed by the GGGTTA canonical motif (figure 4B), but it poorly bound the hairpin structure formed by H70cta compared to the contiguous G4 structure formed by its canonical counterpart H70tta (figure 4C). The binding curve of RPA to the H70cta hairpin in sodium can be transposed to the H70cta hairpin fraction in potassium: if RPA poorly binds the H70cta hairpin in sodium, it will also poorly bind the H70cta hairpin fraction in potassium, since *(i)* the stability of duplex structures is not affected by the nature of the cation (K^+, Na^+ or Li^+) present in solution (as shown in figure S5 for an oligonucleotide mimicking the H70cta hairpin but unable to fold into a G4) and *(ii)* also the affinity of RPA for unstructured DNA does not depend on the nature of the cation (K^+, Na^+ and Li^+) (23). Overall, these EMSA results supported that the decrease in binding efficiency of RPA to H70cta compared to H70tta in potassium (figure 4A) comes from the H70cta hairpin fraction and not from the H70cta G4 fraction. More broadly, these EMSA results demonstrated that, compared to canonical GGGTTA repeats structured in contiguous G4s, GGGCTA repeats challenge RPA binding when they fold into hairpins but not when they fold into G4s.

The heterotrimeric RPA binds DNA via four of its six OB-fold domains (A, B and C domains in RPA1 subunit and D domain in RPA2 subunit) in an oriented and sequential mode: RPA1 binds first, followed by RPA2 toward the 3' side of DNA (24). In a telomeric context, GGGCTA variant repeats are flanked by GGGTTA canonical repeats; hence RPA can be faced with GGGCTA hairpins when already partially bound to telomeric G-strand via RPA1 (figure 4D). In order to ascertain whether such an initial partial binding could increase the binding efficiency of RPA to GGGCTA repeats structured into hairpins, we carried out EMSA with GGGCTA repeats bearing a T10 tail at their 5' side, under conditions where the equilibrium was strongly shifted toward the hairpin structures, *i.e.* in sodium (figure 4D). A 10-nt unstructured sequence provides an anchoring site for RPA1 A and B domains (25)(26); its position at the 5' side of the DNA structure can allow, in principle, RPA 5'-to-3' directional

laying-out and unfolding the DNA structure to achieve the stable 30-nt binding mode, as we shown for other DNA structures (24). EMSA revealed that RPA poorly bound the hairpin structures formed by GGGCTA repeats even in the presence of an unstructured 10-nt single-stranded tail at the 5' side of the hairpins structures (figures 4E and 4F).

Overall, EMSA results in figure 4 demonstrated that RPA bound GGGCTA repeats folded into G4s as efficiently as GGGTTA repeats, but poorly bound GGGCTA repeats folded into hairpins, even when it can partially bind "in front" of the hairpin structure (*i.e.* at the 5' side) *via* its A and B domain.

A dozen of CCCTAG variant repeats within the C-strand do not fold into stable structures and are bound by RPA as efficiently as CCCTTA canonical repeats.

GGGCTA variant repeats within the telomeric G-strand correspond to CCCTAG repeats within the C-strand. In principle, CCCTAG repeats might fold into hairpins alternating four Watson-Crick base-pairs and a couple of CC base-pairs, similarly to GGGCTA repeats where four Watson-Crick base-pairs alternate with a couple of GG base-pairs. We investigated whether CCCTAG repeats folded into hairpin structures and challenged RPA binding, as GGGCTA repeats did. While a dozen of GGGCTA repeats (H70cta) were sufficient to fold into stable hairpins that impaired RPA binding, a dozen of CCCTAG repeats did not fold into stable structures (at physiological pH and temperatures) and were bound by RPA as efficiently as CCCTAA canonical repeats (figure S6). We cannot exclude that a higher number of CCCTAG variant repeats might induce a local structuration of the C-strand affecting RPA binding; nevertheless, this number would be beyond the ~10-repeat threshold at which GGGCTA/CCCTAG repeat instability was observed (11). This asymmetry in the behaviour of GGGCTA and CCCTAG repeats suggests that GGGCTA/CCCTAG repeat instability in telomeres comes from GGGCTA repeats within the G-strand and not from CCCTAG repeats within the C-strand and highlights the stabilising role of GG base-pairs in GGGCTA repeat hairpin structures.

GGGTCA stable variant repeats fold into structures that do not impair RPA binding.

We investigated a second variant motif, GGGTCA, that differs from the GGGCTA variant motif only for the position of the cytosine, but does not display the same level of instability (11).

As done for GGGCTA repeats, we investigated the structure of A(GGGTCA)*n=3,7,11,15*GGG oligonucleotides (named H22tca, H46tca, H70tca and H94tca) and their interaction with RPA. As GGGTTA canonical repeats and the very major fraction of GGGCTA variant repeats, also GGGTCA variant repeats folded into intramolecular structures, as shown by non-denaturing PAGE assay at 15 - 60 μM strand concentrations (figure S7).

Overall, TDS, CD spectra and melting profiles of GGGTCA repeats in potassium provided evidence of structuration into G4s (figures 5A and 5B), with melting profiles similar to the ones of canonical GGGTTA repeats (figure 5B inset) (13); nevertheless, CD spectra in potassium markedly varied with the number of repeats (figure 5B). We wondered whether GGGTCA repeats folded not only into G4s but also into a non-G4 structure, similarly to GGGCTA repeats. Experiences in LiCl, where G4s were destabilised, revealed that H70tca and H94tca oligonucleotides folded into a structure characterised by a non-G4 TDS signature, an "inverted" CD spectrum (with a positive band around 254 nm and a negative band around 278 nm) and a low thermal stability (*T*t about 20°C) (figures 5C and 5D). In NaCl, GGGTCA repeats displayed mixed TDS and CD signatures (figures 5E and 5F): the signatures of H46tca (blue spectra) were shifted toward the G4 signature displayed by H22tca (red spectra), while the signatures of H70tca and H94tca (green and black spectra) were shifted toward the "inverted" CD signature; melting profiles further supported the presence of more than one structure in NaCl (figure S8). So far, we do not yet have enough information to propose a model for this structure characterised by an "inverted" CD signature and we cannot exclude that a minor fraction of this structure may be present in potassium solution. Anyway, neither of the two GGGTCA repeats structures (G4 and non-G4) impaired RPA binding compared to GGGTTA canonical repeats, as shown by EMSA (figures 5G and 5H): RPA bound H70tca as efficiently as its canonical counterpart H70tta both in KCl (where the equilibrium of H70tca was shifted toward the G4 form) and in NaCl (where the equilibrium of H70tca was shifted toward the non-G4 form), likely because of the low stability of the non-G4 form at physiological temperatures.

In conclusion, the structuration of GGGTCA stable variant repeats in potassium appears to be not as regular as the structuration of GGGTTA canonical repeats into contiguous G4s; nevertheless, it does not challenge RPA binding, contrary to the hairpin form of GGGCTA unstable variant repeats.

DISCUSSION

Results presented in this study suggest working hypotheses about the origin of the length-dependent instability of GGGCTA telomeric variant motif (11) and inspire a few considerations about telomeric G4 and the evolution of telomeric motifs.

Results of our structural investigation demonstrate that, while the A(GGGCTA)$_3$GGG sequence folds into a G4 (as previously demonstrated (12)), longer sequences (A(GGGCTA)$_{7,11,15}$GGG) fold into G4 and into hairpin structures, alternating $^{5'}$CTAG/$^{3'}$GATC Watson-Crick base-pairs and GG/GG base-pairs (figure 1H). It is noteworthy that this kind of hairpin structure may be the intermediate structural state toward the folding of the A(GGGCTA)$_3$GGG sequence into the non-canonical chair-type G4 resolved in potassium solution (figures S1). In potassium, G4 and hairpin structures coexist and the hairpin fraction increases with increasing the number of GGGCTA repeats (figure 3); the presence of a fraction of duplex structures was evident only for A(GGGCTA)$_n$GGG with $n \geq 7$ (*i.e.* starting from H46cta sequence) (data not shown). The existence of an equilibrium between G4 and hairpin structures, transposed into a cellular context, means that sufficiently long GGGCTA arrays ((GGGCTA)$_{n \geq 8}$) can fold either in G4s or in hairpins (figure 6, left side of the diagram). In a simplified view (assuming that folding into G4 or into hairpins is stochastic), the probability *p* of folding into a hairpin corresponds to the fraction *f* of hairpin structure in solution (figure 3).

The importance of an efficient binding of RPA for G-strand synthesis (highlighted in the Introduction section), prompted us to test the binding of RPA to GGGCTA repeats structures. EMSA results show that RPA binds GGGCTA variant repeats as efficiently as GGGTTA canonical repeats when the GGGCTA repeats are structured into G4s, but less efficiently when they are structured into hairpins. The higher stability of GGGCTA hairpins compared to GGGTTA G4s (figure S10) may be one of the factors that make RPA less efficient in binding to hairpin structured GGGCTA repeats.

Overall our *in vitro* results suggest that GGGCTA hairpins (that, in potassium, we unambiguously observed only starting from eight GGGCTA repeats) may be the structures at the origin of the length-dependent instability of GGGCTA arrays (observed starting from a ten of repeats (11)); they also suggest that these structures may persist in G-strand because of poor RPA binding (figure 6). A link between telomeric variant repeat (in)stability, G-strand secondary structures and RPA binding is further suggested by the different behaviours of unstable GGGCTA variant repeats and stable GGGTCA variant repeats. Studies about the

instability of structured Short Tandem Repeats (STR) (such as trinucleotide repeats involved in a variety of neurodegenerative diseases) generally address a "downstream" question: how do STR structures generate instability? Our study provides a working hypothesis to answer to an "upstream" question: why STR structures are not resolved? which are the proteins involved in unfolding DNA secondary structures that fail in resolving STR structures? Our results suggest, as a working hypothesis, that RPA might be one of these proteins: failure of efficient binding of RPA to structured STR, might be the upstream factor in STR replication-dependent and transcription-dependent instability.

Along with RPA (17), other proteins involved in telomere maintenance can be faced with telomeric secondary structures; in particular, the single-strand binding trimeric complex CST and the helicases WRN, BLM, RTEL1 are supposed to play a major role in unfolding G4s at telomeres (27)(28)(29)(30)(31)(32)(33). *In vitro*, binding of RPA to telomeric sequences does not need previous unfolding of G4 by helicase activities: RPA binds telomeric sequences structured into contiguous G4, unfolding them, with an efficiency that is independent of the number of G4 units (18); RPA alone allows progression of polymerase through a telomeric G4, whereas helicase activities are needed in the presence of more stable G4s (34)(35). However, in cells, unfolding of telomeric G4 by helicases might improve the rate of binding of RPA to telomeric G-strand. Testing the ability of helicases involved in telomeres maintenance (WRN, BLM, RTEL1) and of CST to unwind GGGCTA structures will allow identifying which of these proteins may be potential "upstream" factors involved in GGGCTA repeats instability, besides RPA.

Our study inspires a few speculative considerations about telomeric G4 and the evolution of telomeric motif. Telomeric G4 are often considered as problematic structures for the replication machinery (28). This point of view relies on evidence of troubles in telomere replication (especially in G-strand) in cells deficient in proteins supposed to unfold telomeric G4s (the helicase WRN (27), BLM (28)(29) and RTEL1 (28)(30)(31), the SSB proteins RPA (17) and CST (32)(33)). But, when cellular machineries are functional, GGGTTA repeats in lagging strand are stable through replication cycles (36). Are telomeric G4 so problematic structures? With the exception of budding yeast motifs (that have the most variable set of telomeric repeats), telomeric motifs mostly consists of a short motif of the type $G_{2\text{-}4}T_{1\text{-}4}A_{0,1}$, exhibit a paucity in cytosines (compiled in the Telomerase Database (2)) and are prone to fold into G4 (3). The study of Royle and coll. (11) and our present study, together, highlight a link

between the instability of a telomeric variant motif and its potential to fold into hairpins. Overall, telomeric motif features (G4 prone) and the behavior of GGGCTA variant motif (unstable and hairpin-prone) suggest that evolution of telomeric motifs might have been driven by selection of G4-prone motifs and counter-selection of hairpin-prone motif (hence the paucity of cytosines).

On the basis of our previous and present results, we propose a different point of view about telomeric G4s; they are not so challenging structures, but rather "judicious" structures (easily manageable by DNA interacting proteins). Indeed, the contiguous G4 units in which telomeric sequences may fold have similar stability and do not interact with each other (13). Therefore, proteins dealing with such structures are faced with identical and independent structural units; this makes the interaction (the efficiency of binding or unfolding) independent of the number of G4 units, *i.e.* independent of the length of the tract of G-strand exposed as single-strand, as we showed for RPA (18). A recent study support, somehow, this point of view. Analysis of 20.000 Short Tandem Repeats showed that STR structure impact the frequency of expansion/contraction (37). Notably, unstructured STR are the most frequently expanded/contracted; whereas, among structured STR (hairpin-like, i-motif and G4), STR folding into G4 are the less frequently expanded/contracted. Hence, when considering STR, such as telomeric sequences, it seems that, to limit instability, "structured" is better that "unstructured" and "G4 structured" is better that "hairpin-structured". In conclusion, it may be that telomeric G4s are the secondary structures allowing the highest level of telomere stability. Maybe, telomeric G4s are telomeres' best friend.

Funding

This work was supported by Muséum National d'Histoire Naturelle (MNHN), Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM). J.C. was funded by a MESRI PhD fellowship awarded by MNHN Doctoral School. Funding for open access charge: INSERM.

Acknowledgment

We thank all the members of the *Nucleic Acid Structure, Telomeres and Evolution* team (present and past members), in particular Pauline Lejault for motivating us in our first steps in NMR measurements, Anthony Bugaut and Laureline Roger for enriching discussions. We thank Anne De Cian for gel staining and Phong Lan Thao Tran for testing one of our sequences in an original single molecule assay (to be continued ...). NMR spectra reported in this manuscript were acquired at the NMR facility of the Museum of National History, Paris (Plateau technique de Résonance Magnétique Nucléaire, UMR 7245 CNRS/MNHN, Molécules de Communication et d'Adaptation des Micro-organismes, Muséum national d'Histoire naturelle), directed by Alain Blond.

REFERENCES

1. de Lange,T. (2018) Shelterin-Mediated Telomere Protection. *Annu Rev Genet*, **52**, 223–247.
2. Podlevsky,J.D., Bley,C.J., Omana,R. V, Qi,X. and Chen,J.J. (2008) The telomerase database. *Nucleic Acids Res*, **36**, D339-43.
3. Tran,P.L., Mergny,J.L. and Alberti,P. (2011) Stability of telomeric G-quadruplexes. *Nucleic Acids Res*, **39**, 3282–3294.
4. Allshire,R.C., Dempster,M. and Hastie,N.D. (1989) Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res*, **17**, 4611–4627.
5. Baird,D.M., Jeffreys,A.J. and Royle,N.J. (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *EMBO J*, **14**, 5433–5443.
6. Coleman,J., Baird,D.M. and Royle,N.J. (1999) The plasticity of human telomeres demonstrated by a hypervariable telomere repeat array that is located on some copies of 16p and 16q. *Hum Mol Genet*, **8**, 1637–1646.
7. Baird,D.M., Coleman,J., Rosser,Z.H. and Royle,N.J. (2000) High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: implications for telomere biology and human evolution. *Am J Hum Genet*, **66**, 235–250.
8. Conomos,D., Stutz,M.D., Hills,M., Neumann,A.A., Bryan,T.M., Reddel,R.R. and Pickett,H.A. (2012) Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. *J Cell Biol*, **199**, 893–906.
9. Lee,M., Hills,M., Conomos,D., Stutz,M.D., Dagg,R.A., Lau,L.M., Reddel,R.R. and Pickett,H.A. (2014) Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. *Nucleic Acids Res*, **42**, 1733–1746.
10. Lee,M., Teber,E.T., Holmes,O., Nones,K., Patch,A.M., Dagg,R.A., Lau,L.M.S., Lee,J.H., Napier,C.E., Arthur,J.W., *et al.* (2018) Telomere sequence content can be used to determine ALT activity in tumours. *Nucleic Acids Res*, **46**, 4903–4918.
11. Mendez-Bermudez,A., Hills,M., Pickett,H.A., Phan,A.T., Mergny,J.L., Riou,J.F. and Royle,N.J. (2009) Human telomeres that contain (CTAGGG)n repeats show replication dependent instability in somatic cells and the male germline. *Nucleic Acids Res*, **37**, 6225–6238.
12. Lim,K.W., Alberti,P., Guedin,A., Lacroix,L., Riou,J.F., Royle,N.J., Mergny,J.L. and Phan,A.T. (2009) Sequence variant (CTAGGG)n in the human telomere favors a G-

quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res*, **37**, 6239–6248.

13. Bugaut,A. and Alberti,P. (2015) Understanding the stability of DNA G-quadruplex units in long human telomeric strands. *Biochimie*, **113**, 125–133.
14. Chen,R. and Wold,M.S. (2014) Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. *Bioessays*, **36**, 1156–1161.
15. Verdun,R.E. and Karlseder,J. (2006) The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres. *Cell*, **127**, 709–720.
16. Moser,B.A., Subramanian,L., Chang,Y.T., Noguchi,C., Noguchi,E. and Nakamura,T.M. (2009) Differential arrival of leading and lagging strand DNA polymerases at fission yeast telomeres. *EMBO J*, **28**, 810–820.
17. Audry,J., Maestroni,L., Delagoutte,E., Gauthier,T., Nakamura,T.M., Gachet,Y., Saintome,C., Geli,V. and Coulon,S. (2015) RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *EMBO J*, **34**, 1942–1958.
18. Lancrey,A., Safa,L., Chatain,J., Delagoutte,E., Riou,J.F., Alberti,P. and Saintome,C. (2018) The binding efficiency of RPA to telomeric G-strands folded into contiguous G-quadruplexes is independent of the number of G4 units. *Biochimie*, **146**, 68–72.
19. Mergny,J.L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res*, **33**, e138.
20. Saintome,C., Amrane,S., Mergny,J.L. and Alberti,P. (2016) The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. *Nucleic Acids Res*, **44**, 2926–2935.
21. Masiero,S., Trotta,R., Pieraccini,S., De Tito,S., Perone,R., Randazzo,A. and Spada,G.P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org Biomol Chem*, **8**, 2683–2692.
22. Bhattacharya,P.K., Cha,J. and Barton,J.K. (2002) 1H NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches. *Nucleic Acids Res*, **30**, 4740–4750.
23. Safa,L., Delagoutte,E., Petruseva,I., Alberti,P., Lavrik,O., Riou,J.F. and Saintome,C. (2014) Binding polarity of RPA to telomeric sequences and influence of G-quadruplex stability. *Biochimie*, **103**, 80–88.
24. Safa,L., Gueddouda,N.M., Thiebaut,F., Delagoutte,E., Petruseva,I., Lavrik,O.,

Mendoza,O., Bourdoncle,A., Alberti,P., Riou,J.F., *et al.* (2016) 5' to 3' Unfolding Directionality of DNA Secondary Structures by Replication Protein A: G-QUADRUPLEXES AND DUPLEXES. *J Biol Chem*, **291**, 21246–21256.

25. Blackwell,L.J. and Borowiec,J.A. (1994) Human replication protein A binds single-stranded DNA in two distinct complexes. *Mol Cell Biol*, **14**, 3993–4001.
26. Bochkarev,A., Pfuetzner,R.A., Edwards,A.M. and Frappier,L. (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature*, **385**, 176–181.
27. Crabbe,L., Verdun,R.E., Haggblom,C.I. and Karlseder,J. (2004) Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science (80-.).*, **306**, 1951–1953.
28. Sfeir,A., Kosiyatrakul,S.T., Hockemeyer,D., MacRae,S.L., Karlseder,J., Schildkraut,C.L. and de Lange,T. (2009) Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*, **138**, 90–103.
29. Zimmermann,M., Kibe,T., Kabir,S. and de Lange,T. (2014) TRF1 negotiates TTAGGG repeat-associated replication problems by recruiting the BLM helicase and the TPP1/POT1 repressor of ATR signaling. *Genes Dev*, **28**, 2477–2491.
30. Vannier,J.B., Pavicic-Kaltenbrunner,V., Petalcorin,M.I., Ding,H. and Boulton,S.J. (2012) RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell*, **149**, 795–806.
31. Vannier,J.B., Sandhu,S., Petalcorin,M.I., Wu,X., Nabi,Z., Ding,H. and Boulton,S.J. (2013) RTEL1 is a replisome-associated helicase that promotes telomere and genome-wide replication. *Science (80-.).*, **342**, 239–242.
32. Huang,C., Dai,X. and Chai,W. (2012) Human Stn1 protects telomere integrity by promoting efficient lagging-strand synthesis at telomeres and mediating C-strand fill-in. *Cell Res*, **22**, 1681–1695.
33. Zhang,M., Wang,B., Li,T., Liu,R., Xiao,Y., Geng,X., Li,G., Liu,Q., Price,C.M., Liu,Y., *et al.* (2019) Mammalian CST averts replication failure by preventing G-quadruplex accumulation. *Nucleic Acids Res*, **47**, 5243–5259.
34. Dahan,D., Tsirkas,I., Dovrat,D., Sparks,M.A., Singh,S.P., Galletto,R. and Aharoni,A. (2018) Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Res*, **46**, 11847–11857.
35. Sparks,M.A., Singh,S.P., Burgers,P.M. and Galletto,R. (2019) Complementary roles of Pif1 helicase and single stranded DNA binding proteins in stimulating DNA replication through G-quadruplexes. *Nucleic Acids Res*, **47**, 8595–8605.

36. Damerla,R.R., Knickelbein,K.E., Kepchia,D., Jackson,A., Armitage,B.A., Eckert,K.A. and Opresko,P.L. (2010) Telomeric repeat mutagenicity in human somatic cells is modulated by repeat orientation and G-quadruplex stability. *DNA Repair*, **9**, 1119–1129.
37. Murat,P., Guilbaud,G. and Sale,J.E. (2020) DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome Biol*, **21**, 209.
38. Nguyen,H.D., Yadav,T., Giri,S., Saez,B., Graubert,T.A. and Zou,L. (2017) Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. *Mol Cell*, **65**, 832-847 e4.
39. Freudenreich,C.H. (2018) R-loops: targets for nuclease cleavage and repeat instability. *Curr Genet*, **64**, 789–794.

FIGURE LEGENDS

Figure 1. **UV-absorption investigation of GGGCTA repeat structures**

(A-F) TDS and CD spectra of H22cta 6 μM (red), H46cta 3 μM (blue), H70cta 2 μM (green) and H94cta 1.5 μM (black) in 100 mM KCl **(A, B)**, LiCl **(C, D)** and NaCl **(E, F)**.

Panel D inset: normalised melting profiles at 245 nm of H46cta 3 μM (blue), H70cta 2 μM (green) and H94cta 1.5 μM (black) in LiCl 100 mM (*T*t values range from 54°C for H46cta to 57°C for H94cta.

Panel B and panel F insets: normalised melting profiles at 295 nm of the G4 structures formed by H22cta 6 μM in 100 mM KCl and NaCl (*T*t 64°C and 55°C, respectively); in LiCl, H22cta does not fold into any stable structure.

Melting profiles of H46cta, H70cta and H94cta in KCl and NaCl are not shown: they are more complex and wavelength-dependent, consistently with the presence of more than one structure.

(G) CD spectra of H46cta 3 μM in KCl 100 mM (red), KCl 50 mM (green), KCl 10 mM (blue) and in LiCl 100 mM (dotted black line).

(H) H94cta putative hairpin structure and dxGG duplex.

(I-K) CD spectra of the G4 formed by H22cta 6 μM (red), of the duplex formed by dxGG 10 μM (blue) and of H94cta 1.5 μM (black bold line) in 100 mM KCl **(I)**, LiCl **(J)** and NaCl **(K).**

Figure 2. **NMR investigation of GGGCTA repeat structures**

NMR spectra of: H22cta 160 μM in KCl 10 mM **(A)**; H46cta 100 μM in KCl 10 mM **(B)**, KCl 50 mM **(C)**, NaCl 50 mM **(D)** and LiCl 50 mM **(E)**; dxGG 250 μM in KCl 50 mM **(F)**.

WC/H is the ratio between the integral of Watson-Crick peaks and the integral of Hoogsteen peaks.

In **panel A**, the folding of H22cta in KCl from (12) is schematically illustrated: the two G-tetrads are depicted in black, the non-canonical Watson-Crick GCGC tetrad is depicted in red; the peak around 13.5 ppm had been assigned to C11•G14 Watson-Crick base-pairing in the central loop of the G4, the two peaks around 12.9 ppm had been assigned to G8•C17 and G20•C5 Watson-Crick base-pairing in the GCGC tetrad (12).

Figure 3. **Estimated GGGCTA repeats duplex fractions**

Duplex fractions *f* of oligonucleotides H46cta, H70cta and H94cta in NaCl and KCl estimated from CD spectra in in figure 1 (recorded at 5°C), as detailed in figure S4 (similar fractions were obtained from CD spectra recorded at 20°C and 37°C).

Figure 4. **Binding efficiency of RPA to GGGCTA variant repeats versus GGGTTA canonical repeats**

(A-C) Percentage of radiolabelled oligonucleotide bound to RPA calculated from EMSA: **(A)** H70tta (black circles) and H70cta (red squares) in KCl 100 mM; **(B)** H22tta (black circles) and H22cta (red squares) in KCl 100 mM; **(C)** H70tta (black circles) and H70cta (red squares) in NaCl 100 mM. Error bars are standard deviations from six assays. For each oligonucleotide, one representative EMSA gel is shown above the corresponding plot.

RPA, in its extended binding-mode, covers 25-30 nt; for 70mer oligonucleotides, the shifted bands correspond to RPA:DNA complexes of increasing stoichiometry (1:1, 2:1 and 3:1).

GGGTTA canonical repeat structures from our previous studies (13)(20) and GGGCTA variant repeat structures from the present study are schematically illustrated. The Watson-Crick GCGC tetrad in H22cta G4 structure (12) is depicted in red.

(D) RPA can be faced with GGGCTA hairpins when already partially bound to telomeric DNA *via* its RPA1 subunit. Considering the oriented and sequential binding mode of RPA to DNA, in order to mimic this situation, we designed GGGCTA repeat hairpins bearing a T10 tail at their 5' side.

(E, F) Percentage of radiolabelled oligonucleotide bound to RPA from EMSA experiments in NaCl 100 mM: **(E)** T10-H46tta (black circles) and T10-H46cta (red squares), **(F)** T10-H70tta (black circles) and T10-H70cta (red squares). Error bars are standard deviations from three assays.

Representative gels are show in figures S9A and S9B.

Figure 5. **UV-absorption investigation of GGGTCA repeat structures**

(A-F) TDS and CD spectra of H22tca 6 μM (red), H46tca 3 μM (blue), H70tca 2 μM (green) and H94tca 1.5 μM (black) in 100 mM KCl **(A, B)**, LiCl **(C, D)** and NaCl **(E, F)**.

Panel B inset: normalised melting profiles of H46cta 3 μM (blue), H70cta 2 μM (green) and H94cta 1.5 μM (black) in KCl 100 mM.

Panel D inset: normalised melting profiles of H70cta 2 μM (green) and H94cta 1.5 μM (black) in LiCl 100 mM (H22tca and H46tca in LiCl do not fold into stable structures).
Melting profiles in NaCl are shown in figure S8: they are more complex and indicative of the presence of a G4 and a non G4-structure.
(G, H) Percentage of oligonucleotide bound to RPA from EMSA experiments: **(G)** H70tta (black circles) and H70tca (red squares) in KCl 100 mM; **(H)** H70tta (black circles) and H70tca (red squares) in NaCl 100 mM. Error bars are standard deviations from six assays.
Representative gels are show in figure S9C.

Figure 6

Hypothetical origin of GGGCTA repeat length-dependent instability

When tracts of the G-strand are transiently exposed as single-strands, (GGGCTA)n≥8 arrays can fold into G4s or into hairpins (diagram left side). G4s do not challenge RPA binding, whereas GGGCTA hairpins impairs RPA binding (diagram right side).
The structuration of GGGCTA arrays into hairpins and the poor binding of RPA to these structures may result into unresolved hairpin structures, that can then lead to to gains or losses of GGGCTA repeats *via* a variety of possible mechanisms.
GGGCTA instability has been suggested to be replication-dependent (11). The poor binding of RPA to GGGCTA hairpins might leads to the persistence of such structures also in R-loops generated during the transcription of *telomere repeat-containing RNA* (TERRA) from the C-strand Indeed, RPA colocalises with R-loops (38) and processing of unresolved structures in R-loops can lead to transcription-dependent instability of Short Tandem Repeats (39).

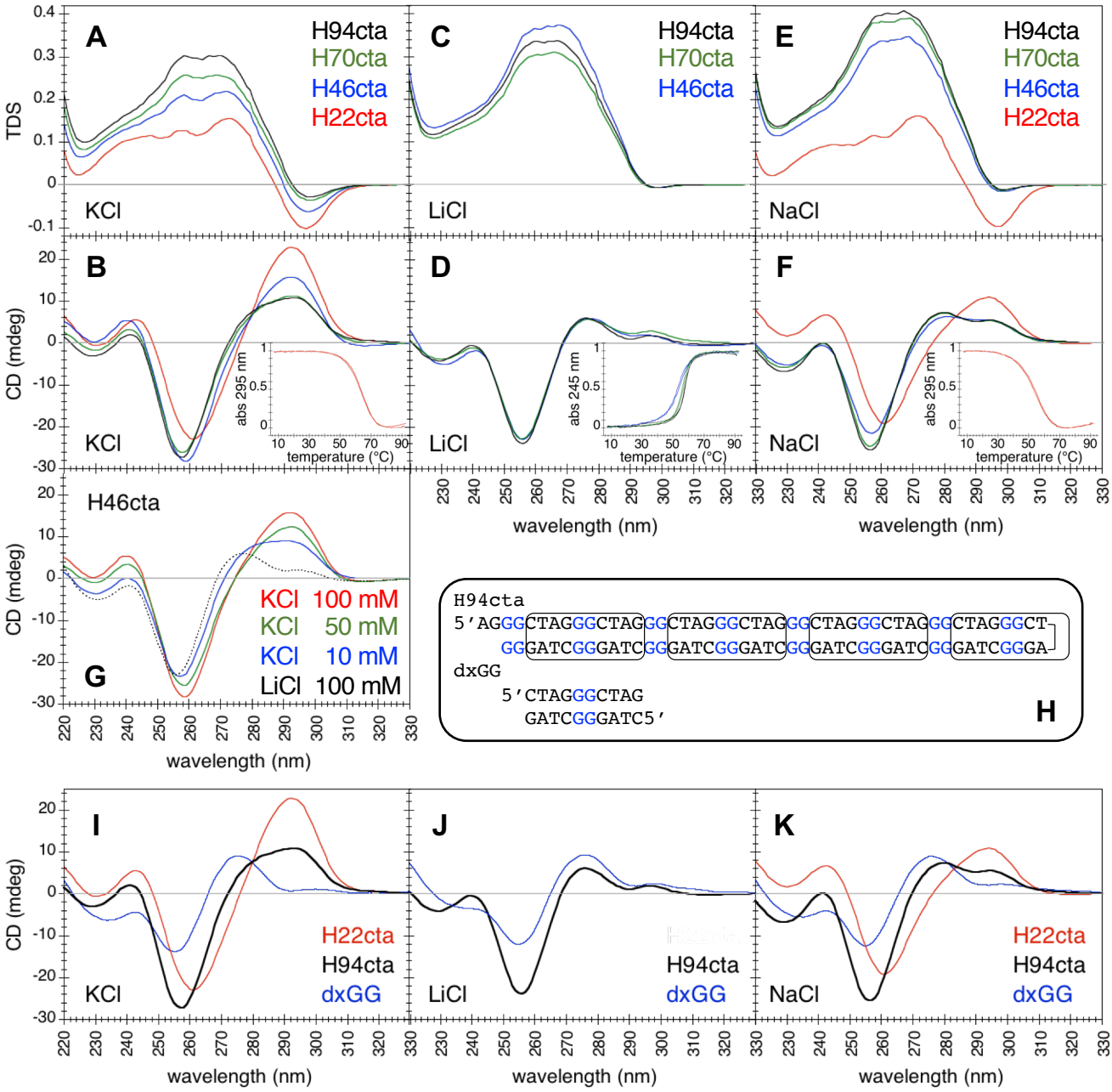


Figure 1
UV-absorption investigation of GGGCTA repeat structures

(A-F) TDS and CD spectra of H22cta 6 μM (red), H46cta 3 μM (blue), H70cta 2 μM (green) and H94cta 1.5 μM (black) in 100 mM KCl **(A, B)**, LiCl **(C, D)** and NaCl **(E, F)**.

Panel D inset: normalised melting profiles at 245 nm of H46cta 3 μM (blue), H70cta 2 μM (green) and H94cta 1.5 μM (black) in LiCl 100 mM (*T*t values range from 54°C for H46cta to 57°C for H94cta.

Panel B and panel F insets: normalised melting profiles at 295 nm of the G4 structures formed by H22cta 6 μM in 100 mM KCl and NaCl (*T*t 64°C and 55°C, respectively); in LiCl, H22cta does not fold into any stable structure.

Melting profiles of H46cta, H70cta and H94cta in KCl and NaCl are not shown: they are more complex and wavelength-dependent, consistently with the presence of more than one structure.

(G) CD spectra of H46cta 3 μM in KCl 100 mM (red), KCl 50 mM (green), KCl 10 mM (blue) and in LiCl 100 mM (dotted black line).

(H) H94cta putative hairpin structure and dxGG duplex.

(I-K) CD spectra of the G4 formed by H22cta 6 μM (red), of the duplex formed by dxGG 10 μM (blue) and of H94cta 1.5 μM (black bold line) in 100 mM KCl **(I)**, LiCl **(J)** and NaCl **(K).**

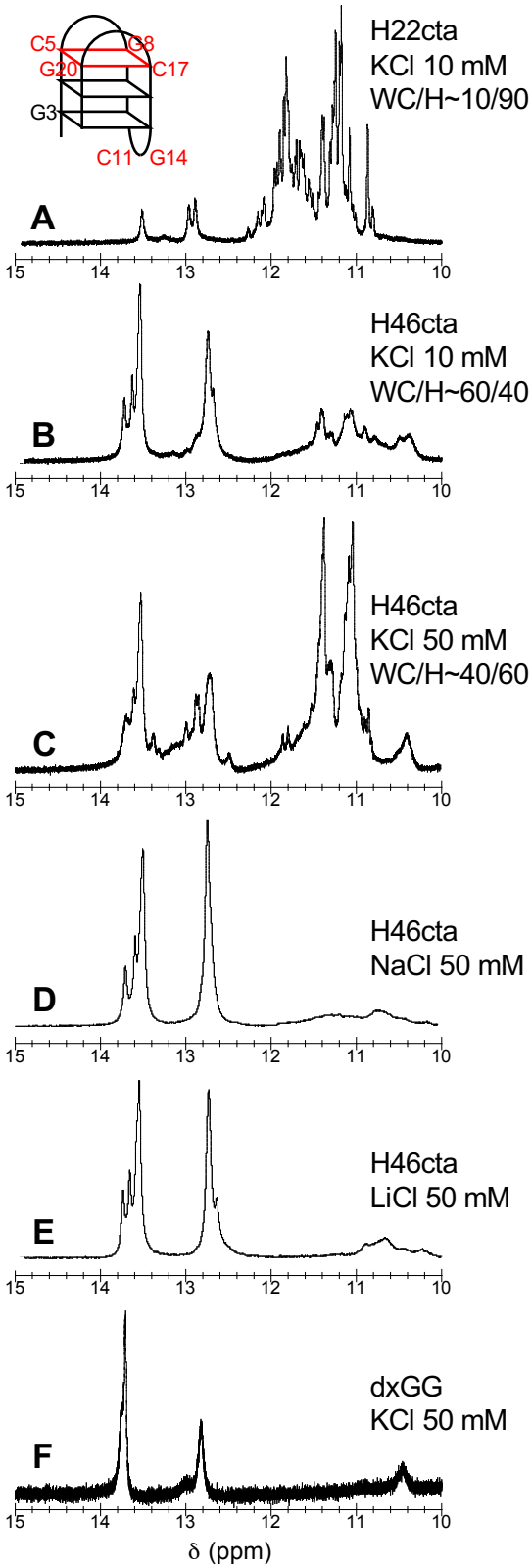


Figure 2
NMR investigation of GGGCTA repeat structures

NMR spectra of: H22cta 160 μM in KCl 10 mM **(A)**; H46cta 100 μM in KCl 10 mM **(B)**, KCl 50 mM **(C)**, NaCl 50 mM **(D)** and LiCl 50 mM **(E)**; dxGG 250 μM in KCl 50 mM **(F)**.
WC/H is the ratio between the integral of Watson-Crick peaks and the integral of Hoogsteen peaks.
In **panel A**, the folding of H22cta in KCl from (12) is schematically illustrated: the two G-tetrads are depicted in black, the non-canonical Watson-Crick GCGC tetrad is depicted in red; the peak around 13.5 ppm had been assigned to C_{11}•G_{14} Watson-Crick base-pairing in the central loop of the G4, the two peaks around 12.9 ppm had been assigned to G_8•C_{17} and G_{20}•C_5 Watson-Crick base-pairing in the GCGC tetrad (12).

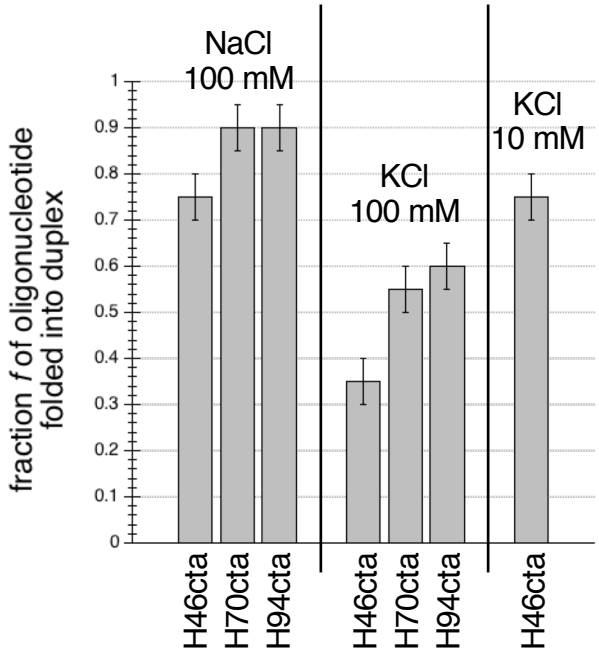


Figure 3
Estimated GGGCTA repeats duplex fractions

Duplex fractions *f* of oligonucleotides H46cta, H70cta and H94cta in NaCl and KCl estimated from CD spectra in in figure 1 (recorded at 5°C), as detailed in figure S4 (similar fractions were obtained from CD spectra recorded at 20°C and 37°C).

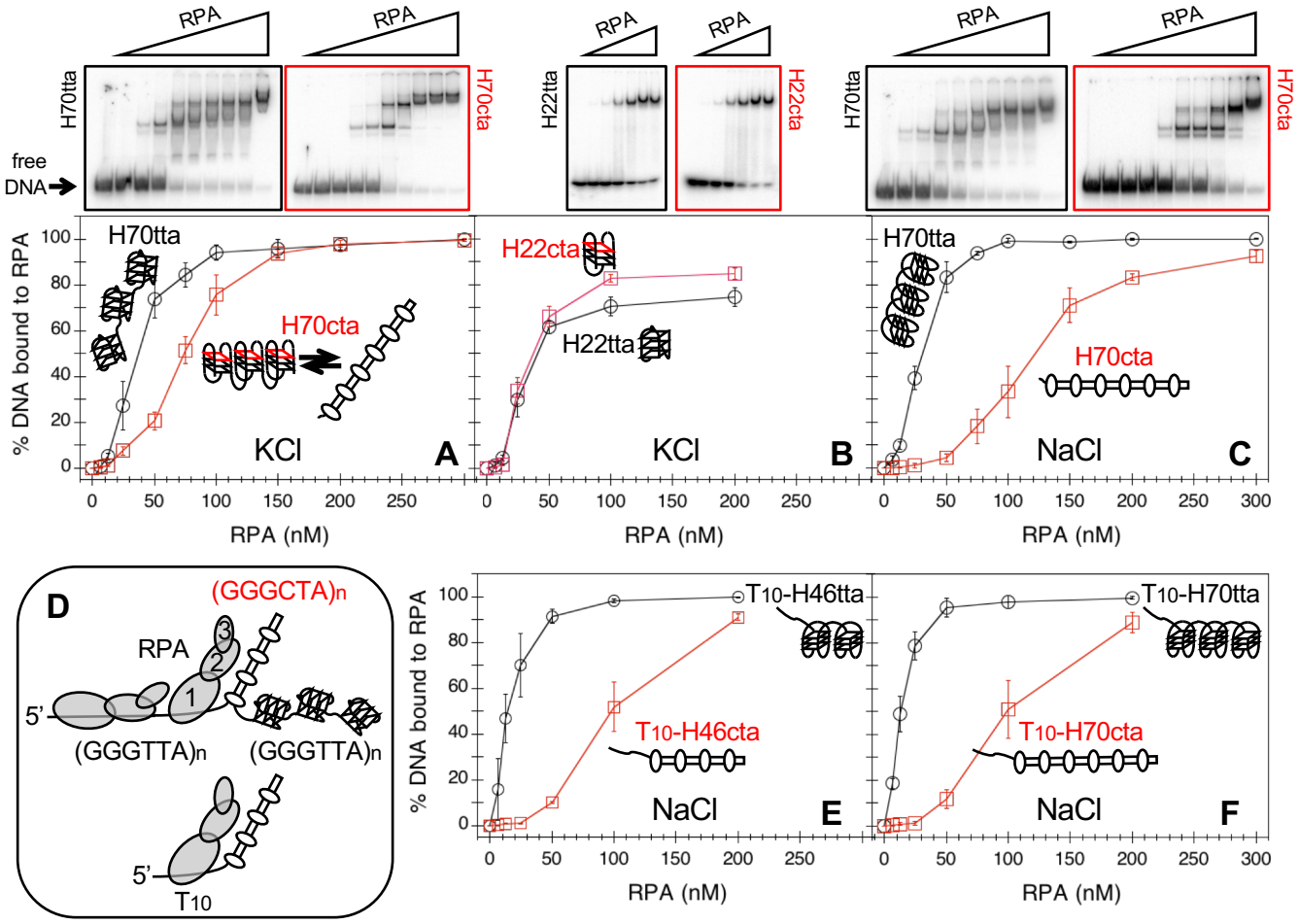


Figure 4

Binding efficiency of RPA to GGGCTA variant repeats versus GGGTTA canonical repeats

(A-C) Percentage of radiolabelled oligonucleotide bound to RPA calculated from EMSA: **(A)** H70tta (black circles) and H70cta (red squares) in KCl 100 mM; **(B)** H22tta (black circles) and H22cta (red squares) in KCl 100 mM; **(C)** H70tta (black circles) and H70cta (red squares) in NaCl 100 mM. Error bars are standard deviations from six assays. For each oligonucleotide, one representative EMSA gel is shown above the corresponding plot.

RPA, in its extended binding-mode, covers 25-30 nt; for 70mer oligonucleotides, the shifted bands correspond to RPA:DNA complexes of increasing stoichiometry (1:1, 2:1 and 3:1).

GGGTTA canonical repeat structures from our previous studies (13)(20) and GGGCTA variant repeat structures from the present study are schematically illustrated. The Watson-Crick GCGC tetrad in H22cta G4 structure (12) is depicted in red.

(D) RPA can be faced with GGGCTA hairpins when already partially bound to telomeric DNA *via* its RPA1 subunit. Considering the oriented and sequential binding mode of RPA to DNA, in order to mimic this situation we designed GGGCTA repeat hairpins bearing a T10 tail at their 5' side.

(E, F) Percentage of radiolabelled oligonucleotide bound to RPA from EMSA experiments in NaCl 100 mM: **(E)** T10-H46tta (black circles) and T10-H46cta (red squares), **(F)** T10-H70tta (black circles) and T10-H70cta (red squares). Error bars are standard deviations from three assays.

Representative gels are show in figures S7A and S7B.

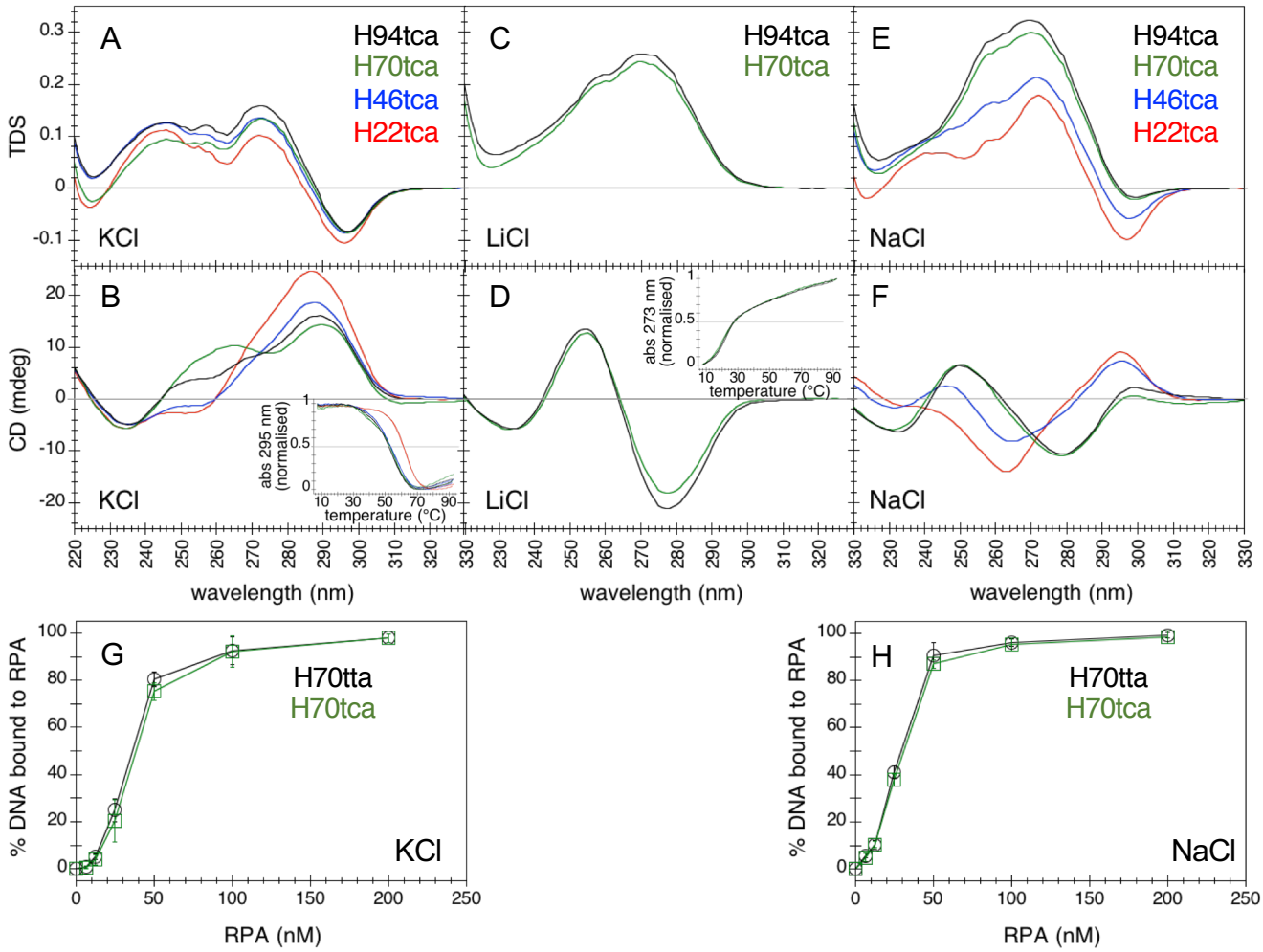


Figure 5
UV-absorption investigation of GGGTCA repeat structures

(A-F) TDS and CD spectra of H22tca 6 μM (red), H46tca 3 μM (blue), H70tca 2 μM (green) and H94tca 1.5 μM (black) in 100 mM KCl **(A, B)**, LiCl **(C, D)** and NaCl **(E, F)**.
Panel B inset: normalised melting profiles of H46cta 3 μM (blue), H70cta 2 μM (green) and H94cta 1.5 μM (black) in KCl 100 mM.
Panel D inset: normalised melting profiles of H70cta 2 μM (green) and H94cta 1.5 μM (black) in LiCl 100 mM (H22tca and H46tca in LiCl do not fold into stable structures).
Melting profiles in NaCl are shown in figure S9: they are more complex and indicative of the presence of a G4 and a non G4-structure.
(G, H) Percentage of oligonucleotide bound to RPA from EMSA experiments: **(G)** H70tta (black circles) and H70tca (red squares) in KCl 100 mM; **(H)** H70tta (black circles) and H70tca (red squares) in NaCl 100 mM. Error bars are standard deviations from six assays.
Representative gels are show in figure S7C.

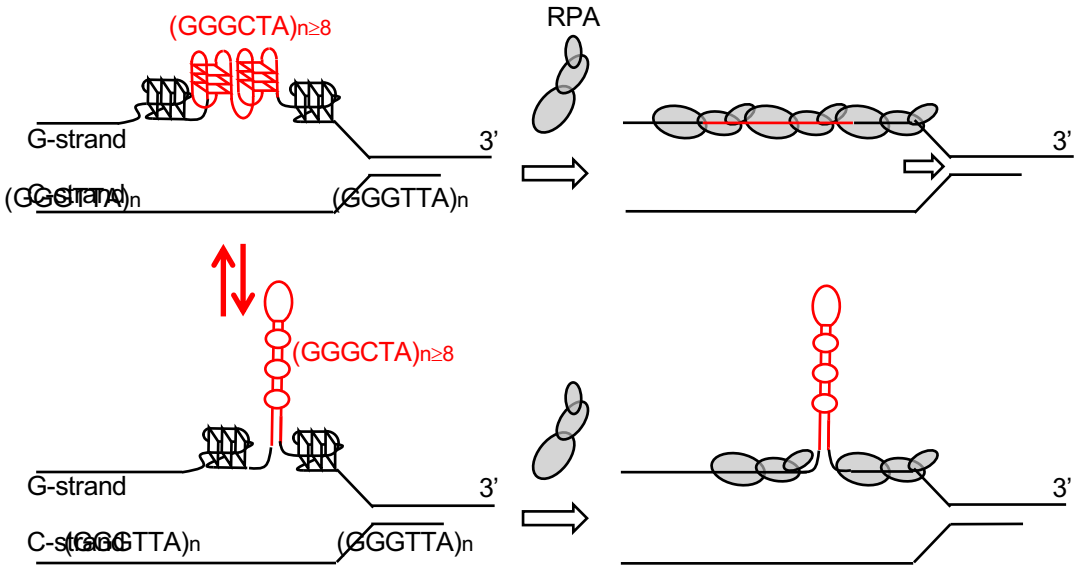


Figure 6
Hypothetical origin of GGGCTA repeat length-dependent instability

When tracts of the G-strand are transiently exposed as single-strands, $(GGGCTA)_{n\geq 8}$ arrays can fold into G4s or into hairpins (diagram left side). G4s do not challenge RPA binding, whereas GGGCTA hairpins impairs RPA binding (diagram right side).
The structuration of GGGCTA arrays into hairpins and the poor binding of RPA to these structures may result into unresolved hairpin structures, that can then lead to to gains or losses of GGGCTA repeats *via* a variety of possible mechanisms.
GGGCTA instability has been suggested to be replication-dependent (11). The poor binding of RPA to GGGCTA hairpins might leads to the persistence of such structures also in R-loops generated during the transcription of *telomere repeat-containing RNA* (TERRA) from the C-strand Indeed, RPA colocalises with R-loops (38) and processing of unresolved structures in R-loops can lead to transcription-dependent instability of Short Tandem Repeatss (39).

SUPPLEMENTARY DATA

$(GGGCTA)_{n \geq 8}$ repeats can fold into hairpins poorly unfolded by Replication Protein A: a possible origin of the length-dependent instability of GGGCTA variant repeats in human telomeres.

Jean Chatain [1], Alain Blond [3], Anh Tuân Phan [4,5], Carole Saintomé [1,2], Patrizia Alberti [1]*

[1] Laboratoire Structure et Instabilité des Génomes (StrInG), Muséum national d'Histoire naturelle, CNRS, Inserm, Paris 75005, France

[2] Sorbonne Université, UFR927, F-75005, Paris, France

[3] Laboratoire Molécules de Communication et Adaptation des Microorganismes (MCAM), Muséum national d'Histoire naturelle, CNRS, 75005 Paris, France.

[4] School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore;

[5] NTU Institute of Structural Biology, Nanyang Technological University, Singapore 636921, Singapore

* To whom correspondence should be addressed: patrizia.alberti@mnhn.fr

Table S1. Names and sequences of oligonucleotides used in this study

Name	*Sequence from 5′ to 3′*
H22cta	A(GGGCTA)$_3$GGG
H46cta	A(GGGCTA)$_7$GGG
H70cta	A(GGGCTA)$_{11}$GGG
H94cta	A(GGGCTA)$_{15}$GGG
H22tta	A(GGGTTA)$_3$GGG
H70tta	A(GGGTTA)$_{11}$GGG
T10-H46cta	T$_{10}$A(GGGCTA)$_7$GGG
T10-H70cta	T1$_0$A(GGGCTA)$_{11}$GGG
T10-H46tta	T$_{10}$A(GGGTTA)$_7$GGG
T10-H70tta	T$_{10}$A(GGGTTA)$_{11}$GGG
H22tca	A(GGGTCA)$_3$GGG
H46tca	A(GGGTCA)$_7$GGG
H70tca	A(GGGTCA)$_{11}$GGG
H94tca	A(GGGTCA)$_{15}$GGG
H21tta	(GGGTTA)$_3$GGG
H45tta	(GGGTTA)$_7$GGG
H69tta	(GGGTTA)$_{11}$GGG
H93tta	(GGGTTA)$_{15}$GGG
dxGG	CTAGGGCTAG (autocomplementary)
dx22-strand1 dx22-strand2	AGTGCTAGGGATCGGGCTACGGAT ATGGGTAGGGCGATGGCTAGCA
dx44-strand1 dx44-strand2	AGTGCTAGGGATCGGGCTACGGATAGGGATCGGGCTACGGATCGAC AGGTCGATGGGTAGGGCGATGGCTATGGGTAGGGCGATGGCTAGCA
GTTCTA	TTCTA(GTTCTA)$_{10}$GTT
H22cta-complementary sequence	(CCCTAG)$_3$CCCT
H46cta-complementary sequence	(CCCTAG)$_7$CCCT
H70cta-complementary sequence	(CCCTAG)$_{11}$CCCT
H70tta-complementary sequence	(CCCTAA)$_{11}$CCCT

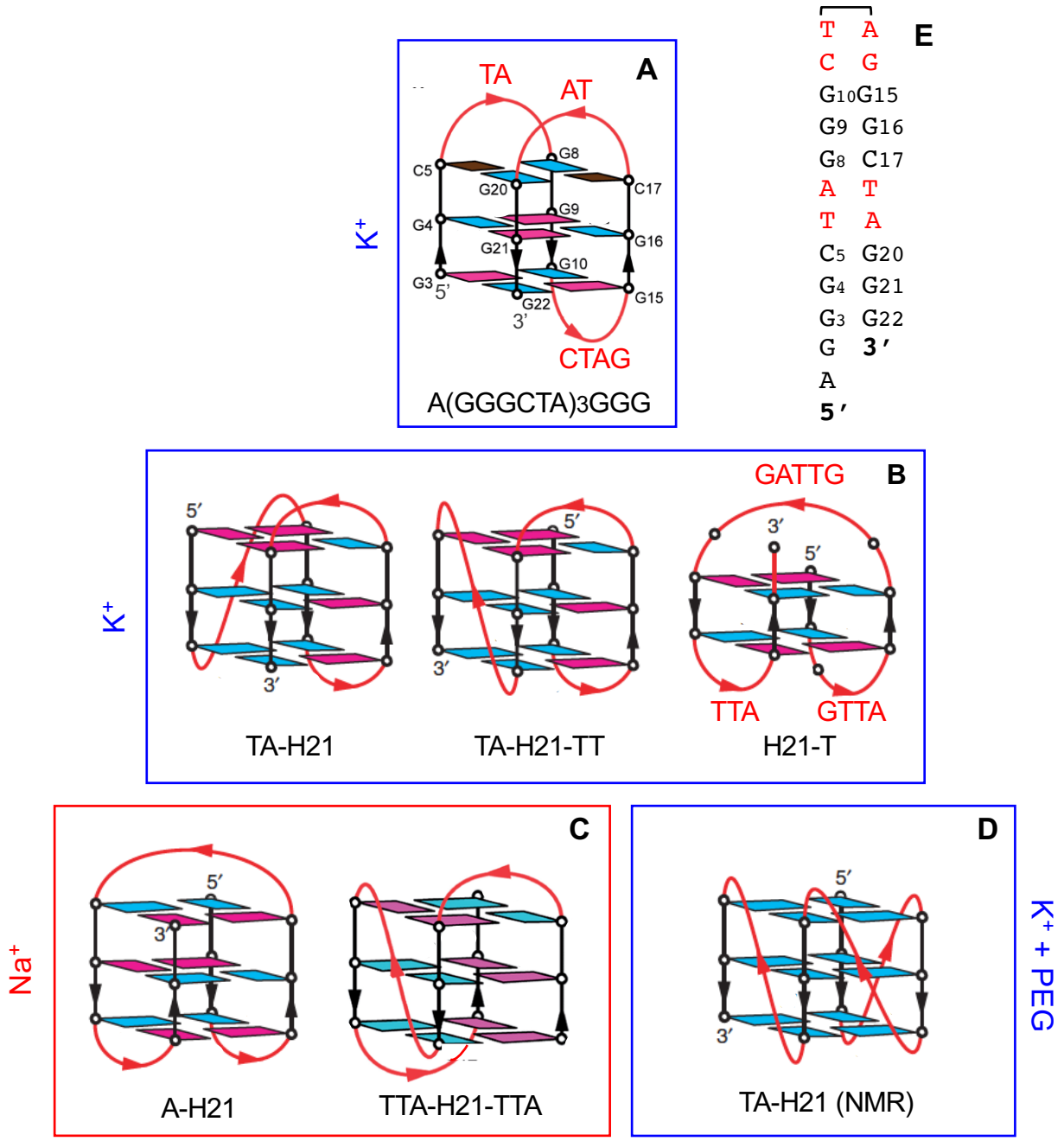


Figure S1
Schematic structures of the human telomeric G4s formed in solution by the GGGCTA variant motif and the GGGTTA canonical motif.

(A) NMR structure of the chair-type G4 formed by A(G3CTA)3G3 in K^+ (1). This chair-type G4 has two G- tetrads and a GCGC tetrad, relying on G20•C5 and G8•C17 Watson-Crick base pairs.
(B, C) NMR structures of the G4 formed by the (GGGTTA)3G3 core sequence (H21) (with different flanking nucleotides) in K^+ (2-4) **(B)** and in Na^+ (5, 6) **(C)**.
(D) NMR structure of the G4 formed by the (GGGTTA)3G3 core sequence (H21) in K^+ and polyethylene glycol (PEG) (7). The same folding was resolved by X-ray crystallography for the A-H21 sequences in K^+ (where initial drop conditions included PEG and BRACO19 ligand) (8).
(E) Putative hairpin structure that might be the intermediate structure toward the folding into the G4 formed by GGGCTA repeats.
Anti and *syn* guanines are depicted in cyan and magenta respectively.
G4 schematic illustrations are adapted from (9, 6).

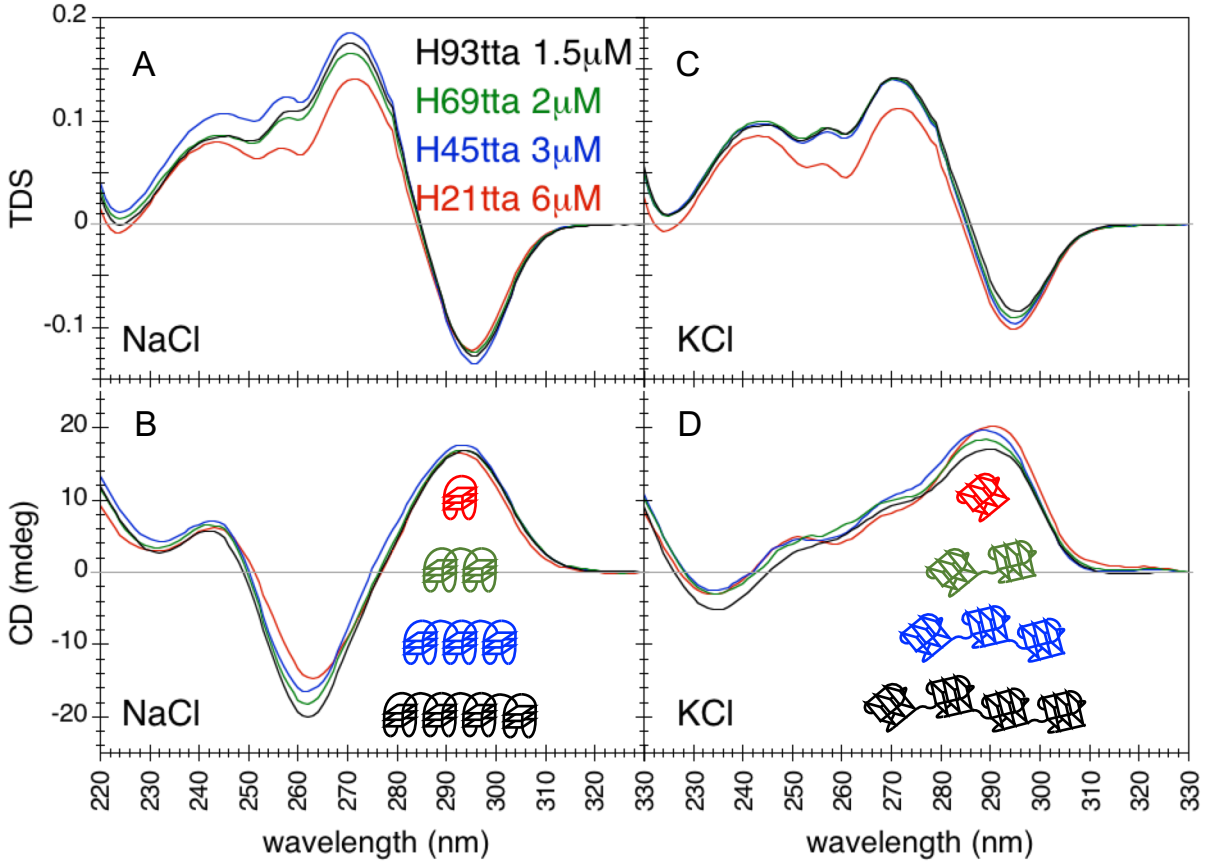


Figure S2
TDS and CD spectra of GGGTTA canonical repeats

TDS and CD spectra of (GGGTTA)*n=3,7,11,15*GGG oligonucleotides (named H21tta, H45tta, H69tta and H93tta, respectively) in 100 mM NaCl **(A, B)** and KCl **(C, D)** at the following strand concentrations: 6 μM H22tta, 3 μM H46tta, 2 μM H70tta and 1.5 μM H94tta (corresponding to 6 μM G4 units for each oligonucleotide).
Spectra are from our previous studies (10, 11).

H46cta putative 22-bp hairpin:
5′AGGGCTAGGGCTAGGGCTAGGGCT
GGGATCGGGATCGGGATCGGGA

H46cta putative 44-bp intermolecular duplex:
5′AGGGCTAGGGCTAGGGCTAGGGCTAGGGCTAGGGCTAGGGCTAGGG
GGGATCGGGATCGGGATCGGGATCGGGATCGGGATCGGGATCGGGA5′

dx22:
5′AGTGCTAGGGATCGGGCTACGGAT
ACGATCGGTAGCGGGATGGGTA5′

dx44:
5′AGTGCTAGGGATCGGGCTACGGATAGGGATCGGGCTACGGATCGAC
ACGATCGGTAGCGGGATGGGTATCGGTAGCGGGATGGGTAGCTGGA5′

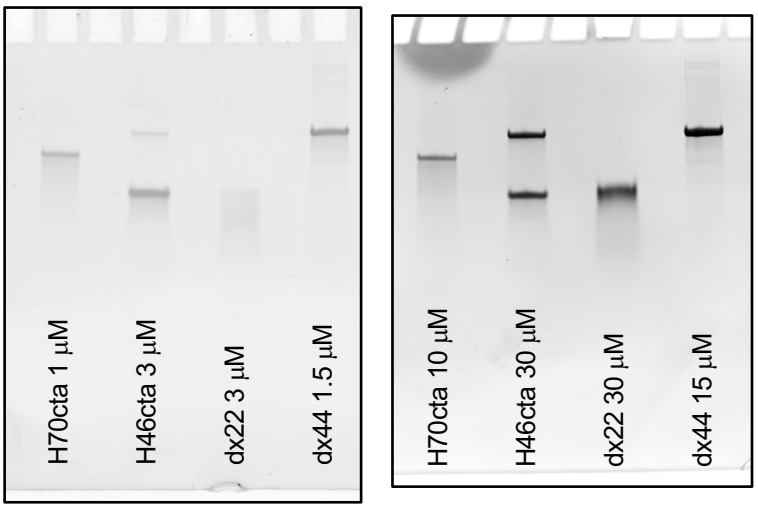


Figure S3
Molecularity of duplex structures formed by GGGCTA variant repeats

PAGE of H70cta, H46cta, dx22 and dx44 in LiCl 100 mM at the concentrations indicated in gel panels. Gel in the left panel was stained with ethidium bromide, gel in the right panel was stained with methylene bleu.

In principle H46cta can fold into a 22-bp hairpin or into a 44-bp intermolecular duplex, alternating four Watson-Crick bp and two GG bp. Similarly, H70cta can fold into a 34-bp hairpin or into a 68-bp intermolecular duplex. We compared H46cta and H70cta migration to the migration of a 22-bp and a 44-bp duplex (dx22 and dx44) mimicking H46cta putative hairpin and intermolecular duplex structures.
H46cta displayed a two-band pattern and migrated as dx22 and dx44, supporting the formation of an intramolecular and an intermolecular structure; the intramolecular structure was the major fraction (at 30 µM strand concentration, the two bands has similar intensity, this indicates that the intramolecular structure is two-fold the intermolecular structure). H70 displayed a single-band pattern and migrated faster than dx44, supporting the formation of an intramolecular structure.

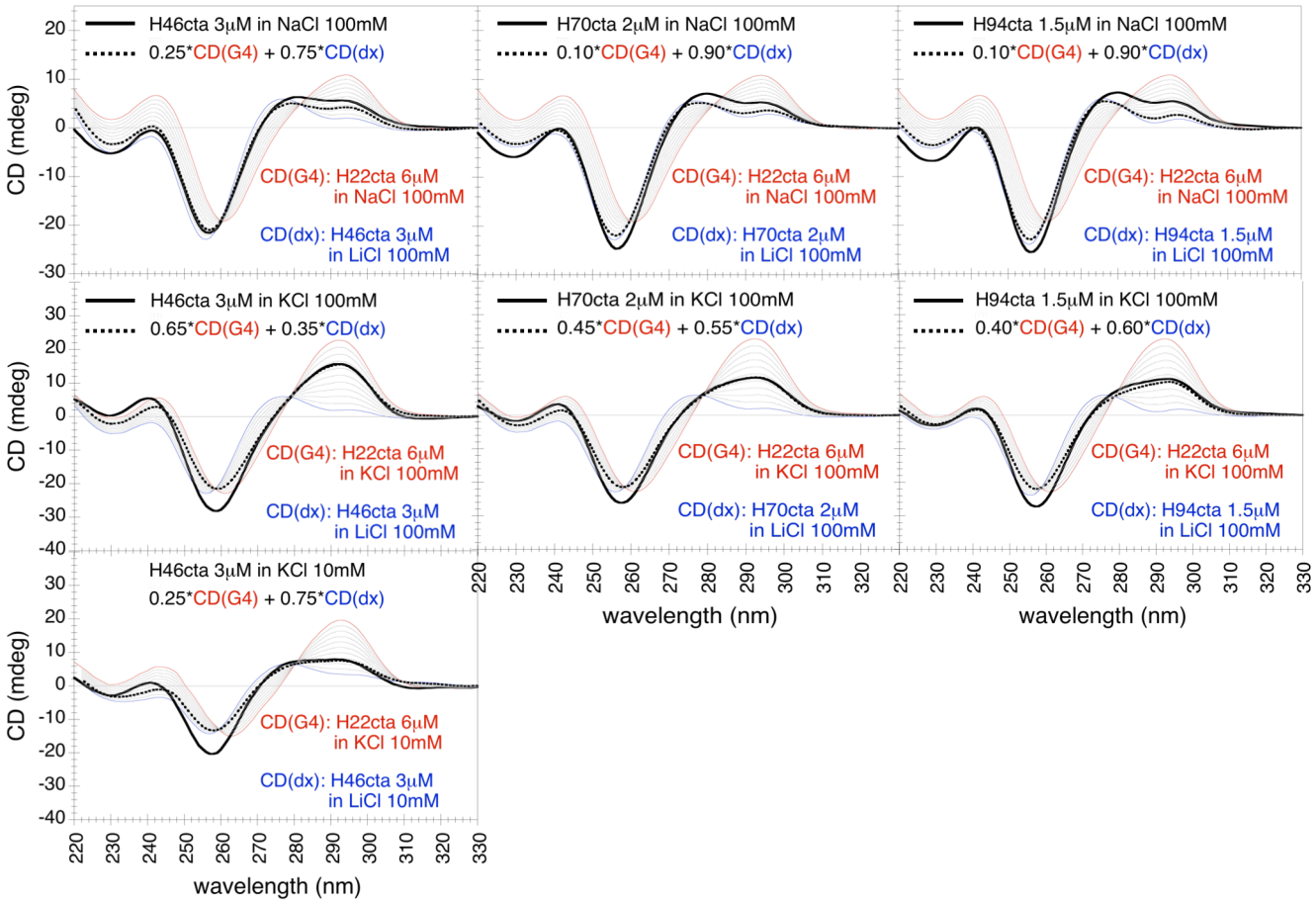


Figure S4

Estimation of G4 and duplex structure fractions of GGGCTA variant repeats

In order to roughly estimate the fraction *f* of duplex structure and the fraction *(1-f)* of G4 for H46cta, H70cta and H94cta oligonucleotides in KCl and NaCl, we assumed that:
- the CD spectrum of H22cta in KCl (or NaCl) corresponded to the CD spectrum of the G4 component in KCl (or NaCl) (defined as "CD(G4)"),
- the CD spectra of H46cta, H70cta and H94cta in LiCl corresponded to the CD spectra of their duplex component (defined "CD(dx)").
We hence searched for the factor *f* that provided the best approximation of the measured CD spectra of H46cta, H70cta and H94cta in KCl (or NaCl) as the weighted sum of the G4 and of the duplex spectra: *(1-f)*CD(G4) + f*CD(dx)*.

In each panel,
- the red spectrum is the CD spectrum of H22cta in NaCl or KCl ("CD(G4)");
- the blue spectrum is the CD spectrum of H46cta or H70cta or H94cta in LiCl ("CD(dx)");
- the black bold spectrum is the measured CD spectrum of H46cta or H70cta or H94cta in NaCl or KCl;
- grey spectra are the weighted sums of the red and of the blue spectra: *(1-f)*CD(G4) + f*CD(dx)*, where *f* increases from 0.1 to 0.9 in steps of 0.1;
- the black bold dotted spectrum is the weighted sum that best approximates the measured spectrum of H46cta or H70cta or H94cta (*i.e.* the black bold spectrum); its factor *f* provides a rough estimation of the fraction of H46cta or H70cta or H94cta folded into the duplex structure.

The duplex fractions *f* calculated from CD spectra shown in this figure are represented in the histogram of figure 3 (where we fixed an error bar of +/- 0.05 since we moved by steps of 0.1 for the estimation of *f*).

GTTCTA oligonucleotide:

5'TTCTAGTTCTAGTTCTAGTTCTAGTTCTAGTTCT
TTGATCTTGATCTTGATCTTGATCTTGATCTTGA

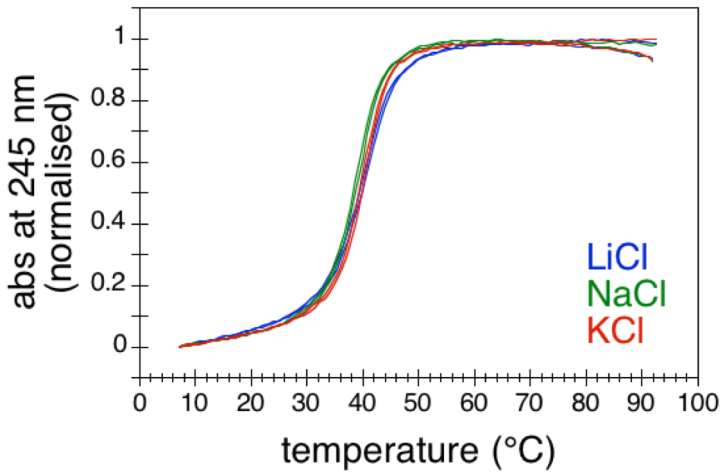


Figure S5

Melting profiles of a DNA duplex structure mimicking H70cta duplex in LiCl, NaCl and KCl

Absorbance at 245 nm as a function of temperature of the duplex-folding GTTCTA oligonucleotide composed of GTTCTA repeats (sequence: 5'TTCTA(GTTCTA)$_{10}$GTT), at 2 μM strand concentration, in LiCl, NaCl and KCl 100 mM.

GTTCTA oligonucleotide mimics H70cta hairpin: it alternates four Watson-Crick base pairs with a couple of TT base-pairs instead of a couple of GG base-pairs. The replacement of GG by TT allows comparing the stability of the duplex structure in potassium, sodium and lithium.

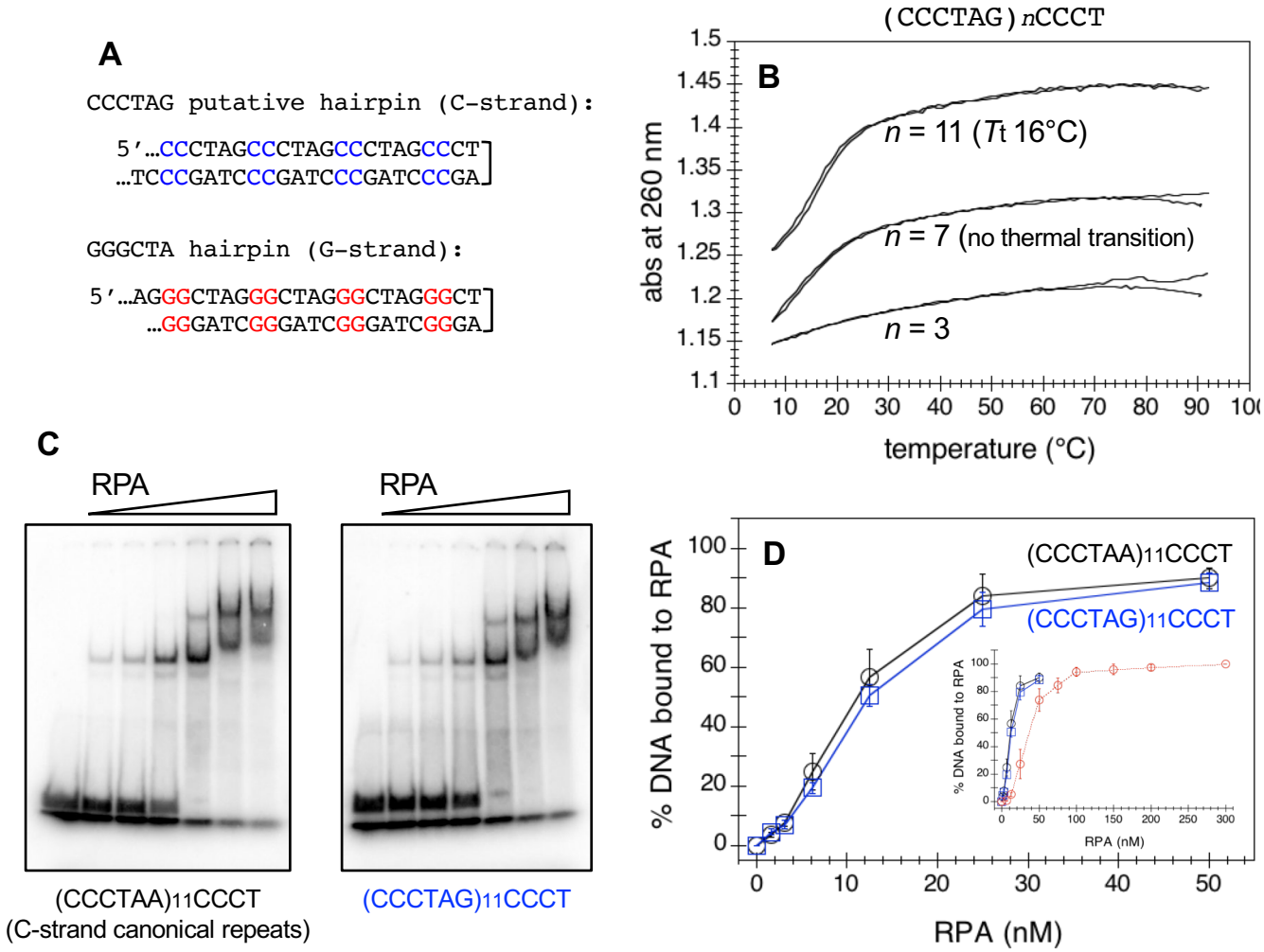


Figure S6
CCCTAG repeats: melting profiles and RPA binding

(A) As GGGCTA variant repeats in the G-strand, so CCCTAG complementary repeats in the C-strand may fold, in principle, into hairpin structures where four Watson-Crick base-pairs alternate with a couple of non-canonical base-pairs (a GG couple in the G-strand and a CC couple in the C-strand).

(B) Absorbance at 260 nm as a function of temperature of (CCCTAG)*n*CCCT oligonucleotides at 6 µM (*n* = 3), 3 µM (*n* = 7) and 2 µM (*n* =11) strand concentrations, in KCl 100 mM at pH 7.2 (identical melting profiles were obtained in NaCl and LiCl).
Any thermal transition was observed for (CCCTAA)3,7,11CCCT oligonucleotides (C-strand canonical repeats) in the same buffer conditions (not shown).

(C) Representative EMSA of radiolabelled (CCCTAA)11CCCT and (CCCTAG)11CCCT oligonucleotides incubated with increasing concentrations of RPA (0, 1.5, 3, 6.25, 12.5, 25, 50 nM) in KCl 100 mM.
Retarded bands in gels corresponds to RPA:DNA complexes on increasing stoichiometry.

(D) Percentage of oligonucleotides bound to RPA obtained by EMSA: (CCCTAA)11CCCT (black square) and (CCCTAG)11CCCT (blue squares). Error bars are standard deviation from 3 assays.
For comparison of RPA binding to G-strand and C-strand, in panel D inset the same curves are plotted on a different x-scale along with the binding curve of RPA to A(GGGTTA)11GGG in KCl 100 mM (red circles, dashed line, from figure 4A).

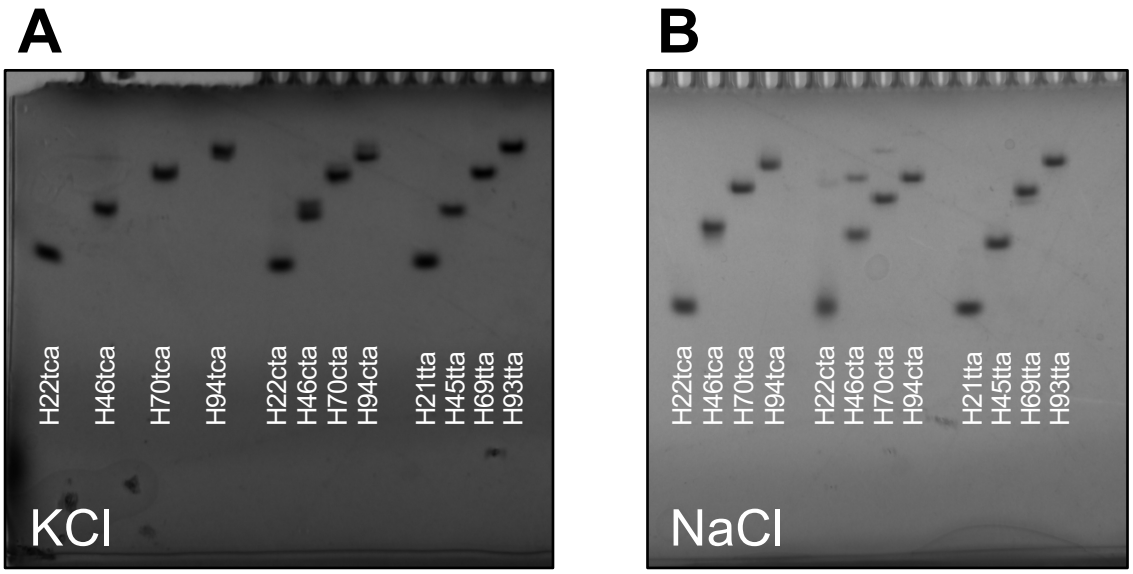


Figure S7
Molecularity of structures formed by GGGCTA and GGGTCA variant repeats.

Non-denaturing PAGE assays of GGGCTA and GGGTCA variant repeats in 100mM KCl **(A)** and NaCl **(B),** at the following strand concentrations: 60 µM for H22tca/cta/tta µM , 30 µM for H46tca/cta/tta, 20 µM for H70tca/cta/tta and 15 µM for H94tca/cta/tta. Oligonucleotides were detected by UV-shadow. GGGTTA canonical repeats were run as reference intramolecular structures.

In potassium, GGGCTA and GGGTCA variant repeats migrate as single bands, with migration rates similar to those of GGGTTA canonical repeats, supporting intramolecular folding. Nevertheless, while GGGTTA and GGGTCA repeats migrate as well defined single bands, GGGCTA repeats migrate as larger single bands, consistently with the presence of a mix of conformation (G4 and hairpins) demonstrated by UV-spectroscopy and NMR.

In sodium, for GGGTCA repeats, only intramolecular structures were detected; for GGGCTA repeats, a very minor fraction of intermolecular structures were detected for H46cta, H70cta and H94cta oligonucleotides (that decreased with increasing the number of repeats) consistently with non-denaturing PAGE assay in LiCl (figure S3).

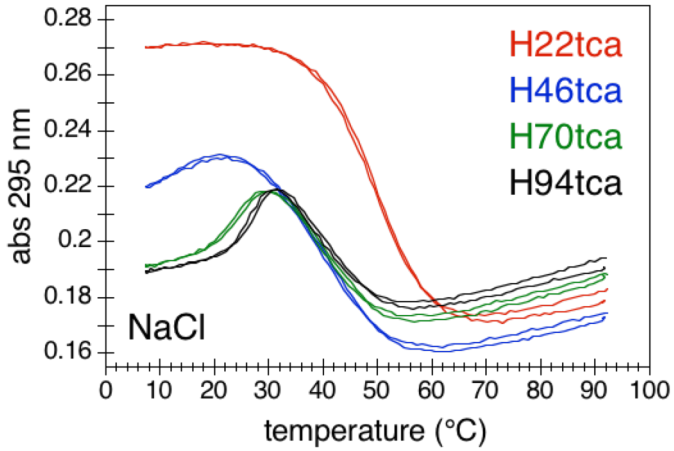


Figure S8
Melting profiles of GGGTCA variant repeats in NaCl

Absorbance at 295 nm as a function of temperature of H22tca 6 μM (red), H46tca 3 μM (blue), H70tca 2 μM (green) and H94tca 1.5 μM (black) in NaCl 100 mM.

Melting profiles of H46tca, H70tca and H94tca support the presence of more than one structure:
the increase of the absorbance at 295 nm with decreasing temperature, observed in the high temperature range, is indicative of the presence of G4 structures; while the decrease of the absorbance at 295 nm with decreasing temperature, observed in the low temperature range, is indicative of the presence of non-G4 structures, as shown by TDS spectra of a variety of DNA structures (12).

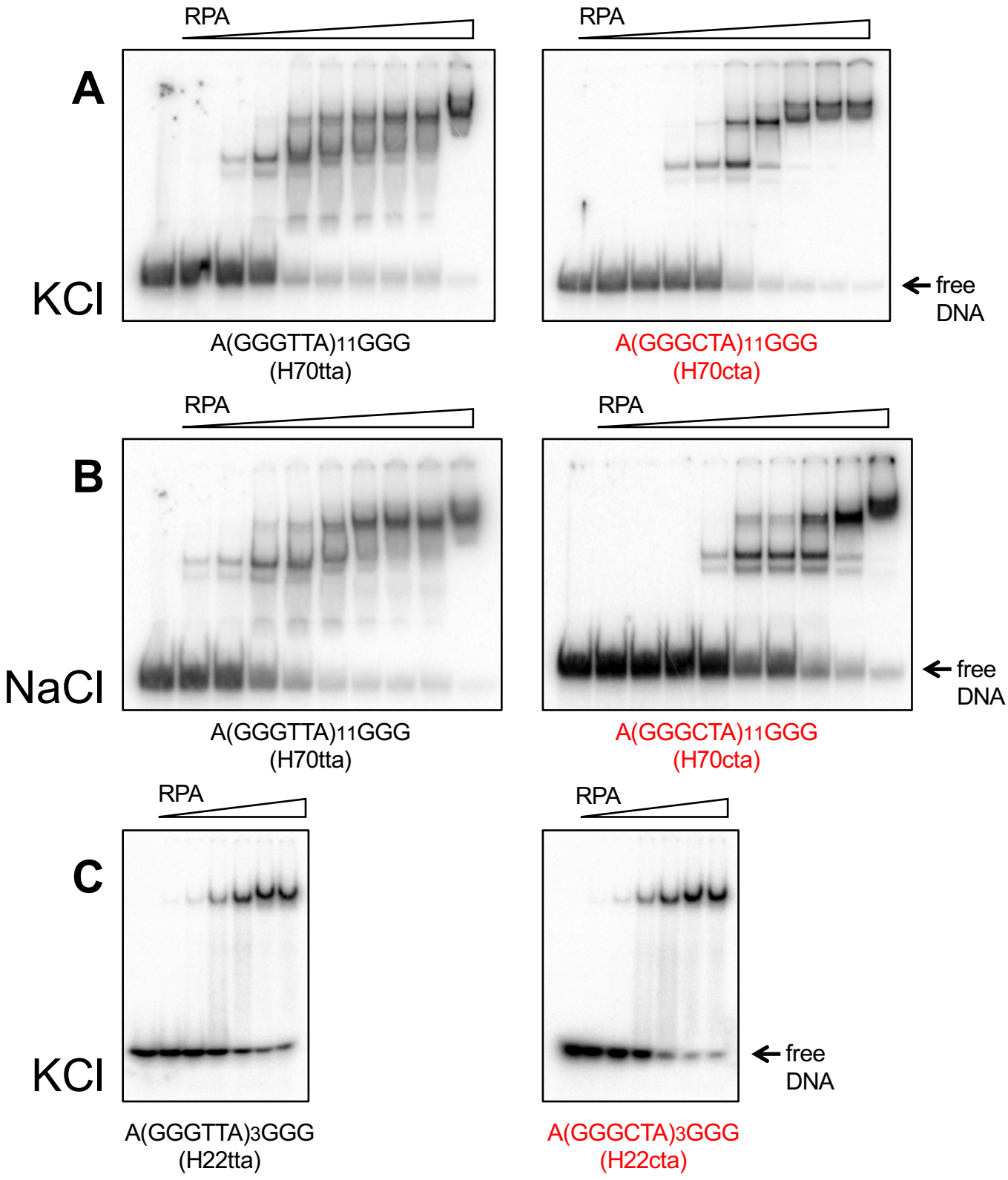


Figure S9A
Representative EMSA relative to figures 4A, 4B and 4C

(A, B) EMSA of radiolabelled A(GGGTTA)11GGG (H70tta) and A(GGGCTA)11GGG (H70cta) oligonucleotides incubated with increasing concentrations of RPA (0, 6.25, 12.5, 25, 50, 75, 100, 150, 200, 300 nM) in 100 mM KCl **(A)** and in NaCl **(B)**.
(C) EMSA of radiolabelled A(GGGTTA)3GGG (H22tta) and A(GGGCTA)3GGG (H22cta) oligonucleotides incubated with increasing concentrations of RPA (0, 6.25, 12.5, 25, 50, 100, 200 nM) in 100 mM KCl .
Retarded bands in H70tta and H70cta gels corresponds to 1:1, 2:1 and 3:1 RPA:DNA complexes (one RPA, in its extended mode of binding, covers about 25-30 nt).

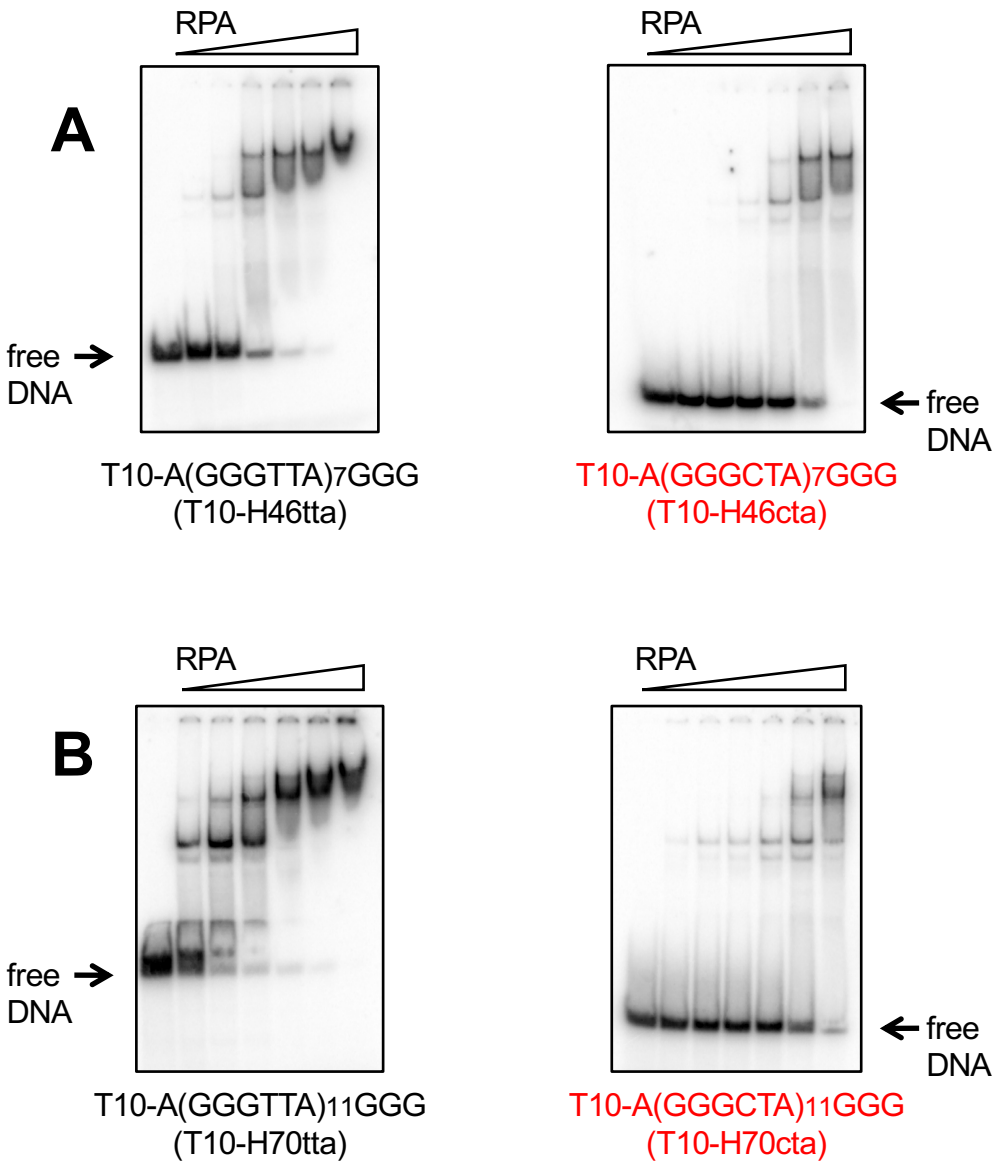


Figure S9B
Representative EMSA relative to figures 4E and 4F

EMSA of radiolabelled oligonucleotides
T10-A(GGGTTA)7GGG (T10-H46tta) and T10-A(GGGCTA)7GGG (T10-46cta) **(A)** and T10-A(GGGTTA)11GGG (T10-H70tta) and T10-A(GGGCTA)11GGG (T10-70cta) **(B)** incubated with increasing concentrations of RPA (0, 6.25, 12.5, 25, 50, 100, 200 nM) in 100 mM NaCl.
Retarded bands in gels corresponds RPA:DNA complexes on increasing stoichiometry.

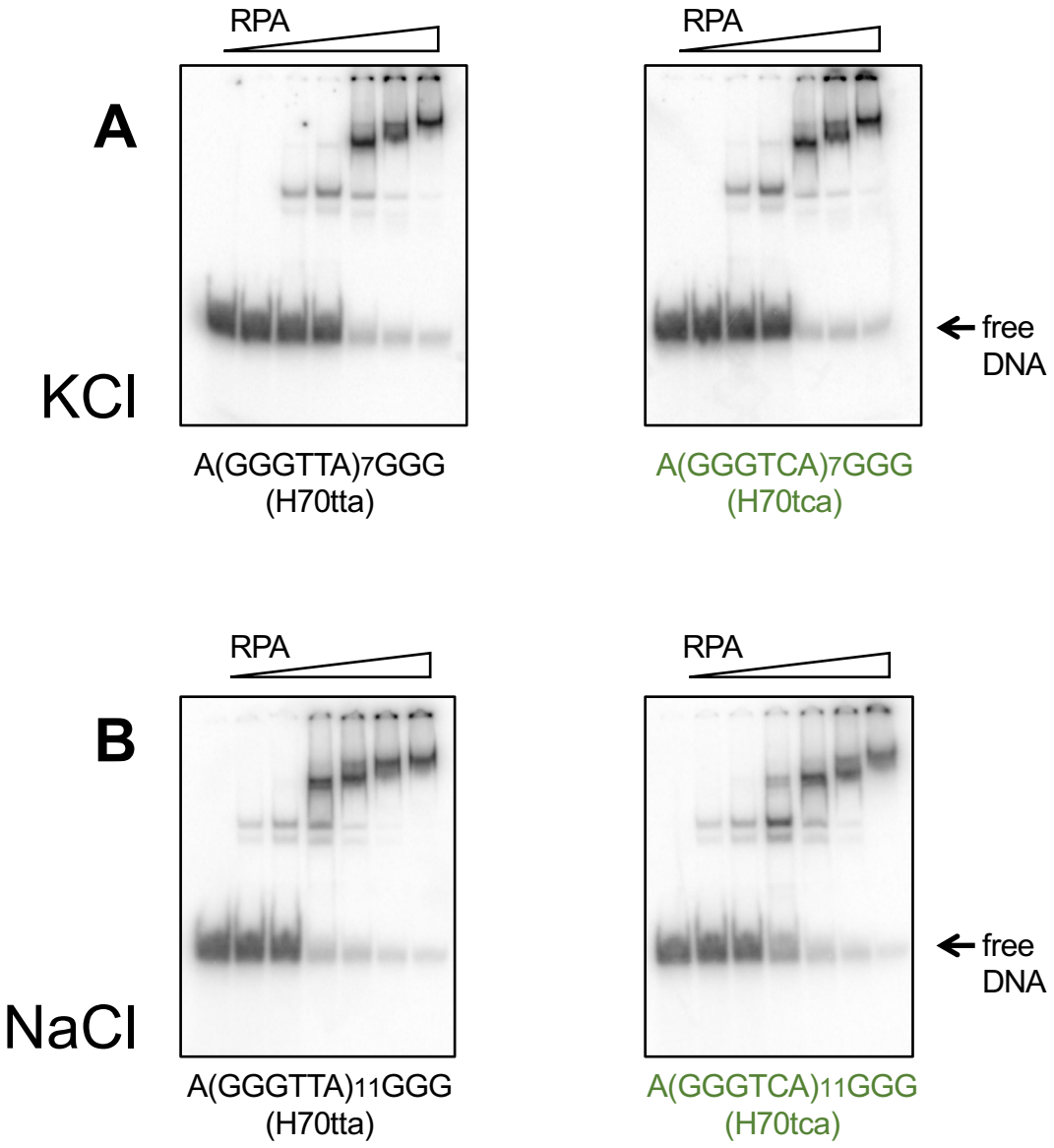


Figure S9C
Representative EMSA relative to figures 5G and 5H

EMSA of radiolabelled A(GGGTTA)11GGG (H70tta) and A(GGGTCA)11GGG (H70tca) oligonucleotides incubated with increasing concentrations of RPA (0, 6.25, 12.5, 25, 50, 100, 200 nM) in 100 mM KCl **(A)** and NaCl **(B)**.
Retarded bands in gels corresponds RPA:DNA complexes on increasing stoichiometry.

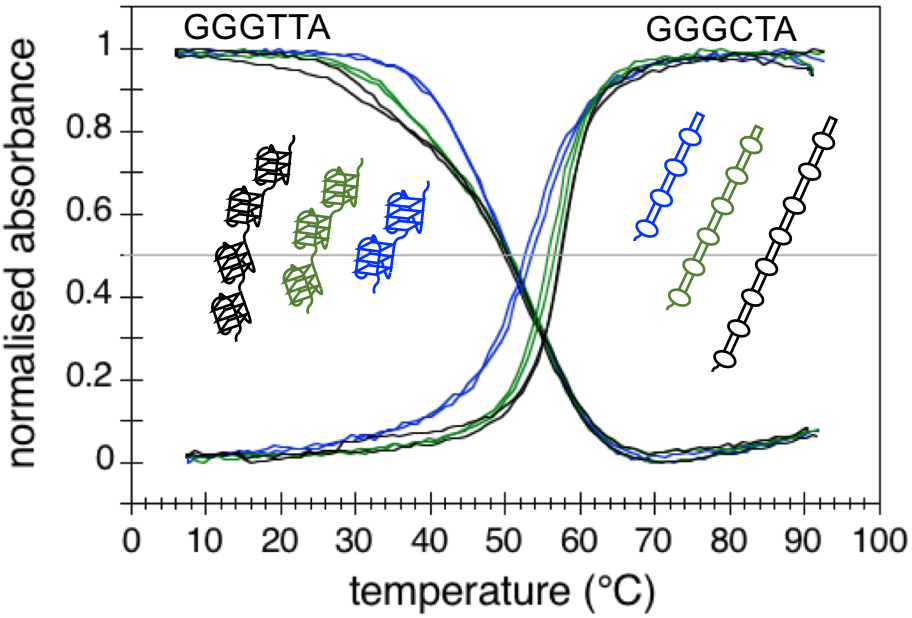


Figure S10
Melting profiles of GGGTCA hairpins and GGGTTA G4s

Normalised absorbance as a function of temperature of:
GGGTTA repeats in KCl 100 mM (at 295 nm) and of GGGCTA repeats in LiCl 100 mM (at 245 nm).
Under these salt conditions, GGGTTA are structured into G4, GGGCTA into hairpins.

The GGGTTA profile is inverted with respect to the GGGCTA profile, because the abs of G4s at 295 nm increase upon folding, contrary to duplex structures for which the abs (at any wavelength) decrease upon folding.

GGGTTA sequences: TTA(GGGTTA)$_{n=7,11,15}$GGGTTA ,
GGGCTA sequences: A(GGGCTA)$_{n=7,11,15}$GGG,
blu: n=7 (3 μM), green: n=11 (2 μM) , black: n=15 (1.5 μM).

These melting profiles show that GGGCTA hairpin structures ($T_{1/2}$ 50°C) are more stable than GGGTTA G4 structures ($T_{1/2}$ 53, 54 and 57°C).

Melting profiles of GGGTTA repeats are from our previous study (11).

Supplementary references

1. Lim,K.W., Alberti,P., Guedin,A., Lacroix,L., Riou,J.F., Royle,N.J., Mergny,J.L. and Phan,A.T. (2009) Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res*, **37**, 6239–6248.
2. Luu,K.N., Phan,A.T., Kuryavyi,V., Lacroix,L. and Patel,D.J. (2006) Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J Am Chem Soc*, **128**, 9963–9970.
3. Phan,A.T., Luu,K.N. and Patel,D.J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. *Nucleic Acids Res*, **34**, 5715–5719.
4. Lim,K.W., Amrane,S., Bouaziz,S., Xu,W., Mu,Y., Patel,D.J., Luu,K.N. and Phan,A.T. (2009) Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J Am Chem Soc*, **131**, 4301–4309.
5. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263–282.
6. Lim,K.W., Ng,V.C., Martin-Pintado,N., Heddi,B. and Phan,A.T. (2013) Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res*, **41**, 10556–10562.
7. Heddi,B. and Phan,A.T. (2011) Structure of human telomeric DNA in crowded solution. *J Am Chem Soc*, **133**, 9824–9833.
8. Parkinson,G.N., Lee,M.P. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
9. Phan,A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J*, **277**, 1107–1117.
10. Saintome,C., Amrane,S., Mergny,J.L. and Alberti,P. (2016) The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. *Nucleic Acids Res*, **44**, 2926–2935.
11. Bugaut,A. and Alberti,P. (2015) Understanding the stability of DNA G-quadruplex units in long human telomeric strands. *Biochimie*, **113**, 125–133.
12. Mergny,J.L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res*, **33**, e138.

4.4 Conclusions, perspectives et résultats complémentaires

4.4.1 Le motif instable TCAGGG : des structures en tiges-boucles qui affectent la liaison de *RPA*

En conclusion, notre travail suggère que des structures en tige-boucle seraient à l'origine de l'instabilité du motif CTAGGG. La formation de structures en tige-boucle stables à partir de 8 répétitions expliquerait pourquoi l'instabilité du motif CTAGGG ne se manifeste qu'à partir d'une dizaine de répétions. Pourquoi ces structures en tige-boucle seraient-elles problématiques ? À qui nuisent-elles ? Les résultats de notre étude démontrent que la liaison de *RPA* au brin G télomérique est perturbée en présence de répétitions GGGCTA structurées en tige-boucle. Ces résultats permettent d'émettre une nouvelle hypothèse de travail : que la difficulté de *RPA* à se lier aux structures en tige-boucle formées par des répétitions CTAGGG soit un des facteurs ou le facteur à l'origine de l'instabilité de ce motif non-canonique. Cette hypothèse de travail mérite d'être explorée de façon plus approfondie. Les événements de perte ou de gain des répétitions CTAGGG à l'origine de l'instabilité auraient lieu lors de la réplication des télomères (Mendez-Bermudez et al., 2009). En plus de *RPA*, plusieurs protéines sont impliquées dans la réplication des télomères,parmi lesquelles le complexe *CST* et les hélicases *WRN*, *BLM* et *RTEL1*. Le complexe *CST* participe à la réplication des télomères (Miyake et al., 2009; Chen et al., 2012), notamment au redémarrage de la fourche de réplication (Stewart et al., 2012; Zhang et al., 2019). En défaut de *WRN*, *BLM* et *RTEL1* on observe des aberrations télomériques, notamment à niveau du brin G (Crabbe, 2004; Sfeir et al., 2009; Vannier et al., 2012; Zimmermann et al., 2014). Comme *RPA*, toutes ces protéines (*CST*, *WRN*, *BLM* et *RTEL1*) déroulent *in vitro* les structures G4 formées par les séquences télomériques ; sont-elles capables de dérouler aussi efficacement les structures en tige-boucle formées par les répétitions du motif CTAGGG ? Dans les perspectives de cette étude, afin de comprendre quelles sont les protéines affectées par la structuration en tige-boucle des répétitions CTAGGG, il serait donc intéressant d'étudier l'interaction de ces protéines avec ces structures.

4.4.2 *POT1-TPP1*

L'étude de Royle and coll. a montré que *POT1* et *TRF2* se lient aussi bien aux répétitions CTAGGG qu'aux répétitions canoniques TTAGGG (en simple brin pour *POT1*, en double-brin pour *TRF2*) (Mendez-Bermudez et al., 2009). Au sein de la cellule la protéine *TPP1* est importante pour le recrutement de *POT1* aux télomères (Liu et al., 2004; Ye, 2004; Hockemeyer et al., 2007; Kibe et al., 2009). En effet si *TPP1* est délétée ou si l'interaction entre *POT1* et *TPP1* est perturbée, alors *POT1* ne se lie plus aux télomères. Nous avons comparé la liaison du complexe *POT1-TPP1* aux répétitions du motif CTAGGG et aux répétitions canoniques TTAGGG. Nous avons montré que, contrairement à *POT1* seule, le complexe *POT1-TPP1* ne se lie pas au motif CTAGGG (Figure 4.4). Cependant il est improbable que *POT1-TPP1* soit à l'origine de l'instabilité du motif CTAGGG. En effet l'instabilité induite par le motif CTAGGG ne se manifeste qu'à partir d'une dizaine de répétitions. Si l'instabilité du motif CTAGGG était liée à *POT1-TPP1*, nous devrions observer cette instabilité du motif variant CTAGGG quelque-soit le nombre de répétitions, et non seulement à partir d'une dizaine de répétitions de ce motif, puisque POT1-TPP1 ne se lie pas du tout au motif CTAGGG.

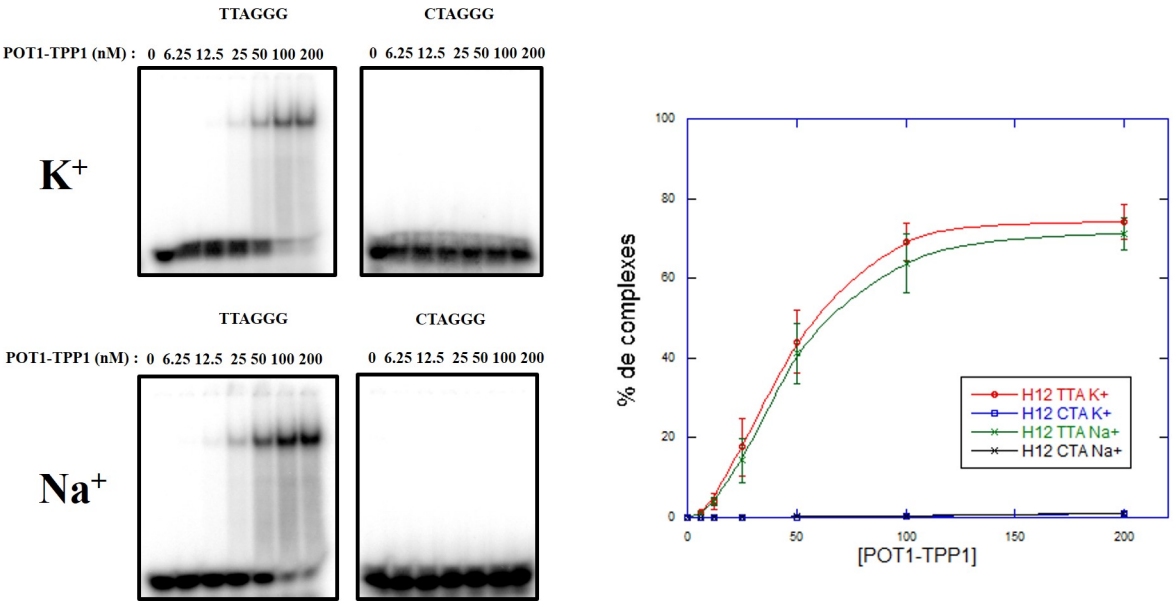


FIGURE 4.4 – **Expérience *EMSA* avec *POT1* en présence du motif TTAGGG et du motif CTAGGG.** *EMSA* en présence de TTAGGG(10nt) et CTAGGG(10nt) [*POT1-TPP1*] : 0 / 6.25 / 12.5 / 25 / 50 / 100 / 200 nM. **A.** en KCl ; **B.** en NaCl ; **C.** Quantification des expériences *EMSA*.

Nous avons décidé de ne pas insérer ce résultat dans l'article sur le motif instable CTA, mais de le présenter dans le cadre de l'article sur l'interaction de *POT1-TPP1* avec son site de liaison, présenté dans le chapitre précèdent. En effet ce résultat nous livre une information supplémentaire originale sur l'apport de *TPP1* dans le complexe *POT1-TPP1* : le fait que *POT1* se lie au site de motif CTAGGG aussi bien qu'au site de motif canonique TTAGGG, tandis que *POT1-TTP1* ne se lie pas du tout au site de motif CTAGGG, signifie que la présence de *TPP1* permet la discrimination entre le motif canonique et le motif non-canonique CTAGGG. De façon plus générale ce résultat montre que*TPP1* confère au complexe *POT1-TPP1* une spécificité de liaison aux séquences télomériques plus accrue que *POT1* toute seule.

Il est intéressant de noter que ce rôle joué par *TPP1* n'est pas manifeste quand un seul nucléotide est muté dans le site. J'ai effectué des expériences de liaison de *POT1-TPP1* au site de liaison muté en un seul nucléotide à la fois (Figure 4.5.A). Les résultats que j'ai obtenus avec *POT1* seule (Figure 4.5.B) sont similaires à ceux publiés dans l'article de Lei et al. (Lei et al., 2004) (Figure 4.5.C) : quand on introduit une seule mutation à la fois dans le 12mer site de liaison 5'GGTTAGGGTTAG, les affinité de *POT1* seule et de *POT1-TPP1* sont modulées de la même façon : Ce n'est que le travail sur le motif non-canonique CTAGGG, où le site de liaison porte deux mutations (5'GGCTAGGGCTAG) qui nous a dévoilé ce rôle de *TPP1* au sein du complexe *POT1-TPP1*.

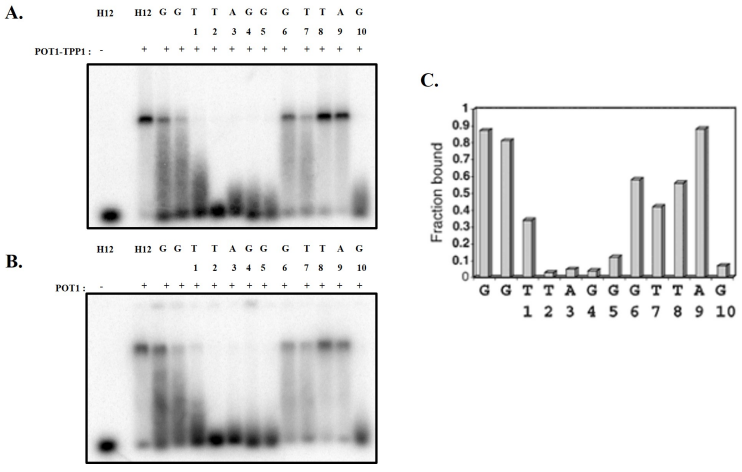


FIGURE 4.5 – **Expérience d'*EMSA* de liaison de *POT1* et *POT1-TPP1* à des séquences mutées.** Les séquences décrivent la mutation d'un seul nucléotide par son complémentaires. **A.** en présence de 25nM de *POT1-TPP1* ; **B.** en présence de 85nM de *POT1* ; **C.** (Lei et al., 2004)

Cependant il est à relever que les expériences *EMSA* en introduisant une seule mutation on été réalisé en présence de 25nM de *POT1-TPP1* et 85nm*POT1*. Une étude réalisé par J. Nandakumar et al. montre que *TPP1* améliore la discrimination de *POT1* contre l'ARN *TERRA*, en améliorant la sensibilité de reconnaissance de *POT1* pour la position 2 du motif de reconnaissance de *POT1* (Nandakumar et al., 2009). En effet dans cette étude, les auteurs observent en remplaçant la thymine en position 2 par une uracile que *TPP1* n'améliore plus l'efficacité de liaison de *POT1* pour ce substrat(Figure 4.6). Pour approfondir l'effet de la mutation de la séquence télomérique sur le complexe *POT1-TPP1*, il faudrait comparer l'efficacité de liaison de *POT1-TPP1* à *POT1* sur ces séquences à la même concentration.

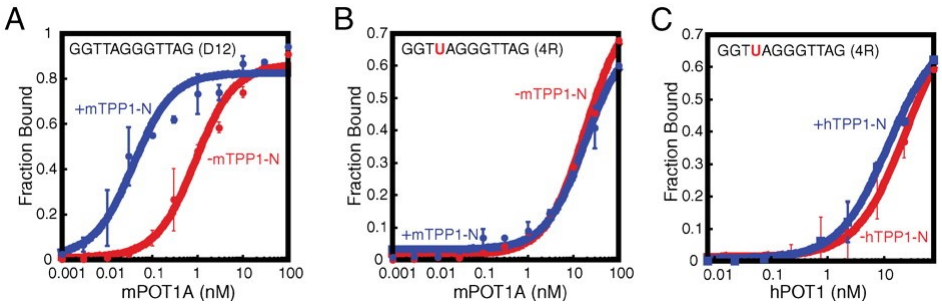


FIGURE 4.6 – **La substitution de T2 par U2 empêche *TPP1* d'améliorer l'affinité de *POT1* à son substrat.** Courbes de liaison pour les complexes *POT1* en présence (bleu) et absence (rouge) de *TPP1* : (**A.**) chez la souris en présence H12 ; (**B.**) chez la souris en présence H12-4U ; (**C.**) chez l'homme en présence H12-4U

4.5 Annexes méthodes

4.5.1 Dichroïsme circulaire

Le dichroïsme circulaire, ou en anglais Circular Dichroism (**CD**), est un phénomène de spectroscopie qui dépend de la polarisation de la lumière Hecht (2002) . La polarisation décrit la propriété physique d'une onde vectorielle à osciller dans une orientation privilégiée (Annexe **??**). Dans le cas de la technique de dichroïsme circulaire, la polarisation qui nous intéresse est la polarisation circulaire. Il existe deux types de polarisation circulaire : la polarisation circulaire droite (PCD) qui traduit la propagation d'une radiation polarisée circulairement dans le sens horaire (Annexe 8.6.A) ; et à l'inverse la polarisation circulaire gauche (PCG) qui décrit une radiation polarisée circulairement dans le sens anti-horaire (Annexe 8.6.B).

Le phénomène de dichroïsme circulaire repose sur l'absorbance d'une radiation PCG et d'une radiation PCD telle que la grandeur physique mesurée est $CD = A_G - A_D$ (Figure 4.7). A_G correspond à l'absorbance en présence de lumière polarisée circulaire gauche, et réciproquement A_D en présence de lumière polarisée circulaire droite. L'absorbance est calculée à partir de l'équation de Beer-Lambert : $A = \epsilon\, l\, c$. Dans notre étude, la longueur du trajet parcouru par la lumière l et la concentration molaire c de l'échantillon sont des constantes. On peut alors exprimer cette différence d'absorbance en une différence de coefficient d'extinction molaire $\Delta\epsilon = \epsilon_G - \epsilon_D$ dépendant de la longueur d'onde λ.

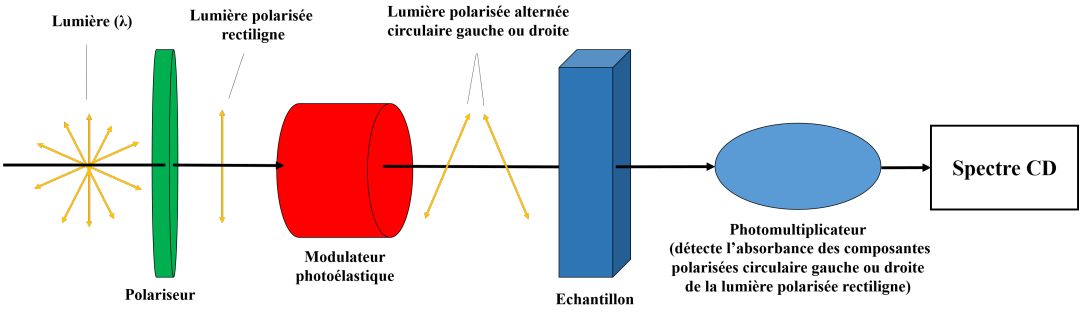


FIGURE 4.7 – **Schéma du chemin d'un appareil de spectroscopie par dichroïsme circulaire.**

La somme vectorielle de deux ondes lumineuses polarisées circulairement, l'une PCG et l'autre PCD, donne une onde lumineuse polarisée elliptiquement (Figure 4.8). On définit la valeur de l'ellipticité $\theta = \arctan(minor\ axis/Major\ axis)$. Dans le cas où $\Delta\epsilon << 1$, il est possible de faire la conversion de la mesure CD en valeur d'ellipticité θ (eq 4.1) :

$$\theta\,(degres) = \frac{(ln10)(\Delta\epsilon\, l\, c)}{4} \times \frac{180}{\pi} \tag{4.1}$$

La variation de la valeur CD, qui peut être exprimée en variation de l'ellipticité, est due au pouvoir rotatoire. Le pouvoir rotatoire est la propriété qu'ont certains matériaux à dévier le vecteur de la lumière les traversant, en modifiant l'angle de polarisation de cette dernière. Ces matériaux peuvent induire une déviation de la propagation de l'onde dans le sens horaire ou anti-horaire.

Dans le cadre d'étude de biomolécules, le pouvoir rotatoire provient de la propriété chimique appelée chiralité traduisant l'absence de symétrie intrinsèque de certaines molécules (Fasman, 1996). Une molécule est dite chirale (du grec $\chi\epsilon\iota\rho$: la main) lorsqu'elle peut se présenter sous deux formes, qui

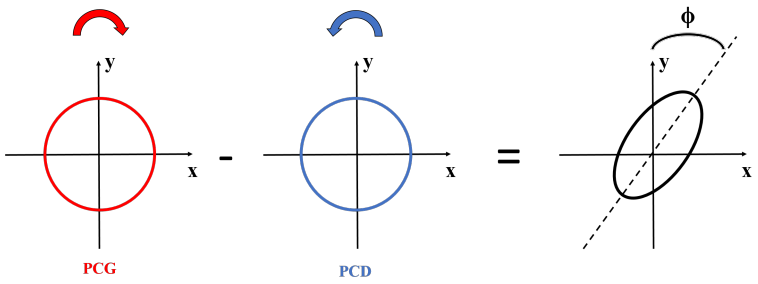


FIGURE 4.8 – **Schéma d'une onde elliptique résultant de deux ondes polarisées circulairement gauche et droite.**

sont images l'une de l'autre dans un miroir mais non superposable. La chiralité des biomolécules est donnée par la présence de certains acides aminés ou sucres ; mais également par la présence de structures secondaires comme l'hélice α ou le feuillet β qui ont une signature spectrale caractéristique (Annexe 8.7).

Le phénomène de dichroïsme circulaire s'applique également aux acides nucléiques. Ainsi, les structures secondaires comme la double hélice d'ADN, le I-motif (Školáková et al., 2019) ou encore le G-quadruplex, qui nous intéressent dans notre étude, peuvent être caractérisées par *CD*. Dans le cas de la structure G-quadruplex, le signal *CD* est donné par l'agencement des plateaux de guanines qui le composent (Masiero et al., 2010). En effet, chaque plateau de guanines a une orientation horaire ou anti-horaire, qui nous permettent de définir trois types d'empilements différents de ces plateaux au sein d'un G4 :

1. **Empilement de plateaux avec la même orientation.** Ce qui donne des spectres *CD* du type Figure 4.9 courbe noire : présence d'un pic "positif" autour de 265 nm. Cet agencement de plateaux est caractéristique des G4 où les brins sont parallèles.
2. **Empilement de plateaux avec des orientations alternées.** Ce qui donne des spectres *CD* du type Figure 4.9 courbe rouge : présence d'un pic "négatif" autour de 265 nm et d'un pic "positif" vers 295 nm. Cet agencement de plateaux est caractéristique des G4 où les brins sont anti-parallèles, forme pouvant être adoptées par la séquence télomérique humaine dans le sodium.
3. **Empilement de plateaux avec des orientations mixtes.** Ce qui donne des spectre *CD* du type Figure 4.9 courbe bleue : présence d'un pic "positif" autour de 295 nm. Cet agencement de plateaux est ca-

ractéristique des formes dites hybrides (présentant des brins parallèles et antiparallèles) pouvant être adoptées par la séquence télomérique humaine dans le potassium.

L'intérêt de cette analyse spectroscopique dans le cadre de biomolécules, outre l'aspect non destructif, est de révéler la présence de certaines structures secondaires au sein de la structure tertiaire de ces biomolécules. De même, le phénomène *CD* est adapté pour mettre en évidence la dépendance de ces structures secondaires à des variations de température, de pH ou encore de concentration de sels, tel que cela est le cas pour les G4s (Tran et al., 2010). En revanche, le dichroïsme circulaire ne donne aucune information sur l'emplacement de la structure secondaire au sein de la structure tertiaire de la biomolécule, il se limite aux structures secondaires dont les spectres *CD* sont caractérisés. Toutefois, l'analyse *CD* présente rapidement des limites d'interprétation des spectres dans le cas d'hétérogénéité de structures secondaires au sein de l'échantillon.

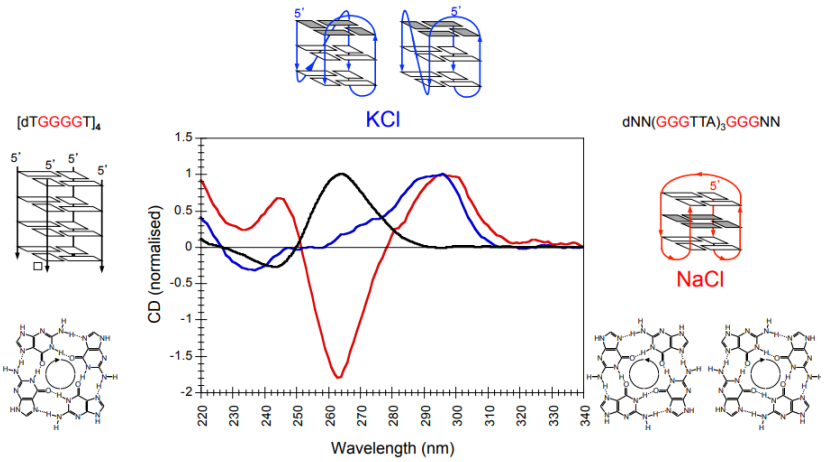


FIGURE 4.9 – **Spectre *CD* des différents empilements de plateaux guanines dans le cas d'un G4** Résultats publiés avec la permission de Patrizia Alberti

4.5.2 Résonance magnétique nucléaire

La résonance magnétique nucléaire Abragam and Abragam (1961) (**RMN**) comme son nom l'indique porte sur des propriétés magnétiques des noyaux atomiques, et plus précisément sur les noyaux possédant un spin nucléaire non nul : ^{1}H, ^{13}C, ^{17}O, ^{19}F, ^{31}P, ... Le principe de la technique de RMN

repose sur l'utilisation d'un champs magnétique B_0 de sorte que les moments magnétiques nucléaires de chaque atome décrivent un mouvement de précession appelé précession de Larmor. La somme des moments magnétiques nucléaires donne une aimantation globale qu'on définit à l'aide d'un vecteur orienté selon B_0 (Figure 4.10.A). On cherche à basculer cette aimantation globale dans le plan Ox, y à l'aide du phénomène de résonance. En effet, on utilise un champs pulsé radiofréquence B_1 défini tel que la fréquence de ce dernier corresponde à la fréquence de résonance des moments magnétiques nucléaires des atomes étudiés (Figure 4.10.B). Enfin en arrêtant le champs pulsé B_1, les moments magnétiques retournent à l'équilibre en "precessant" suivant le champs B_0 (Figure 4.10.C). On mesure alors le signal de précession libre (Free Induction Decay en anglais) ainsi que le temps de relaxation des moments magnétiques, c'est à dire le temps de retour de la magnétisation globale selon B_0. Il est important de noter que le moment magnétique nucléaire d'un spin est sensible à son environnement électronique. En effet, l'interaction magnétique dipôle-dipôle, appelée également couplage dipolaire, contribue de manière significative au spectre de RMN. C'est pourquoi dans le cas d'expérience de RMN, on parle de déplacement chimique δ pour décrire la dépendance des niveaux d'énergie magnétique du noyau avec l'environnement électronique de la molécule. Ce déplacement chimique δ, donné en ppm (partie par million), est défini par l'équation 4.2 :

$$\delta = \frac{\upsilon - \upsilon_{ref}}{\upsilon_{travail}} \tag{4.2}$$

Avec

υ_{ref} : fréquence de référence.
$\upsilon_{travail}$: fréquence de travail donné par la relation de Larmor $\omega_0 = \gamma B_0$

La technique de RMN donne des informations structurales sur la molécule en raison de la dépendance du phénomène de résonance magnétique nucléaire à l'environnement électronique. Cette technique permet de déterminer notamment les régions de la molécule en contact avec le solvant, ou encore les repliements spatiaux internes de la molécule. Par conséquence, la RMN est qualifiée de technique d'analyse structurale pouvant même déterminer la structure tridimensionnelle de la molécule dans le cas d'expérience RMN à plusieurs dimensions. On parle ici de dimension pour définir le nombre de spins nucléaires actifs en RMN, c'est-à-dire le suivi des déplacements chimiques des n noyaux où n défini la dimension de RMN.

Dans notre étude, nous nous limitons à des expériences de RMN à une seule dimension en suivant les spins nucléaires ^{1}H de nos acides nucléiques.

Le but de nos expériences est de déterminer la présence de structures G-quadruplexe ou tige-boucle au sein de nos séquences télomériques, en complément des expériences de dichroïsme circulaire vue précédemment (partie 4.5.1). En effet, l'environnement électronique d'une guanine impliquée dans la formation d'un G-quadruplex est différent de celui d'une guanine en interaction Watson-Crick au sein d'une structure tige-boucle.

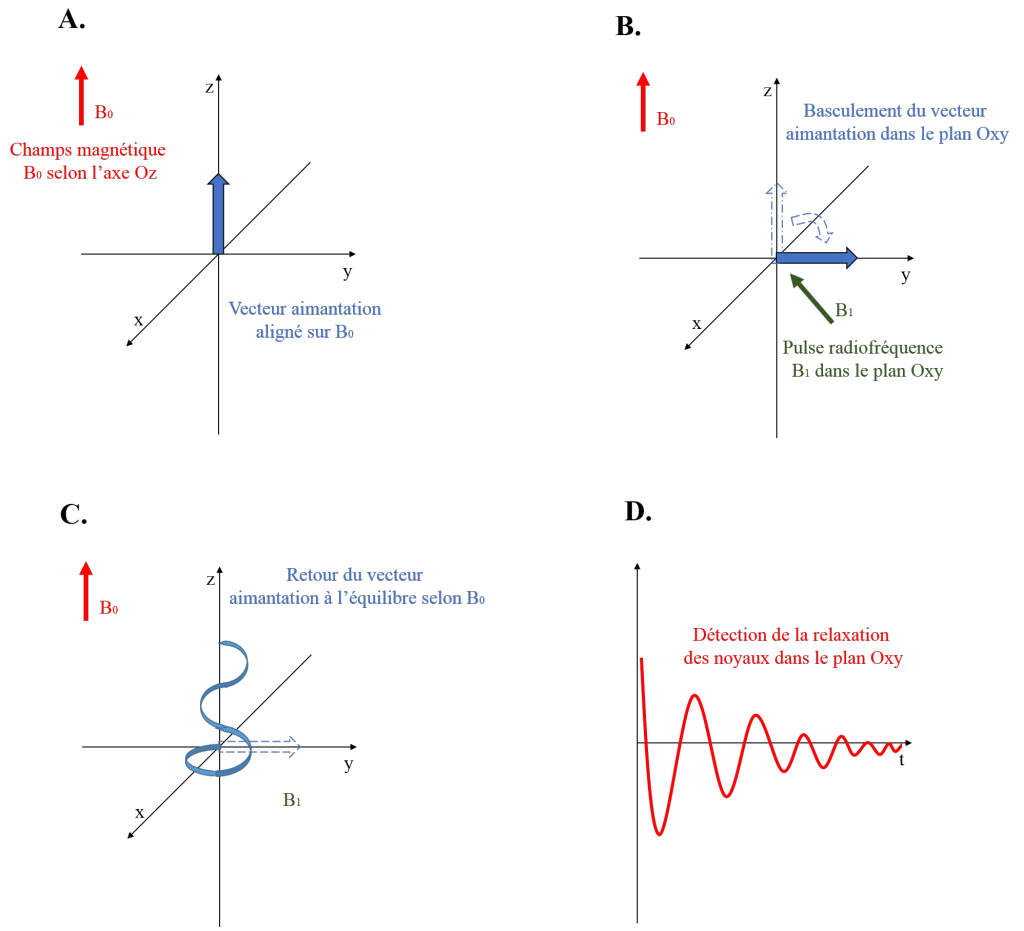


FIGURE 4.10 – **Schéma des différentes étapes du principe de la résonance magnétique nucléaire.**

Dans le cas de nos expériences de RMN, nous avons choisi le TMS (tétraméthylsilane) comme référence ; c'est à dire que nous exprimons les déplacements chimiques δ en fonction de ce dernier. Afin de déterminer si nos oligonucléotides forment des structures G-quadruplexe ou tige-boucle, nous suivons le proton situé sur l'azote N_1 des guanines, noté H_1 (Figure 4.11.A). Ce dernier possède un déplacement chimique δ autour de 13-14 ppm lors-

qu'il est impliqué dans une liaison hydrogène de type Watson-Crick (Figure 4.11.B) ; à l'inverse lorsqu'il est impliqué dans une liaison hydrogène de type Hoogsteen la valeur du déplacement chimique δ est autour de 10-12 ppm Wijmenga et al. (1997) (Figure 4.11.C). Le double intérêt de suivre les protons iminos H_1 est que ces derniers sont moins échangeables avec le solvant lorsqu'ils sont impliqués dans un G-quartet ((Figure 4.11.C) ; de sorte que les signaux obtenus par résonance magnétique nucléaire sont plus résolutifs (Adrian et al., 2012). Il est intéressant de noter que les signaux obtenus sont proportionnels au nombre de protons H_1 impliqués dans les structures G-quartet. Dans notre cas, nous pouvons évaluer la contribution des structures G-quadruplex et tige-boucle révélant une hétérogénéité structurale au sein de l'échantillon. De ce fait nous pouvons quantifier la distribution de ces structures de manière relative l'une par rapport à l'autre. La technique de RMN permet d'étudier des molécules en solution, c'est à dire dans un environnement proche de la réalité ; en opposition à la cristallographie. Cependant, la RMN présente rapidement des limites notamment sur la taille des molécules étudiées, inférieure à 100 aa pour les protéines et 50 nt pour les acides nucléiques. De même, la technique de RMN demande des concentrations assez élevées, de l'ordre du μM , pouvant induire dans notre cas la formation de structures intermoléculaires.

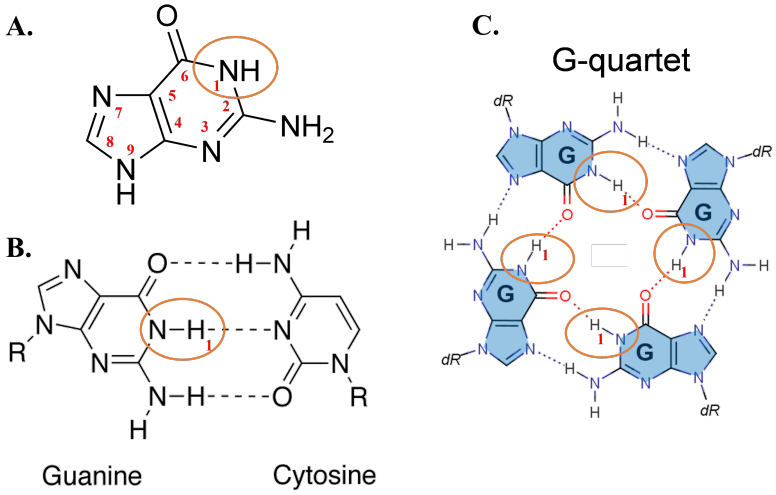


FIGURE 4.11 – **Schéma des interactions des protons H_1 des guanines.** **A.** Schéma d'une guanine. **B.** Schéma d'une liaison Watson-Crick entre une guanine et une cytosine. ©Wikipédia **C.** Schéma d'un G-quartet

Chapitre 5

Conclusion générale

5.1 Les structures G4 aux télomères ne posent pas de problème à la liaison de *RPA* et de *POT1-/+TPP1*

5.1.1 Les protéines *SSB-OB fold RPA* et *POT1-/+TPP1* : des modes de liaison différents aux séquences télomériques

Nous avons vu dans le chapitre 2, Mécanisme de liaison des *SSB* à des séquences télomériques structurées en G4 contigus, que la protéine *RPA* lie et ouvre les longues séquences télomériques structurées en G4 contigus avec une faible coopérativité (Lancrey et al., 2018) ; dans cette étude sur *RPA*, nous avions émis l'hypothèse que cette faible coopérativité était due à la structuration en G4. Cette étude montre que *RPA* n'a pas de difficulté à prendre en charge un, deux, trois ou quatre G-quadruplexes, et par conséquent que ces séquences ne sont pas un problème pour la liaison de *RPA* au séquences télomériques. Même si, en principe, la liaison de *RPA* aux séquences télomériques structurées en G4 ne nécessite pas d'hélicases, quelques hélicases impliquées dans le maintien des télomères pourrait permettre une liaison plus efficace (plus rapide) aux séquences télomériques.

Pour aller plus loin dans l'étude des *SSB-OB fold* en présence de multimère de G4, j'ai étudié la liaison du complexe *POT1-TPP1* et de la protéine *POT1* aux séquences télomériques structurées en G4s contigus. Mes expériences ont montré que le complexe *POT1-TPP1* lie et ouvre coopérativement les longues séquences télomériques formant des multimères de G4. Cette coopérativité du complexe *POT1-TPP1* est due à la structuration de

l'ADN, puisqu'elle n'est observée uniquement qu'en présence de G4 (à savoir en présence de K^+). J'ai également montré que la liaison de plusieurs *POT1-TPP1* progresse avec une directionnalité d'ouverture des G4 contigus de 3' vers 5', cela en raison de la préférence de liaison de *POT1-TPP1* à l'extrémité 3' et de la coopérativité due à la structuration en G4. Cette directionnalité est propre à *POT1-TPP1* ; pour *RPA*, j'ai pu montrer que la liaison à un oligonucléotide se structurant en trois G4 commence par le G4 le moins stable.

Pour *POT1* seule, j'ai observé un comportement similaire à celui du complexe *POT1-TPP1*, mais plus complexe. En effet nous avons vu que le mode de liaison de *POT1* varie en fonction de la concentration en protéine. *POT1* lie de manière "classique" à faible concentration, c'est-à-dire à l'aide de ces deux domaines *OB fold* la séquence 5'-TTAGGGTTAG-3' (Lei et al., 2004). En revanche à haute concentration on observe une liaison partielle de la protéine *POT1*, c'est-à-dire reconnaissant la séquence 5'-TTAGGG-3' par l'intermédiaire d'un unique *OB fold*, probablement *OB1* (par analogie avec *Pot1* de *S. pombe*). Ces modes de liaison de *POT1* n'ont jamais été décrit.

En conclusion nos études montrent que la protéine *RPA*, le complexe *POT1-TPP1* et la protéine *POT1* lient et ouvrent les G4s contigus présents sur de longues séquences télomériques : *RPA* de façon très faiblement coopérative et à partir du G4 moins stable, *POT1-/+TPP1* avec une directionnalité de 3' vers 5' qui découle d'un préférence de liaison pour l'extrémité 3' et d'un certain dégrée de coopérativité due à la structuration de l'ADN télomérique en G4. En perspective, il sera intéressant d'étudier le troisième complexe SSB impliqué dans la stabilité des télomères, le complexe *CST*.

5.1.2 Une instabilité télomérique observée en présence d'autres structures que des G4s

Dans le chapitre 4, des répétitions non-canonique au sein des télomères : structure et interaction avec *RPA*, nous avons présenté une étude portant sur un motif télomérique humain instable, le motif CTAGGG, et visant à la compréhension de l'origine de cette instabilité. Cette instabilité génomique, à savoir des gains ou des pertes de répétitions CTAGGG, est observée à partir d'une dizaine de répétition CTAGGG (Mendez-Bermudez et al., 2009). Des études structurales réalisées par Patrizia Alberti ont montré que, à partir de huit répétitions, le motif CTAGGG se structure également en de longues tige-boucles, contrairement au motif canonique TTAGGG qui se structure uniquement en G4 Mes expériences d'*EMSA* montrent que *RPA* ne lie pas ces séquences CTAGGG lorsqu'elles forment ces structures en tige-boucle. Ces ré-

sultats démontrent que la présence de structures G4s contigus aux télomères ne posent pas de problème à *RPA* contrairement à la présence de structures tige-boucle suffisamment longues (plusieurs dizaines de nucléotides). Il est à relever que *POT1-TPP1* ne reconnaît pas la séquence CTAGGG, pouvant décrire un nouveau rôle de *TPP1* à savoir d'augmenter la discrimination de *POT1* contre les motifs non-canonique. Ces résultats, couplés avec ceux obtenus en présence de *RPA*, suggèrent que l'évolution des motifs télomériques a, peut-être, été guidée par une sélection de motif se structurant en G4 (facilement pris en charge pas *RPA* et *POT1-TPP1*) et une contre-sélection de motif se structurant en tige-boucle, mal pris en charge par une protéine comme *RPA*, essentielle pour le métabolisme de l'ADN.

5.2 Conservation des mécanismes et des structures

5.2.1 Un contact *TPP1*/ADN conservé au sein des protéines de type *POT1*

Au sein du chapitre 3, Structure du complexe ternaire *POT1-TPP1*-ADN, des expériences de pontage du complexe *POT1-TPP1* à l'ADN ont mis en évidence un contact entre *TPP1* et l'ADN. Ce contact a lieu uniquement quand *TPP1* est complexé avec son partenaire *POT1*. Ce résultat correspond au mécanisme décrit chez le cilié *Oxytricha nova*, à savoir que la sous-unité *TEBP*β, homologue de *TPP1* (Wang et al., 2007), lie l'ADN télomérique en présence de *TEBP*α (Horvath et al., 1998). Ce résultat suggère une conservation du mécanisme de liaison des protéines spécifiques des télomères, de type *POT1* (*POT1-like* ou *TEBP-like*). Pour aller plus loin, il serait intéressant d'approfondir cette étude en étudiant d'autre protéines de type *POT1* d'autres organismes modèles, notamment *Schizosaccharomyces pombe* : il serait intéressant de mettre en évidence (ou non) l'existence d'un contact entre *Tpz1*, la protéine partenaire de *Pot1* chez *S. pombe*, et l'ADN télomérique.

Il est cependant étonnant que le domaine *OB fold* de *TEBP*β, et peut-être de *TPP1*, lie l'ADN télomérique uniquement en présence de son partenaire respectif, *TEBP*β, *POT1*. On peut se demander : "Pourquoi la protéine *TEBP*β seule (et *TPP1* seule) ne lie pas l'ADN télomérique ? Nous avons vu dans l'introduction générale section 1.5.1 que les domaines *OB fold* possèdent une polarité de liaison à l'ADN à savoir : l'acide nucléique se lie de sorte que l'extrémité 5' est dirigée vers les brins β_4 et β_5 et l'extrémité 3' vers β_2 (Theo-

bald et al., 2003). Les deux exceptions à cette règle sont les domaines *OB fodl* de *TEBPβ* chez *O. nova* et *SSB* chez *E. coli*. La *SSB* de *E. coli*, qui possède un unique domaine *OB fold*, lie l'ADN sous la forme d'un homotétramère (Bujalowski and Lohman, 1986), suggérant que le domaine *OB fold* seul de *E. coli* ne lie pas seul l'ADN. Cette polarité de liaison inversée pourrait expliquer pourquoi *TEBPβ* seul ne lie pas l'ADN. Une autre hypothèse pouvant expliqué que *TEBPβ* seul ne lie pas l'ADN est l'emplacement du domaine *OB fold* dans la séquence d'acides aminées. Dans une étude sur *Pot1* chez (*S. pombe*), où les deux domaines *OB1* et *OB2* de *Pot1* ont pu être produits et donc étudiés séparément (Croy et al., 2009), les auteurs indiquent avoir ajouté quelques acides aminés en N-term du domaine *OB2* pour que ce dernier lie l'ADN. Ce résultat laisse à penser qu'un domaine *OB fold* seul situé directement en N-term ne puisse lié l'ADN, comme c'est le cas pour *TEBPβ* ou encore *TPP1*. Il est important de noter qu'au sein des cellules, il existe une isoforme minoritaire de *TPP1* qui pour rappel possède 81 a.a en N-term. Par conséquent, on peut se demander si cette isoforme de *TPP1* peut lier l'ADN télomérique. Si cette isoforme de *TPP1* peut lier l'ADN télomérique, sa présence pourrait expliquer pourquoi dans les cellules cancéreuses où *POT1* est muté, soit dans ses domaines de liaison à l'ADN (*OB1-OB2*), soit dans son domaines de liaison à *TPP1* (*OB3*), on observe une activité de la télomérase, qui est pourtant recrutée par le complexe *POT1-TPP1* (Calvete et al., 2017). En effet, si l'isoforme minoritaire de *TPP1* lie l'ADN télomérique seule, alors même si minoritaire au sein de la cellule, cette isoforme pourrait assurer la processivité de la télomérase indépendamment de *POT1*.

5.2.2 Une structure conservée au sein des protéines de type *POT1*

Basé sur une étude *SAXS*, Chen et al. avaient proposé un modèle de structure de *POT1-TPP1* en complexe avec l'ADN plutôt allongé, où les domaines *OB1* - *OB2* de *POT1* sont situés à une extrémité et le domaine *OB* de *TPP1* est situé à l'autre extrémité (Chen et al., 2017). Les résultats de notre étude par microscopie électronique (donc sur des molécules uniques), suggèrent plutôt que *POT1-TPP1* lié à l'ADN adopte une structure en forme de cœur comme celle observée pour le complexe *TEBPα-TEBPβ* des ciliés (Horvath et al., 1998). Notre niveau de résolution ne nous permet pas encore d'affirmer ou d'infirmer cette hypothèse.

En perspective à cette étude nous proposons de résoudre la structure du complexe *POT1-TPP1* à haute résolution par cryo-microscopie électronique. En effet mon étude peut servir de base pour se lancer dans une étude

structurale plus approfondie. Mes résultats montrent que le complexe *POT1-TPP1* en absence d'ADN présente une flexibilité pouvant expliquer pourquoi aucune structure haute résolution du complexe *POT1-TPP1* n'a été résolue à ce jour. Cependant en présence d'ADN, à savoir du 12mer (H12) 5'-GGTTAGGGTTAG-3', les complexes *POT1-TPP1*-H12 formés sont homogènes, ce qui est encourageant pour la détermination de la structure haute résolution du complexe *POT1-TPP1*.

Si les données à haute résolution confirment une structure de *POT1-TPP1* "en forme de cœur" autour de l'ADN, nous pourrons proposer un modèle de structure conservée pour les complexes *POT1-TPP1* et *TEBP*α*-TEBP*β de protection de l'extrémité 3' simple-brin.

5.2.3 Un domaine de résolution de jonction de Holliday fonctionnelle au sein de *POT1* ?

Au sein de la protéine *POT1*, il existe un domaine de résolution de jonction de Holliday (*HJRL*, *Holliday junction resolvase-like*) . Une étude menée par Chen et al. a montré que ce domaine *HJRL* ne lie pas des *jonctions de Holliday* (Chen et al., 2017). Ce qu'on peut reprocher à cette étude est qu'ils ont travaillé uniquement en présence du domaine *HJRL* et non en présence de toute la protéine *POT1*. J'ai voulu faire un tentative d'étude de l'interaction de la protéine *POT1* entière avec la jonction de Holliday proposée dans l'article de Chen et al.. Cependant, au vu de la migration des ADN, je n'ai pas réussi à former à partir des quatre ADN simple brin cette jonction de Hollidays. Cette étape aurait demandé plus de temps afin d'être optimisée. Il serait en effet intéressant de savoir si ce domaine *HJRL* est fonctionnel. Dans le cas d'un résultat positif, cela pourrait laisser envisager d'autres rôles de *POT1*.

Même si je n'ai pas la réponse à la question : "Est-ce que *POT1* peut lier et ouvrir les jonctions de Holliday ?" Je me permets de formuler des hypothèses en me demandant pourquoi il existe un domaine *HJRL* au sein de *POT1*. Est-ce que ce domaine pourrait correspond à un reliquat d'une protéine ancestrale de *POT1* pouvant prendre en charge l'extrémité des chromosomes ? J'ai voulu m'intéresser aux autres modes de protection des extrémités des chromosomes n'impliquant pas des télomères. J'ai donc cherché s'il existait des cellules procaryotes possédant des chromosomes linéaires. C'est ainsi que j'ai trouvé *Borrelia burgdorferi* une bactérie qui possède un chromosome linéaire de 910725 pb et 17 plasmides linéaires et circulaires (Fraser et al., 1997; Casjens et al., 2002). Les extrémités des chromosomes linéaires au sein de *B. burgdorferi* sont constituées de structures tige-boucle covalentes (Figure 5.1)

(Barbour and Garon, 1987; Baril et al., 1989; Ferdows and Barbour, 1989). Lors de la réplication du chromosome linéaire (et des plasmides linéaires) la présence de la structure tige-boucle à l'extrémité nécessite l'intervention de la protéine *ResT Resolvase Telomere* (Huang et al., 2016) qui permet de maintenant la linéarité en résolvant une structure de jonction de Holliday (Figure 5.1). De même, Huang et al. montre dans leur étude que *ResT* est capable de défaire la structure D-loop formée par l'invasion d'un simple brin dans une région double brin. Cette structure *D-loop* n'est pas sans rappeler celle présente lors de l'invasion du *G-overhang* en présence du *Shelterin*.

Il n'existe probablement aucun lien évolutif entre *ResT* et *POT1*, d'autant plus que *ResT* est une enzyme ATP-dépendant. Toutefois, je trouve intéressant qu'au sein d'une bactérie, qui possède un chromosome et des réplicons linéaires, la protéine en charge de la protection de cette extrémité et de son maintien soit une protéine capable de défaire des structures semblables à des jonctions de Holliday, et que, la *POT1* humaine contienne un domaine de résolution de jonction de Holliday en C-term.

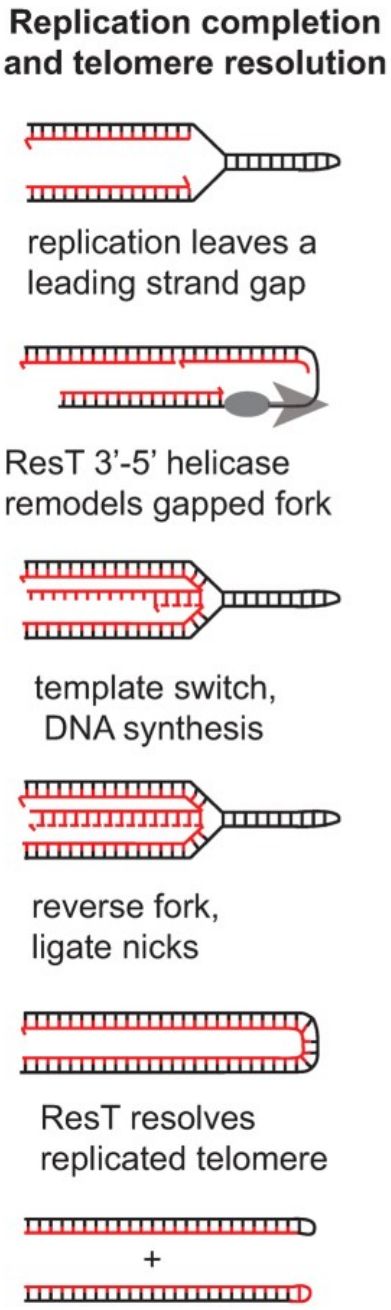


FIGURE 5.1 – **Modèle de la réplication des télomères chez *Borrelia burgdorferi*** (Huang et al., 2016)

Bibliographie

Abragam, A. and Abragam, A. (1961). *The principles of nuclear magnetism*. Number 32. Oxford university press.

Adrian, M., Heddi, B., and Phan, A. T. (2012). NMR spectroscopy of g-quadruplexes. *Methods*, 57(1) :11–24.

Aklilu, B. B., Peurois, F., Saintomé, C., Culligan, K. M., Kobbe, D., Leasure, C., Chung, M., Cattoor, M., Lynch, R., Sampson, L., Fatora, J., and Shippen, D. E. (2020). Functional diversification of replication protein a paralogs and telomere length maintenance in arabidopsis. *Genetics*, 215(4) :989–1002.

Allshire, R. C., Dempster, M., and Hastie, N. D. (1989). Human telomeres contain at least three types of g–rich repeat distributed non-randomly. *Nucleic Acids Research*, 17(12) :4611–4627.

Audry, J., Maestroni, L., Delagoutte, E., Gauthier, T., Nakamura, T. M., Gachet, Y., Saintomé, C., Géli, V., and Coulon, S. (2015). RPA prevents g-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *The EMBO Journal*, 34(14) :1942–1958.

Autexier, C. and Lue, N. F. (2006). The structure and function of telomerase reverse transcriptase. *Annual Review of Biochemistry*, 75(1) :493–517.

Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, 318(5851) :798–801.

Baird, D. M., Coleman, J., Rosser, Z. H., and Royle, N. J. (2000). High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q : Implications for telomere biology and human evolution. *The American Journal of Human Genetics*, 66(1) :235–250.

Baird, D. M., Jeffreys, A. J., and Royle, N. J. (1995). Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human xp/yp telomere. *The EMBO Journal*, 14(21) :5433–5443.

Barbour, A. and Garon, C. (1987). Linear plasmids of the bacterium borrelia burgdorferi have covalently closed ends. *Science*, 237(4813) :409–411.

Baril, C., Richaud, C., Baranton, G., and Girons, I. (1989). Linear chromosome of borrelia burgdorferi. *Research in Microbiology*, 140(7) :507–516.

Bates, P., Mergny, J.-L., and Yang, D. (2007). Quartets in g-major. *EMBO reports*, 8(11) :1003–1010.

Bauer, L., Tlucvkova, K., Tothova, P., and Viglasky, V. (2011). G-quadruplex motifs arranged in tandem occurring in telomeric repeats and the insulin-linked polymorphic region. *Biochemistry*, 50(35) :7484–7492.

Baumann, P. and Cech, T. R. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*, 292(5519) :1171–1175.

Beaume, N., Pathak, R., Yadav, V. K., Kota, S., Misra, H. S., Gautam, H. K., and Chowdhury, S. (2012). Genome-wide study predicts promoter-g4 DNA motifs regulate selective functions in bacteria : radioresistance of d. radiodurans involves g4 DNA-mediated regulation. *Nucleic Acids Research*, 41(1) :76–89.

Bekaert, S., Derradji, H., and Baatout, S. (2004). Telomere biology in mammalian germ cells and during development. *Developmental Biology*, 274(1) :15–30.

Bhattacharjee, A., Wang, Y., Diao, J., and Price, C. M. (2017). Dynamic DNA binding, junction recognition and g4 melting activity underlie the telomeric and genome-wide roles of human CST. *Nucleic Acids Research*, 45(21) :12311–12324.

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA g-quadruplex structures in human cells. *Nature Chemistry*, 5(3) :182–186.

Biroccio, A., Rizzo, A., Elli, R., Koering, C.-E., Belleville, A., Benassi, B., Leonetti, C., Stevens, M. F., D'Incalci, M., Zupi, G., and Gilson, E. (2006). TRF2 inhibition triggers apoptosis and reduces tumourigenicity of human melanoma cells. *European Journal of Cancer*, 42(12) :1881–1888.

Blackburn, E. H. (1992). Telomerases. *Annual Review of Biochemistry*, 61(1) :113–129.

Blackburn, E. H. and Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in tetrahymena. *Journal of Molecular Biology*, 120(1) :33–53.

Blackwell, L. J. and Borowiec, J. A. (1994). Human replication protein a binds single-stranded DNA in two distinct complexes. *Molecular and Cellular Biology*, 14(6) :3993–4001.

Bochkarev, A. (1999). The crystal structure of the complex of replication protein a subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *The EMBO Journal*, 18(16) :4498–4504.

Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein a bound to DNA. *Nature*, 385(6612) :176–181.

Bochkareva, E. (2002). Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *The EMBO Journal*, 21(7) :1855–1863.

Bochkareva, E., Frappier, L., Edwards, A. M., and Bochkarev, A. (1998). The RPA32 subunit of human replication protein a contains a single-stranded DNA-binding domain. *Journal of Biological Chemistry*, 273(7) :3932–3936.

Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S., and Reddel, R. R. (1995). Telomere elongation in immortal human cells without detectable telomerase activity. *The EMBO Journal*, 14(17) :4240–4248.

Bugaut, A. and Alberti, P. (2015). Understanding the stability of DNA g-quadruplex units in long human telomeric strands. *Biochimie*, 113 :125–133.

Bujalowski, W. and Lohman, T. M. (1986). Escherichia coli single-strand binding protein forms multiple, distinct complexes with single-stranded DNA. *Biochemistry*, 25(24) :7799–7802.

Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M., and Murzin, A. G. (1997). The solution structure of the s1 RNA binding domain : A member of an ancient nucleic acid–binding fold. *Cell*, 88(2) :235–242.

Cai, L., Roginskaya, M., Qu, Y., Yang, Z., Xu, Y., and Zou, Y. (2007). Structural characterization of human RPA sequential binding to single-stranded DNA using ssDNA as a molecular ruler†. *Biochemistry*, 46(28) :8226–8233.

Calvete, O., Garcia-Pavia, P., Domínguez, F., Bougeard, G., Kunze, K., Braeuninger, A., Teule, A., Lasa, A., y Cajal, T. R., Llort, G., Fernández, V., Lázaro, C., Urioste, M., and Benitez, J. (2017). The wide spectrum of POT1 gene variants correlates with multiple cancer types. *European Journal of Human Genetics*, 25(11) :1278–1281.

Cardullo, R. A., Agrawal, S., Flores, C., Zamecnik, P. C., and Wolf, D. E. (1988). Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences*, 85(23) :8790–8794.

Casjens, S., Palmer, N., Vugt, R. V., Huang, W. M., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J., Dodson, R. J., Haft, D., Hickey, E., Gwinn, M., White, O., and Fraser, C. M. (2002). A bacterial genome in flux : the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the lyme disease spirochete borrelia burgdorferi. *Molecular Microbiology*, 35(3) :490–516.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., and Mechali, M. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Research*, 21(9) :1438–1449.

Cayrou, C., Grégoire, D., Coulombe, P., Danis, E., and Méchali, M. (2012). Genome-scale identification of active DNA replication origins. *Methods*, 57(2) :158–164.

Cesare, A. J. and Reddel, R. R. (2010). Alternative lengthening of telomeres : models, mechanisms and implications. *Nature Reviews Genetics*, 11(5) :319–330.

Chaires, J. B., Gray, R. D., Dean, W. L., Monsen, R., DeLeeuw, L. W., Stribinskis, V., and Trent, J. O. (2020). Human POT1 unfolds g-quadruplexes by conformational selection. *Nucleic Acids Research*, 48(9) :4976–4991.

Chariker, J. H., Miller, D. M., and Rouchka, E. C. (2016). Computational analysis of g-quadruplex forming sequences across chromosomes reveals high density patterns near the terminal ends. *PLOS ONE*, 11(10) :e0165101.

Chen, C., Gu, P., Wu, J., Chen, X., Niu, S., Sun, H., Wu, L., Li, N., Peng, J., Shi, S., Fan, C., Huang, M., Wong, C. C. L., Gong, Q., Kumar-Sinha, C., Zhang, R., Pusztai, L., Rai, R., Chang, S., and Lei, M. (2017). Structural insights into POT1-TPP1 interaction and POT1 c-terminal mutations in human cancer. *Nature Communications*, 8(1).

Chen, L.-Y., Redon, S., and Lingner, J. (2012). The human CST complex is a terminator of telomerase activity. *Nature*, 488(7412) :540–544.

Chiang, Y. J., Kim, S.-H., Tessarollo, L., Campisi, J., and Hodes, R. J. (2004). Telomere-associated protein TIN2 is essential for early embryonic development through a telomerase-independent pathway. *Molecular and Cellular Biology*, 24(15) :6631–6634.

Choi, K. H., Lakamp-Hawley, A. S., Kolar, C., Yan, Y., Borgstahl, G. E., and Ouellette, M. M. (2015). The OB-fold domain 1 of human POT1 recognizes both telomeric and non-telomeric DNA motifs. *Biochimie*, 115 :17–27.

Cohen, S., Jacob, E., and Manor, H. (2004). Effects of single-stranded DNA binding proteins on primer extension by telomerase. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1679(2) :129–140.

Coleman, J., Baird, D. M., and Royle, N. J. (1999). The plasticity of human telomeres demonstrated by a hypervariable telomere repeat array that is located on some copies of 16p and 16q. *Human Molecular Genetics*, 8(9) :1637–1646.

Conomos, D., Stutz, M. D., Hills, M., Neumann, A. A., Bryan, T. M., Reddel, R. R., and Pickett, H. A. (2012). Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. *The Journal of Cell Biology*, 199(6) :893–906.

Corriveau, M., Mullins, M. R., Baus, D., Harris, M. E., and Taylor, D. J. (2013). Coordinated interactions of multiple POT1-TPP1 proteins with telomere DNA. *Journal of Biological Chemistry*, 288(23) :16361–16370.

Crabbe, L. (2004). Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science*, 306(5703) :1951–1953.

Croy, J. E., Altschuler, S. E., Grimm, N. E., and Wuttke, D. S. (2009). Nonadditivity in the recognition of single-stranded DNA by theSchizosaccharomyces pombeProtection of telomeres 1 DNA-binding domain, pot1-DBD. *Biochemistry*, 48(29) :6864–6875.

de Lange, T. (2005). Shelterin : the protein complex that shapes and safeguards human telomeres. *Genes & Development*, 19(18) :2100–2110.

de Lange, T. (2009). How telomeres solve the end-protection problem. *Science*, 326(5955) :948–952.

Déjardin, J. and Kingston, R. E. (2009). Purification of proteins associated with specific genomic loci. *Cell*, 136(1) :175–186.

Delagoutte, E., Heneman-Masurel, A., and Baldacci, G. (2011). Single-stranded DNA binding proteins unwind the newly synthesized double-stranded DNA of model miniforks. *Biochemistry*, 50(6) :932–944.

Denchi, E. L. and de Lange, T. (2007). Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature*, 448(7157) :1068–1071.

Dickey, T. H. and Wuttke, D. S. (2014). The telomeric protein pot1 from schizosaccharomyces pombe binds ssDNA in two modes with differing 3′ end availability. *Nucleic Acids Research*, 42(15) :9656–9665.

Else, T., Theisen, B. K., Wu, Y., Hutz, J. E., Keegan, C. E., Hammer, G. D., and Ferguson, D. O. (2007). Tpp1/acd maintains genomic stability through a complex role in telomere protection. *Chromosome Research*, 15(8) :1001–1013.

Fan, J. and Pavletich, N. P. (2012). Structure and conformational change of a replication protein a heterotrimer bound to ssDNA. *Genes & Development*, 26(20) :2337–2347.

Fan, J.-H., Bochkareva, E., Bochkarev, A., and Gray, D. M. (2009). Circular dichroism spectra and electrophoretic mobility shift assays show that human replication protein a binds and melts intramolecular g-quadruplex structures. *Biochemistry*, 48(5) :1099–1111.

Fang, G. and Cech, T. R. (1993). The beta subunit of oxytricha telomere-binding protein promotes g-quartet formation by telomeric DNA. *Cell*, 74(5) :875–885.

Fasman, G. D. (1996). *Circular Dichroism and the Conformational Analysis of Biomolecules*. Springer US.

Feng, X., Hsu, S.-J., Kasbek, C., Chaiken, M., and Price, C. M. (2017). CTC1-mediated c-strand fill-in is an essential step in telomere length maintenance. *Nucleic Acids Research*, 45(8) :4281–4293.

Ferdows, M. S. and Barbour, A. G. (1989). Megabase-sized linear DNA in the bacterium borrelia burgdorferi, the lyme disease agent. *Proceedings of the National Academy of Sciences*, 86(15) :5969–5973.

Fernandes, C. A., Morea, E. G. O., dos Santos, G. A., da Silva, V. L., Vieira, M. R., Viviescas, M. A., Chatain, J., Vadel, A., Saintomé, C., Fontes, M. R. M., and Cano, M. I. N. (2020). A multi-approach analysis highlights the relevance of RPA-1 as a telomere end-binding protein (TEBP) in leishmania amazonensis. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1864(7) :129607.

Flynn, R. L., Chang, S., and Zou, L. (2012). RPA and POT1 : Friends or foes at telomeres ? *Cell Cycle*, 11(4) :652–657.

Flynn, R. L. and Zou, L. (2010). Oligonucleotide/oligosaccharide-binding fold proteins : a growing family of genome guardians. *Critical Reviews in Biochemistry and Molecular Biology*, 45(4) :266–275.

Flynn, R. L. and Zou, L. (2011). ATR : a master conductor of cellular responses to DNA replication stress. *Trends in Biochemical Sciences*, 36(3) :133–140.

Forstemann, K., Höss, M., and Lingner, J. (2000). Telomerase-dependent repeat divergence at the 3' ends of yeast telomeres. *Nucleic Acids Research*, 28(14) :2690–2694.

Förster, T. (1965). *Delocalized excitation and excitation transfer*. Florida State University.

Frank, A. K., Tran, D. C., Qu, R. W., Stohr, B. A., Segal, D. J., and Xu, L. (2015). The shelterin TIN2 subunit mediates recruitment of telomerase to telomeres. *PLOS Genetics*, 11(7) :e1005410.

Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M. D., Gocayne, J., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O., and Venter, J. C. (1997). Genomic sequence of a lyme disease spirochaete, borrelia burgdorferi. *Nature*, 390(6660) :580–586.

Frees, S., Menendez, C., Crum, M., and Bagga, P. S. (2014). QGRS-conserve : a computational method for discovering evolutionarily conserved g-quadruplex motifs. *Human Genomics*, 8(1).

Gellert, M., Lipsett, M. N., and Davies, D. R. (1962). HELIX FORMATION BY GUANYLIC ACID. *Proceedings of the National Academy of Sciences*, 48(12) :2013–2018.

Gomes, X. V., Henricksen, L. A., and Wold, M. S. (1996). Proteolytic mapping of human replication protein a : evidence for multiple structural domains and a conformational change upon interaction with single-stranded DNA†. *Biochemistry*, 35(17) :5586–5595.

Gong, Y. and de Lange, T. (2010). A shld1-controlled POT1a provides support for repression of ATR signaling at telomeres through RPA exclusion. *Molecular Cell*, 40(3) :377–387.

Gray, J. T., Celander, D. W., Price, C. M., and Cech, T. R. (1991). Cloning and expression of genes for the oxytricha telomere-binding protein : Specific subunit interactions in the telomeric complex. *Cell*, 67(4) :807–814.

Greider, C. W. and Blackburn, E. H. (1985). Identification of a specific telomere terminaltransferase activity in tetrahymena extracts. *Cell*, 43(2) :405–413.

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., and de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell*, 97(4) :503–514.

Guardino, K., Sheftic, S., Slattery, R., and Alexandrescu, A. (2009). Relative stabilities of conserved and non-conserved structures in the OB-fold superfamily. *International Journal of Molecular Sciences*, 10(5) :2412–2430.

Guo, X., Deng, Y., Lin, Y., Cosme-Blanco, W., Chan, S., He, H., Yuan, G., Brown, E. J., and Chang, S. (2007). Dysfunctional telomeres activate an ATM-ATR-dependent DNA damage response to suppress tumorigenesis. *The EMBO Journal*, 26(22) :4709–4719.

Hayflick, L. and Moorhead, P. (1961). The serial cultivation of human diploid cell strains. *Experimental Cell Research*, 25(3) :585–621.

Hecht, E. (2002). *Optics*. Addison-Wesley, fourth edition.

Heddi, B. and Phan, A. T. (2011). Structure of human telomeric DNA in crowded solution. *Journal of the American Chemical Society*, 133(25) :9824–9833.

Henderson, A., Wu, Y., Huang, Y. C., Chavez, E. A., Platt, J., Johnson, F. B., Brosh, R. M., Sen, D., and Lansdorp, P. M. (2013). Detection of g-quadruplex DNA in mammalian cells. *Nucleic Acids Research*, 42(2) :860–869.

ho Kim, S., Kaminker, P., and Campisi, J. (1999). TIN2, a new regulator of telomere length in human cells. *Nature Genetics*, 23(4) :405–412.

Hockemeyer, D., Palm, W., Else, T., Daniels, J.-P., Takai, K. K., Ye, J. Z.-S., Keegan, C. E., de Lange, T., and Hammer, G. D. (2007). Telomere protection by mammalian pot1 requires interaction with tpp1. *Nature Structural & Molecular Biology*, 14(8) :754–761.

Hom, R. A. and Wuttke, D. S. (2017). Human CST prefers g-rich but not necessarily telomeric sequences. *Biochemistry*, 56(32) :4210–4218.

Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A., and Schultz, S. C. (1998). Crystal structure of the oxytricha nova telomere end binding protein complexed with single strand DNA. *Cell*, 95(7) :963–974.

Houghtaling, B. R., Cuttonaro, L., Chang, W., and Smith, S. (2004). A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Current Biology*, 14(18) :1621–1631.

Hu, C., Rai, R., Huang, C., Broton, C., Long, J., Xu, Y., Xue, J., Lei, M., Chang, S., and Chen, Y. (2017). Structural and functional analyses of the mammalian TIN2-TPP1-TRF2 telomeric complex. *Cell Research*, 27(12) :1485–1502.

Huang, S. H., Cozart, M. R., Hart, M. A., and Kobryn, K. (2016). The borrelia burgdorferi telomere resolvase, ResT, possesses ATP-dependent DNA unwinding activity. *Nucleic Acids Research*, 45(3) :1319–1329.

Huppert, J. L. (2005). Prevalence of quadruplexes in the human genome. *Nucleic Acids Research*, 33(9) :2908–2916.

Hwang, H., Buncher, N., Opresko, P. L., and Myong, S. (2012). POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure*, 20(11) :1872–1880.

Iftode, C., Daniely, Y., and Borowiec, J. A. (1999). Replication protein a (RPA) : The eukaryotic SSB. *Critical Reviews in Biochemistry and Molecular Biology*, 34(3) :141–180.

Jacob, N. K., Skopp, R., and M.Price, C. (2001). G-overhang dynamics at tetrahymena telomeres. *The EMBO Journal*, 20(15) :4299–4308.

Jansson, L. I., Hentschel, J., Parks, J. W., Chang, T. R., Lu, C., Baral, R., Bagshaw, C. R., and Stone, M. D. (2019). Telomere DNA g-quadruplex folding within actively extending human telomerase. *Proceedings of the National Academy of Sciences*, 116(19) :9350–9359.

Karlseder, J. (2002). Senescence induced by altered telomere state, not telomere loss. *Science*, 295(5564) :2446–2449.

Kibe, T., Osawa, G. A., Keegan, C. E., and de Lange, T. (2009). Telomere protection by TPP1 is mediated by POT1a and POT1b. *Molecular and Cellular Biology*, 30(4) :1059–1066.

Kim, C., Snyder, R. O., and Wold, M. S. (1992). Binding properties of replication protein a from human and yeast cells. *Molecular and Cellular Biology*, 12(7) :3050–3059.

Kim, C. and Wold, M. S. (1995). Recombinant human replication protein a binds to polynucleotides with low cooperativity. *Biochemistry*, 34(6) :2058–2064.

Kim, N., Piatyszek, M., Prowse, K., Harley, C., West, M., Ho, P., Coviello, G., Wright, W., Weinrich, S., and Shay, J. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science*, 266(5193) :2011–2015.

Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proceedings of the National Academy of Sciences*, 78(5) :3015–3019.

König, S. L., Evans, A. C., and Huppert, J. L. (2010). Seven essential questions on g-quadruplexes. *BioMolecular Concepts*, 1(2) :197–213.

KOBAYASHI, Y., SATO, K., KIBE, T., SEIMIYA, H., NAKAMURA, A., YUKAWA, M., TSUCHIYA, E., and UENO, M. (2010). Expression of mutant RPA in human cancer cells causes telomere shortening. *Bioscience, Biotechnology, and Biochemistry*, 74(2) :382–385.

Kolpashchikov, D. M. (2001). Polarity of human replication protein a binding to DNA. *Nucleic Acids Research*, 29(2) :373–379.

Lancrey, A., Safa, L., Chatain, J., Delagoutte, E., Riou, J.-F., Alberti, P., and Saintomé, C. (2018). The binding efficiency of RPA to telomeric g-strands folded into contiguous g-quadruplexes is independent of the number of g4 units. *Biochimie*, 146 :68–72.

Lee, M., Hills, M., Conomos, D., Stutz, M. D., Dagg, R. A., Lau, L. M., Reddel, R. R., and Pickett, H. A. (2013). Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. *Nucleic Acids Research*, 42(3) :1733–1746.

Lee, M., Teber, E. T., Holmes, O., Nones, K., Patch, A.-M., Dagg, R. A., Lau, L. M. S., Lee, J. H., Napier, C. E., Arthur, J. W., Grimmond, S. M., Hayward, N. K., Johansson, P. A., Mann, G. J., Scolyer, R. A., Wilmott, J. S., Reddel, R. R., Pearson, J. V., Waddell, N., and Pickett, H. A. (2018). Telomere sequence content can be used to determine ALT activity in tumours. *Nucleic Acids Research*, 46(10) :4903–4918.

Lei, M., Podell, E. R., and Cech, T. R. (2004). Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature Structural & Molecular Biology*, 11(12) :1223–1229.

Lim, C. J., Barbour, A. T., Zaug, A. J., Goodrich, K. J., McKay, A. E., Wuttke, D. S., and Cech, T. R. (2020). The structure of human CST reveals a decameric assembly bound to telomeric DNA. *Science*, 368(6495) :1081–1085.

Lim, K. W., Alberti, P., Guédin, A., Lacroix, L., Riou, J.-F., Royle, N. J., Mergny, J.-L., and Phan, A. T. (2009). Sequence variant (CTAGGG)n in the human telomere favors a g-quadruplex structure containing a g·c·g·c tetrad. *Nucleic Acids Research*, 37(18) :6239–6248.

Lim, K. W., Ng, V. C. M., Martín-Pintado, N., Heddi, B., and Phan, A. T. (2013). Structure of the human telomere in na solution : an antiparallel g-quadruplex scaffold reveals additional diversity. *Nucleic Acids Research*, 41(22) :10556–10562.

Lipps, H. J., Gruissem, W., and Prescott, D. M. (1982). Higher order DNA structure in macronuclear chromatin of the hypotrichous ciliate oxytricha nova. *Proceedings of the National Academy of Sciences*, 79(8) :2495–2499.

Liu, D., Safari, A., O'Connor, M. S., Chan, D. W., Laegeler, A., Qin, J., and Songyang, Z. (2004). PTOP interacts with POT1 and regulates its localization to telomeres. *Nature Cell Biology*, 6(7) :673–680.

Lloyd, N. R., Dickey, T. H., Hom, R. A., and Wuttke, D. S. (2016). Tying up the ends : Plasticity in the recognition of single-stranded DNA at telomeres. *Biochemistry*, 55(38) :5326–5340.

Lombardi, E. P. and Londoño-Vallejo, A. (2019). A guide to computational methods for g-quadruplex prediction. *Nucleic Acids Research*, 48(1) :1–15.

Luciano, P., Coulon, S., Faure, V., Corda, Y., Bos, J., Brill, S. J., Gilson, E., Simon, M.-N., and Géli, V. (2012). RPA facilitates telomerase activity at chromosome ends in budding and fission yeasts. *The EMBO Journal*, 31(8) :2034–2046.

Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L., and Patel, D. J. (2006). Structure of the human telomere in kcl solution an intramolecular g-quadruplex scaffold. *Journal of the American Chemical Society*, 128(30) :9963–9970.

Maestroni, L., Matmati, S., and Coulon, S. (2017). Solving the telomere replication problem. *Genes*, 8(2) :55.

Makarov, V. L., Hirose, Y., and Langmore, J. P. (1997). Long g tails at both ends of human chromosomes suggest a c strand degradation mechanism for telomere shortening. *Cell*, 88(5) :657–666.

Masiero, S., Trotta, R., Pieraccini, S., Tito, S. D., Perone, R., Randazzo, A., and Spada, G. P. (2010). A non-empirical chromophoric interpretation of CD spectra of DNA g-quadruplex structures. *Organic & Biomolecular Chemistry*, 8(12) :2683.

McClintock, B. (1939). The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proceedings of the National Academy of Sciences*, 25(8) :405–416.

McElligott, R. and Wellinger, R. J. (1997). The terminal DNA structure of mammalian chromosomes. *The EMBO Journal*, 16(12) :3705–3714.

Mendez-Bermudez, A., Hills, M., Pickett, H. A., Phan, A. T., Mergny, J.-L., Riou, J.-F., and Royle, N. J. (2009). Human telomeres that contain (CTAGGG)n repeats show replication dependent instability in somatic cells and the male germline. *Nucleic Acids Research*, 37(18) :6225–6238.

Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C., Edwards, A. M., and Chazin, W. J. (2000). Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell*, 103(3) :449–456.

Métifiot, M., Amrane, S., Litvak, S., and Andreola, M.-L. (2014). G-quadruplexes in viruses : function and potential therapeutic applications. *Nucleic Acids Research*, 42(20) :12352–12366.

Mirman, Z., Lottersberger, F., Takai, H., Kibe, T., Gong, Y., Takai, K., Bianchi, A., Zimmermann, M., Durocher, D., and de Lange, T. (2018). 53bp1 shieldin counteracts resection through CST- and pol-dependent fill-in. *Nature*, 560(7716) :112–116.

Miyake, Y., Nakamura, M., Nabetani, A., Shimamura, S., Tamura, M., Yonehara, S., Saito, M., and Ishikawa, F. (2009). RPA-like mammalian ctc1-stn1-ten1 complex binds to single-stranded DNA and protects telomeres independently of the pot1 pathway. *Molecular Cell*, 36(2) :193–206.

Morin, G. B. (1989). The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell*, 59(3) :521–529.

Moser, B. A. and Nakamura, T. M. (2009). Protection and replication of telomeres in fission yeastThis paper is one of a selection of papers published in this special issue, entitled 30th annual international asilomar chromatin and chromosomes conference, and has undergone the journal's usual peer review process. *Biochemistry and Cell Biology*, 87(5) :747–758.

Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L., and Wu, J. R. (1988). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences*, 85(18) :6622–6626.

Muller, H. J. (1938). The remaking of chromosomes. *Collecting Net.*, (13) :181–195.

Mullins, M. R., Rajavel, M., Hernandez-Sanchez, W., de la Fuente, M., Biendarra, S. M., Harris, M. E., and Taylor, D. J. (2016). POT1–TPP1 binding and unfolding of telomere DNA discriminates against structural polymorphism. *Journal of Molecular Biology*, 428(13) :2695–2708.

Murzin, A. G. (1993). Ob(oligonucleotide/oligosaccharide binding)-fold : common structural and functional solution for non-homologous sequences. *The EMBO journal*, 12(3) :861–867.

Nandakumar, J., Podell, E. R., and Cech, T. R. (2009). How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA. *Proceedings of the National Academy of Sciences*, 107(2) :651–656.

Nguyen, T. H. D., Tam, J., Wu, R. A., Greber, B. J., Toso, D., Nogales, E., and Collins, K. (2018). Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature*, 557(7704) :190–195.

Olovnikov, A. (1973). A theory of marginotomy. *Journal of Theoretical Biology*, 41(1) :181–190.

Ono, Y. (2003). A novel allele of fission yeast rad11 that causes defects in DNA repair and telomere length regulation. *Nucleic Acids Research*, 31(24) :7141–7149.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H. J. (2005). Telomere end-binding proteins control the formation of g-quadruplex DNA structures in vivo. *Nature Structural & Molecular Biology*, 12(10) :847–854.

Palm, W. and de Lange, T. (2008). How shelterin protects mammalian telomeres. *Annual Review of Genetics*, 42(1) :301–334.

Panier, S. and Boulton, S. J. (2013). Double-strand break repair : 53bp1 comes into focus. *Nature Reviews Molecular Cell Biology*, 15(1) :7–18.

Parkinson, G. N., Lee, M. P. H., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, 417(6891) :876–880.

Petraccone, L., Spink, C., Trent, J. O., Garbett, N. C., Mekmaysy, C. S., Giancola, C., and Chaires, J. B. (2011). Structure and stability of higher-order human telomeric quadruplexes. *Journal of the American Chemical Society*, 133(51) :20951–20961.

Pfuetzner, R. A., Bochkarev, A., Frappier, L., and Edwards, A. M. (1997). Replication protein a. *Journal of Biological Chemistry*, 272(1) :430–434.

Phan, A. T., Luu, K. N., and Patel, D. J. (2006). Different loop arrangements of intramolecular human telomeric g-quadruplexes in k solution. *Nucleic Acids Research*, 34(19) :5715–5719.

Pokhrel, N., Caldwell, C. C., Corless, E. I., Tillison, E. A., Tibbs, J., Jocic, N., Tabei, S. M. A., Wold, M. S., Spies, M., and Antony, E. (2019). Dynamics and selective remodeling of the DNA-binding domains of RPA. *Nature Structural & Molecular Biology*, 26(2) :129–136.

Polman, A. and Atwater, H. A. (2005). Plasmonics : optics at the nanoscale. *Materials Today*, 8(1) :56.

Prakash, A., Natarajan, A., Marky, L. A., Ouellette, M. M., and Borgstahl, G. E. O. (2011). Identification of the DNA-binding domains of human replication protein a that recognize g-quadruplex DNA. *Journal of Nucleic Acids*, 2011 :1–14.

Punnoose, J. A., Cui, Y., Koirala, D., Yangyuoru, P. M., Ghimire, C., Shrestha, P., and Mao, H. (2014). Interaction of g-quadruplexes in the full-length 3′ human telomeric overhang. *Journal of the American Chemical Society*, 136(52) :18062–18069.

Qureshi, M. H., Ray, S., Sewell, A. L., Basu, S., and Balci, H. (2012). Replication protein a unfolds g-quadruplex structures with varying degrees of efficiency. *The Journal of Physical Chemistry B*, 116(19) :5588–5594.

Rajavel, M., Mullins, M. R., and Taylor, D. J. (2014). Multiple facets of TPP1 in telomere maintenance. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1844(9) :1550–1559.

Ray, S., Bandaria, J. N., Qureshi, M. H., Yildiz, A., and Balci, H. (2014). G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proceedings of the National Academy of Sciences*, 111(8) :2990–2995.

Ray, S., Qureshi, M. H., Malcolm, D. W., Budhathoki, J. B., Çelik, U., and Balci, H. (2013). RPA-mediated unfolding of systematically varying g-quadruplex structures. *Biophysical Journal*, 104(10) :2235–2245.

Rhodes, D. and Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research*, 43(18) :8627–8637.

Rice, C., Doukov, T., and Skordalakes, E. (2017a). Structure of the human POT1-TPP1 telomeric complex.

Rice, C., Shastrula, P. K., Kossenkov, A. V., Hills, R., Baird, D. M., Showe, L. C., Doukov, T., Janicki, S., and Skordalakes, E. (2017b). Structural and functional analysis of the human POT1-TPP1 telomeric complex. *Nature Communications*, 8(1).

Saenger, W. (1984). Principles of nucleic acid structure.

Safa, L., Delagoutte, E., Petruseva, I., Alberti, P., Lavrik, O., Riou, J.-F., and Saintomé, C. (2014). Binding polarity of RPA to telomeric sequences and influence of g-quadruplex stability. *Biochimie*, 103 :80–88.

Safa, L., Gueddouda, N. M., Thiébaut, F., Delagoutte, E., Petruseva, I., Lavrik, O., Mendoza, O., Bourdoncle, A., Alberti, P., Riou, J.-F., and Saintomé, C. (2016). 5′ to 3′ unfolding directionality of DNA secondary structures by replication protein a. *Journal of Biological Chemistry*, 291(40) :21246–21256.

Salas, T. R. (2005). *Etude in vitro des interactions entre la protéine de réplication A humaine (hRPA) et l' ADN*. PhD thesis, Université Pierre et Marie Curie (UPMC) Paris 6.

Salas, T. R., Petruseva, I., Lavrik, O., Bourdoncle, A., Mergny, J.-L., Favre, A., and Saintomé, C. (2006). Human replication protein a unfolds telomeric g-quadruplexes. *Nucleic Acids Research*, 34(17) :4857–4865.

Salas, T. R., Petruseva, I., Lavrik, O., and Saintome, C. (2008). Evidence for direct contact between the RPA3 subunit of the human replication protein a and single-stranded DNA. *Nucleic Acids Research*, 37(1) :38–46.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Pluckthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with stylonychia lemnae macronuclei. *Proceedings of the National Academy of Sciences*, 98(15) :8572–8577.

Schramke, V., Luciano, P., Brevet, V., Guillot, S., Corda, Y., Longhese, M. P., Gilson, E., and Géli, V. (2003). RPA regulates telomerase action by providing est1p access to chromosome ends. *Nature Genetics*, 36(1) :46–54.

Setiaputra, D. and Durocher, D. (2019). Shieldin the protector of DNA ends. *EMBO reports*, 20(5).

Sfeir, A., Kosiyatrakul, S. T., Hockemeyer, D., MacRae, S. L., Karlseder, J., Schildkraut, C. L., and de Lange, T. (2009). Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*, 138(1) :90–103.

Shiotani, B. and Zou, L. (2009). Single-stranded DNA orchestrates an ATM-to-ATR switch at DNA breaks. *Molecular Cell*, 33(5) :547–558.

Školáková, P., Renčiuk, D., Palacký, J., Krafčík, D., Dvořáková, Z., Kejnovská, I., Bednářová, K., and Vorlíčková, M. (2019). Systematic investigation of sequence requirements for DNA i-motif formation. *Nucleic Acids Research*, 47(5) :2177–2189.

Smith, J., Zou, H., and Rothstein, R. (2000). Characterization of genetic interactions with RFA1 : the role of RPA in DNA replication and telomere maintenance. *Biochimie*, 82(1) :71–78.

Stewart, J. A., Wang, F., Chaiken, M. F., Kasbek, C., Chastain, P. D., Wright, W. E., and Price, C. M. (2012). Human CST promotes telomere duplex replication and general replication restart after fork stalling. *The EMBO Journal*, 31(17) :3537–3549.

Sui, J., Lin, Y.-F., Xu, K., Lee, K.-J., Wang, D., and Chen, B. P. (2015). DNA-PKcs phosphorylates hnRNP-a1 to facilitate the RPA-to-POT1 switch and telomere capping after replication. *Nucleic Acids Research*, 43(12) :5971–5983.

Sundquist, W. I. and Heaphy, S. (1993). Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proceedings of the National Academy of Sciences*, 90(8) :3393–3397.

Takai, K. K., Kibe, T., Donigian, J. R., Frescas, D., and de Lange, T. (2011). Telomere protection by TPP1/POT1 requires tethering to TIN2. *Molecular Cell*, 44(4) :647–659.

Tan, J. and Lan, L. (2020). The DNA secondary structures at telomeres and genome instability. *Cell & Bioscience*, 10(1).

Taylor, D. J., Podell, E. R., Taatjes, D. J., and Cech, T. R. (2011). Multiple POT1–TPP1 proteins coat and compact long telomeric single-stranded DNA. *Journal of Molecular Biology*, 410(1) :10–17.

Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003). Nucleic acid recognition by OB-fold proteins. *Annual Review of Biophysics and Biomolecular Structure*, 32(1) :115–133.

Tran, P. L. T., Mergny, J.-L., and Alberti, P. (2010). Stability of telomeric g-quadruplexes. *Nucleic Acids Research*, 39(8) :3282–3294.

Tuesuwan, B., Kern, J. T., Thomas, P. W., Rodriguez, M., Li, J., David, W. M., and Kerwin, S. M. (2008). Simian virus 40 large t-antigen g-quadruplex DNA helicase inhibition by g-quadruplex DNA-interactive agents†. *Biochemistry*, 47(7) :1896–1909.

Vannier, J.-B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I., Ding, H., and Boulton, S. J. (2012). RTEL1 dismantles t loops and counteracts telomeric g4-DNA to maintain telomere integrity. *Cell*, 149(4) :795–806.

Verdun, R. E. and Karlseder, J. (2006). The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres. *Cell*, 127(4) :709–720.

Vorlickova, M. (2005). Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Research*, 33(18) :5851–5860.

Walther, A. P., Gomes, X. V., Lao, Y., Lee, C. G., and Wold, M. S. (1999). Replication protein a interactions with DNA. 1. functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit†. *Biochemistry*, 38(13) :3963–3973.

Wang, F., Podell, E. R., Zaug, A. J., Yang, Y., Baciu, P., Cech, T. R., and Lei, M. (2007). The POT1–TPP1 telomere complex is a telomerase processivity factor. *Nature*, 445(7127) :506–510.

Wang, F., Stewart, J., and Price, C. M. (2014). Human CST abundance determines recovery from diverse forms of DNA damage and replication stress. *Cell Cycle*, 13(22) :3488–3498.

Wang, H., Nora, G. J., Ghodke, H., and Opresko, P. L. (2010). Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting g-quadruplex unfolding. *Journal of Biological Chemistry*, 286(9) :7479–7489.

Wang, Y. and Patel, D. J. (1993). Solution structure of the human telomeric repeat d[AG3(t2ag3)3] g-tetraplex. *Structure*, 1(4) :263–282.

Watson, J. D. (1972). Origin of concatemeric t7dna. *Nature New Biology*, 239(94) :197–201.

Wellinger, R. and Sen, D. (1997). The DNA structures at the ends of eukaryotic chromosomes. *European Journal of Cancer*, 33(5) :735–749.

Wellinger, R. J., Ethier, K., Labrecque, P., and Zakian, V. A. (1996). Evidence for a new step in telomere maintenance. *Cell*, 85(3) :423–433.

Wellinger, R. J., Wolf, A. J., and Zakian, V. A. (1993). Saccharomyces telomeres acquire single-strand TG1–3 tails late in s phase. *Cell*, 72(1) :51–60.

Wijmenga, S. S., Kruithof, M., and Hilbers, C. W. (1997). Analysis of ^{1}h chemical shifts in dna : Assessment of the reliability of ^{1}h chemical shift calculations for use in structure refinement. *Journal of Biomolecular NMR*, 10(4) :337–350.

Wold, M. S. and Kelly, T. (1988). Purification and characterization of replication protein a, a cellular protein required for in vitro replication of simian virus 40 DNA. *Proceedings of the National Academy of Sciences*, 85(8) :2523–2527.

Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D., and Shay, J. W. (1997). Normal human chromosomes have long g-rich telomeric overhangs at one end. *Genes & Development*, 11(21) :2801–2809.

Wu, L., Multani, A. S., He, H., Cosme-Blanco, W., Deng, Y., Deng, J. M., Bachilo, O., Pathak, S., Tahara, H., Bailey, S. M., Deng, Y., Behringer, R. R., and Chang, S. (2006). Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell*, 126(1) :49–62.

Wynford-Thomas, D. and Kipling, D. (1997). The end-replication problem. *Nature*, 389(6651) :551–551.

Xin, H., Liu, D., Wan, M., Safari, A., Kim, H., Sun, W., O'Connor, M. S., and Songyang, Z. (2007). TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature*, 445(7127) :559–562.

Xu, M., Kiselar, J., Whited, T. L., Hernandez-Sanchez, W., and Taylor, D. J. (2019). POT1-TPP1 differentially regulates telomerase via POT1 his266 and as a function of single-stranded telomere DNA length. *Proceedings of the National Academy of Sciences*, 116(47) :23527–23533.

Xu, Y., Ishizuka, T., Kurabayashi, K., and Komiyama, M. (2009). Consecutive formation of g-quadruplexes in human telomeric-overhang DNA : A protective capping structure for telomere ends. *Angewandte Chemie International Edition*, 48(42) :7833–7836.

Ye, J. Z.-S. (2004). POT1-interacting protein PIP1 : a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes & Development*, 18(14) :1649–1654.

Yu, E. Y., Sun, J., Lei, M., and Lue, N. F. (2011). Analyses of candida cdc13 orthologues revealed a novel OB fold dimer arrangement, dimerization-assisted DNA binding, and substantial structural differences between cdc13 and RPA70. *Molecular and Cellular Biology*, 32(1) :186–198.

Yu, H.-Q., Miyoshi, D., and Sugimoto, N. (2006). Characterization of structure and stability of long telomeric DNA g-quadruplexes. *Journal of the American Chemical Society*, 128(48) :15461–15468.

Zaug, A. J., Podell, E. R., and Cech, T. R. (2005). Human POT1 disrupts telomeric g-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences*, 102(31) :10864–10869.

Zernik-Kobak, M., Vasunia, K., Connelly, M., Anderson, C. W., and Dixon, K. (1997). Sites of UV-induced phosphorylation of the p34 subunit of replication protein a from HeLa cells. *Journal of Biological Chemistry*, 272(38) :23896–23904.

Zhang, M., Wang, B., Li, T., Liu, R., Xiao, Y., Geng, X., Li, G., Liu, Q., Price, C. M., Liu, Y., and Wang, F. (2019). Mammalian CST averts replication failure by preventing g-quadruplex accumulation. *Nucleic Acids Research*, 47(10) :5243–5259.

Zhong, F. L., Batista, L. F., Freund, A., Pech, M. F., Venteicher, A. S., and Artandi, S. E. (2012). TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell*, 150(3) :481–494.

Zhou, G., Liu, X., Li, Y., Xu, S., Ma, C., Wu, X., Cheng, Y., Yu, Z., Zhao, G., and Chen, Y. (2016). Telomere targeting with a novel g-quadruplex-interactive ligand BRACO-19 induces t-loop disassembly and telomerase displacement in human glioblastoma cells. *Oncotarget*, 7(12) :14925–14939.

Zimmermann, M., Kibe, T., Kabir, S., and de Lange, T. (2014). TRF1 negotiates TTAGGG repeat-associated replication problems by recruiting the BLM helicase and the TPP1/POT1 repressor of ATR signaling. *Genes & Development*, 28(22) :2477–2491.

Chapitre 6

Annexes Articles

6.1 The binding efficiency of RPA to telomeric G-strands folded into contiguous G-quadruplexes is independent of the number of G4 units

Biochimie 146 (2018) 68–72

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Short communication

The binding efficiency of RPA to telomeric G-strands folded into contiguous G-quadruplexes is independent of the number of G4 units

Astrid Lancrey [a], Layal Safa [a], Jean Chatain [a], Emmanuelle Delagoutte [a], Jean-François Riou [a], Patrizia Alberti [a, **], Carole Saintomé [a, b, *]

[a] Structure et Instabilité des Génomes, Sorbonne Universités, Muséum National d'Histoire Naturelle, CNRS UMR 7196, INSERM U1154, 43 rue Cuvier, F-75005, Paris, France
[b] Sorbonne Universités, UPMC Univ Paris 06, UFR927, F-75005, Paris, France

ARTICLE INFO

Article history:
Received 9 October 2017
Accepted 24 November 2017

Keywords:
G-quadruplex
Telomere
Replication protein A
Fluorescence resonance energy transfer (FRET)
Gel electrophoresis

ABSTRACT

Replication protein A (RPA) is a single-stranded DNA binding protein involved in replication and in telomere maintenance. During telomere replication, G-quadruplexes (G4) can accumulate on the lagging strand template and need to be resolved. It has been shown that human RPA is able to unfold a single G4. Nevertheless, the G-strand of human telomeres is prone to fold into higher-order structures formed by contiguous G-quadruplexes. To understand how RPA deals with these structures, we studied its interaction with telomeric G-strands folding into an increasing number of contiguous G4s. The aim of this study was to determine whether the efficiency of binding/unfolding of hRPA to telomeric G-strands depends on the number of G4 units. Our data show that the number n of contiguous G4 units ($n \geq 2$) does not affect the efficiency of hRPA to coat transiently exposed single-stranded telomeric G-strands. This feature may be essential in preventing instability due to G4 structures during telomere replication.
© 2017 Elsevier B.V. and Société Française de Biochimie et Biologie Moléculaire (SFBBM). All rights reserved.

1. Introduction

The G-rich telomeric sequence can adopt non-canonical DNA conformations known as G-quadruplexes (G4s). These four-stranded structures are based on guanine quartets stabilized by Hoogsteen hydrogen bonds and coordinated by central cations. Evidence is accumulating that G4s can form at telomeres during lagging strand DNA replication and at the 3′ G-overhang [1–4]. Several helicases (such as the Werner helicase, WRN) able to unwind G4s are essential for telomere stability and replication [5–7]. In addition, several non-helicase single-stranded DNA binding proteins, involved in telomere maintenance (replication and capping), have been shown to unfold G4s. These proteins are the POT1-TPP1 complex [8,9] (two Shelterin components [10]), RPA [11–14] (a key player in DNA replication, recombination and repair, that is highly conserved in eukaryotes [15]), and the CST complex [16] (a RPA-like complex [17]). The recruitment of these proteins is tightly coordinated through the cell cycle, ensuring telomere replication and telomere protection [18,19]. Here, we focus on human Replication Protein A (hRPA). In human cells, RPA associates with telomeres [20]; it has been proposed that it binds telomeric single-stranded DNA at replication forks (or telomere ends) during early to middle S-phase [18]. A recent study shows that yeast RPA ensures correct replication of the telomeric lagging strand, and suggests a role of RPA in preventing DNA secondary structures on the telomeric lagging strand template [21]. So far, studies concerning the interaction of hRPA with G4s have focused on a single quadruplex formed by four repetitions of the human telomeric motif (GGGTTA) [11,22–24]. Results obtained with a single G4 may not necessarily reflect the binding properties of hRPA to long telomeric G-strands, which are prone to fold into contiguous (or multimeric) G4s, like beads-on-a-string [25–29]. In the present work, we address, for the first time, the question of how hRPA deals with contiguous (or multimeric) telomeric G4s, with the aim of determining whether its efficiency of binding/unfolding long telomeric G-strands folded into contiguous G4 depends on the number of G4 units.

* Corresponding author. Structure et Instabilité des Génomes, Sorbonne Universités, Muséum National d'Histoire Naturelle, CNRS UMR 7196, INSERM U1154, 43 rue Cuvier, F-75005, Paris, France.
** Corresponding author.
E-mail addresses: alberti@mnhn.fr (P. Alberti), carole.saintome@mnhn.fr (C. Saintomé).

https://doi.org/10.1016/j.biochi.2017.11.017
0300-9084/© 2017 Elsevier B.V. and Société Française de Biochimie et Biologie Moléculaire (SFBBM). All rights reserved.

2. Materials and methods

2.1. Materials

Recombinant hRPA was expressed in the *Escherichia coli* BL21 (DE3) strain transformed with the plasmid pET$_{11a}$hRPA that permits the co-expression of RPA1, RPA2 and RPA3. hRPA was purified as previously described [21]. Oligonucleotides were purchased from Eurogentec.

2.2. Electrophoretic mobility shift assays

Oligonucleotides were labeled with γ[^{32}P]ATP using T4 polynucleotide kinase. For all EMSA experiments, hRPA was diluted and pre-incubated (20 min at 4 °C) in a buffer containing 50 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM DTT, 10% glycerol, 0.2 mg/mL BSA and 0.1 mM EDTA. In a standard reaction, radiolabeled oligonucleotide (2 nM) (slowly annealed from high to low temperature) was incubated with various amounts of protein in 10 μL reaction buffer [50 mM HEPES pH 7.9, 0.1 mg/ml BSA, 100 mM KCl or NaCl and 2% glycerol] for 20 min at 20 °C. Samples were then loaded on a native 1% agarose in 0.5× TBE buffer. After electrophoresis, the gel was dried and exposed on a phosphorimager screen, scanned with a Typhoon instrument (Molecular Dynamics). Quantification was made with ImageQuant version 5.1. For each hRPA concentration, the fraction of radiolabelled oligonucleotide bound to hRPA was calculated as follow: $I_{\text{DNA bound to hRPA}}/(I_{\text{free DNA}} + I_{\text{DNA bound to hRPA}})$.

2.3. Fluorescence titrations

Fluorescence titrations were carried out on a SPEX Fluorolog spectrofluorimeter (Jobin Yvon) equipped with a circulating water bath to regulate the temperature of the cell holder. Fluorescence spectra of 100 nM H69-Q1,2,3 oligonucleotides (doubly-labeled with a FAM and a TAMRA) were recorded at 20 °C in a buffer containing 50 mM KCl, 2 mM $MgCl_2$, and 5 mM lithium cacodylate (pH 7.2). H69-Q1,2,3 were pre-incubated 10 min at 20 °C. The protein (from 100 to 1000 nM) was directly added to the solutions containing the oligonucleotides. The spectra were collected after 2 min of incubation while exciting at 470 nm. The fluorescence intensity of the donor FAM (I_D) was measured at 518 nm, the fluorescence intensity of the acceptor TAMRA (I_A) was measured at 586 nm; the percentage of opened labeled G4 units at 20 °C was calculated as:

$$(P-P_f) \times 100/(P_u-P_f),$$

where:

$P = I_D/(I_D + I_A)$ at a given hRPA concentration at 20 °C,
$P_f = I_D/(I_D + I_A)$ in the absence of hRPA at 20 °C,
$P_u = I_D/(I_D + I_A)$ in the absence of hRPA at 80 °C.

We assumed that, in the absence of hRPA, at 20 °C all the labeled G4 units are folded, while at 80 °C all the labeled G4 units are unfolded. Under these assumptions, P_f is the P value corresponding to 0% of unfolded labeled G4 units, while P_u is the P value corresponding to 100% of unfolded labeled G4 units. The expression $(P - P_f)\ x\ 100/(P_u - P_f)$ provides, hence, the percentage of unfolded labeled G4 units for a given hRPA concentration.

3. Results and discussion

We carried out our study in the presence of the biological relevant potassium cation. We first investigated if hRPA is able to unfold contiguous G4s. To this purpose, we carried out fluorescent resonance energy transfer (FRET) titration experiments using the telomeric sequence H69 ($(G_3T_2A)_3G_3$), which folds into three contiguous G4 units. Three modified H69 oligonucleotides, doubly labeled with a 6-carboxyfluorescein (FAM) and a tetramethylrhodamine (TAMRA) at different positions, were used. In H69-Q1, the fluorophores are placed at both sides of the first G4 unit; in H69-Q2, they are placed at both sides of the second G4 unit; in H69-Q3, they are placed at both sides of the third G4 unit (Supplementary Information, Table S1). The doubly labeled G4 units of these three constructs have temperatures of half-denaturation in the range of 50–60 °C, as previously assessed by melting experiments followed by fluorescence [26]. For all H69-Q1,2,3, fluorimetric titration with hRPA revealed an increase of FAM fluorescence and a decrease in TAMRA fluorescence upon hRPA addition, indicating that hRPA is able to open each of the three G4 units, independently of its position (Fig. 1). These results demonstrate that hRPA is able to unfold all the G4s present in a multimeric structure.

Having shown that hRPA unfolds multimeric G4s, we next investigated whether its binding efficiency depended on the number of G4 units. To this purpose, we carried out electrophoretic mobility shift assays (EMSA) using four oligonucleotides: H27 ($T_2A(G_3T_2A)_3G_3T_2A$), H51 ($T_2A(G_3T_2A)_7G_3T_2A$) H75 ($T_2A(G_3T_2A)_{11}G_3T_2A$) and H99 ($T_2A(G_3T_2A)_{15}G_3T_2A$). We previously characterized the structures formed by these oligonucleotides: we showed that H51, H75 et H99 fold into two, three and four contiguous G4 units, respectively, which are non-interacting and have identical stabilities; while the single G4 formed by the shortest sequence H27 has a higher stability [26] (Supplementary Information, Table S1). We point out that the binding constant of hRPA to dT_{50}, dT_{60} and dT_{70} oligonucleotides (longer than the size of the 30-nt RPA extended-binding mode and able in principle to bind more than one RPA, up to three RPA for dT_{70}) does not depend on their length [30]. Hence, any significant difference in the binding of RPA to H51, H75 and H99 should not be due to oligonucleotide length, but solely on the number of G4 units. H27, H51, H75, H99 oligonucleotides in the presence of increasing hRPA concentrations showed retarded bands on agarose gel (Fig. 2A). The number of bands increased with the length of the oligonucleotide, revealing hRPA:DNA complexes of increasing stoichiometry, where one, two, three, four (or more) hRPA are bound. These results show that hRPA is able to bind telomeric strands structured in multimeric G4s and to form complexes where several proteins are present along the DNA strand. For each oligonucleotide, we quantified the percentage of DNA bound to hRPA as a function of hRPA concentration and determined the hRPA concentration to obtain 50% of bound oligonucleotide (Fig. 2B and C). The binding curves showed two main features: *(i)* hRPA bound more efficiently the multimeric G4 sequences (H51, H75 and H99) than the monomeric G4 sequence (H27); *(ii)* the efficiency of hRPA binding to the multimeric G4 sequences (H51, H75 and H99) was independent of the number of formed G4 units (two, three or four). In addition, the non-sigmoidal shape of the binding curves, the coexistence of complexes of different stoichiometries (CI, CII, CIII and CIII+) up to 25–50 nM hRPA concentrations (quantification of the percentage of each complex from Fig. 1 is shown in Supplementary Information, Fig. S1) and data fitting according to a model of binding to a one-dimensional homogeneous lattice [31] (Supplementary Information, Fig. S2) further suggest a poorly cooperative binding mode of hRPA to H51, H75 and H99, as previously observed with long non-structured single-stranded oligonucleotides [30,32]. The higher hRPA binding efficiency ($[hRPA]_{50}$ around 5 nM) toward multimeric G4s compared to the single G4 ($[hRPA]_{50}$ around 20 nM) can be explained by the lower stability of these higher-order structures

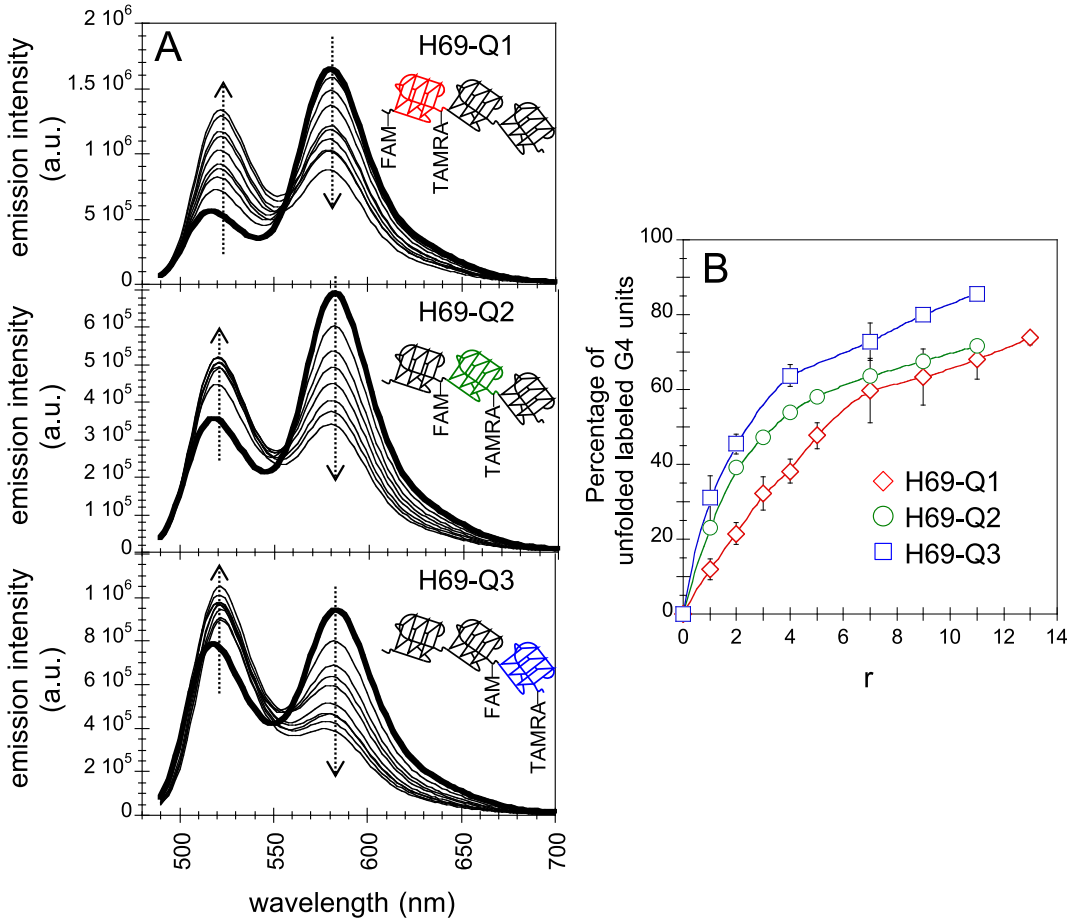


Fig. 1. Fluorescence titrations. (A) Fluorescence emission spectra of H69-Q1, H69-Q2 and H69-Q3 with increasing amounts of hRPA in K^+ solution (excitation wavelength 470 nm). Spectra of the doubly labeled oligonucleotides in the absence of hRPA are in bold line. Dotted arrows indicate the evolution of emission spectra with increasing hRPA concentrations. (B) Percentages of unfolded labeled G4 units in H69-Q1 (diamonds), H69-Q2 (circles) and H69-Q3 (squares) as a function of hRPA/oligonucleotide concentration ratio *r*. The percentage of unfolded labeled G4 units was calculated as detailed in Methods section. The oligonucleotide concentration was 100 nM and the protein concentration varied from 100 to 1100 or 1300 nM.

(temperature of thermal transition T_t 50 °C [26], Supplementary Information, Table S1) compared to the single G4 formed by H27 (T_t 55 °C [26], Supplementary Information, Table S1). Indeed, we previously reported an inverse relationship between the efficiency of binding of hRPA to a G4 sequence and the stability of the formed G4 [23]. In addition, the lower binding efficiency of hRPA to H27 may also be due to its length (27 nt), which is 3 nt shorter than the optimal length to obtain the more stable hRPA/DNA complex (30 nt-binding mode) [33]. The independence of hRPA binding efficiency of the number of contiguous G4 units may result from the poorly cooperative binding mode of hRPA. We asked if this feature is also due to the peculiar character of multimeric G4 structures formed by telomeric G-strands in potassium, where the contiguous G4 units are non-interacting and have similar stabilities, disregarding of their number [26]. To answer to this question, we studied the binding of hRPA to H21 ($(G_3T_2A)_3G_3$), H45 ($(G_3T_2A)_7G_3$), H69 ($(G_3T_2A)_{11}G_3$) and H93 ($(G_3T_2A)_{15}G_3$) sequences in sodium, folding into one, two, three and four G4 units, respectively [34], but lacking the T_2A tails at their extremities. We previously showed that, among contiguous G4 structures formed by telomeric DNA, the structure formed by H45 in sodium is unique, since its two G4 units interact with each other [34]. In sodium, H45 behaves as a single structural entity more stable than the beads-on-a-string structures formed by the longer H69 and H93 sequences [34] (Supplementary Information, Table S1). hRPA displayed a low binding efficiency for the sequence H21, compared to H69 and H93 (Fig. 2D–E); this is due to the higher stability of the single G4 formed by H21 (T_t 58 °C for H21, T_t 43/45 °C for H69/H93 [34], Supplementary Information, Table S1) and to its length (9 nt shorter than the optimal length to obtain the more stable hRPA/DNA complex), similarly to what observed in potassium with H27. Among contiguous G4s, the binding curves obtained by EMSA showed that, in sodium, hRPA had similar binding efficiencies for the two sequences H69 and H93, but a lower binding efficiency for the sequence H45 (Fig. 2E–D). The lower binding efficiency of hRPA to H45 in sodium might results not solely from its higher stability (T_t 53 °C [34], Supplementary Information, Table S1), but also from conformational factors related to the interaction between its two G4 units. Overall, EMSA on H51, H75 and H99 oligonucleotides in potassium demonstrate that the number of contiguous G4s does not affect the efficiency of binding of hRPA to telomeric G-strands. The result obtained with the H45 counterexample in sodium supports that this feature is related to the peculiar character of telomeric G4 units in potassium: non-interacting G4 units with similar stabilities, disregarding of their number.

We point out that, based on results obtained with the shortest oligonucleotide H27 (and H21), it would be wrong to argue that hRPA unfolds less efficiently a single G4 formed within a telomeric G-strand than contiguous G4s. Indeed, the single G4s formed by the 21mer sequence $(G_3T_2A)_3G_3$ (H21) or the extended 27mer sequence

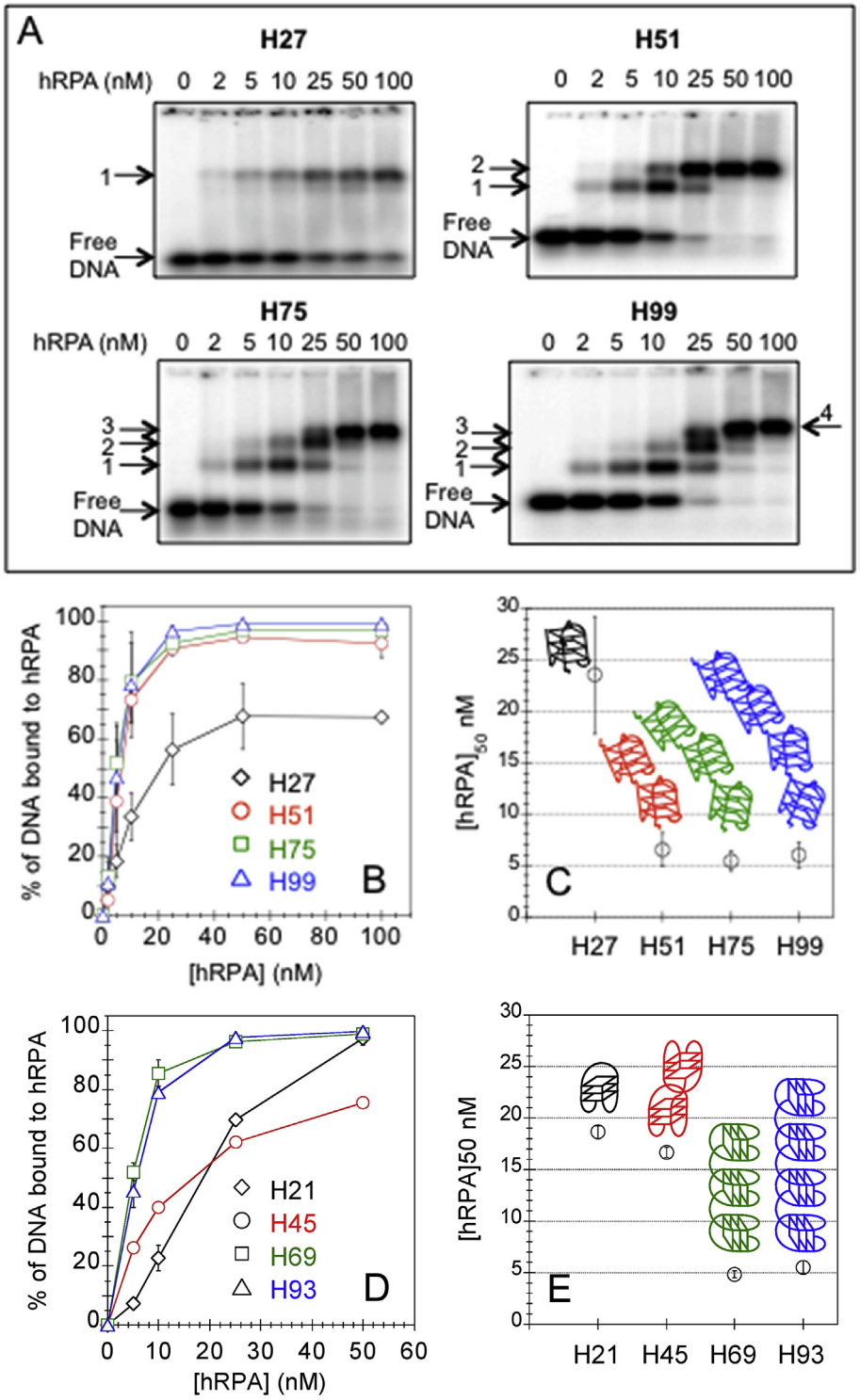


Fig. 2. Electrophoretic mobility shift assays. P^{32}-oligonucleotides (2 nM) were incubated with increasing amounts of hRPA in the presence of 100 mM KCl (A, B, C) or 100 mM NaCl (D, E). Electrophoresis was run on a native 1% agarose gel. (A) EMSA gels in KCl. (B) Percentages of oligonucleotide bound to hRPA in KCl: H27 (diamonds), H51 (circles), H75 (squares), H99 (triangles) as a function of hRPA concentration. Error bars correspond to standard deviations calculated from four independent experiments for points corresponding to 5, 10, 25 and 50 nM of hRPA, and from two independent experiments for points 2 and 100 nM hRPA. (C) Mean concentration of hRPA to obtain 50% of bound oligonucleotide ($[hRPA]_{50}$) from four independent experiments in KCl. (D) Percentages of oligonucleotide bound to hRPA in NaCl: H21 (diamonds), H45 (circles), H69 (squares), H93 (triangles) as a function of hRPA concentration. Error bars correspond to standard deviations calculated from two independent experiments. (E) Mean concentration of hRPA to obtain 50% of bound oligonucleotide ($[hRPA]_{50}$) from two independent experiments in NaCl. EMSA gels in NaCl are shown in Supplementary Information, Fig. S3.

TTA-$(G_3T_2A)_3G_3$-TTA (H27) are not suitable models for the binding of RPA to a single G4 unit within a telomeric G-strand, both for their length and for their higher stability [26].

4. Conclusion

In this study, we addressed for the first time, the question of how a single-stranded DNA binding protein involved in telomere replication, hRPA, deals with multimeric G4s that may form on the lagging strand template during telomere replication. We showed that hRPA binds telomeric G-strands structured into contiguous G4s, unfolding them. Importantly, we showed that the efficiency of binding of hRPA to telomeric G-strands is independent of the number of contiguous G4 units. In other words, if contiguous G4s form during replication, their number does not affect the efficiency of hRPA to coat transiently exposed single-stranded telomeric G-strands. This feature of hRPA may be crucial during telomere replication in suppressing or preventing multiple G4 accumulation on the lagging strand template and in ensuring telomere stability.

Author agreement

All the co-authors have contributed collectively to this study in order named as full participants. All the co-authors approved the final version of the manuscript and its submission to BIOCHIMIE. This work has not previously been published and is not being considered for publication elsewhere.

Acknowledgments

We thank the members of the laboratory team “Nucleic Acid Structure, Telomeres and Evolution”, in particular Anthony Bugaut, for helpful discussions. This work was supported by the “Muséum National d’Histoire Naturelle” (MNHN), the “Centre National de la Recherche Scientifique” (CNRS) and the “Institut National de la Santé et de la Recherche Médicale” (Inserm). LS was supported by a doctoral fellowship from the “Ligue Nationale contre le Cancer”.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.biochi.2017.11.017.

References

[1] L. Crabbe, R.E. Verdun, C.I. Haggblom, J. Karlseder, Defective telomere lagging strand synthesis in cells lacking WRN helicase activity, Science 306 (2004) 1951–1953.
[2] D. Gomez, R. Paterski, T. Lemarteleur, K. Shin-Ya, J.L. Mergny, J.F. Riou, Interaction of telomestatin with the telomeric single-strand overhang, J. Biol. Chem. 279 (2004) 41487–41494.
[3] A. Sfeir, S.T. Kosiyatrakul, D. Hockemeyer, S.L. MacRae, J. Karlseder, C.L. Schildkraut, T. de Lange, Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication, Cell 138 (2009) 90–103.
[4] J.B. Vannier, V. Pavicic-Kaltenbrunner, M.I. Petalcorin, H. Ding, S.J. Boulton, RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity, Cell 149 (2012) 795–806.
[5] P. Mohaghegh, J.K. Karow, R.M. Brosh Jr., V.A. Bohr, I.D. Hickson, The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases, Nucleic Acids Res. 29 (2001) 2843–2849.
[6] C. Ribeyre, J. Lopes, J.B. Boule, A. Piazza, A. Guedin, V.A. Zakian, J.L. Mergny, A. Nicolas, The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo, PLoS Genet. 5 (2009) e1000475.
[7] J.B. Vannier, G. Sarek, S.J. Boulton, RTEL1: functions of a disease-associated helicase, Trends Cell Biol. 24 (2014) 416–425.
[8] A.J. Zaug, E.R. Podell, T.R. Cech, Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 10864–10869.
[9] H. Hwang, N. Buncher, P.L. Opresko, S. Myong, POT1-TPP1 regulates telomeric overhang structural dynamics, Structure 20 (2012) 1872–1880.
[10] I. Schmutz, T. de Lange, Shelterin, Curr. Biol. CB 26 (2016) R397–R399.
[11] T.R. Salas, I. Petruseva, O. Lavrik, A. Bourdoncle, J.L. Mergny, A. Favre, C. Saintome, Human replication protein A unfolds telomeric G-quadruplexes, Nucleic acids Res. 34 (2006) 4857–4865.
[12] T. Kibe, Y. Ono, K. Sato, M. Ueno, Fission yeast Taz1 and RPA are synergistically required to prevent rapid telomere loss, Mol. Biol. Cell 18 (2007) 2378–2387.
[13] V. Schramke, P. Luciano, V. Brevet, S. Guillot, Y. Corda, M.P. Longhese, E. Gilson, V. Geli, RPA regulates telomerase action by providing Est1p access to chromosome ends, Nat. Genet. 36 (2004) 46–54.
[14] J. Smith, H. Zou, R. Rothstein, Characterization of genetic interactions with RFA1: the role of RPA in DNA replication and telomere maintenance, Biochimie 82 (2000) 71–78.
[15] M.S. Wold, Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism, Annu. Rev. Biochem. 66 (1997) 61–92.
[16] N.F. Lue, R. Zhou, L. Chico, N. Mao, O. Steinberg-Neifach, T. Ha, The telomere capping complex CST has an unusual stoichiometry, makes multipartite interaction with G-Tails, and unfolds higher-order G-tail structures, PLoS Genet. 9 (2013), e1003145.
[17] C. Rice, E. Skordalakes, Structure and function of the telomeric CST complex, Comput. Struct. Biotechnol. J. 14 (2016) 161–167.
[18] R.L. Flynn, R.C. Centore, R.J. O'Sullivan, R. Rai, A. Tse, Z. Songyang, S. Chang, J. Karlseder, L. Zou, TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA, Nature 471 (2011) 532–536.
[19] L.Y. Chen, S. Redon, J. Lingner, The human CST complex is a terminator of telomerase activity, Nature 488 (2012) 540–544.
[20] J. Dejardin, R.E. Kingston, Purification of proteins associated with specific genomic Loci, Cell 136 (2009) 175–186.
[21] J. Audry, L. Maestroni, E. Delagoutte, T. Gauthier, T.M. Nakamura, Y. Gachet, C. Saintome, V. Geli, S. Coulon, RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends, EMBO J. 34 (2015) 1942–1958.
[22] M.H. Qureshi, S. Ray, A.L. Sewell, S. Basu, H. Balci, Replication protein A unfolds G-quadruplex structures with varying degrees of efficiency, J. Phys. Chem. B 116 (2012) 5588–5594.
[23] L. Safa, E. Delagoutte, I. Petruseva, P. Alberti, O. Lavrik, J.F. Riou, C. Saintome, Binding polarity of RPA to telomeric sequences and influence of G-quadruplex stability, Biochimie 103 (2014) 80–88.
[24] L. Safa, N.M. Gueddouda, F. Thiebaut, E. Delagoutte, I. Petruseva, O. Lavrik, O. Mendoza, A. Bourdoncle, P. Alberti, J.F. Riou, C. Saintome, 5' to 3' unfolding directionality of DNA secondary structures by replication protein A: G-quadruplexes and duplexes, J. Biol. Chem. 291 (2016) 21246–21256.
[25] J. Abraham Punnoose, Y. Cui, D. Koirala, P.M. Yangyuoru, C. Ghimire, P. Shrestha, H. Mao, Interaction of G-quadruplexes in the full-length 3' human telomeric overhang, J. Am. Chem. Soc. 136 (2014) 18062–18069.
[26] A. Bugaut, P. Alberti, Understanding the stability of DNA G-quadruplex units in long human telomeric strands, Biochimie 113 (2015) 125–133.
[27] L. Petraccone, C. Spink, J.O. Trent, N.C. Garbett, C.S. Mekmaysy, C. Giancola, J.B. Chaires, Structure and stability of higher-order human telomeric quadruplexes, J. Am. Chem. Soc. 133 (2011) 20951–20961.
[28] Y. Xu, T. Ishizuka, K. Kurabayashi, M. Komiyama, Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends, Angew. Chem. 48 (2009) 7833–7836.
[29] H.Q. Yu, D. Miyoshi, N. Sugimoto, Characterization of structure and stability of long telomeric DNA G-quadruplexes, J. Am. Chem. Soc. 128 (2006) 15461–15468.
[30] C. Kim, B.F. Paulus, M.S. Wold, Interactions of human replication protein A with oligonucleotides, Biochemistry 33 (1994) 14197–14206.
[31] J.D. McGhee, P.H. von Hippel, Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice, J. Mol. Biol. 86 (1974) 469–489.
[32] C. Kim, M.S. Wold, Recombinant human replication protein A binds to polynucleotides with low cooperativity, Biochemistry 34 (1995) 2058–2064.
[33] I.M. Wyka, K. Dhar, S.K. Binz, M.S. Wold, Replication protein A interactions with DNA: differential binding of the core domains and analysis of the DNA interaction surface, Biochemistry 42 (2003) 12909–12918.
[34] C. Saintome, S. Amrane, J.L. Mergny, P. Alberti, The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium, Nucleic acids Res. 44 (2016) 2926–2935.

6.2 Binding properties of mono- and dimeric pyridine dicarboxamide ligands to human telomeric higher-order G-quadruplex structures

Please do not adjust margins

View Article Online
DOI: 10.1039/C7CC07048A

Published on 30 January 2018. Downloaded by Thompson Rivers University (formerly University College of the Cariboo) on 31/01/2018 01:50:37.

Journal Name

COMMUNICATION

Binding properties of mono- and dimeric pyridine dicarboxamide ligands to human telomeric higher-order G-quadruplex structures

C. Saintomé,[a,b]† P. Alberti,[a]† N. Guinot,[a] P. Lejault,[a] J. Chatain,[a] P. Mailliet,[a] J.-F. Riou[a] and A. Bugaut*[a]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Here, we report on the *in vitro* binding properties of the known pyridine dicarboxamide G-quadruplex ligand 360A and a new dimeric analogue (360A)$_{2A}$ to human telomeric DNA higher-order G-quadruplex (G4) structures. This study points to original binding features never reported for G4 ligands, and reveals a greater efficiency for the dimeric ligand to displace RPA (a ssDNA binding protein involved in telomere replication) from telomeric DNA.

In normal human somatic cells, telomeric DNA typically spans a 4-15 kb double-stranded region composed of 5'-TTAGGG-3'/3'-AATCCC-5' tandem repeats that ends with an about 150-300 nt single-stranded 3'-overhang.[1] The G-rich strand of telomeric DNA is prone to fold into G4 structures, which have been visualised in human cells using a structure-specific antibody.[2] Over the past two decades, a large number of small molecules have been synthesized and tested for their potential to interact with human telomeric G4 structures.[3] Some of these G4 ligands have been reported as telomere-targeting agents able to interfere with the binding of telomere-associated proteins.[4]

So far, most of the studies on the interactions between human telomeric DNA and G4 ligands have focused on short sequences able to fold into a single G4 unit. However, others and we have shown that longer telomeric DNA sequences can form higher-order structures consisting of several contiguous G4 units joined by TTA linkers.[5,6] A few recent studies have indicated that the binding properties of a ligand to multimeric G4 structures can significantly differ from its binding to a single G4 unit.[7]

Pyridine dicarboxamide derivatives are synthetic organic molecules that are amongst the most potent and selective G4 ligands. A lead molecule from this family is 360A (Fig. 1A), which we previously developed and was the first molecule of this family to be evaluated for its G4 binding properties.[8] This molecule was shown to display a good selectivity for G4 structures as compared with double-stranded DNA, and it was demonstrated that 360A preferentially localises at telomeres in human living cells.[9] The use of dimeric G4 ligands, composed of two G4 ligand moieties linked together by a chemical arm, has been proposed as a possible means to enforce the

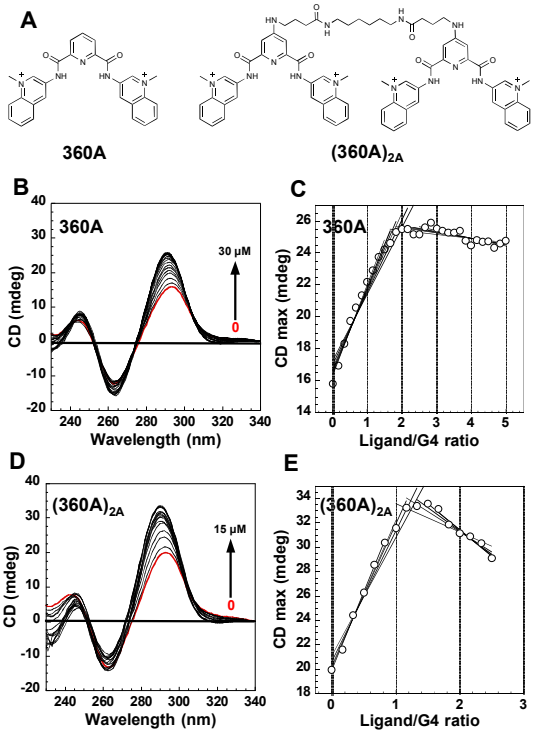


Fig. 1 (A) Chemical structures of the G4 ligands used in this study. Counter ions are I^-. (B and D) Changes in the CD spectrum and (C and E) changes in the maximum CD value of H21 (6 µM in 10 mM lithium cacodylate pH 7.2, 100 mM NaCl) upon addition of increasing amounts of 360A and (360A)$_{2A}$.

[a] "Structure and Instability of Genomes" laboratory, Sorbonne Universités, Muséum National d'Histoire Naturelle (MNHN), Inserm U1154, CNRS UMR 7196, 75005 Paris, France. E-mail: abugaut@mnhn.fr
[b] UPMC Paris 6 Université, 75005 Paris, France.
† These authors contributed equally to the work.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

ChemComm Accepted Manuscript

Please do not adjust margins

View Article Online
DOI: 10.1039/C7CC07048A

Published on 30 January 2018. Downloaded by Thompson Rivers University (formerly University College of the Cariboo) on 31/01/2018 01:50:37.

targeting of contiguous telomeric G4s.[10] In this work, we have synthesized a dimer of 360A (see Materials and methods in ESI†), named $(360A)_{2A}$ (Fig. 1A); and we have characterised the binding of 360A and its newly synthesised dimeric form $(360A)_{2A}$ to human telomere-mimicking oligonucleotides forming an increasing number of contiguous G4 units.

In order to assess the stabilisation and selectivity of $(360A)_{2A}$ towards human telomeric G4s, we performed competitive FRET-melting experiments in the presence of a large excess of unlabelled double-stranded DNA (ds26, Table S1, ESI†). We used either a fluorescent-labelled 21-nt oligonucleotide (F21T), which forms a single telomeric G4 structure, or three oligonucleotides (H69Q1, H69Q2 and H69Q3) that form fluorescent-labelled contiguous G4s (see Materials and methods in ESI†). The results demonstrated that the dimer ligand is a strong stabilizer of telomeric G4s and exhibits a great G4 *vs.* duplex selectivity (Fig. S1, ESI†).

Next, we used CD spectroscopy to evaluate the binding stoichiometry.[11] We first carried out CD titration experiments using H21 oligonucleotide that folds into a single G4 unit (Table S1, ESI†). We titrated a 6-µM solution of H21 in a sodium containing buffer with increasing amounts of 360A or its dimer analogue $(360A)_{2A}$. The changes in the CD spectrum of H21 upon increasing the concentration of 360A ligand up to 30 µM by increments of 1 µM are shown in Fig. 1B. In the absence of ligand, the CD spectrum of H21 displayed the characteristic signature of an antiparallel G-quadruplex structure, with a major positive band around 295 nm and a negative band at about 265 nm. Addition of 360A led to an increase in the 295 nm band up to about a 12 µM ligand concentration in solution, at which a maximum amplitude was reached. The spectral changes occurred immediately after each addition, and did not progress over time. This result indicated a 2:1 molar ratio of 360A to H21, presumably

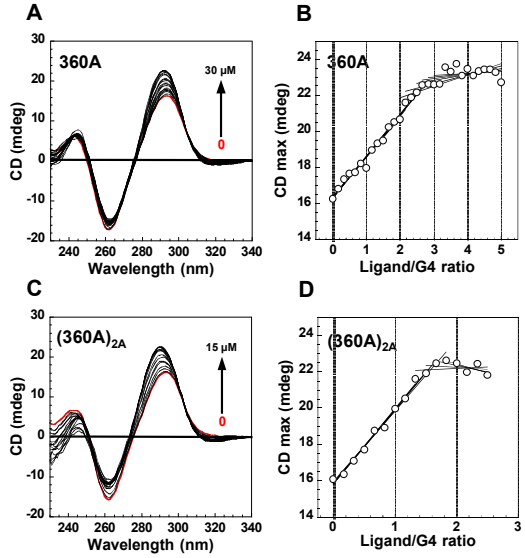


Fig. 2 (A and C) Changes in the CD spectrum and (B and D) changes in the maximum CD value of H69 (2 µM in 10 mM lithium cacodylate pH 7.2, 100 mM NaCl) upon addition of increasing amounts of 360A and $(360A)_{2A}$.

involving interactions with the two terminal G-quartets of the G4 structure. Stoichiometry values and associated errors were also determined by multiple linear regression analyses (Fig. 1C and Table S2, ESI†), with systematically varying the range of fitting area. Titration of a 6-µM solution of H21 with $(360A)_{2A}$ showed again an increase of the positive band around 295 nm (Fig. 1D); however, a maximum value was reached at around 7 µM ligand concentration, indicating a 1:1 binding stoichiometry of $(360A)_{2A}$ to H21 (Fig. 1E).‡ This suggests that the chemical arm between the two aromatic G4-binding moieties is long enough to afford them interacting with the top and the bottom G-quartet surfaces of the G4 structure in a "sandwich-like" manner. Such result is comparable to other studies from the Nagasawa's laboratory on dimers of macrocyclic hexaoxazole ligands.[12]

We then investigated the binding stoichiometry of 360A and $(360A)_{2A}$ to telomeric oligonucleotides able to form two and three contiguous intramolecular G4 structures, named H45 and H69, respectively (Table S1, ESI†).[6] In these experiments, DNA strand concentrations were adjusted in order to maintain an identical G4 unit concentration (6 µM) in solution. Consequently, in the absence of ligand, the CD spectra of H45 and H69 have similar shapes and amplitudes (Fig. 2 and Fig. S3, ESI†).[6] As for H21, we also observed an increase in the positive CD signal around 295 nm upon increasing the concentration of ligand in solution, until it reached a maximum value (Fig. 2 and Fig. S3, ESI†). However, from these experiments, we determined ligand to G4 unit molar ratios higher than the one obtained for H21 (see Table S2, ESI†). For 360A, the determined binding stoichiometry was around 2.5 with H45 (Fig. S3A and B, ESI†), and 2.7 with H69 (Fig. 2A and B); while for $(360A)_{2A}$, it was around 1.5 and 1.7 for H45 and H69, respectively (Fig. 2C and D; Fig. S3C and D, ESI†).

These binding stoichiometries that exceeded what could be conceived by taking into account the results obtained with H21 suggested that new binding sites became available when telomeric G4 units are arranged contiguously. Noteworthily, the obtained binding stoichiometries seemed to increase with the length of the oligonucleotide used, matching with the number of TTA linkers anchoring the G4 units together. If a TTA linker gives rise to one extra binding site for the ligands, the calculated 360A:G4 unit molar ratios would be of 2.50 for H45 and of 2.67 for H69 (Table S2, ESI†). These calculated values are very close to our experimental values (Fig. 3 and Table S2,ESI†). This has also been verified for H93, an oligonucleotide capable of forming four contiguous G4 units (Fig. 3; Table S1 and S2, ESI†). For the dimeric ligand $(360A)_{2A}$, the experimental ligand:G4 ratio values also followed the trend of the calculated values (Fig. 3 and Table S2, ESI†). These results indicate that, for both 360A and $(360A)_{2A}$, a binding site exists between two adjacent G4 units in a multimeric telomeric G4 structure.

To further investigate this possibility, we next wondered whether the observed stoichiometry values depended on the length of the TTA nucleotidic linker that joins the G4 units. We performed CD titrations with oligonucleotides bearing longer

ChemComm Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins
View Article Online
DOI: 10.1039/C7CC07048A

Published on 30 January 2018. Downloaded by Thompson Rivers University (formerly University College of the Cariboo) on 31/01/2018 01:50:37.

TTA linkers between the G4 units ($H45(TTA)_3$ and $H69(TTA)_3$, Table S1, ESI†). From these experiments, we determined a binding stoichiometry of 2:1 for 360A and 1:1 for its dimeric analogue $(360A)_{2A}$ for both $H45(TTA)_3$ and $H69(TTA)_3$

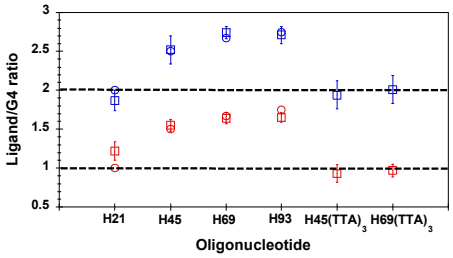


Fig. 3 Calculated (open circle) versus experimental (open square) ligand to G4 ratio values (Table S2, ESI†) for the telomeric oligonucleotides H21, H45, H69 and H93 (see Fig. S4 in ESI†) and the extended oligonucleotides $H45(TTA)_3$ and $H69(TTA)_3$. Data for 360 and $(360A)_{2A}$ are in blue and red, respectively.

oligonucleotides (Fig. 3 and Fig. S5, ESI†). These values are similar to the ones obtained for the single G4-forming sequence H21, indicating that the extra interstitial binding sites within the multimeric G4 structures are lost when the G4 units are moved away by additional TTA sequences.

Overall, our results point to a binding model whereby two aromatic G4-binding moieties interact with the external G-quartets of a G4 unit and a third one is accommodated in a binding pocket involving the TTA linker between two contiguous G4s. On the basis of molecular modelling simulations, Haider *et al.* proposed the existence of possible binding sites at the interface between two adjacent telomeric G4 units.[13] Subsequently, few G4 ligands have been studied experimentally for their binding to multimeric G4 structures.[7] Among them, a chiral cyclic helicene has been proposed to bind in the cleft pocket located between two adjacent human telomeric G4s.[7b] Cummaro *et al.* reported a spectroscopic study on the binding of a triazatruxene molecule to a telomeric higher-order structure formed by two G4 units, which revealed that two molecules could bind by end-stacking at the 5' and 3' extremities of the structure and a third molecule at the interface between the G4s.[7c] Bai et *al.* also proposed a similar binding mode for sanguinarine to a two-tandem G4-forming sequence.[7a] Recently, a triaryl-substituted imidazole derivative was also found to selectively bind to telomeric G4 multimers likely through intercalating into the pocket formed between two G4 units.[7k] By contrast no intersitial binding sites between adjacent G4s were observed for a macrocyclic hexaoxazole ligand.[7g] Regarding the binding of ligand dimers to higher-order G4 structures, there has been only very few studies.[10] All these studies were carried on telomeric sequences forming two contiguous G4 structures, and no interstitial binding site was observed. Thus, when compared to previous works, our results indicate unprecedented modes of binding to multimeric G4 for 360A and its dimeric analogue $(360A)_{2A}$, which combine end-stacking with both external G-quartets of the G4 units and additional binding sites between the G4s. Detailed information on how these extra binding sites occur, and whether they pre-exist or are generated upon binding of the ligands to the G-quartets will require in-depth structural studies.

We next sought to assess the potential of our G4 ligands to interfere with the binding of human replication protein A (RPA) to telomeric DNA. RPA is a single-stranded DNA (ssDNA) binding protein that is essential to whole genome replication, including telomeres.[14] We previously showed that the efficiency of RPA binding to G4-forming telomeric sequences displays an inverse relationship with the stability of the formed structure, and that G4 ligands can impair RPA binding to single telomeric G4-forming sequences.[15] We carried out gel shift experiments with RPA and the telomere-mimicking oligonucleotides used in this study. For these experiments, RPA concentration in the reaction mixture was set to obtain about 95-100% DNA bound to RPA for all the sequences used. First, we incubated the G4 ligand and the radiolabelled DNA substrate for 5 min, and then we added RPA to the reaction mixture and incubated for further 20 min. We observed a concentration-dependent inhibitory effect of the ligands for the association between RPA and the DNA substrates (Fig. 4A and Fig. S6A, ESI†). The ability of the ligands to inhibit RPA binding was maintained even in the presence of a large excess of the double-stranded competitor ds26 (Fig. S7, ESI†). The level of RPA/DNA complexes formed was inhibited by about 75%, 85% and 65% for H21, H45 and H69, respectively, in the presence of 2.5 µM of 360A. Raising the concentration of ligand to 5 µM resulted only in a slight increase in RPA binding inhibition by about 5% for all binding substrates. The dimer $(360A)_{2A}$ exhibited a greater inhibitory effect on RPA binding to the telomeric sequences, reaching about 95% inhibition for all binding substrates at 2.5 µM ligand concentration in solution (Fig. 4A). By contrast, we did not observe such inhibition of

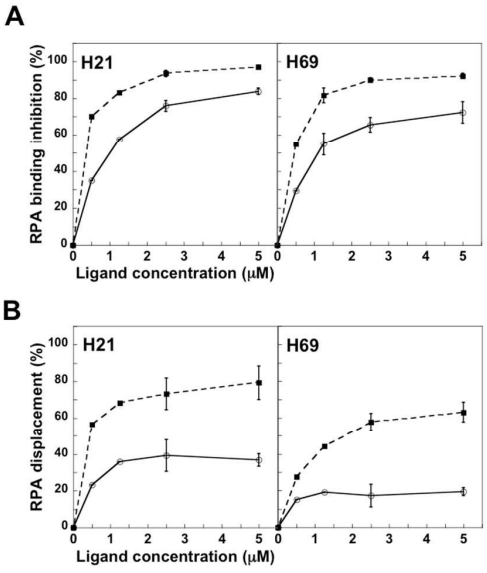


Fig. 4 (A) Percentage of inhibition of RPA binding to telomeric oligonucleotide substrates H21 and H69 in the presence of increasing concentrations of 360A (open circles) or $(360A)_{2A}$ (black squares) (B) Percentage of displacement of RPA from the telomeric oligonucleotide substrates H21 and H69 (20 min incubation) upon addition of increasing amounts of 360A (open circles) or $(360A)_{2A}$ (black squares).

ChemComm Accepted Manuscript

Please do not adjust margins

View Article Online
DOI: 10.1039/C7CC07048A

Published on 30 January 2018. Downloaded by Thompson Rivers University (formerly University College of the Cariboo) on 31/01/2018 01:50:37.

RPA binding to control oligonucleotides that do not fold into G4s (Table S1 and Fig. S8, ESI†). These results suggested that more stable small molecule/DNA complexes are formed between the dimeric ligand and the higher-order G4 structures than with the monomeric ligand. This was further corroborated by UV-melting experiments that showed higher thermal stabilities of the complexes formed with $(360A)_{2A}$ as compared with 360A for an identical aromatic core to G4 unit ratio (Table S3, ESI†).

In the nucleus of a cell, RPA may already be associated with telomeric DNA. We then asked whether our ligands were able to displace RPA from telomeric DNA. We hence performed experiments where the DNA binding substrates were pre-coated with RPA (20 min incubation) before the addition of the G4 ligand in the mixture for 5 min incubation. We observed that, under these experimental conditions, both G4 ligands were able to displace RPA from the DNA substrates in a concentration-dependent manner (Fig. 4B). We also performed control experiments, in the presence of a large excess of double-stranded competitor or with mutated telomeric oligonucleotide substrates, demonstrating the G4 selectivity of the ligands (Fig. S7 and Fig. S8, ESI†). Importantly, the dimer $(360A)_{2A}$ displayed a greater efficiency in displacing RPA from telomeric DNA. When considering the longer sequence H69, 5 µM of $(360A)_{2A}$ dissociated about 60% of the RPA/DNA complex, while at the same concentration the 360A molecule dissociated only 20% of the complex. It is noteworthy than even for the same concentration in G4-binding aromatic moiety (*i.e.* 2.5 µM in $(360A)_{2A}$ versus 5 µM in 360A), the dimer molecule was more efficient in displacing RPA than the monomer (Fig. 4B).

In conclusion, we undertook here to evaluate the binding properties of a known G4 ligand and its new dimeric analogue towards human telomeric higher-order G4 structures. Our results propose original binding modes for these ligands, which have never been reported before with other ligands. Furthermore, we showed that the dimeric ligand is more efficient than the monomer for impairing the formation and for inducing the dissociation of the complex formed RPA and human telomeric DNA, thus highlighting the potential of dimeric G4 ligands to efficiently interfere with protein binding at telomeres in order to interrogate telomere biology.

This work was supported by Inserm, CNRS and the MNHN. P. L. was supported by a doctoral fellowship from the MNHN.

Notes and references

‡ *The choice of a sodium buffer was privileged because of the less polymorphic nature of the telomeric G4 structure in this condition, similar results were obtained in a potassium-containing buffer (Figure S2, ESI†).*

1 (*a*) R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff and J. R. Wu, *Proc. Natl Acad. Sci. USA*, 1988, **85**, 6622-6626; (*b*) R. McElligott and R. J. Wellinger, *EMBO J.*, 1997, **16**, 3705-3714.

2 G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182-186.

3 (*a*) D. Monchaud and M. P. Teulade-Fichou, *Org. Biomol. Chem.*, 2008, **6**, 627-636; (*b*) S. Neidle, *FEBS J.*, 2010, **277**, 1118-1125.

4 (*a*) D. Gomez, M. F. O'Donohue, T. Wenner, C. Douarre, J. Macadre, P. Koebel, M. J. Giraud-Panis, H. Kaplan, A. Kolkes, K. Shin-ya and J. F. Riou, *Cancer Res.*, 2006, **66**, 6908-6912; (*b*) R. Rodriguez, S. Muller, J. A. Yeoman, C. Trentesaux, J. F. Riou and S. Balasubramanian, *J. Am. Chem. Soc.*, 2008, **130**, 15758-15759; (*c*) H. Tahara, K. Shin-Ya, H. Seimiya, H. Yamada, T. Tsuruo and T. Ide, *Oncogene*, 2006, **25**, 1955-1966.

5 (*a*) H. Q. Yu, D. Miyoshi and N. Sugimoto, *J. Am. Chem. Soc.*, 2006, **128**, 15461-15468; (*b*) Y. Xu, T. Ishizuka, K. Kurabayashi and M. Komiyama, *Angew. Chem. Int. Ed.*, 2009, **48**, 7833-7836; (*c*) L. Petraccone, C. Spink, J. O. Trent, N. C. Garbett, C. S. Mekmaysy, C. Giancola and J. B. Chaires, *J. Am. Chem. Soc.*, 2011, **133**, 20951-20961; (*d*) V. Singh, M. Azarkh, M. Drescher and J. S. Hartig, *Chem. Commun.*, 2012, **48**, 8258-8260; (*e*) R. Hansel, F. Lohr, L. Trantirek and V. Dotsch, *J. Am. Chem. Soc.*, 2013, **135**, 2816-2824.

6 A. Bugaut and P. Alberti, *Biochimie*, 2015, **113**, 125-133.

7 (*a*) L. P. Bai, M. Hagihara, Z. H. Jiang and K. Nakatani, *ChemBioChem*, 2008, **9**, 2583-2587; (*b*) K. Shinohara, Y. Sannohe, S. Kaieda, K. Tanaka, H. Osuga, H. Tahara, Y. Xu, T. Kawase, T. Bando and H. Sugiyama, *J. Am. Chem. Soc.*, 2010, **132**, 3778-3782; (*c*) A. Cummaro, I. Fotticchia, M. Franceschin, C. Giancola and L. Petraccone, *Biochimie*, 2011, **93**, 1392-1400; (*e*) C. Q. Zhao, L. Wu, J. S. Ren, Y. Xu and X. G. Qu, *J. Am. Chem. Soc.*, 2013, **135**, 18786-18789; (*f*) L. N. Zhu, B. Wu and D. M. Kong, *Nucleic Acids Res.*, 2013, **41**, 4324-4335; (*g*) K. Iida, G. Tsubouchi, T. Nakamura, S. Majima, H. Seimiya and K. Nagasawa, *MedChemComm*, 2013, **4**, 260-264; (*h*) A. R. O. Cousins, D. Ritson, P. Sharma, M. F. G. Stevens, J. E. Moses and M. S. Searle, *Chem. Commun.*, 2014, **50**, 15202-15205; (*i*) X. X. Huang, L. N. Zhu, B. Wu, Y. F. Huo, N. N. Duan and D. M. Kong, *Nucleic Acids Res.*, 2014, **42**, 8719-8731; (*j*) Q. Zhang, Y. C. Liu, D. M. Kong and D. S. Guo, *Chem. Eur. J.*, 2015, **21**, 13253-13260; (*k*) M. H. Hu, S. B. Chen, B. Wang, T. M. Ou, L. Q. Gu, J. H. Tan and Z. S. Huang, *Nucleic Acids Res.*, 2017, **45**, 1606-1618.

8 G. Pennarun, C. Granotier, L. R. Gauthier, D. Gomez, F. Hoffschir, E. Mandine, J. F. Riou, J. L. Mergny, P. Mailliet and F. D. Boussin, *Oncogene*, 2005, **24**, 2917-2928.

9 C. Granotier, G. Pennarun, L. Riou, F. Hoffschir, L. R. Gauthier, A. De Cian, D. Gomez, E. Mandine, J. F. Riou, J. L. Mergny, P. Mailliet, B. Dutrillaux and F. D. Boussin, *Nucleic Acids Res.*, 2005, **33**, 4182-4190.

10 (*a*) A. Paul, A. K. Jain, S. K. Misra, B. Maji, K. Muniyappa and S. Bhattacharya, *Plos One*, 2012, **7,** e39467; (*b*) K. Iida, S. Majima, T. Nakamura, H. Seimiya and K. Nagasawa, *Molecules*, 2013, **18**, 4328-4341; (*c*) C. Q. Zhou, J. W. Yang, C. Dong, Y. M. Wang, B. Sun, J. X. Chen, Y. S. Xu and W. H. Chen, *Org. Biomol. Chem.*, 2016, **14**, 191-197; (*d*) C. Q. Zhou, T. C. Liao, Z. Q. Li, J. Gonzalez-Garcia, M. Reynolds, M. Zou and R. Vilar, *Chem. Eur. J.*, 2017, **23**, 4713-4722; (*e*) Z. Q. Li, T. C. Liao, C. Dong, J. W. Yang, X. J. Chen, L. Liu, Y. Luo, Y. Y. Liang, W. H. Chen and C. Q. Zhou, Org. Biomol. Chem., 2017, **15**, 10221-10229.

11 P. Murat, Y. Singh and E. Defrancq, *Chem. Soc. Rev.*, 2011, **40**, 5293-5307.

12 K. Iida, M. Tera, T. Hirokawa, K. Shin-ya and K. Nagasawa, *Chem. Commun.*, 2009, 6481-6483.

13 S. M. Haider and S. Neidle, *Biochem. Soc. Trans.*, 2009, **37**, 583-588.

14 J. Audry, L. Maestroni, E. Delagoutte, T. Gauthier, T. M. Nakamura, Y. Gachet, C. Saintome, V. Geli and S. Coulon, *EMBO J.*, 2015, **34**, 1942-1958.

15 (*a*) L. Safa, E. Delagoutte, I. Petruseva, P. Alberti, O. Lavrik, J. F. Riou and C. Saintome, *Biochimie*, 2014, **103**, 80-88; (*b*) L. Safa, N. M. Gueddouda, F. Thiebaut, E. Delagoutte, I. Petruseva, O. Lavrik, O. Mendoza, A. Bourdoncle, P. Alberti, J. F. Riou and C. Saintome, *J. Biol. Chem.*, 2016, **291**, 21246-21256.

ChemComm Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx

6.3 A multi-approach analysis highlights the relevance of RPA-1 as a telomere end-binding protein (TEBP) in *Leishmania amazonensis*

BBA - General Subjects 1864 (2020) 129607

Contents lists available at ScienceDirect

BBA - General Subjects

journal homepage: www.elsevier.com/locate/bbagen

A multi-approach analysis highlights the relevance of RPA-1 as a telomere end-binding protein (TEBP) in *Leishmania amazonensis*

Carlos A.H. Fernandes[a,b,1], Edna Gicela O. Morea[c,1], Gabriel A. dos Santos[c], Vitor L. da Silva[c], Marina Roveri Vieira[c], Maria Alejandra Viviescas[c], Jean Chatain[d], Aurélie Vadel[d], Carole Saintomé[d,e], Marcos Roberto M. Fontes[a], Maria Isabel Nogueira Cano[c,*]

[a] Department of Biophysics and Pharmacology, Biosciences Institute, São Paulo State University (UNESP) – Botucatu, SP, Brazil
[b] Laboratoire de Biologie et Pharmacologie Appliquée, École Normale Supérieure Paris-Saclay, Cachan, France
[c] Department of Chemical and Biological Sciences, São Paulo State University (UNESP) – Botucatu, SP, Brazil
[d] MNHN CNRS UMR 7196, INSERM U1154, 43 rue Cuvier, 75005 Paris, France
[e] Sorbonne Université, UFR927, 4 place Jussieu, 75005 Paris, France

ARTICLE INFO

Keywords:
Leishmania sp.
OB-fold
Telomere end-binding protein
RPA-1
G-quadruplex

ABSTRACT

Background: Telomeres are chromosome end structures important in the maintenance of genome homeostasis. They are replenished by the action of telomerase and associated proteins, such as the OB (oligonucleotide/oligosaccharide-binding)-fold containing telomere-end binding proteins (TEBP) which plays an essential role in telomere maintenance and protection. The nature of TEBPs is well known in higher and some primitive eukaryotes, but it remains undetermined in trypanosomatids. Previous *in silico* searches have shown that there are no homologs of the classical TEPBs in trypanosomatids, including *Leishmania* sp. However, Replication Protein A subunit 1 (RPA-1), an OB-fold containing DNA-binding protein, was found co-localized with trypanosomatids telomeres and showed a high preference for the telomeric G-rich strand.

Methods and results: We predicted the absence of structural homologs of OB-fold containing TEBPs in the *Leishmania* sp. genome using structural comparisons. We demonstrated by molecular docking that the ssDNA binding mode of LaRPA-1 shares features with the higher eukaryotes POT1 and RPA-1 crystal structures ssDNA binding mode. Using fluorescence spectroscopy, protein-DNA interaction assays, and FRET, we respectively show that LaRPA-1 shares some telomeric functions with the classical TEBPs since it can bind at least one telomeric repeat, protect the telomeric G-rich DNA from 3′-5′ Exonuclease I digestion, and unfold telomeric G-quadruplex.

Conclusions: Our results suggest that RPA-1 emerges as a TEBP in trypanosomatids, and in this context, we present two possible evolutionary landscapes of trypanosomatids RPA-1 that could reflect upon the evolution of OB-fold containing TEBPs from all eukaryotes.

General significance

Leishmania RPA-1 is a strong TEBP candidate.

1. Introduction

Telomeres are protective DNA-protein complexes found at the end of eukaryotic chromosomes and essential for maintaining genome integrity and stability, as well as cell proliferation. They are composed of G-rich repetitive DNA sequences in double and single-strand forms, with the G-rich ssDNA (single-stranded DNA) forming a protrusion towards the end of the chromosomes, known as the 3′-G-overhang. The length of telomeres is maintained by telomerase, a ribonucleoprotein reverse transcriptase that adds telomeric repeats to the ends of chromosomes using an RNA sequence as the template [1].

Proteins that specifically associate with telomeres are regulators of

Abbreviations: ATR, Ataxia-Telangiectasia Mutated; DDR, DNA damage repair; OB-fold, oligonucleotide/oligosaccharide-binding fold; EMSA, Electrophoretic Mobility Shift Assay; POT1, Protection Of Telomeres 1; RAP1, Repressor-Activator Protein 1; RPA, Replication Protein A; RPA-1, Replication Protein A subunit 1; ssDNA, single-stranded DNA; TEBP, telomere end-binding protein; TIN2, TRF1-interacting nuclear factor 2; TRF1 and TRF2, Telomere Repeat binding Factor 1 and 2

* Corresponding author at: Department of Genetics, Biosciences Institute, São Paulo State University (UNESP), Botucatu, SP, Brazil.
E-mail address: maria.in.cano@unesp.br (M.I.N. Cano).

[1] These authors contributed equally to the paper

https://doi.org/10.1016/j.bbagen.2020.129607
Received 4 December 2019; Received in revised form 21 February 2020; Accepted 24 March 2020
Available online 25 March 2020
0304-4165/ © 2020 Elsevier B.V. All rights reserved.

telomere length maintenance and among them are the OB (oligonucleotide/oligosaccharide-binding)-fold containing telomere end-binding proteins (TEBP), which preferentially associate with the 3′ G-overhang [2]. TEBPs are crucial for chromosome end protection; they act as telomere shortening sensors and help cells to distinguish telomeres from DNA double-stranded breaks thereby avoiding a local DNA damage response (DDR) [2–4]. They are also involved with the recruitment of telomerase and the regulation of telomere length maintenance [2,5].

In most vertebrates, fungi, and plants, telomeric DNA is found associated with the shelterin and CST complexes, both being essential to proper telomere length regulation and maintenance, and also critical for genome stability [5,6]. Shelterin is mainly composed of the well-known proteins TRF1 and TRF2 (Telomere Repeat binding Factors 1 and 2), RAP1 (Repressor-Activator Protein 1), TPP1, TIN2 (TRF1-interacting nuclear factor 2) and POT1 (Protection Of Telomeres 1) [3,6–8], although depending on the organism, it can also have a slightly different composition and form diverse configurations [3,9]. In all cases, some of the shelterin components can both associate with each other and with the telomeric DNA, showing very dynamic and cell cycle-dependent interactions [1,3]. In mammals, for example, TRF1 and TRF2 interact directly with the double-stranded telomeric repeats, whereas RAP1 does not associate with DNA but with TRF2 and TIN2 which bridges TRF1 and TRF2 to POT1 *via* TPP1. POT1 is a TEBP that has high affinity and specificity for the telomeric G-rich ssDNA and preferentially associates with the 3′ G-overhang [2] providing telomere capping that helps cells to distinguish telomeres from DNA double-stranded breaks and to avoid a local ATR (Ataxia-Telangiectasia and Rad3 related)-dependent DDR [4,10–12]. When telomeres are shortened, POT1-TPP1 regulates telomere replication by giving telomerase access to telomeres [2,5] and by interacting with proteins of the CST complex responsible for the C-strand fill-in synthesis [3,6–8,10].

The CST complex was first identified in budding yeast, where it is composed of the heterotrimer CDC13, STN1, and TEN1, three OB-fold containing proteins. Budding yeasts lack the shelterin components, and CDC13 is the CST component that specifically associates with the telomeric G-rich ssDNA, sharing with POT1 a similar binding mode to telomeric DNA. CDC13 is thus, considered a TEBP that plays a dual role in telomere length regulation by protecting telomeres from exonuclease attack [13,14], and by controlling the access of telomerase to the ends of the chromosomes [15,16]. In fungi, plants, and vertebrates, CTC1 replaces CDC13 in the CST complex, which also contains homologs of the budding yeast proteins STN1 and TEN1 [5]. However, because CDC13 and CTC1 differ greatly in ssDNA sequence affinity, specificity, and binding mode, they are not considered as orthologues [15,17]. In most of these organisms, CST plays conserved roles and is also involved with the telomeric C-strand fill-in synthesis during telomere replication [5,18]. In humans, CST is also responsible for telomere protection independently of the POT1 pathway [17] and can bind to ssDNA. However, CST has also an extra-telomeric function in mammals, playing important roles in general stalled fork recovery during replicative stress [19].

Telomeric proteins with similar roles to the TEBPs, as mentioned earlier, were also described in some primitive eukaryotes, such as *Tetrahymena thermophila* [20,21] and *Oxytricha nova* [22]. The proteins described in these organisms (*i.e.* TEB1-3, p75, p45 and p19 in *Tetrahymena thermophila* [20,21] and TEBP α/β in *Oxytricha nova* [22]) also share with known TEBPs (*i.e.* POT1, TPP1, and CST components) structural similarities, such as the OB-fold domains, and a high affinity for the telomeric G-rich ssDNA [23–25]. These data indicate that TEBPs and telomeric proteins may vary in composition between eukaryotic lineages, especially at lower eukaryotes; however, the occurrence of OB-fold domains in all of these proteins indicates that they share conserved structural and functional features [26].

OB-fold domains are also present in proteins of the heterotrimeric complex RPA (Replication Protein A) [27,28], which is also involved with telomere length maintenance in budding and fission yeast [29–31]. In the last years, several studies have been pointing out that the CST and RPA complexes share a common origin due to close structural and functional similarities [17,18,25,32]. The CST complex may likely have given birth to the POT1-TPP1 complex and are exclusively found in eukaryotes [32] in contrast to RPA, which has archaeal homologs [33].

Some homologs of the shelterin proteins, such as TRF2, RAP1, and TIN2, were described in protozoa belonging to the Trypanosomatidae Family, including the *Leishmania* and *Trypanosoma* genera [34–37]. In contrast, no homologs for POT1, TPP1, and the components of the CST complex (CDC13/CTC1, STN1, and TEN1) were found after a thorough search in the trypanosomatids genomes [38]. Interestingly, RPA-1 alone (without RPA-2 and RPA-3) was one of the proteins captured from *L. amazonensis* telomerase positive nuclear extracts in an affinity chromatography purification using the telomeric G-rich ssDNA as bait [39]. *L. amazonensis* RPA-1 (LaRPA-1), as well as all RPA-1 from trypanosomatids, are smaller than other eukaryotes RPA-1 (approximately 52 kDa) because they lack the N-terminal RPA70N domain. In *L. amazonensis*, for example, the absence of the RPA70N does not prevent LaRPA-1 participation in different DDRs at parasite telomeres [40,41]. Besides, LaRPA-1 co-localizes with telomeric DNA and has a preference to bind *in vitro* and *in vivo* the telomeric G-rich ssDNA, although it can bind with less affinity other GT-rich and C-rich ssDNA sequences [38,42]. Unlike other RPA-1 that interacts with ssDNA using three OB-fold domains [43,44], LaRPA-1, due to differences in its tertiary structure in comparison to RPA-1 from higher eukaryotes, can bind the telomeric G-rich ssDNA using only its N-terminal OB-fold domain [38,42]. RPA-1 from *T. cruzi* also presents a higher affinity for the telomeric G-rich ssDNA and can also localize at telomeres, even outside the replication phase [45]. These data are consistent with a telomeric role for trypanosomatids RPA-1 [38,42,45]. Using structural comparisons, we predicted the absence of classical TEBPs in *Leishmania* sp. telomeres and provided pieces of experimental evidence that shows LaRPA-1 shares important telomeric functions with known TEBPs. For example, LaRPA-1 can protect the telomeric G-rich ssDNA from 3′-5′ exonucleolytic degradation, and similar to CST and POT1-TTP1, it unfolds telomeric G-quadruplex, suggesting that in the possible absence of other OB-fold containing telomeric proteins, RPA-1 emerges as a trypanosomatid TEBP.

2. Materials and methods

2.1. Expression and purification of recombinant RPA-1 from L. amazonensis (LaRPA-1)

Recombinant LaRPA-1 was obtained, as previously described [46]. Briefly, the LaRPA-1 gene cloned in a pET28a + vector was expressed in *Escherichia coli* BL21 (DE3) codon plus with a 6xHis-tag at its N-terminal. The recombinant protein was obtained in the insoluble form and was further solubilized in 7 M urea and purified in denaturing conditions using two chromatographic steps (Q Sepharose FF and His-Trap FF – GE Healthcare). The protein was refolded by dialysis in the presence of 100 $\mu g \cdot mL^{-1}$ of sodium heparin (Sigma) to aid the correct folding of the protein. The purity of LaRPA-1 was checked in Coomassie blue-stained 12% SDS-PAGE (Supplementary Fig. 1).

2.2. Binding reactions and EMSA (electrophoretic mobility shift assay)

Binding assays were done incubating 1.5 μg (28.6 pmol) of purified LaRPA-1 with 1.55 pmol of TELG-DIG DIG (5′-DIG $(TTAGGG)_5$–3′) or with 1.55 pmol of TELC-DIG (5'-DIG $(AATCCC)_5$ – 3′) in 1× binding buffer (20 mM HEPES pH 6, 1 mM EDTA, 10 mM $(NH_4)SO_4$, 1 mM DTT, 0,2% Tween 20, 30 mM KCl) in the presence of 0.1 μg of poly-L-lysine and 1 μg of poly[d(I-C)] for 30 min at room temperature. Protein/DNA complexes were fractionated on a 10% or a 6% non-denaturing PAGE

(37,5 acrylamide: 1 bis-acrylamide) or on a 1% agarose gel in 0.25× TBE (Tris-borate-EDTA) buffer (25 mM Tris, 25 mM Boric acid, and 0.5 mM EDTA pH 8) at 110 V for 1.5 h at room temperature. Gels were transferred by capillarity to the Hybond-N+ membrane (Amersham) using a 0.25× TBE buffer. The membranes were blocked for 30 min, incubated for 30 min with anti-DIG AP antibody (Roche), and revealed with CSPD (Roche). Images were captured using G:BOX Chemi XR5 and Gene Tools software from Syngene.

2.3. Exonuclease I protection assays

We performed three different Exo I protection experiments using *E. coli* 3′-5′ exonuclease I. It is well known that one unit of *E. coli* Exo I hydrolyzes 10 nmol of a single-stranded nucleic acid in 30 min at 37 °C, meaning that 10 U of bacterial Exo I can degrade 90% of 1 μg of a 100mer single-stranded oligonucleotide in 20 min at 37 °C [47]. Thus, extrapolating to our system, 1.8-3 U of bacterial Exo I would be enough to respectively degrade about 90% of 1 μg (172 pmol) of the 18 mer TEL3 and 1 μg (102.72 pmol) of the 30 mer TELG-DIG, in 20 min at 37 °C. In the first assay, binding reactions containing 1.5 μg (28.85 μM or 28.85 pmol) of purified LaRPA-1 and 1.55 pmol of TELG-DIG in 1× binding buffer, were prior incubated for 30 min at room temp and subsequently treated with various concentrations (1 U, 2 U, 3 U, 6 U, 12 U, 25 U and 50 U) of 3′-5′ *E. coli* Exonuclease I (Exo I) (New England Biolabs) for 30 min at 37 °C in 1× Exonuclease I Reaction Buffer. The second assay was done using the same binding conditions used in the first assay, and 30 min after incubation, protein/DNA complexes formed with 1.55 pmol of TELG-DIG or 1.55 pmol of TELC-DIG were individually treated at room temperature with 2 U of Exo I at different time points (0, 5, 10, 15, 30, 45 min). In control reactions, binding assays were done in the absence of protein, which was replaced by an equivalent amount of LaRPA-1 storage buffer (Tris-HCl 20 mM, NaCl 20 mM pH 8.0) or in the presence of heat-denatured (5 min at 95 °C) LaRPA-1. In the third experiment, binding reactions were done with different amount of LaRPA-1 (13.46 pmol, and 40.39 pmol) and 1.55 pmol TELG-DIG in 1× binding buffer followed by incubation with 2 U Exo I for 30 min at 37 °C in 1× Exo I Reaction Buffer and subsequently heat denatured for 5 min at 95 °C. For the control experiments, binding assays were done in the absence of protein, and TELG-DIG alone was treated with 2 U Exo I. In all, three different assays, binding complexes and Exo I-treated samples were further fractionated on a non-denaturing 6% PAGE (37.5 acrylamide: 1 bis acrylamide) or 1% agarose gel in 0.25× TBE.

2.4. Fluorescence spectroscopy measurements

Static fluorescence spectroscopy measurements were recorded using a Hitachi F-7000 spectrofluorimeter with a 1 cm path length cuvette and bandwidths of excitation and emission of 5 and 10 nm, respectively. The fluorescence emission spectra were measured from 300 to 450 nm at 1200 $nm{\cdot}s^{-1}$ with excitation wavelengths of 280 and 290 nm. For these assays, LaRPA-1 in 1× storage buffer (Tris-HCl 20 mM, NaCl 20 mM pH 8.0) was used at a final concentration of 28.85 pmol (28.85 μM). Oligonucleotides TEL1 (5′- TTAGGG- 3′), TEL2 (5′- $(TTAGGG)_2$–3′), TEL3 (5′- $(TTAGGG)_3$–3′) and TELG-DIG were also diluted in 1× storage buffer at a final amount of 32 pmol (32 μM). Binding reactions were done in the presence of a 1× storage buffer for 30 min at room temperature. Protection assays were done by adding 3 U of bacterial Exo I (New England Biolabs) in binding reactions containing LaRPA-1 and TEL3. After the addition of Exo I, these reactions were incubated in 1× Exonuclease I Reaction Buffer for 30 min at 37 °C. Exo I alone (3 U) gave a low fluorescent signal (Supplementary Fig. 2), which was subtracted from the LaRPA-1 + Exo I and LaRPA-1 + TEL3 + Exo I spectra. The fluorescence spectra of buffers and oligonucleotides did not present fluorescence signals for the concentrations and wavelengths tested.

2.5. G-quadruplex unfolding assay using fluorescence resonance energy transfer (FRET)

For this assay, the modified oligonucleotide F-htelo-T purchased from Eurogentec (Belgium), and purified recombinant LaRPA-1 (as described above) were used. Fluorescence titrations were carried out on a SPEX Fluorolog spectrofluorimeter (JobinYvon) equipped with a circulating water bath to regulate the temperature of the cell holder. Fluorescence spectra of 100 nM FhteloT oligonucleotides (double-labeled with FAM and TAMRA) were recorded at 20 °C in a buffer containing 50 mM NaCl, and 5 mM lithium cacodylate (pH 7.2). F-htelo-T was pre-incubated for 10 min at 20 °C. 100 to 1200 nM of LaRPA-1 was directly added to the solution containing the oligonucleotide. The spectra were collected after 5 min of incubation while exciting at 470 nm. The fluorescence intensity of the donor FAM (I_D) was measured at 518 nm, the fluorescence intensity of the acceptor TAMRA (I_A) was measured at 586 nm; the percentage of opened labeled G4 units at 20 °C was calculated as:

$$(P - P_f) \times 100/(P_u - P_f)$$

where:

$P = I_D/(I_D + I_A)$ at a given LaRPA-1 concentration at 20 °C,
$P_f = I_D/(I_D + I_A)$ in the absence of LaRPA-1 at 20 °C,
$P_u = I_D/(I_D + I_A)$ in the absence of LaRPA-1 at 80 °C.

We assumed that, in the absence of LaRPA-1, at 20 °C, all the labeled G4 units are folded, while at 80 °C, all the labeled G4 units are unfolded. Under these assumptions, P_f is the *P*-value corresponding to 0% of unfolded labeled G4 units, while P_u is the *P*-value corresponding to 100% of unfolded labeled G4 units. The expression $(P - P_f) \times 100/(P_u - P_f)$ provides, hence, the percentage of unfolded labeled G4 for a given LaRPA-1 concentration.

2.6. Search for homologous OB-fold domains in the Leishmania major genome

L. major genome was used to identify in *Leishmania* sp., proteins containing OB-fold domains homologous to the TEBPs OB-folds. We choose *L. major* genome because it is the best-annotated genome compared to the genomes of all other species of *Leishmania* currently available at TriTrypDB, the genomic database for protozoa belonging to the Trypanosomatidae family (https://tritrypdb.org/tritrypdb/). We predicted protein domains of all sequences (~ 9000 sequences) from the Annotated Proteins file from the *L. major* genome (strain Friendlin) obtained at TriTrypDB (release 40) using the jackhammer algorithm [48] available at the PFAM database [49] in batch mode. Subsequently, we searched for protein families and clans that are represented by OB-fold domains in prediction results towards their PFAM (version 32) identification codes using a minimum E-value of 1.0 as a cut-off.

2.7. Molecular docking with telomeric G-rich ssDNA

The molecular docking between the previously obtained *in silico* model of LaRPA-1-ΔCter [38] available at Model Archive database under the DOI https://doi.org/10.5452/ma-aydn8 and the crystal structure of telomeric G-rich ssDNA (TTAGGG) from PDB 1XJV (crystal structure of human POT1 bound to telomeric single-stranded DNA – TTAGGGTTAG) [10] was performed using the HADDOCK 2.2 software [50]. Since LaRPA-1 is able to bind the telomeric ssDNA only by its N-terminal OB-fold domain (OBF1) [38], we have chosen as active residues to drive the docking, the residues involved at ssDNA binding in human RPA70 that are conserved in OBF1 from LaRPA-1, according to a previous analysis [38]: Arg36, Trp38, Arg60, Phe64, Lys89, Phe95, and Glu103.

The protein/DNA docking protocol consisted in three stages [51]:

rigid-body docking; (ii) semi-flexible refinement stage, including a rigid-body torsion dynamics, a two semi-flexible simulated annealing refinements, where the side chains and backbone of the protein residues at the interface with DNA and the DNA molecule are allowed to mode; and (iii) a final refinement in explicit solvent. The final docking solutions were clustered through the pairwise backbone root mean square deviation (I-RMSD) at the interface, and the cluster was defined as an ensemble that displays I-RMSD smaller than 1.0 Å [50]. The resulting clusters ranked on the HADDOCK score, summarize the average interaction energies (electrostatic interaction energy; van der Waals interactions and restraints violation energy) and their average buried surface area [50].

2.8. Phylogenetic analysis

Homologous amino acid sequences of Replication Protein A1 (RPA-1) and OB-fold containing telomeric proteins CTC1 and POT1 were obtained from the NCBI database to construct two phylogenetic trees. First, we constructed a phylogenetic tree containing only RPA-1 homologs and posteriorly, a second phylogenetic tree containing RPA-1, CTC1, and POT1 sequences. Multiple sequence alignments were executed with MUSCLE [52], and the phylogenetic trees were initially built based on the Bayesian Method MCMC (Monte Carlo-Markov chain) using MrBayes v.3.2.6 [53] at CIPRES Science Gateway [54]. The phylogenetic trees were built using two concurrent MCMC runs (nruns) and four chains (nchains), a heating parameter of 0.2 (temp), swaps number (nswaps), and frequency (swapfreq) of 1, and a burn-in fraction of 0.25. RPA-1 and RPA-1/POT1/CTC1 phylogenetic trees reached a convergence value (average standard deviation of split frequencies) below 0.01 after 255,000 and 160,000 generations, respectively. Then, the topologies obtained by the Bayesian method were submitted as starting trees to a maximum-likelihood evaluation in PhyML [55] using NNI moves topology optimization. However, the topology obtained in ML after NNI optimization was the same obtained previously by the Bayesian method. Posterior probability values (in a 0–1 scale) and bootstrap values (1000 replicates in 0–100 scale) are respectively shown both beneath and above the internodes of the trees. Homologs of RPA-1 from the Archaea *Methanocaldococcus jannaschii* (RPA-1-M_jan) [33] and other two RPA-1-M_jan Archaea homologous sequences were used as outgroups. All sequences used in phylogenetic analysis are shown in Supplementary Table 1. Phylogenetic trees were drawn using Mesquite v.3.51 [56].

3. Results

3.1. Structural homologs of classical telomeric OB-fold domains were not found in Leishmania major genome

We predicted the domains of all sequences from the Annotated Proteins file from the *L. major* genome (strain Friendlin) obtained at TriTrypDB (version 30) to search for OB-fold containing proteins in *L. major* genome. The structural domain prediction analysis using jackhammer revealed proteins involved in DNA replication, repair, transcription, and translation, but found no real homologs of known TEBPs involved in telomeric functions (Table 1). Moreover, at the protein families (Pfam) database, no OB-fold domains registered in proteins involved with telomeric functions were found in any protein annotated in the *L. major* genome database (Supplementary Table 2). These data may indicate the absence of classical TEBPs at *Leishmania* sp. telomeres, highlighting previous results [38]. Similar results were obtained for the *Trypanosoma cruzi* CL-Brenner genome using comparable approaches [45]. Furthermore, orthologues of the CST proteins (CDC13/CTC1, STN1, and TEN1) and the shelterin members POT1 and TPP1 were also not found in any trypanosomatid genome at the TriTrypDB using algorithms based on protein sequence similarity [38]. However, it is important to point out that although bioinformatic search algorithms have greatly improved over the years, it is still not trivial to predict OB-fold domains from a sequence. A good example is the absence of an efficient algorithm to identify the C-terminal POT1 OB-fold domain (POT1C), which has an HJR domain insertion within it, which prevented its identification as an OB-fold domain until its structure was solved by two separate groups [57,58].

To reinforce our hypothesis, we also did a test showing that jackhammer can predict the presence of telomeric OB folds even among evolutionary divergent organisms. We first performed blastp searches using *Homo sapiens* and *Schizosaccharomyces pombe* POT1 (both present crystal structures) as queries. POT1 homologs were identified in higher and lower eukaryotes, and these POT1 homologs were further submitted to jackhammer domain prediction. Jackhammer algorithm could predict telomeric OB-fold domains in the N-terminal region of all studied sequences (Supplementary Table 3), independent of the conservation level shared among them. This result suggests that if there were POT1 homologs or other OB-fold containing telomeric proteins in trypanosomatids, we would find them using this approach.

3.2. LaRPA-1 can interact with one telomeric repeat and shares similar ssDNA binding structural characteristics with eukaryotes RPA-1 and with POT1

As previously shown, *Leishmania* sp. telomeres are composed of the conserved 5′-TTAGGG-3′ repeated sequence in double and single-stranded form, with the G-rich ssDNA forming a protrusion, the 3′ G-overhang, towards the end of the chromosome [59]. The 3′ G-overhang serves as the substrate for telomere addition by telomerase and as the binding site for LaRPA-1 and other already characterized parasite-specific single-stranded telomere-binding proteins [39,46,60].

To better understand the interaction between LaRPA-1 and the telomeric G-rich single-stranded DNA, we performed docking predictions using HADDOCK 2.2 [50] between the previously obtained OBF1 from LaRPA-1 *in silico* model (available at Model Archive database under the DOI https://doi.org/10.5452/ma-aydn8) and a telomeric ssDNA containing a single 5′ TTAGGG 3′ repeat (extracted from PDB 1XJV). The best docking solution (Fig. 1) obtained a HADDOCK score of -77.2 ± 8.7, a root mean square fluctuation (r.m.s.d.) from the overall lowest-energy structure of 0.9 ± 0.5 Å; a van der Waals energy of -58.0 ± 10.2 kcal·mol^{-1}; an electrostatic energy of -158.1 ± 47.4 kcal·mol^{-1}; a desolvation energy of 11.8 ± 10.0 kcal·mol^{-1}; a restraints violation energy of 6.0 ± 10.13 kcal·mol^{-1}; a buried surface area of the protein interface bound to ssDNA of 1266.5 ± 93.7 Å^2 and a *Z*-score of -1.6. The best docking solution (Fig. 1A) showed that the telomeric ssDNA fits into the DNA-binding channel of LaRPA-1 OBF1 formed by residues Arg36, Trp38, Arg60, Phe64, Lys89, Phe95, and Glu103, which are involved with ssDNA binding in human RPA-1 (RPA70) [43] and are conserved in LaRPA-1, according to our previous analysis [38]. Also, besides Trp38, two other tryptophan residues (Trp23 and Trp24) are present in the LaRPA-1 DNA-binding channel and are located in the DNA vicinity (Fig. 1A).

The nature of all hydrogen-bonds and the several hydrophobic interactions between OBF1 from LaRPA-1 and the telomeric ssDNA (Fig. 1B), showed that the ssDNA/LaRPA-1 complex presents similar ssDNA binding mode to that respectively observed for the crystal structures of ssDNA/human POT1 and ssDNA/fungi and human RPA-1 [10,43,44,61]. The OB-fold domains of these crystal structures present a structural ssDNA binding motif that is characterized by, proceeding from 5′ to 3′, a base stacked on an aromatic residue and the next two stacked on each other with a second aromatic residue [10,43,44,61]. The same ssDNA binding motif is observed in the LaRPA-1/ssDNA complex, where Phe95 stacks on the base positioned at the 5′ end of the ssDNA and Phe64 stacks on the two subsequent bases (Fig. 1).

Table 1
Sequences from the Annotated Proteins file from the *Leishmania major* genome (strain Friendlin, version 40) with OB-fold domains predicted in their amino acid sequences by jackhammer algorithm available at the PFAM database (version 32). The access number of these sequences in the Annotated Proteins file (Transcription Tryp), the region predicted as an OB-fold domain, and the respective E-values are also shown.

Description PFAM	Transcript Tryp	Product	Protein region	Full-length protein (residues)	E-value
OB-fold nucleic acid binding domain	LmjF.34.2340.1	asparaginyl-tRNA synthetase, putative	239–310	667	3.0e-14
	LmjF.34.2340.2	asparaginyl-tRNA synthetase, putative	462–533	890	4.4e-14
	LmjF.30.0130	lysyl-tRNA synthetase, putative	1–68	536	7.8e-11
	LmjF.30.0460	aspartyl-tRNA synthetase, putative	81–164	550	6.0e-13
	LmjF.28.1820	Replication factor A protein 1	23–104	467	1.6e-12
	LmjF.21.0895	aspartyl-tRNA synthetase, putative	137–214	641	1.9e-12
	LmjF.15.0230	lysyl-tRNA synthetase, putative	113–193	586	3.3e-17
	LmjF.15.0270	Replication Factor A 28 kDa subunit, putative	62–138	257	8.8e-13
S1 RNA binding domain	LmjF.03.0980	Eukaryotic translation initiation factor 2 subunit alpha	128–202	413	6.0e-16
Oligonucleotide/oligosaccharide (OB)-fold	LmjF.36.5670	RNA helicase, putative	612–690	697	1.8e-18
	LmjF.36.2830	pre-mRNA splicing factor ATP-dependent RNA helicase, putative	968–1046	1088	7.7e-20
	LmjF.35.1200	pre-mRNA splicing factor ATP-dependent RNA helicase, putative	589–665	704	2.9e-17
	LmjF.34.2260	ATP-dependent RNA helicase-like protein	636–730	768	3.4e-11
	LmjF.24.1100	pre-mRNA-splicing factor ATP-dependent RNA helicase, putative	976–1055	1138	1.2e-14
	LmjF.09.0070	RNA helicase, putative	1062–1152	1317	8.3e-17
Eukaryotic elongation factor 5A hypusine, DNA-binding OB fold	LmjF.25.0720	eukaryotic initiation factor 5a, putative	86–161	166	1.7e-30
	LmjF.25.0730	eukaryotic initiation factor 5a, putative	86–161	166	1.7e-30
RNA polymerase Rpb8	LmjF.28.0810	RNA polymerase B subunit RPB8, putative	9–146	149	3.3e-40
Domain of unknown function (DUF4539)	LmjF.28.0705	Domain of unknown function (DUF4539), putative	33–117	250	1.5e-17
RecQ mediated genome instability protein	LmjF.23.0090	Domain of unknown function (DUF1767), putative	8–180	346	7.7e-29
	LmjF.25.1760	hypothetical protein, conserved	50–253	501	8.5e-16
BRCA2, oligonucleotide/oligosaccharide-binding, domain 1	LmjF.20.0060	DNA repair protein BRCA2, putative	424–534	1165	2.1e-26
MCM OB domain	LmjF.35.4910	DNA replication licensing factor MCM9, putative	126–244	908	4.3e-16
	LmjF.33.2700	DNA replication licensing factor MCM3, putative	134–272	881	1.7e-29
	LmjF.32.2960	DNA replication licensing factor MCM7, putative	143–276	725	3.2e-30
	LmjF.28.0850	DNA replication licensing factor MCM2, putative	314–443	969	2.5e-39
	LmjF.28.2385	DNA replication licensing factor MCM6, putative	156–288	880	1.9e-38
	LmjF.24.0910	DNA replication licensing factor MCM5, putative	138–281	801	4.7e-37
	LmjF.09.0250	DNA replication licensing factor MCM4, putative	250–379	985	2.1e-38
	LmjF.05.0330	DNA replication licensing factor MCM8, putative	278–380	1049	7.9e-13
	LmjF.05.0330.2	DNA replication licensing factor MCM8, putative	222–324	993	9.7e-13
ATP dependent DNA ligase C terminal region	LmjF.30.3440	DNA ligase I, putative	626–736	767	6.5e-27
Ribosomal protein S28e	LmjF.26.1630	40S ribosomal protein S33, putative	20–86	87	1.5e-33
	LmjF.26.1640	40S ribosomal protein S33, putative	20–86	87	1.5e-33
Translation initiation factor 1AIF-1	LmjF.16.0140	eukaryotic translation initiation factor 1A, putative	36–99	168	1.2e-16
Putative tRNA binding domain	LmjF.23.1650	Putative tRNA binding domain containing protein, putative	46–146	157	3.1e-17
	LmjF.22.0470	Putative tRNA binding domain containing protein, putative	268–359	426	4.2e-30
	LmjF.16.1130	Tyrosyl or methionyl-tRNA synthetase-like protein	64–175	237	9.7e-27
Rrp44-like cold shock domain	LmjF.28.0330	rrp44p homolog, putative	236–325	992	8.1e-11
RNA polymerase III subunit Rpc25	LmjF.09.1060	DNA-directed RNA polymerase III subunit, putative	121–190	297	9.7e-11
Ribosomal Proteins L2, RNA binding domain	LmjF.35.1430	60S ribosomal protein L2, putative	11–90	260	1.7e-16
	LmjF.35.1440	60S ribosomal protein L2, putative	11–90	260	1.7e-16
	LmjF.32.3900	60S ribosomal protein L8, putative	11–90	260	1.7e-16
Ribosomal protein S17	LmjF.35.3850	30S Ribosomal protein S17-like protein	96–163	312	1.6e-21
	LmjF.20.1650	ribosomal protein S11 homolog	88–156	173	4.4e-27
	LmjF.21.1550	40S ribosomal protein S11, putative	88–156	173	4.4e-27
Replication factor A protein 3	LmjF.35.3440	replication factor A protein 3, putative	21–117	117	6.1e-11
DNA ligase OB-like domain	LmjF.26.1340	DNA ligase k alpha, putative	549–621	630	6.9e-22
	LmjF.26.1350	mitochondrial DNA ligase homolog, putative	415–485	494	2.5e-19
Replication protein A OB domain	LmjF.28.1820	Replication factor A protein 1	135–230	467	2.4e-24
Replication factor-A C terminal domain	LmjF.28.1820	Replication factor A protein 1	293–444	467	9.8e-43
'Cold-shock' DNA-binding domain	LmjF.31.1340	'Cold-shock' DNA-binding domain containing protein, putative	398–447	461	8.7e-13
	LmjF.28.0825	RNA binding protein rbp16, putative	18–85	142	6.5e-20
	LmjF.14.1140	hypothetical protein, conserved	400–460	515	8.7e-10
mRNA capping enzyme, C-terminal domain	LmjF.22.0090	cap guanylyltransferase-methyltransferase 1	397–505	1037	7.4e-12
Single-strand binding protein family	LmjF.36.0700	MP18 RNA editing complex protein, putative	77–180	229	9.8e-10
Ribosomal protein S12/S23	LmjF.21.1060	40S ribosomal protein S23, putative	31–142	143	1e-52
	LmjF.21.1070	40S ribosomal protein S23, putative	31–142	143	1e-52

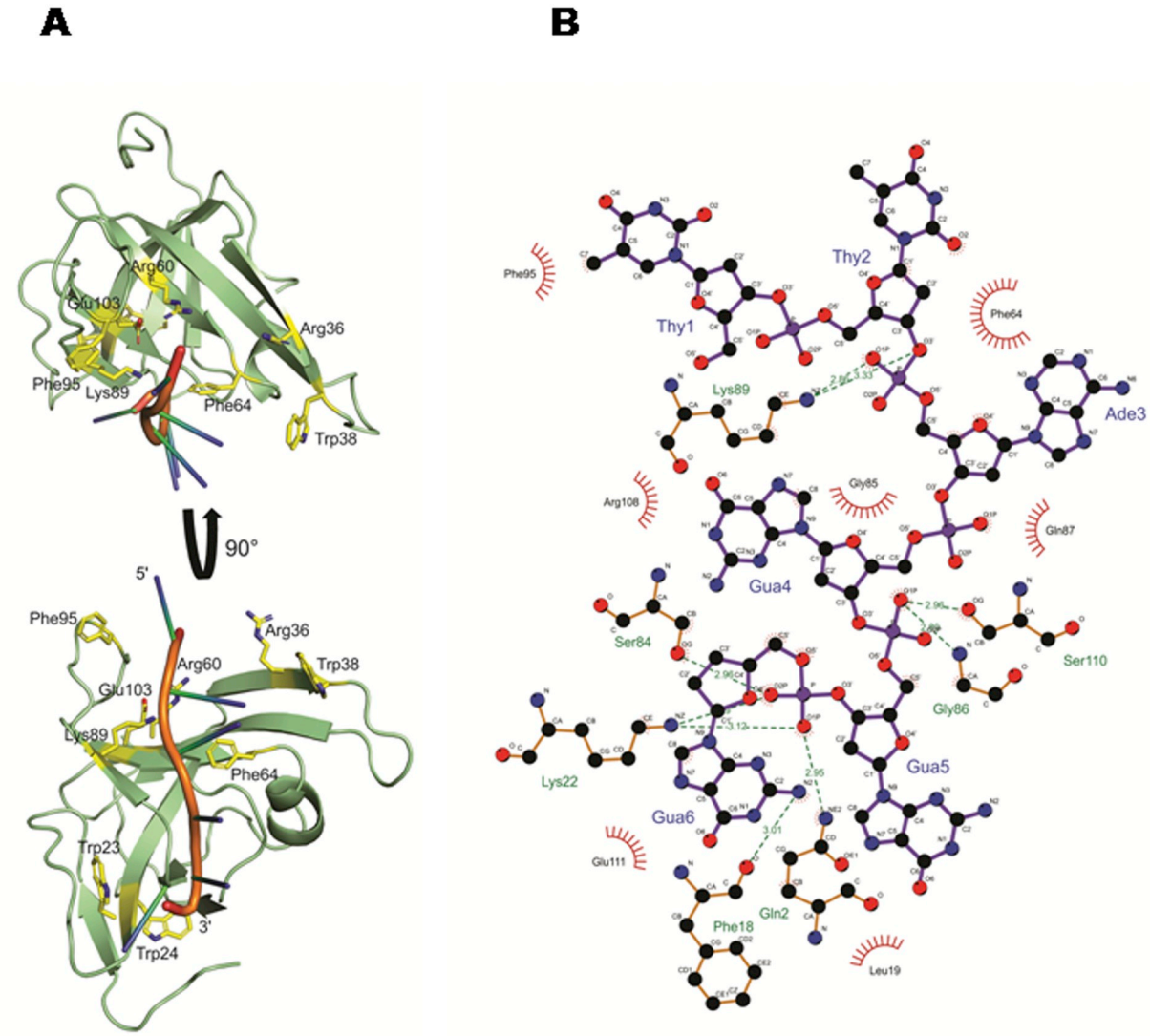


Fig. 1. Molecular docking between LaRPA-1 *in silico* model and telomeric ssDNA. A) Cartoon representation of best docking solution from N-terminal OB-fold domain (OBF1) from LaRPA-1 *in silico* model (available at Model Archive database under the code ma-aydn8) with the telomeric G-rich strand (TTAGGG; from PDB ID 1XJV) provided by HADDOCK v.2.2 software. The residues involved with ssDNA binding in human RPA70 that are conserved in LaRPA-1 (Arg36, Trp38, Arg60, Phe64, Lys89, Phe95, and Glu103) and Trp23 and Trp24 residues are highlighted in yellow sticks. Single-stranded DNA is shown in red-orange and blue cartoons. B) Diagram of interactions established between LaRPA-1 and the telomeric G-rich DNA. The hydrogen bonds established between protein residues and telomeric DNA (in blue) are shown as green dashes with distances indicated in A). The hydrophobic contacts are shown as red arcs with radiation spokes. This figure was obtained using DIMPLOT software [78].

3.3. LaRPA-1 binds minimally one, but preferentially three or more telomeric repeat sequences

To determine the optimal length of the telomeric G-rich sequence for LaRPA-1 binding, we investigated the binding of LaRPA-1 with telomeric oligonucleotides containing one (TEL1), two (TEL2), three (TEL3) or five (TELG-DIG) telomeric TTAGGG repeats. TELG-DIG was studied in this experiment to check if the presence of the digoxigenin (DIG), used as a probe in EMSA assays (see Materials and methods), does not modify LaRPA-1/ssDNA interactions. Static fluorescence spectroscopy was performed (Fig. 2), based on our previous data demonstrating that the interaction with telomeric ssDNA causes significant structural changes in LaRPA-1 [38] and, particularly, an expected environmental change for three tryptophan residues in vicinity of DNA binding site (Trp23, Trp24, and Trp38; Fig. 1A).

Fluorescence spectra of LaRPA-1 alone using 280 and 290 nm as excitation wavelengths showed a fluorescence emission maxima at 339–340 nm whereas, in the presence of telomeric sequences containing one to five G-rich repeats, the fluorescence emission maxima shifted from 339 to 340 nm to 336 nm (280 nm as excitation wavelength) and 331 nm (290 nm as excitation wavelength) (Fig. 2A and B, respectively). These data indicated that the tryptophan residues near the DNA-binding site came to a more hydrophobic environment in the presence of DNA, thus, showing the interaction between LaRPA-1 and the telomeric G-rich ssDNA. However, the suppression of the fluorescent signal was observed when LaRPA-1 was bound to three or more telomeric repeats (TEL3 and TELG-DIG) (Fig. 2A and B). The emission maxima shift, accompanied by a high decrease of the fluorescence emission intensity in LaRPA-1/TEL3 and LaRPA-1/TELG-DIG spectra indicated a stronger interaction between the protein and these oligonucleotides in comparison to shorter telomeric repeats (TEL1 and TEL2). Besides, it is worth mentioning that we already knew that LaRPA-1 binds similarly G-rich telomeric sequences with three or more repeats (Lira & Cano, unpublished data) and here we show that the fluorescence spectrum of LaRPA-1/TELG-DIG was identical to LaRPA-1/TEL3 (Fig. 2A and B), indicating that the presence of DIG did not

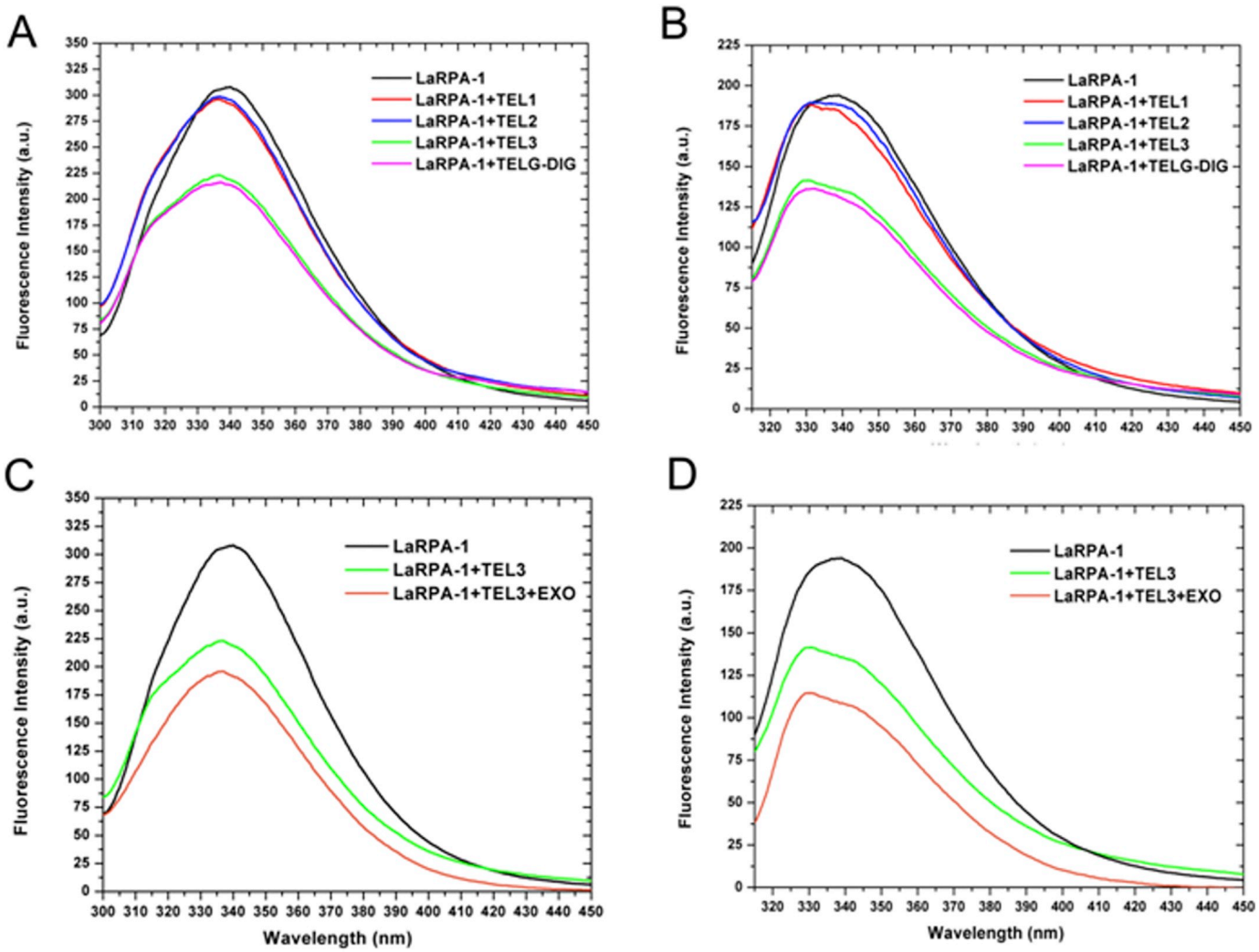


Fig. 2. Spectroscopy fluorescence spectra of LaRPA-1 in the presence of telomeric ssDNA. Spectroscopy fluorescence spectra using 280 nm (A) and 290 nm (B) as excitation wavelengths of LaRPA-1 alone (black line) and in the presence of ssDNA containing one, two or three telomeric repeats (TEL1 - red line; TEL2 - blue line and TEL3 - green line, respectively), and five telomeric repeats tagged with digoxigenin at the 5′ end (TELG-DIG - magenta line). Spectroscopy fluorescence spectra using 280 nm (C) and 290 nm (D) as excitation wavelengths of LaRPA-1 alone (black line) and in the presence of ssDNA with three telomeric repeats (TEL3), in the absence (green line) or in the presence of 3 U of *E. coli* exonuclease I (EXO) (orange line).

disturb LaRPA-1–DNA interactions.

Together, these results show that LaRPA-1, similar to *T. cruzi* RPA-1 (TcRPA-1) [45], can bind a single telomeric repeat but has a preference to bind three or more telomeric repeats.

3.4. Bacterial 3′-5′ exonuclease I does not disturb the interaction between LaRPA-1 and telomeric DNA

To test the ability of LaRPA-1 to protect the telomeric G-rich ssDNA from exonucleolytic degradation, we performed an exonuclease protection assay using *E. coli* exonuclease I (Exo I). Exo I presents a 3′-5′ activity and degrades deoxyribonucleosides releasing 5′-monophosphates and leaving free and intact dinucleotides at the 5′ end [62].

We performed static fluorescence spectroscopy assays to check whether *E. coli* Exo I could dissociate LaRPA-1 from telomeric ssDNA. The LaRPA-1/TEL3 spectra before and after Exo I incubation (Fig. 2C and D) presented the same fluorescence emission maxima at 336 nm using 280 nm as excitation wavelength (Fig. 2C) and at 331 nm using 290 nm as excitation wavelength (Fig. 2D), indicating that in the presence of Exo I, LaRPA-1 was still bound to the DNA. It was also possible to observe a slight difference in fluorescence emission intensity that was maybe attributed to the dilution of the sample after Exo I addition. Thus, the fluorescence spectroscopy analysis indicated that the LaRPA-1/ssDNA complex was not dissociated by the presence of Exo I. Although this data did not provide direct evidence of the absence of DNA degradation, it suggests that LaRPA-1, even in the presence of high amounts of Exo I, can protect the telomeric G-rich ssDNA from exonucleolytic degradation.

3.5. LaRPA-1 protects the telomeric G-rich ssDNA from 3′-5′ exonucleolytic degradation

Initially, we tested the amount of Exo I that would degrade most of TELG-DIG in the absence of LaRPA-1. The results demonstrated that the amount of degraded TELG-DIG was proportional to the enzyme concentration as detected by the decrease in the chemiluminescent signal (100–14.3%), which may correspond to the extension of the hydrolysis (Supplementary Fig. 3A). Concentrations of Exo I ≥ 2 U significantly decreased the chemiluminescent signal (~56%–14.3%), compared to the non-treated TELG-DIG (100%) (Supplementary Fig. 3A, lanes 3–8 and accompanying graph). Thus, we subsequently did binding assays using 1.5 μg (28.85 pmol) of LaRPA-1 and 1.55 pmol of TELG-DIG in the absence or presence of different amounts of Exo I and with heat-denatured (HD) LaRPA-1 (control reactions). The results showed that

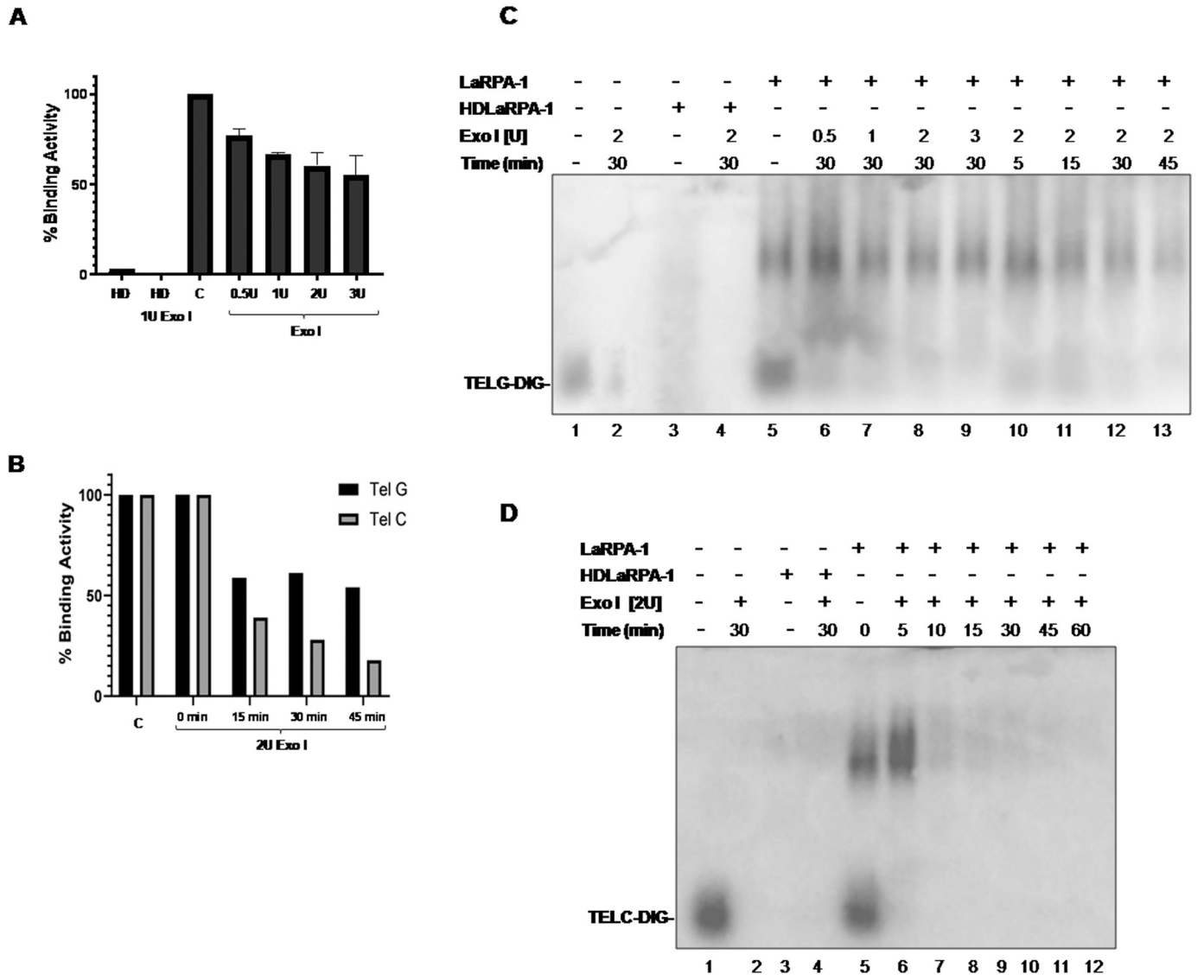


Fig. 3. The presence of bacterial Exo I does not perturb LaRPA-1 binding to telomeric DNA. In A), the quantitative analysis of EMSA assays performed with 1.55 pmol of TELG-DIG and 28.85 pmol LaRPa-1, incubated for 30 min at 37 °C with different concentrations of Exo I (0.5, 1 U, 2 U, and 3 U). Control reactions (C) were done in the absence of LaRPA-1 or Exo I and with heat-denatured LaRPA-1 (HDLaRPA-1) and HDLaRPA-1 + 2 U Exo I. The data shown in the graph represent the average ± SEM of three independent assays. B) The quantitative analysis of EMSA assays performed with 1.55 pmol of TELG-DIG (black bars) and TELC-DIG (gray bars) and 28.85 pmol LaRPa-1, incubated at different time points (0 min, 5 min, 15 min, 30 min, 45 min) at 37 °C with 2 U Exo I. The data shown in the graph represent the average ± SEM of two independent assays. As control (C) binding reactions with TELG-DIG and TELC-DIG were done in the absence of Exo I. C) Representative EMSA assays were done with 1.55 pmol of TELG-DIG as the probe and 28.85 pmol LaRPa-1. Lane 1, the reaction was done in the absence of protein; lane 2, the reaction was done in the absence of protein followed by treatment with 2 U Exo I for 30 min at 37 °C; lane 3, the reaction was done with heat-denatured LaRPA-1 (HDLaRPA-1); lane 4, the reaction was done with HDLaRPA-1 followed by treatment with 2 U Exo I for 30 min at 37 °C; lane 5, binding reaction was done with LaRPA-1 and TELG-DIG and used as the control; lanes 6–9, binding reactions were done with LaRPA-1 and TELG-DIG followed by treatment respectively with 0.5 U, 1 U, 2 U and 3 U Exo I for 30 min at 37 °C; lanes 10–13, binding reactions were done with LaRPA-1 and TELG-DIG were respectively treated with 2 U Exo I for 5 min, 15 min, 30 min, 45 min at 37 °C. D) Representative EMSA assays were done with 1.55 pmol of TELC-DIG as the probe and 28.85 pmol LaRPA-1. Lane 1, the reaction was done in the absence of protein; lane 2, the reaction was done in the absence of protein followed by treatment with 2 U Exo I for 30 min at 37 °C; lane 3, the reaction was done with HDLaRPA-1 OK; lane 4, the reaction was done with HDLaRPA-1 followed by treatment with 2 U Exo I for 30 min at 37 °C; lane 5, binding reaction was done with LaRPA-1 and TELC-DIG and used as the control; lanes 6–12, binding reactions were done with LaRPA-1 and TELC-DIG were respectively treated with 2 U Exo I for 0 min, 5 min, 10 min, 15 min, 30 min, 45 min, and 60 min at 37 °C. In A) and B), Binding complexes were fractionated on a 1% agarose gel in 0.25× TBE followed by transfer to nylon N+ Hybond membrane. The results were detected using an anti-DIG AP antibody (Roche) and were revealed by chemiluminescence with CSPD (Roche).

even in the presence of high amounts of Exo I (2-3 U), it was still possible to detect LaRPA-1 bound to TELG-DIG (Fig. 3A). Besides, it could be observed that the percentage of binding activity decreased proportionally (about 80–60%) to the amount of Exo I in the mixture (Fig. 3A), probably due to a competition between Exo I and LaRPA-1 for TELG-DIG.

The following step was to test the effect of Exo I activity in the stability of LaRPA-1/TELG-DIG (Fig. 3B, black bars) and LaRPA-1/TELC-DIG (Fig. 3B, gray bars) complexes along the time. The results showed that 60% of the complexes were still formed after 15 min incubation with 2 U Exo I and remained stable even after 45 min in the presence of Exo I. Assays using TELC-DIG as the probe (Fig. 3B gray bars) showed that the % of binding activity decreased over time and to 15% after 45 min incubation. These data confirmed that LaRPA1

efficiently protects TELG but not TELC, suggesting specific protection of LaRPA1 to the telomeric G-rich strand. Fig. 3C and D show EMSA assays summarizing the results with TELG and TELC (in both assays, control reactions were done in the absence of Exo I and with heat-denatured LaRPA-1).

To confirm that RPA-1 protects the telomeric G-rich DNA from Exo I degradation, LaRPA1/TELG-DIG complexes obtained from two different LaRPA-1 concentrations (13.46 and 40.39 pmol) and a fixed amount of TELG-DIG (1.55 pmol) were incubated with 2 U Exo I for 30 min at 37 °C. Exo I treated and Exo I non-treated TELG-DIG was used as the control. Exo I-treated LaRPA-1/TELG-DIG complexes were subsequently incubated at 95 °C for 5 min to dissociate the LaRPA-1/TELG-DIG complexes, followed by the analysis of the remaining DNA (TELG-DIG) by PAGE (Supplementary Fig. 3B). The results showed a high reduction of the TELG-DIG signal in the Exo I-treated control reaction (Supplementary Fig. 3B, lane 2), whereas no decrease in the chemiluminescent signal of the Exo I-treated LaRPA-1/TELG-DIG complexes was observed (Supplementary Fig. 3B, lanes 3 and 4), independent on the amount of LaRPA-1 in the formed complex. These data confirm that LaRPA-1 protects TELG-DIG from Exo I degradation.

3.6. FRET experiments show that LaRPA-1 efficiently unfolds telomeric G-quadruplex structures

To determine whether LaRPA-1 can unfold telomeric G-quadruplex structures similar to OB-fold containing telomeric proteins, we carried out FRET experiments with a 21 mer telomeric sequence, htelo ($(G_3T_2A)_3G_3$), that mimics the telomeric G-rich 3′ overhang and may adopt G-quadruplex structures [63,64]. htelo was labeled with two fluorophores, a 6-carboxyfluorescein (FAM) and tetramethyl-rhodamine (TAMRA) at the 5'and 3'termini, respectively. The fluorimetric titration of F-htelo-T with LaRPA-1 revealed an increase of FAM fluorescence and a decrease in TAMRA fluorescence upon LaRPA-1 addition (Fig. 4A), indicating that LaRPA-1 can open the telomeric G-quadruplex. The percentage of unfolded labeled G-quadruplex upon LaRPA-1 concentration was calculated (Fig. 4B). LaRPA-1 ability to resolve telomeric G-quadruplex can be compared to hPOT1, although it has never been tested if single subunits of both hRPA (*e.g.*, hRPA-1) and CST (*e.g.*, CTC1) heterotrimeric complexes can unfold these intramolecular structures without the presence of the other subunits.

3.7. Phylogenetic relationships of trypanosomatids RPA-1

To gain insights into the phylogenetic relationships of trypanosomatids RPA-1, we constructed phylogenetic trees with trypanosomatids RPA-1 and with RPA-1, POT1, and CTC1 amino acid sequences from different organisms (including higher and lower eukaryotes). We first obtained the topologies of the phylogenetic trees by the Bayesian method, which were subsequently submitted to the maximum-likelihood (ML) evaluation by NNI moves topology optimization. The same topology was obtained by the ML method showing the robustness of topologies obtained from the phylogenetic trees.

We started the evolutionary study of trypanosomatids RPA-1 by constructing a phylogenetic tree containing only RPA-1 homologous amino acid sequences (Fig. 5A). The RPA-1 phylogenetic tree shows that RPA-1 from trypanosomatids occupies the most basal clade of the tree and shares a common ancestry with RPA-1 from higher eukaryotes with excellent statistic support (97% of bootstrap and 1 of posterior probability value) (Fig. 5A). Also, RPA-1 from ciliates and apicomplexan nest in a monophyletic group with RPA-1 from plants with moderate statistic support (55% of bootstrap and 0.77 of posterior probability value), as well as fungi RPA-1 and metazoan RPA-1, which also nest in a monophyletic group with good statistic support (70% of bootstrap and 0.83 of posterior probability value) (Fig. 5A).

To estimate the phylogenetic relationships between RPA-1 from trypanosomatids, and POT1 and CTC1 from different organisms, we also constructed a phylogenetic tree (Fig. 5B). This phylogenetic tree shows trypanosomatids RPA-1 nest in a well-supported clade with POT1 and CTC1 from higher eukaryotes (87% of bootstrap and 0.98 of posterior probability value) (Fig. 5B). This data indicates that trypanosomatids RPA-1 may share the same common ancestor with higher eukaryotes OB-fold containing telomeric proteins POT1 and CTC1, highlighting the telomeric functions of RPA-1 from trypanosomatids.

4. Discussion

The trypanosomatid family is composed of earlier-divergent eukaryotes that include protozoa from the *Leishmania* and *Trypanosoma* genera, among which are the causative agents of severe neglected tropical diseases, such as leishmaniasis, African and American trypanosomiasis [65]. In this context, telomeres may be pointed out as a potential target for the development of new therapies to treat these diseases due to their key role in genome maintenance and stability [66].

However, unveiling the composition of trypanosomatids telomeric chromatin has been challenging due to recent evidence showing that these organisms may present TEBPs that are exclusive to the group and thus, not found at other eukaryote telomeres. Recently, it was identified in *L. amazonensis* telomeres, a calmodulin-like protein, LCalA [60], and a protein prior described as mitochondrial, Rbp38 [67] that has no homologs outside the Trypanosomatid family. Also, the present work highlights previous studies predicting the absence of OB-fold containing telomeric proteins in trypanosomatids, raising the possibility that other OB-fold containing proteins such as RPA-1, would be a potential TEBP in these organisms.

4.1. The telomere-end binding functions of LaRPA-1

Previous results from our group suggested that RPA-1 from trypanosomatids is a potential TEBP candidate. For example, LaRPA-1 was shown to pull-down alone and free of RPA-2 and RPA-3, but together with other recently characterized organism-specific telomeric proteins [60,67] and DNA polymerase alpha in affinity chromatography assays using the telomeric G-rich DNA as the ligand [36]. Immunoaffinity purification of *L. amazonensis* telomerase positive extracts using an specific anti-LaRPA-1 serum, followed by mass spectrometry analysis, have also identified among other non-characterized parasite proteins, LaRPA-1 alone and free of LaRPA-2 and LaRPA-3 (Supplementary Table 4). Moreover, LaRPA-1 and TcRPA-1 were shown to co-localize with telomeric DNA, and to preferentially bind *in vitro* and *in vivo* the telomeric G-rich ssDNA [38,42,45]. LaRPA-1 can bind to ssDNA using a single N-terminal OB-fold domain (OBF1) [38,42]; interestingly, a unique OB-fold domain seems also to be enough for the DNA-binding activity of CDC13 [68]. And finally, experimental evidence has shown that RPA-2 from trypanosomatids may not bind ssDNA [69], which also argues in favor of LaRPA-1 protagonism at parasite telomeres.

In this work, we showed that the LaRPA-1 ssDNA binding mode observed by molecular docking studies (Fig. 1) presents many similarities with the ssDNA binding mode shown by POT1 and RPA-1 crystal structures [10,43,44,61], highlighting the existence of a conserved binding motif in an RPA-1 from an earlier-divergent eukaryote. Besides that, LaRPA-1 also shares similar ssDNA binding structural features with RPA-1 and POT1 from higher eukaryotes. Moreover, the structural model and fluorescence spectroscopy analysis of LaRPA-1/ssDNA showed that LaRPA-1 can bind a single telomeric repeat (Figs. 1 and 2), but has a preference to bind three or more telomeric repeats. However, LaRPA-1, differently from POT1 [70] and CDC13 [14] does not show preferences for any subset of bases in the telomeric G-rich repeat since it can bind 18 mer oligonucleotides containing all the 3'end permutations of the TTAGGG sequence [39]. RPA-1 from *T. cruzi* is also able to bind random sequences of ssDNA but, as LaRPA-1, it shows a higher affinity for telomeric G-rich sequences [38,42,45]. Here it is worth to remind that human RPA-1 can also bind telomeric G-rich ssDNA during

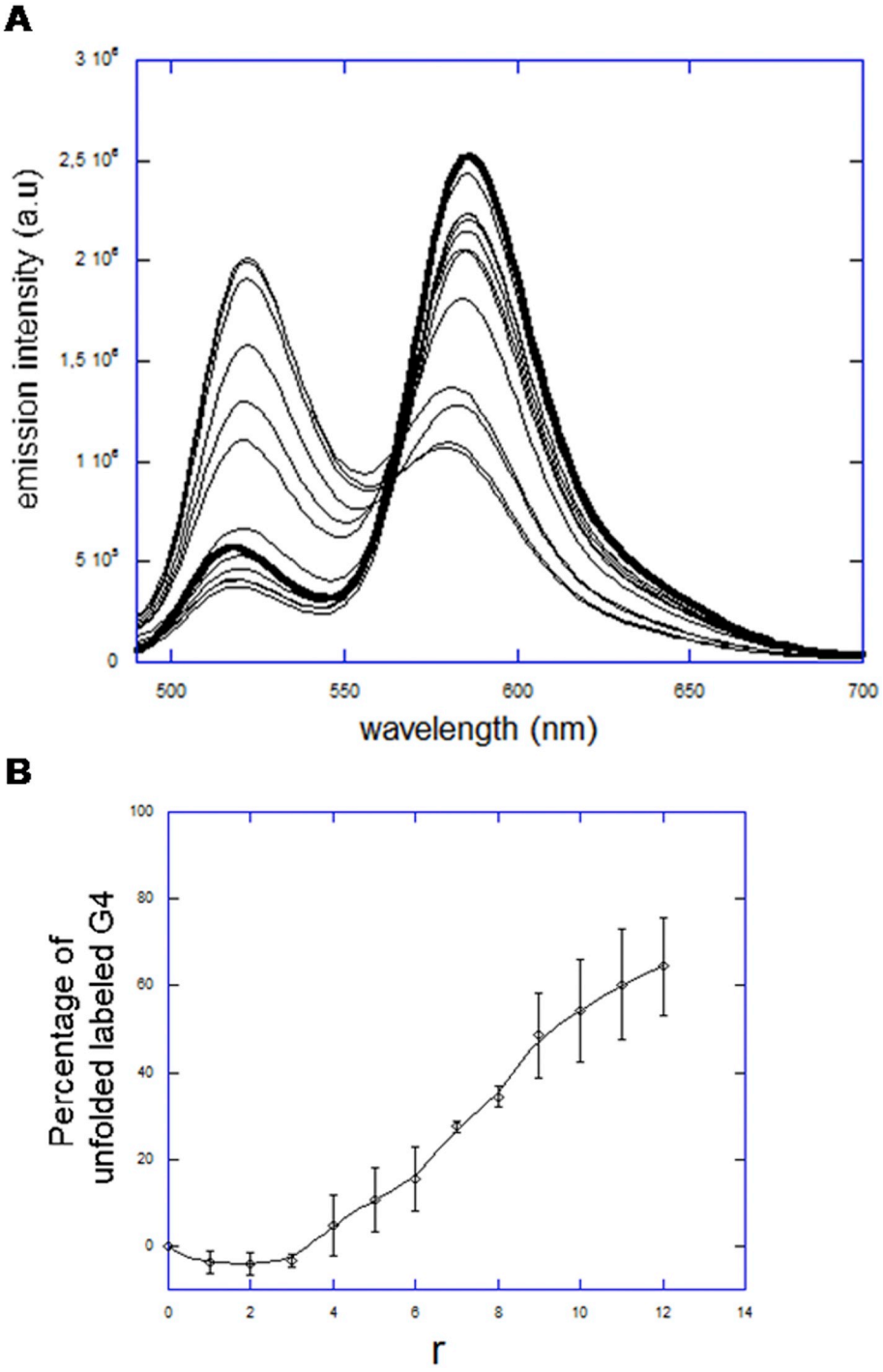


Fig. 4. LaRPA-1 binding leads to G-quadruplex unfolding. A) Fluorimetric titration of 100 nM of F-htelo-T by increasing amounts of LaRPA-1 was performed in the presence of Na+. The spectra were recorded after 5 min of incubation. B) Increasing the percentage of unfolded labeled G-quadruplex as a function of the LaRPA-1/F-htelo-T ratio (r) indicates the unfolding of F-htelo-T by the addition of LaRPA-1.

DNA replication [71] and to interact with similar affinity to other ssDNA [17]. Similarly, CTC1 alone and higher eukaryotes CST complex also show sequence-independent ssDNA-binding activity [17] whereas CDC13 and POT1, present specific and higher affinity to the telomeric G-rich ssDNA [14,17,70].

An important feature shared among the OB-fold containing telomeric proteins is preventing telomeres from being recognized as DNA double-stranded breaks and, from being exposed to exonucleolytic degradation [10,11]. Using different approaches (Figs. 2 and 3), we were able to show that LaRPA-1 can protect the telomeric G-rich ssDNA from *E. coli* 3′-5′ Exo I activity and with a higher efficiency compared to the telomeric C-rich strand (Fig. 3A), indicating that LaRPA-1 protection may be specific to the telomeric G-rich ssDNA. Moreover, Exo I could not destabilize LaRPA-1/ssDNA complexes at increasing time intervals of 15 min, even when the complexes were exposed to a high amount of Exo I (Fig. 3B–C), indicating that LaRPA-1 can have protective functions on telomeres.

LaRPA-1 alone, without RPA-2 and RPA-3 can unfold telomeric G-quadruplex structures (Fig. 4), which is another important biochemical feature played by the CST complex, POT1 alone, and POT1-TPP1

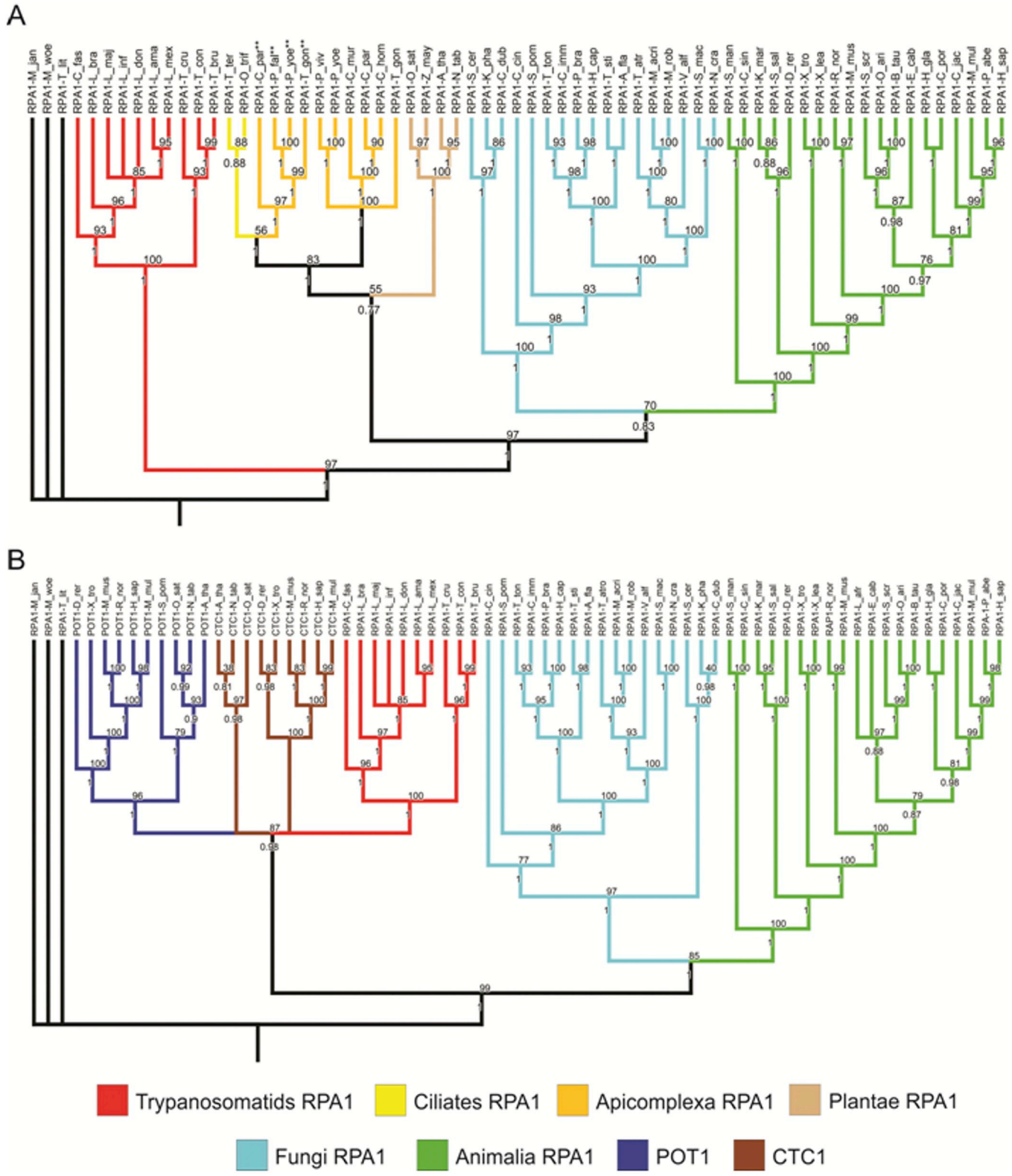


Fig. 5. Phylogenetic trees of telomeric OB-fold containing proteins. A) Phylogenetic tree obtained by Bayesian and maximum-likelihood methods (same topology) of homologous Replication Protein A1 (RPA-1) sequences from eukaryotes. B) Phylogenetic tree obtained by Bayesian and maximum-likelihood methods (same topology) of RPA-1, Conserved Telomere Replication Complex Component 1 (CTC1), and Protection of telomeres protein 1 (POT1) sequences from eukaryotes. Homologous RPA-1 from the archaebacteria *Methanocaldococcus jannaschii* (RPA-1-M_jan), *Methanobrevibacter woesei* (RPA-1-M_woe) and *Thermococcus litoralis* (RPA-1-T_lit) were used as outgroups. The branch color corresponds to each specific sequence. Posterior probability values are indicated in the internodes. Posterior probability values (in 0–1 scale) and bootstrap values (1000 replicates; in 0–100 scale) are shown below and above the internodes, respectively). The identificaSupplementary Table 2. ** indicates RPA-1 sequences from apicomplexan with ~55 kDa.

heterodimer, as well as by the heterotrimeric RPA complex [63,72–74]. Both human RPA heterotrimer and the CST complex share the same mode of disassociation and re-association of individual OB-folds to unfold telomeric G-quadruplex. These similarities probably help them to prevent the formation of G-quadruplex, which will otherwise impede DNA synthesis at the telomeres and elsewhere in the genome [63,74]. In contrast, POT1 binds stably from 3′ to 5′ the telomeric DNA folded as G-quadruplex using one OB-fold at a time. It can also unfold, in a stepwise manner, the human telomeric DNA folded into a quadruplex structure. On the other hand, POT1 associated with TPP1, induces in a sliding motion (back and forth), continuous folding, and unfolding of G-quadruplex structures, allowing telomerase to elongate telomeric DNA primers that have a specific 3′ end sequence [73].

Together, these results show that LaRPA-1 shares functional and structural similarities with classical OB-fold containing telomeric proteins due to its ability to protect telomeric G-rich ssDNA from the exonucleolytic activity and to unfold telomeric G-quadruplex structures. All these data support the idea that RPA-1 could be a TEBP and a substitute to the classical OB-fold containing telomeric proteins at trypanosomatids telomeres.

4.2. Two possible evolutionary landscapes of trypanosomatids RPA-1

The study of telomeric proteins from earlier-divergent eukaryotes can give insights about the evolution of telomeres in eukaryotes, especially in the context of the absence of classical OB-fold containing telomeric protein homologs at trypanosomatids telomeres. For example, it is well known that CST is likely to share a common origin with RPA, being respectively CTC1, STN1, and TEN1 paralogous of RPA-1, RPA2, and RPA3 [17,18,75]. Also, a recent study proposed that CST evolved from an archaeal-eukaryal RPA complex and may have, in turn, given birth to the POT1-TPP1 complex [32]. Thus, based on these observations, we studied the phylogenetic relationships between RPA-1 from trypanosomatids with RPA-1 from other eukaryotes as well as with the classical OB-fold containing telomeric proteins POT1 and CTC1. Here we show for the first time the analyses of all the OB-fold containing proteins using phylogenetic constructions trees, and we compared our results with several previously performed evolutionary propositions that were based mainly on structural and functional similarities [18,32,76].

The RPA-1 phylogenetic tree showed that RPA-1 from trypanosomatids occupies the most basal clade of the tree (Fig. 5A) and that RPA-1 from plants nests in a monophyletic clade with apicomplexan and ciliates RPA-1 with moderate statistic support. Similar results regarding the nest between plants and ciliates in a monophyletic clade were previously obtained [77]; and, interestingly, the apicomplexan presents different forms of RPA-1 with great differences in sequence and size (54 to 135 kDa) in comparison to the conserved RPA-1 (~ 70 kDa) from higher eukaryotes [77]. The smaller forms of apicomplexan RPA-1 (~54 kDa) share similar molecular mass with trypanosomatids RPA-1 (~52 kDa) and also do not present the 70 N domain [77]. In our phylogenetic tree (Fig. 5A) and as previously shown [77], apicomplexan RPA-1 grouped based on their types rather than species probably due to the presence of separated clades of small and large proteins. The basal position of trypanosomatids RPA-1 in our phylogenetic tree (Fig. 5A) and the previous observation that they lack the 70 N domain, which is also absent in archaeal RPA-1 homologs outgroups [33], suggest that the 70 N domain should have arisen lately in the eukaryotic evolution. Indeed, an *in silico* homology search using blastp (data not shown), in the large forms of apicomplexan RPA-1 and ciliates RPA-1, did not indicate that they posses the 70 N domain; suggesting that the 70 N domain may have arisen in higher eukaryotes RPA-1. Remarkably, even homologous RPA-1 sequences from plants presented in the phylogenetic tree display different sizes (57 to 72 kDa) and only the N-terminal of *O. sativa* and *A. thaliana* RPA-1 present similarities with the mammalian and fungi RPA-1 70 N domain (e-values e-8).

The basal position of trypanosomatids RPA-1 in our phylogenetic tree (Fig. 5A) and the previous appointment that CST components and POT1-TPP1 are exclusively found in eukaryotes may suggest that CST, POT1, and TPP1 could have arisen later than trypanosomatids in eukaryotic evolution. Despite little information about RPA-1 from other basal eukaryotes, this observation suggests that, since trypanosomatids RPA-1 have telomeric functions, they could be one of the most ancient OB-fold containing telomeric proteins. In agreement, there is a recent hypothesis [32] proposing that the origin of the most derivative and specialized telomeric proteins occurred after the duplication of primitive RPA genes followed by neofunctionalization.

To deeply explore the evolutionary issues presented by LaRPA-1, a phylogenetic tree containing RPA-1, POT1 and CTC1 was constructed. The results showed that trypanosomatids RPA-1 sequences nest in a well-supported clade with POT1 and CTC1 from higher eukaryotes (Fig. 5B), indicating that they share the same common ancestor, which may explain the telomeric functions previously identified in LaRPA-1. Furthermore, this phylogenetic tree also suggests that trypanosomatids RPA-1 are orthologues of POT1 and CTC1 and paralogues of RPA-1 from other eukaryotes. This result emphasizes the assertion that the 70 N domain must have arisen only in RPA-1 sequences from most derivative eukaryotes since it is also absent in the POT1/CTC1/trypanosomatids RPA-1 clade, but present in the metazoan/fungi RPA-1 clade (Fig. 5B). On the other hand, this result brings an antagonistic point concerning what was observed in the previous RPA-1 phylogenetic tree since it suggests that POT1 and CTC1 were already present among earlier eukaryotes and that they disappeared in the trypanosomatids group. Further studies with other RPA complex components (RPA2 and RPA3) from trypanosomatids and from other primitive eukaryotes as well as with other components of CST complex (STN1 and TEN1) may provide new insights about the evolutionary path of all OB-fold containing telomeric proteins. Finally, it is important to emphasize that some organisms may have adopted a proper evolutive way for telomere end protection by OB-fold containing proteins. For example, the proteins TEB1, TEB2 and TEB3, and p75, p45, and p19 from *Tetrahymena thermophila* are paralogues of RPA and the CST complexes, respectively, and may have arisen from RPA duplications [20,21]. Similarly, CDC13 and TEBP α and β, are proteins from organisms that do not present the shelterin complex (*S. cerevisiae* and *Oxytricha nova*, respectively), which can also exert telomeric protective functions, to regulate telomerase activity and specifically bind the telomeric G-rich ssDNA [14,22].

In summary, we present two opposite evolutionary landscapes of RPA-1 from trypanosomatids that could reflect upon the OB-fold containing telomeric proteins evolution: i) RPA-1 from trypanosomatids as one of the most ancient OB-fold containing telomeric proteins. In this context, telomeric functions of LaRPA-1 demonstrated here are primitive characteristics of this protein and may give rise to other OB-fold containing telomeric proteins. In the light of recent proposition about the evolution of OB-fold containing telomeric proteins [32], the most derivative and specialized telomeric proteins seemed to occur after the duplication of primitive RPA genes, such as trypanosomatids RPA, followed by neofunctionalization; ii) trypanosomatids RPA-1 as orthologues of POT1 and CTC1 and paralogues of RPA-1 from other eukaryotes, suggest that POT1 and CTC1 were already present among eukaryotes and that they have disappeared in the trypanosomatids group. In this context, the presence of RPA-1 at telomeres performing similar functions of OB-fold containing telomeric proteins replaces the absence of these proteins at parasites telomeres. A recent hypothesis [32] suggests that the telomeric functions of RPA-1 would be already present in primitive RPA-1. In this context, the telomeric functions remain conserved at RPA-1 from trypanosomatids, enabling trypanosomatids RPA-1 to act at telomeres due to the possible absence of other OB-fold containing telomeric proteins.

5. Conclusions

The present work suggests the absence of classical TEBPs at trypanosomatids telomeres and a challenge to overcome. We point out that the OB-fold containing protein RPA-1, represented by the *Leishmania amazonensis* RPA-1, can be a strong candidate and a substitute for the classical TEBPs at trypanosomatids telomeres. We have demonstrated that LaRPA-1 besides sharing structural features with higher eukaryotes TEBPs, can protect telomeric G-rich ssDNA from 3′-5′ exonucleolytic attack and unfold telomeric G-quadruplex structures. In the context of LaRPA-1 as a TEBP, we present two opposing evolutionary proposals of RPA-1 from trypanosomatids that could reflect upon the evolution of OB-fold containing telomeric proteins. Although they are conflicting, both evolutionary landscapes support that RPA-1 is a potential trypanosomatid TEBP and can guide further evolutionary analysis of these proteins in future works, especially in other earlier-divergent eukaryotes.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2020.129607.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by São Paulo Research Foundation, FAPESP (grants 2015/18641-5 and 2018/04375-2 to MICANO and 2015/17286-0 to MRMF), and by the «Muséum National d'Histoire Naturelle» (MNHN), the «Centre National de la Recherche Scientifique» (CNRS) and the «Institut National de la Santé et de la Recherche Médicale» (Inserm) to CS. MRMF is also supported by the National Council for Scientific and Technological Development, CNPq (grant 305300/2006-7), Brazil. CAHF is a postdoctoral fellow from the Coordination for the Improvement of Higher Education Personnel, CAPES (grant 88881, 171895/2018-01)-Brazil, EGOM is a FAPESP postdoctoral fellow (grant 2019/11496-3). GAS and MRV were CAPES master degree fellows, MAV was a CNPq doctoral fellow, and VLS is an undergraduate FAPESP fellow (grant 2018/0395-6). JC receives financial support from the doctoral school “ED 227” of the MNHN and AV was supported by a CNRS funding (professional degree). We would like to thanks Dr. Loïc Ponger for a helpful discussion about the phylogenetic trees and Amanda C. Marcolino for the graphical abstract design and Mark Ewusi Shiburah for a careful English editing and proofreading.

References

[1] M.A. Giardini, M. Segatto, M.S. Da Silva, V.S. Nunes, M.I.N. Cano, Telomere and telomerase biology, Prog. Mol. Biol. Transl. Sci. (2014), https://doi.org/10.1016/B978-0-12-397898-1.00001-3.

[2] S.W.R.L. Chan, E.H. Blackburn, Telomeres and telomerase, Philos. Trans. R. Soc. Lond. B 359 (2004) 109–121, https://doi.org/10.1098/rstb.2003.1370.

[3] W. Palm, T. de Lange, How Shelterin protects mammalian telomeres, Annu. Rev. Genet. 42 (2008) 301–334, https://doi.org/10.1146/annurev.genet.41.110306.130350.

[4] J.M. Dewar, D. Lydall, Similarities and differences between “uncapped” telomeres and DNA double-strand breaks, Chromosoma 12 (2012) 117–130, https://doi.org/10.1007/s00412-011-0357-2.

[5] C. Rice, E. Skordalakes, Structure and function of the telomeric CST complex, Comput. Struct. Biotechnol. J. 14 (2016) 161–167, https://doi.org/10.1016/j.csbj.2016.04.002.

[6] T. De Lange, Shelterin: the protein complex that shapes and safeguards human telomeres, Genes Dev. 19 (2005) 2100–2110, https://doi.org/10.1101/gad.1346005.

[7] F. Wang, E.R. Podell, A.J. Zaug, Y. Yang, P. Baciu, T.R. Cech, M. Lei, The POT1-TPP1 telomere complex is a telomerase processivity factor, Nature 445 (2007) 506–510, https://doi.org/10.1038/nature05454.

[8] K.H. Choi, A.S. Lakamp-Hawley, C. Kolar, Y. Yan, G.E.O. Borgstahl, M.M. Ouellette, The OB-fold domain 1 of human POT1 recognizes both telomeric and non-telomeric DNA motifs, Biochimie 115 (2015) 117–127, https://doi.org/10.1016/j.biochi.2015.04.015.

[9] C.A. Armstrong, V. Moiseeva, L.C. Collopy, S.R. Pearson, T.R. Ullah, S.T. Xi, J. Martin, S. Subramaniam, S. Marelli, H. Amelina, K. Tomita, Fission yeast Ccq1 is a modulator of telomerase activity, Nucleic Acids Res. 46 (2018) 704–716, https://doi.org/10.1093/nar/gkx1223.

[10] M. Lei, E.R. Podell, T.R. Cech, Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection, Nat. Struct. Mol. Biol. 11 (2004) 1223–1229, https://doi.org/10.1038/nsmb867.

[11] D. Hockemeyer, A.J. Sfeir, J.W. Shay, W.E. Wright, T. De Lange, POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end, EMBO J. 24 (2005) 2667–2678, https://doi.org/10.1038/sj.emboj.7600733.

[12] E.L. Denchi, T. De Lange, Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1, Nature 448 (2007) 1068–1071, https://doi.org/10.1038/nature06065.

[13] B. Garvik, M. Carson, L. Hartwell, Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint, Mol. Cell. Biol. 15 (1995) 6128–6138, https://doi.org/10.1128/mcb.15.11.6128.

[14] J.-J. Lin, V.A. Zakian, The Saccharomyces CDC13 protein is a single-strand TG1-3 telomeric DNA-binding protein in vitro that affects telomere behavior in vivo, Proc. Natl. Acad. Sci. 93 (1996) 13760–13765, https://doi.org/10.1073/pnas.93.24.13760.

[15] J. Nandakumar, T.R. Cech, Finding the end: recruitment of telomerase to telomeres, Nat. Rev. Mol. Cell Biol. 14 (2013) 69–82, https://doi.org/10.1038/nrm3505.

[16] B. Meier, L. Driller, S. Jaklin, H.M. Feldmann, New function of CDC13 in positive telomere length regulation, Mol. Cell. Biol. 21 (2001) 4233–4245, https://doi.org/10.1128/mcb.21.13.4233-4245.2001.

[17] Y. Miyake, M. Nakamura, A. Nabetani, S. Shimamura, M. Tamura, S. Yonehara, M. Saito, F. Ishikawa, RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway, Mol. Cell 36 (2009) 193–206, https://doi.org/10.1016/j.molcel.2009.08.009.

[18] C.M. Price, K.A. Boltz, M.F. Chaiken, J.A. Stewart, M.A. Beilstein, D.E. Shippen, Evolution of CST function in telomere maintenance, Cell Cycle 9 (2010) 3157–3165, https://doi.org/10.4161/cc.9.16.12547.

[19] J.A. Stewart, F. Wang, M.F. Chaiken, C. Kasbek, P.D. Chastain, W.E. Wright, C.M. Price, Human CST promotes telomere duplex replication and general replication restart after fork stalling, EMBO J. 31 (2012) 3537–3549, https://doi.org/10.1038/emboj.2012.215.

[20] J. Jiang, H. Chan, D.D. Cash, E.J. Miracco, R.R.O. Loo, H.E. Upton, D. Cascio, R.O.B. Johnson, K. Collins, J.A. Loo, Z.H. Zhou, J. Feigon, Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions, Science 350 (2015) aab4070, , https://doi.org/10.1126/science.aab4070.

[21] H.E. Upton, H. Chan, J. Feigon, K. Collins, Shared subunits of *Tetrahymena* telomerase holoenzyme and replication protein a have different functions in different cellular complexes, J. Biol. Chem. 292 (2017) 217–228, https://doi.org/10.1074/jbc.M116.763664.

[22] M.P. Horvath, V.L. Schweiker, J.M. Bevilacqua, J.A. Ruggles, S.C. Schultz, Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA, Cell 95 (1998) 963–974, https://doi.org/10.1016/S0092-8674(00)81720-1.

[23] R.M. Mitton-Fry, E.M. Anderson, T.R. Hughes, V. Lundblad, D.S. Wuttke, Conserved structure for single-stranded telomeric DNA recognition, Science 296 (2002) 145–147, https://doi.org/10.1126/science.1068799.

[24] M.J. Giraud-Panis, S. Pisano, A. Poulet, M.H. Le Du, E. Gilson, Structural identity of telomeric complexes, FEBS Lett. 584 (2010) 3785–3799, https://doi.org/10.1016/j.febslet.2010.08.004.

[25] K.A. Lewis, D.S. Wuttke, Telomerase and telomere-associated proteins: structural insights into mechanism and evolution, Structure 20 (2012) 28–39, https://doi.org/10.1016/j.str.2011.10.017.

[26] B.R. Linger, C.M. Price, Conservation of telomere protein complexes: shuffling through evolution, Crit. Rev. Biochem. Mol. Biol. 44 (2009) 434–446, https://doi.org/10.3109/10409230903307329.

[27] E. Bochkareva, S. Korolev, S.P. Lees-Miller, A. Bochkarev, Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA, EMBO J. 21 (2002) 1855–1863, https://doi.org/10.1093/emboj/21.7.1855.

[28] M.S. Wold, Replication protein a: a Heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism, Annu. Rev. Biochem. 66 (1997) 61–92, https://doi.org/10.1146/annurev.biochem.66.1.61.

[29] J. Smith, H. Zou, R. Rothstein, Characterization of genetic interactions with RFA1: the role of RPA in DNA replication and telomere maintenance, Biochimie 82 (2000) 71–78, https://doi.org/10.1016/S0300-9084(00)00183-8.

[30] V. Schramke, P. Luciano, V. Brevet, S. Guillot, Y. Corda, M.P. Longhese, E. Gilson, V. Géli, RPA regulates telomerase action by providing Est1p access to chromosome ends, Nat. Genet. 36 (2004) 46–54, https://doi.org/10.1038/ng1284.

[31] P. Luciano, S. Coulon, V. Faure, Y. Corda, J. Bos, S.J. Brill, E. Gilson, M.N. Simon, V. Géli, RPA facilitates telomerase activity at chromosome ends in budding and fission yeasts, EMBO J. 18 (2012) 2034–2046, https://doi.org/10.1038/emboj.2012.40.

[32] N.F. Lue, Evolving linear chromosomes and telomeres: a C-Strand-centric view, Trends Biochem. Sci. 43 (2018) 314–326, https://doi.org/10.1016/j.tibs.2018.02.008.

[33] F. Chédin, E.M. Seitz, S.C. Kowalczykowski, Novel homologs of replication protein a in Archaea: implications for the evolution of ssDNA-binding proteins, Trends

Biochem. Sci. 23 (1998) 273–277, https://doi.org/10.1016/S0968-0004(98)01243-2.
[34] B. Li, A. Espinal, G.A.M. Cross, Trypanosome telomeres are protected by a homologue of mammalian TRF2, Mol. Cell. Biol. 25 (2005) 5011–5021, https://doi.org/10.1128/mcb.25.12.5011-5021.2005.
[35] X. Yang, L.M. Figueiredo, A. Espinal, E. Okubo, B. Li, RAP1 is essential for silencing Telomeric variant surface glycoprotein genes in *Trypanosoma brucei*, Cell 137 (2009) 99–109, https://doi.org/10.1016/j.cell.2009.01.037.
[36] M.S. da Silva, A.M. Perez, R.C.V. da Silveira, C.E. de Moraes, J.L. Siqueira-Neto, L.H. Freitas-Junior, M.I.N. Cano, The *Leishmania amazonensis* TRF (TTAGGG repeat binding factor) homologue binds and co-localizes with telomeres, BMC Microbiol. 10 (2010) 136, https://doi.org/10.1186/1471-2180-10-136.
[37] S.E. Jehi, F. Wu, B. Li, Trypanosoma brucei TIF2 suppresses VSG switching by maintaining subtelomere integrity, Cell Res. 24 (2014) 870–885, https://doi.org/10.1038/cr.2014.60.
[38] R.S. Pavani, C. Fernandes, A.M. Perez, E.J.R. Vasconcelos, J.L. Siqueira-Neto, M.R. Fontes, M.I.N. Cano, RPA-1 from *Leishmania amazonensis* (LaRPA-1) structurally differs from other eukaryote RPA-1 and interacts with telomeric DNA via its N-terminal OB-fold domain, FEBS Lett. 588 (2014) 4740–4748, https://doi.org/10.1016/j.febslet.2014.11.005.
[39] M.F. Fernández, R.R. Castellari, F.F. Conte, F.C. Gozzo, A.A. Sabino, H. Pinheiro, J.C. Novello, M.N. Eberlin, M.I.N. Cano, Identification of three proteins that associate in vitro with the *Leishmania* (Leishmania) *amazonensis* G-rich telomeric strand, Eur. J. Biochem. 271 (2004) 3050–3063, https://doi.org/10.1111/j.1432-1033.2004.04237.x.
[40] R.D.C.V. Da Silveira, M.S. Da Silva, V.S. Nunes, A.M. Perez, M.I.N. Cano, The natural absence of RPA1N domain did not impair *Leishmania amazonensis* RPA-1 participation in DNA damage response and telomere protection, Parasitology 140 (2013) 547–559, https://doi.org/10.1017/S0031182012002028.
[41] M.S. da Silva, M. Segatto, R.S. Pavani, F. Gutierrez-Rodrigues, V.S. da Bispo, M.H.G. de Medeiros, R.T. Calado, M.C. Elias, M.I.N. Cano, Consequences of acute oxidative stress in *Leishmania amazonensis*: from telomere shortening to the selection of the fittest parasites, Biochim. Biophys. Acta, Mol. Cell Res. 1864 (2017) 138–150, https://doi.org/10.1016/j.bbamcr.2016.11.001.
[42] J.L.S. Neto, C.B.B. Lira, M.A. Giardini, L. Khater, A.M. Perez, L.A. Peroni, J.R.R. dos Reis, L.H. Freitas-Junior, C.H.I. Ramos, M.I.N. Cano, *Leishmania* replication protein A-1 binds in vivo single-stranded telomeric DNA, Biochem. Biophys. Res. Commun. 358 (2007) 417–423, https://doi.org/10.1016/j.bbrc.2007.04.144.
[43] A. Bochkarev, R.A. Pfuetzner, A.M. Edwards, L. Frappier, Structure of the single-stranded-DNA-binding domain of replication protein a bound to DNA, Nature 385 (1997) 176–181, https://doi.org/10.1038/385176a0.
[44] J. Fan, N.P. Pavletich, Structure and conformational change of a replication protein a heterotrimer bound to ssDNA, Genes Dev. 26 (2012) 2337–2347, https://doi.org/10.1101/gad.194787.112.
[45] R.S. Pavani, M.O. Vitarelli, C.A.H. Fernandes, F.F. Mattioli, M. Morone, M.C. Menezes, M.R.M. Fontes, M.I.N. Cano, M.C. Elias, Replication protein A-1 has a preference for the Telomeric G-rich sequence in *Trypanosoma cruzi*, J. Eukaryot. Microbiol. 65 (2018) 345–356, https://doi.org/10.1111/jeu.12478.
[46] C.B.B. Lira, J.L. Siqueira Neto, M.A. Giardini, F.V. Winck, C.H.I. Ramos, M.I.N. Cano, LaRbp38: a *Leishmania amazonensis* protein that binds nuclear and kinetoplast DNAs, Biochem. Biophys. Res. Commun. 358 (2007) 854–860, https://doi.org/10.1016/j.bbrc.2007.05.005.
[47] D. Lu, A.R. Myers, N.P. George, J.L. Keck, Mechanism of exonuclease i stimulation by the single-stranded DNA-binding protein, Nucleic Acids Res. 39 (2011) 3545–6536. d^{1}oi:https://doi.org/10.1093/nar/gkr315.
[48] S.R. Eddy, Profile hidden Markov models, Bioinformatics 9 (1998) 755–763, https://doi.org/10.1093/bioinformatics/14.9.755.
[49] S. El-Gebali, J. Mistry, A. Bateman, S.R. Eddy, A. Luciani, S.C. Potter, M. Qureshi, L.J. Richardson, G.A. Salazar, A. Smart, E.L.L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S.C.E. Tosatto, R.D. Finn, The Pfam protein families database in 2019, Nucleic Acids Res. 47 (2019), https://doi.org/10.1093/nar/gky995 D427-D432.
[50] G.C.P. Van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. Van Dijk, S.J. de Vries, A.M.J.J. Bonvin, The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes, J. Mol. Biol. 428 (2016) 720–725, https://doi.org/10.1016/j.jmb.2015.09.014.
[51] M. van Dijk, A.M.J.J. Bonvin, A protein-DNA docking benchmark, Nucleic Acids Res. 36 (2008) e88, , https://doi.org/10.1093/nar/gkn386.
[52] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.
[53] F. Ronquist, M. Teslenko, P. Van Der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, J.P. Huelsenbeck, Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space, Syst. Biol. 61 (2012) 539–542, https://doi.org/10.1093/sysbio/sys029.
[54] M.A. Miller, W. Pfeiffer, T. Schwartz, The CIPRES science gateway: Enabling high-impact science for phylogenetics researchers with limited resources, Proc. 1st Conf. Extrem. Sci. Eng. Discov. Environ. Bridg. From Extrem. To Campus beyond, 2012, https://doi.org/10.1145/2335755.2335836.
[55] S. Guindon, J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (2010) 307–321, https://doi.org/10.1093/sysbio/syq010.
[56] W.P. Maddison, D. Maddison, Mesquite: A Modular System for Evolutionary Analysis, Version 3.51, 2018. http://www.mesquiteproject.org.
[57] C. Chen, P. Gu, J. Wu, X. Chen, S. Niu, H. Sun, L. Wu, N. Li, J. Peng, S. Shi, C. Fan, M. Huang, C.C.L. Wong, Q. Gong, C. Kumar-Sinha, R. Zhang, L. Pusztai, R. Rai, S. Chang, M. Lei, Structural insights into POT1-TPP1 interaction and POT1 C-terminal mutations in human cancer, Nat. Commun. 14929 (2017), https://doi.org/10.1038/ncomms14929.
[58] C. Rice, P.K. Shastrula, A.V. Kossenkov, R. Hills, D.M. Baird, L.C. Showe, T. Doukov, S. Janicki, E. Skordalakes, Structural and functional analysis of the human POT1-TPP1 telomeric complex, Nat. Commun. 14928 (2017), https://doi.org/10.1038/ncomms14928.
[59] F.F. Conte, M.I.N. Cano, Genomic organization of telomeric and subtelomeric sequences of *Leishmania* (Leishmania) *amazonensis*, Int. J. Parasitol. 35 (2005) 1435–1443, https://doi.org/10.1016/j.ijpara.2005.05.011.
[60] E.G.O. Morea, M.A. Viviescas, C.A.H. Fernandes, F.F. Matioli, C.B.B. Lira, M.F. Fernandez, B.S. Moraes, M.S. da Silva, C.B. Storti, M.R.M. Fontes, M.I.N. Cano, A calmodulin-like protein (LCALA) is a new *Leishmania amazonensis* candidate for telomere end-binding protein, Biochim. Biophys. Acta, Gen. Subj. 1861 (2017) 2583–2597, https://doi.org/10.1016/j.bbagen.2017.08.011.
[61] M. Lei, E.R. Podell, P. Baumann, T.R. Cech, DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA, Nature 426 (2003) 198–203, https://doi.org/10.1038/nature02092.
[62] R.S. Brody, K.G. Doherty, P.D. Zimmerman, Processivity and kinetics of the reaction of exonuclease I from Escherichia coli with polydeoxyribonucleotides, J. Biol. Chem. 261 (1986) 7136–7143.
[63] T.R. Salas, I. Petruseva, O. Lavrik, A. Bourdoncle, J.L. Mergny, A. Favre, C. Saintomé, Human replication protein a unfolds telomeric G-quadruplexes, Nucleic Acids Res. 34 (2006) 4857–4865, https://doi.org/10.1093/nar/gkl564.
[64] M. Vorlíčková, J. Chládková, I. Kejnovská, M. Fialová, J. Kypr, Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats, Nucleic Acids Res. 33 (2005) 5851–5860, https://doi.org/10.1093/nar/gki898.
[65] A.H. Lopes, T. Souto-Padrón, F.A. Dias, M.T. Gomes, G.C. Rodrigues, L.T. Zimmermann, T.L. Alves e Silva, A.B. Vermelho, Trypanosomatids: Odd organisms, devastating diseases, Open Parasitol. J. 4 (2010) 30–59, https://doi.org/10.2174/1874421401004010030.
[66] V. Sekaran, J. Soares, M.B. Jarstfer, Telomere maintenance as a target for drug discovery, J. Med. Chem. 57 (2014) 521–538, https://doi.org/10.1021/jm400528t.
[67] C.A.H. Fernandes, A.M. Perez, A.C. Barros, T.R. Dreyer, M.S. da Silva, E.G.O. Morea, M.R.M. Fontes, M.I.N. Cano, Dual cellular localization of the *Leishmania amazonensis* Rbp38 (LaRbp38) explains its affinity for telomeric and mitochondrial DNA, Biochimie 162 (2019) 15–25, https://doi.org/10.1016/j.biochi.2019.03.017.
[68] T.R. Hughes, R.G. Weilbaecher, M. Walterscheid, V. Lundblad, Identification of the single-strand telomeric DNA binding domain of the *Saccharomyces cerevisiae* Cdc13 protein, Proc. Natl. Acad. Sci. 97 (2002) 6457–6462, https://doi.org/10.1073/pnas.97.12.6457.
[69] R.S. Pavani, M.S. da Silva, C.A.H. Fernandes, F.S. Morini, C.B. Araujo, M.R.M. de Fontes, O.A. Sant'Anna, C.R. Machado, M.I. Cano, S.P. Fragoso, M.C. Elias, Replication protein A presents canonical functions and is also involved in the differentiation capacity of *Trypanosoma cruzi*, PLoS Negl. Trop. Dis. 10 (2016) e0005181, , https://doi.org/10.1371/journal.pntd.0005181.
[70] D. Loayza, H. Parsons, J. Donigian, K. Hoke, T. De Lange, DNA binding features of human POT1: a nonamer 5′-TAGGGTTAG-3′ minimal binding site, sequence specificity, and internal binding to multimeric sites, J. Biol. Chem. 279 (2004) 13241–13248, https://doi.org/10.1074/jbc.M312309200.
[71] R.E. Verdun, J. Karlseder, Replication and protection of telomeres, Nature 447 (2007) 924–931, https://doi.org/10.1038/nature05976.
[72] A.J. Zaug, E.R. Podell, T.R. Cech, Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro, Proc. Natl. Acad. Sci. 102 (2005) 10864–10869, https://doi.org/10.1073/pnas.0504744102.
[73] H. Hwang, N. Buncher, P.L. Opresko, S. Myong, POT1-TPP1 regulates telomeric overhang structural dynamics, Structure 20 (2012) 1872–1880, https://doi.org/10.1016/j.str.2012.08.018.
[74] A. Bhattacharjee, Y. Wang, J. Diao, C.M. Price, Dynamic DNA binding, junction recognition and G4 melting activity underlie the telomeric and genome-wide roles of human CST, Nucleic Acids Res. 45 (2017) 12311–12324, https://doi.org/10.1093/nar/gkx878.
[75] J. Sun, E.Y. Yu, Y. Yang, L.A. Confer, S.H. Sun, K. Wan, N.F. Lue, M. Lei, Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres, Genes Dev. 23 (2009) 2900–2914, https://doi.org/10.1101/gad.1851909.
[76] T. De Lange, A loopy view of telomere evolution, Front. Genet. 6 (2015) 321, https://doi.org/10.3389/fgene.2015.00321.
[77] S.D. Rider, X. Cai, W.J. Sullivan, A.T. Smith, J. Radke, M. White, G. Zhu, The protozoan parasite *Cryptosporidium parvum* possesses two functionally and evolutionarily divergent replication protein a large subunits, J. Biol. Chem. 280 (2005) 31460–31469, https://doi.org/10.1074/jbc.M504466200.
[78] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery, J. Chem. Inf. Model. 51 (2011) 2778–2786, https://doi.org/10.1021/ci200227u.

Chapitre 7

Annexe technique : purification des protéines

7.1 Vecteurs et cultures cellulaires

7.1.1 Les vecteurs *POT1* et *TPP1*

Les protéines recombinantes *POT1* et *TTP1* humaines ont été exprimées en utilisant le système vecteur d'expression baculovirus (BEVS). Les bacmides (pFastBac HT-A) nous ont été généreusement donnés par l'équipe du Dr. Ming Lei. La protéine *POT1*(1-634 résidus) est taguée avec une étiquette 6 histidines (6His-tag) en N-term, tandis que la protéine *TTP1* (forme tronquée contenant les résidus 87-334) avec une étiquette GST en N-term. Le site de coupure 3C (LGVLFGQ/G) a été introduit afin de pouvoir supprimer les étiquettes des protéines, à l'aide de la PreScission protéase.

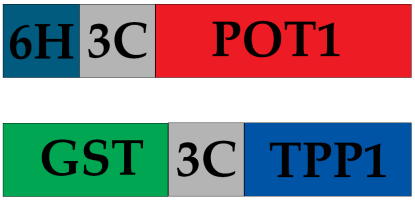


FIGURE 7.1 – **Représentation schématique des protéines recombinantes taguées *POT1* et *TPP1***

7.1.2 Cultures cellulaires

La transfection des virus, la co-infection des cellules d'insecte et les cultures cellulaires ont été réalisées par le Dr Nathalie Troffer-Charlier à l'IGBMC-CERBM de Strasbourg. Les protéines *POT1* et *TPP1* sont co-exprimées par le baculovirus dans des cellules sf9, cellules d'insectes *Spodoptera frugiperda*.

7.2 Étapes de purification des protéines et contrôles qualités

Les protocoles de purification des protéines sont détaillés dans les parties Matériels et Méthodes des différents articles. L'objectif de cette partie est de présenter les grandes lignes des étapes de purifications et de mettre en avant les contrôles qualité de celles-ci.

7.2.1 Le complexe *POT1-TPP1*

Le protocole de purification détaillé du complexe *POT1-TPP1* est décrit dans la partie Matériels et Méthodes de l'article 1 (Chapitre 2).

Étape 1 : Lyse des cellules

Nous partons de culots de cultures cellulaires contenant les protéines *POT1-TPP1*. Pour réaliser la lyse des cellules, nous re-suspendons les culots cellulaires dans un tampon Lyse (Tris HCl pH 7.8 25mM, NaCl 300mM, DTT 2mM) contenant un cocktail d'inhibiteurs de protéase, à 4 °C et sous agitation, puis nous effectuons une étape de sonication pendant 20 min à 60% (1 sec ON / 1 sec OFF) à 4 °C. Pour éliminer les débris cellulaires, nous effectuons une centrifugation à 17.000 rpm pendant 1h15 à 4 °C.

Étape 2 : Colonne His-Select

La protéine *POT1* possèdant un tag 6 histines en N-term, nous réalisons une étape de purification par chromatographie d'affinité à l'aide d'une colonne de nickel (colonne His-Select). Pour ce faire, nous chargeons l'éluat obtenu après centrifugation sur une colonne de His-Select préalablement équilibrée à l'aide du tampon de Lyse. Puis nous effectuons une étape de lavage à l'aide du tampon Wash Ni (Tris HCl pH 7.8 25mM, NaCl 150mM, DTT 1mM et Imidazole 20nM) avant de réaliser l'étape d'élution à l'aide du tampon Elution Ni (Tris HCl pH 7.8 25mM, NaCl 150mM, DTT 1mM et Imidazole 300nM). Les étapes de lavage et d'élution sont suivies grâce au logiciel

BioRad et permet de suivre en temps réel l'élution de la protéine (Figure 7.2).

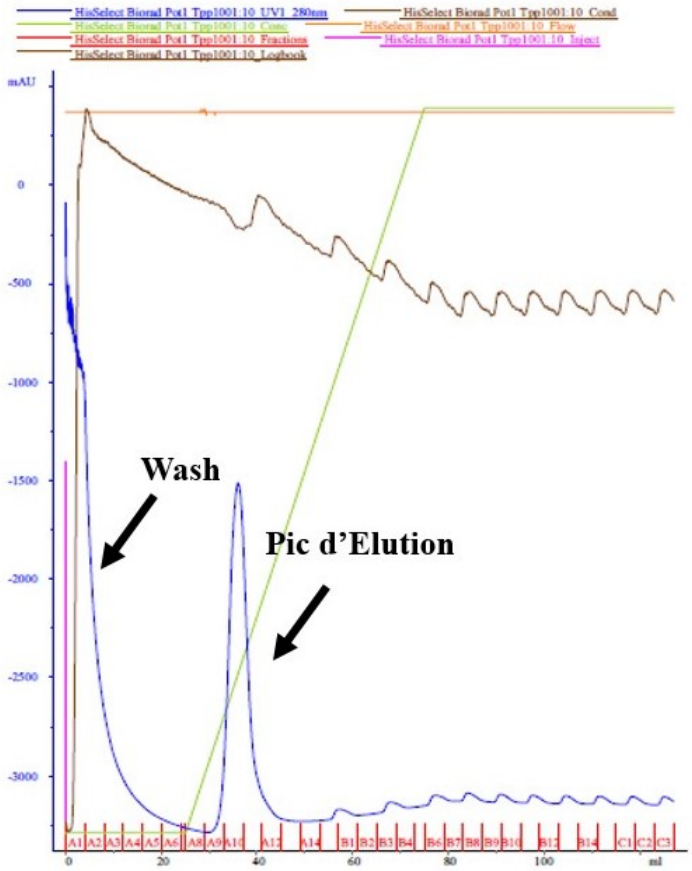


FIGURE 7.2 – **Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne His-Select.** les courbes correspondent : en bleu à l'absorbance à 280 nm, en marron à la conductivité, en vert à la concentration en tampon, en orange à la vitesse du flux.

Les différentes fractions récoltées sont analysées à l'aide d'un gel SDS-Page suivie de la coloration au bleu de Coomassie du gel (Figure 7.3). Nous observons la présence des deux protéines *POT1* et *TPP1* dans les fractions éluées.

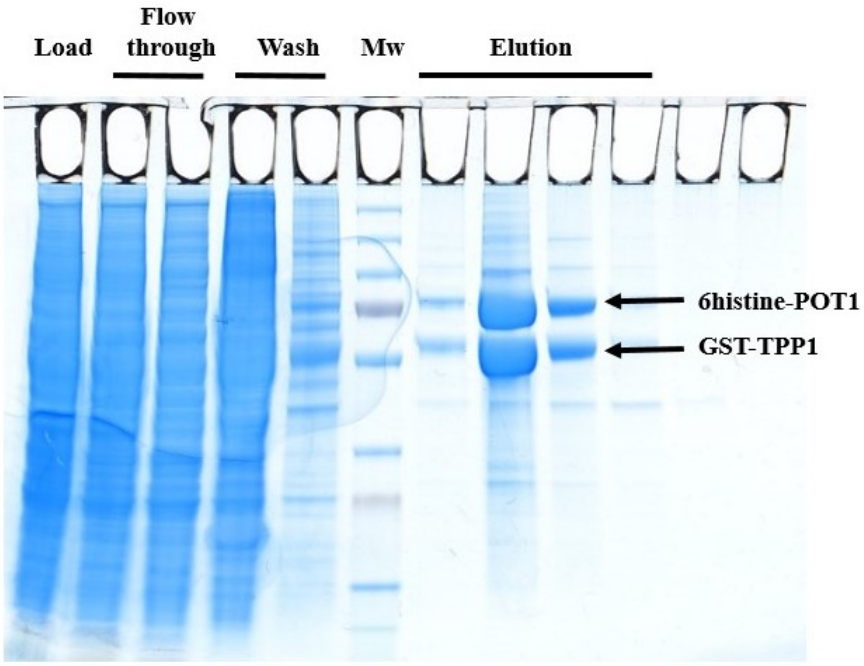


FIGURE 7.3 – **Gel SDS de la purification *POT1-TPP1* à l'issue de l'étape de chromatographie par affinité sur colonne His-Select.** Gel SDS-Page gradient acrylamide 8%-16% révélé par coloration au bleu de Coomassie.

Étape 3 : Colonne GSH

La protéine *TPP1* possédant un tag GST en N-term, nous réalisons ensuite une étape de purification par chromatographie d'affinité à l'aide d'une colonne gutathion réduit(GSH). Cette étape nous permet en autre d'éliminer la protéine *POT1* qui ne serait pas complexée à *TPP1*. Pour ce faire, nous chargeons les fractions obtenues lors de l'élution de la colonne His-Select sur une colonne GSH préalablement équilibrée à l'aide du tampon Elution Ni, puis nous effectuons un lavage de la colonne à l'aide du tampon Wash GST (Tris HCl pH 7.8 25mM, NaCl 75mM, DTT 1mM) avant de réaliser l'élution à l'aide du tampon Elution GST (Tris HCl pH 7.8 25mM, NaCl 150mM, DTT 1mM et GSH 300nM). Le chromatogramme présenté sur la figure 7.4 montre un pic d'absorbance élevé, dans la phase d'élution, qui correspond aux protéines *POT1* et *TPP1*, comme observé dans gel SDS (Figure 7.5).

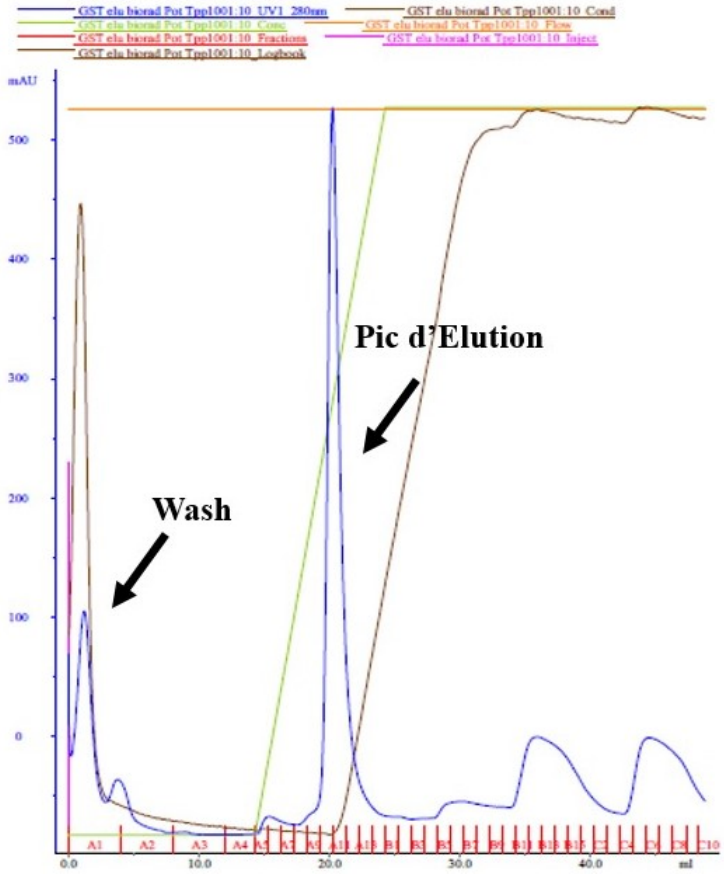


FIGURE 7.4 – **Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne GSH.** les courbes correspondent : en bleu à l'absorbance à 280 nm, en marron à la conductivité, en vert à la concentration en tampon, en orange la vitesse du flux.

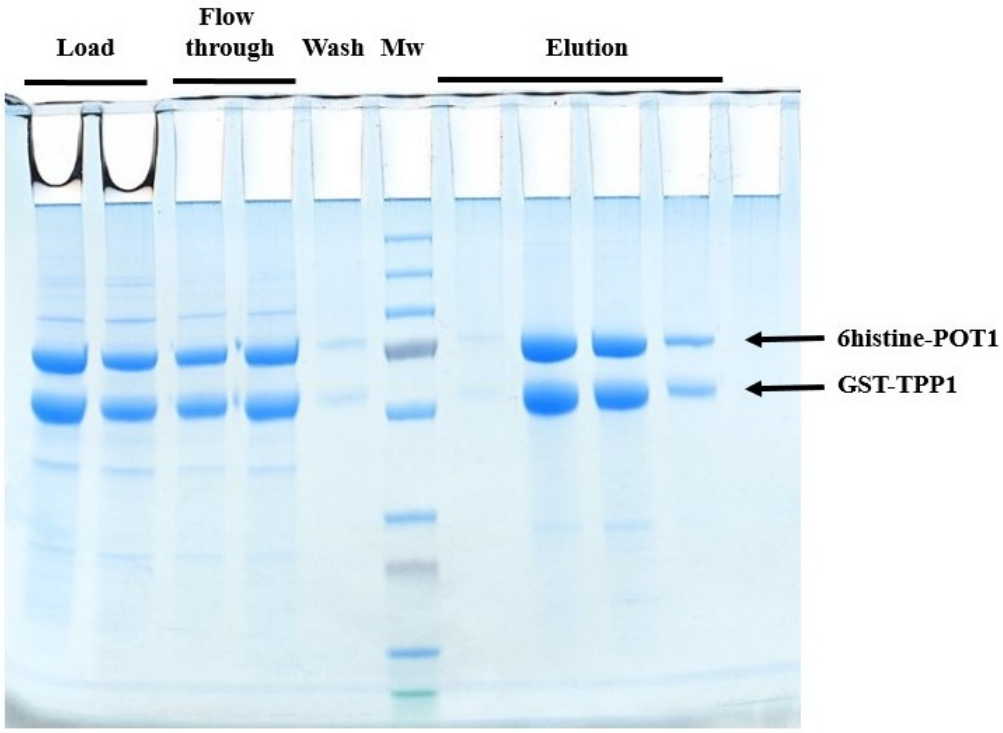


FIGURE 7.5 – **Gel SDS de la purification *POT1-TPP1* à l'issue de l'étape de chromatographie par affinité sur colonne GST** Gel SDS-Page gradient acrylamide 8%-16% révélé par coloration au bleu de Coomassie.

Étape 4 : Digestion des tags

Les protéines *POT1* et *TPP1* possèdent un site de coupure 3C en C-term de leur tag. Dans le but d'éliminer les tags, les fractions éluées de la colonne GST sont incubées avec l'enzyme PreScission sur la nuit à 4 °C. Nous réalisons un gel SDS-Page (Figure 7.6) afin de contrôler l'étape de coupure des tags. La protéine *POT1* a une masse moléculaire de 71 kDa, il est donc difficile de conclure si la coupure du tag 6 histines (environ 800 Dalton) a bien eu lieu. En revanche, la masse moléculaire du tag GST étant de 34 kDa et celle de *TPP1* de 35kDa, on peut clairement voir que la digestion du tag s'est faite correctement.

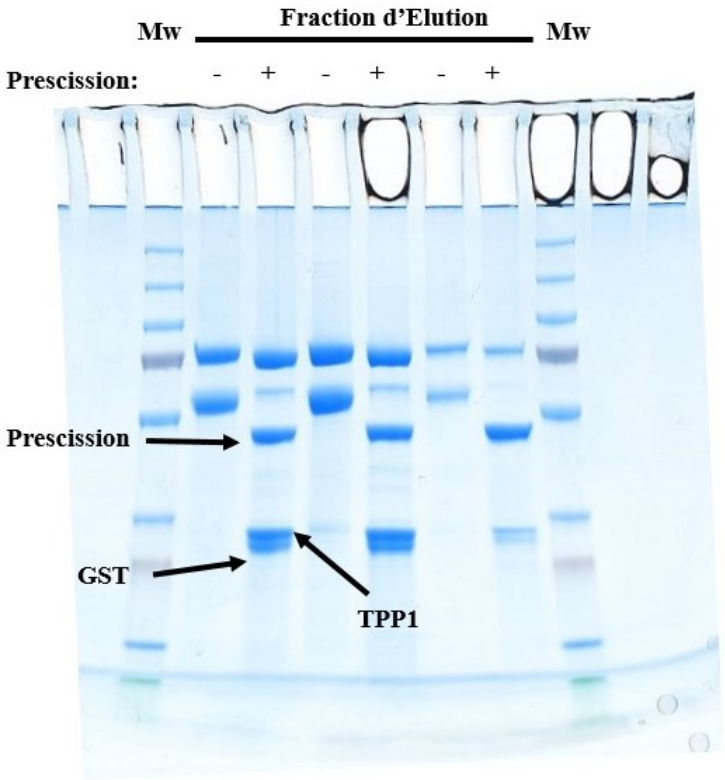


FIGURE 7.6 – **Gel SDS des fractions *POT1-TPP1* après l'étape de digestion par l'enzyme Prescission.** Gel SDS-Page gradient acrylamide 8%-16% révélé par coloration au bleu de Coomassie.

Étape 5 : Colonne Mono-Q

Nous réalisons une étape de purification à l'aide d'une résine échangeuse d'anions (colonne Mono-Q) dans le but d'éliminer la présence des tags, la protéine PreScission, mais également la potentielle présence ADN lié à nos protéines. Pour ce faire, nous chargeons les fractions digérées sur une colonne Mono-Q préalablement équilibrée à l'aide du tampon Low Salt(Tris HCl pH 7.8 25mM, NaCl 150mM, DTT 1mM), puis nous effectuons un lavage de la colonne à l'aide du même tampon Low Salt avant de réaliser l'élution à l'aide d'un gradient en NaCl de 150 mM à 300 mM. Nous visualisons sur la figure 7.8 la présence de nos fractions *POT1* et *TPP1*, cependant il reste encore des contaminants à l'issu de cette étape.

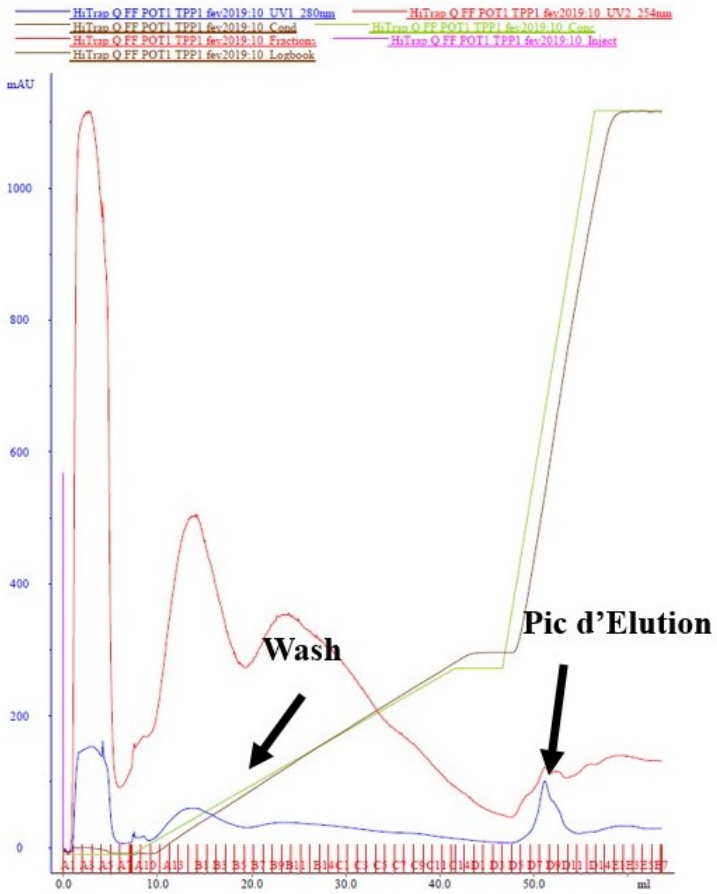


FIGURE 7.7 – **Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne Mono-Q.** les courbes correspondent : en bleu à l'absorbance a 280 nm, en rouge à l'absorbance à 260nm, en marron à la conductivité, en vert à la concentration en tampon, en orange la vitesse du flux.

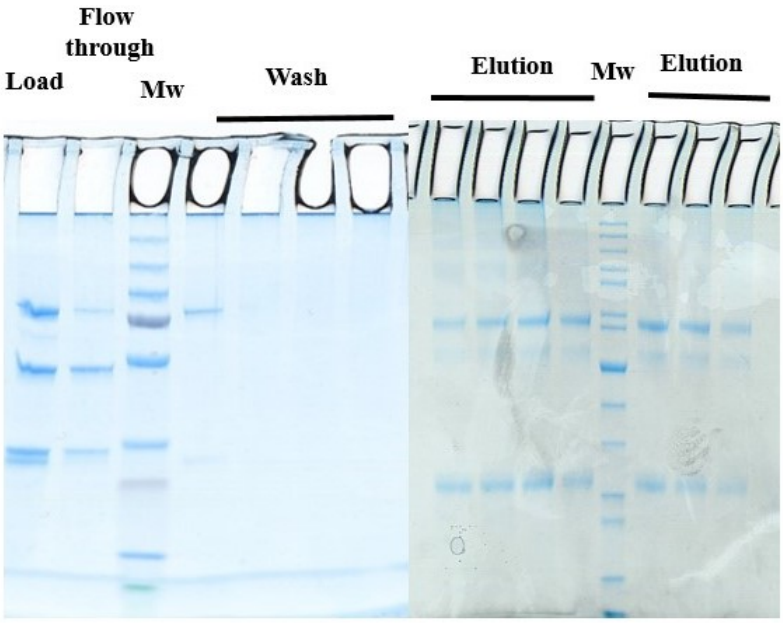


FIGURE 7.8 – **Gel SDS de la purification *POT1-TPP1* à l'issue de l'étape de chromatographie par affinité sur une colonne Mono-Q.** Gel SDS-Page gradient acrylamide 8%-16% révélé par coloration au bleu de Coomassie.

Étape 6 : Colonne d'exclusion

Dans le cadre de notre étude en microscopie électronique, nous réalisons une étape de purification par colonne d'exclusion (SuperDex 200) dans le but de nous assurer que le complexe *POT1-TPP1* est formé. De plus cette étape nous permet d'éliminer l'enzyme PreScission pouvant encore être présente dans notre échantillon. Pour ce faire, nous chargeons les fractions éluées de la Mono-Q sur une colonne SurperDex 200 préalablement équilibrée à l'aide du tampon Buffer (Tris HCl pH 7.8 25mM, NaCl 150mM, DTT 1mM).

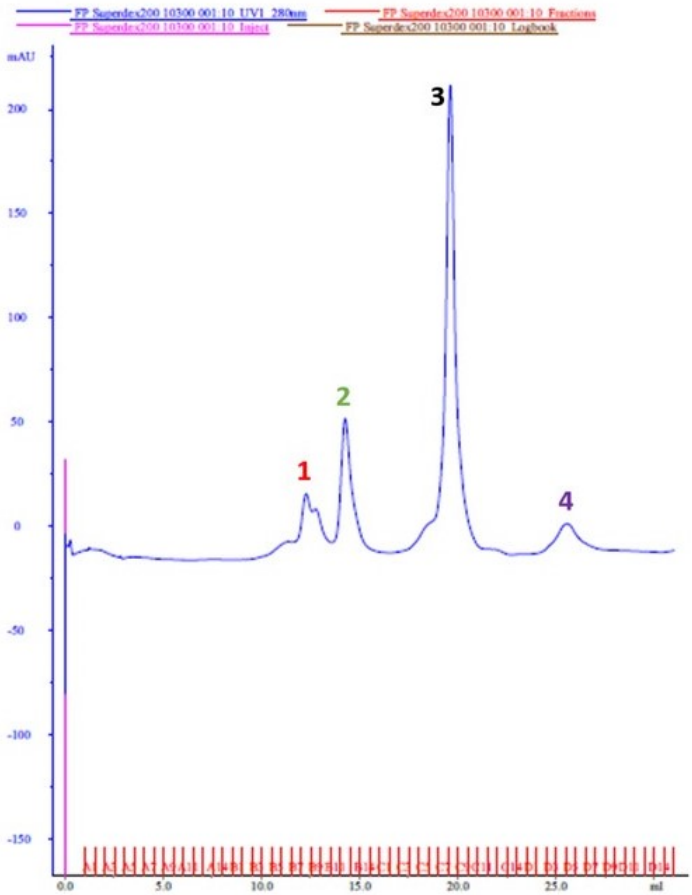


FIGURE 7.9 – **Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne d'exclusion.** les courbes correspondent : en bleu à l'absorbance a 280 nm, en marron à la conductivité, en vert à la concentration en tampon, en orange la vitesse du flux.

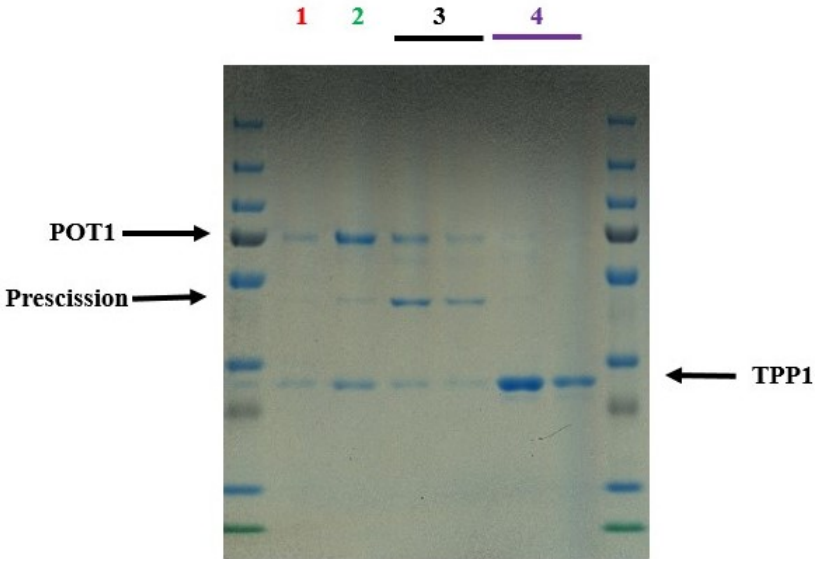


FIGURE 7.10 – **Gel SDS de la purification *POT1-TPP1* à l'issue de l'étape de chromatographie par affinité sur une colonne Superdex 200.** Gel SDS-Page gradient acrylamide 8%-16% révélé par coloration au bleu de Coomassie.

D'après les résultats obtenus sur le gel SDS (Figure 7.10), avons retenu les fractions issues des pics 1 et 2 pour nos études biochimiques, car elles comportent les deux protéines relativement pures. Les fractions du pic 3 présentant une quantité importante de l'enzyme PreScission, n'ont pas pu être utilisées. Enfin les fractions du pic 4 correspondent à *TPP1* seule.

7.2.2 La protéine *POT1*

Étape 1 : Lyse des cellules

Nous utilisons le même protocole que pour le complexe *POT1-TPP1*.

Étape 2 : Colonne His-Select

Nous utilisons le même protocole que pour le complexe *POT1-TPP1*. Nous suivons les étapes de lavage et d'élution de la protéine *POT1* (Figure 7.2).

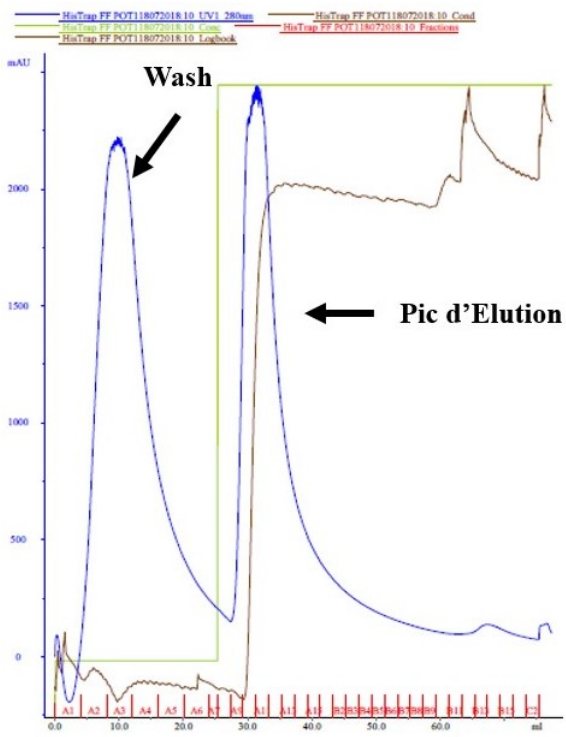


FIGURE 7.11 – **Chromatogramme de purification du complexe *POT1* sur une colonne His-Select.** les courbes correspondent : en bleu à l'absorbance a 280 nm, en marron à la conductivité, en vert à la concentration en tampon, en orange la vitesse du flux.

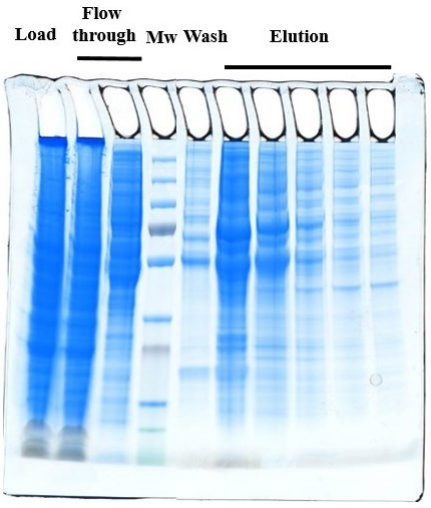


FIGURE 7.12 – **Gel SDS de la purification *POT1* à l'issue de l'étape de chromatographie par affinité sur colonne His-Select.** Gel SDS-Page gradient acrylamide 8%-16% révélé par coloration au bleu de Coomassie.

Étape 3 : Colonne His-Select (2)

Après l'élimination du tag 6 histidines de *POT1*, nous avons choisi de répéter une autre fois l'étape de purification par chromatographie d'affinité à l'aide d'une colonne His-Select. En effet, la majorité des contaminants présents dans notre échantillon se sont liés sur la colonne His-Select. Nous avons réalisé une étape de dialyse pour éliminer l'imidazole présent dans les fractions éluées, puis nous avons chargé les fractions sur une colonne His-Select préalablement équilibrée à l'aide du tampon Low Salt (Tris HCl pH 7.8 25nM, NaCl 150nM, DTT 1mM). Enfin nous avons effectué un lavage de la colonne à l'aide du tampon Wash Ni (Tris HCl pH 7.8 25nM, NaCl 150nM, DTT 1mM et Imidazole 20nM) avant de réaliser l'étape d'élution à l'aide du tampon Elution Ni (Tris HCl pH 7.8 25mM, NaCl 150mM, DTT 1mM et Imidazole 300nM).

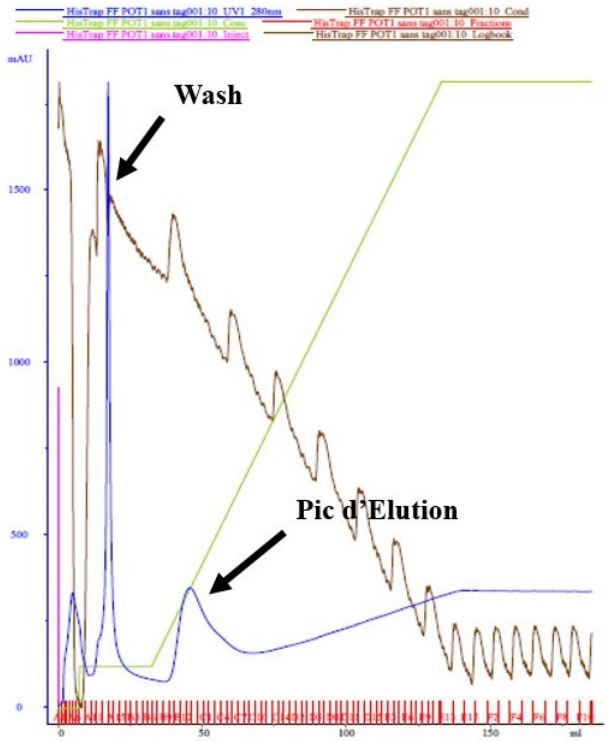


FIGURE 7.13 – **Chromatogramme de purification du complexe *POT1* sur une colonne His-Select(2).** les courbes correspondent : en bleu à l'absorbance a 280 nm, en marron à la conductivité, en vert à la concentration en tampon, en orange la vitesse du flux.

A cette étape nous visualisons dans le gel SDS-Page (Figure 7.14) qu'il reste encore beaucoup de contaminant qui ne sont pas retenu par la colonne His-Select. Cependant on constate un enrichissement en *POT1*

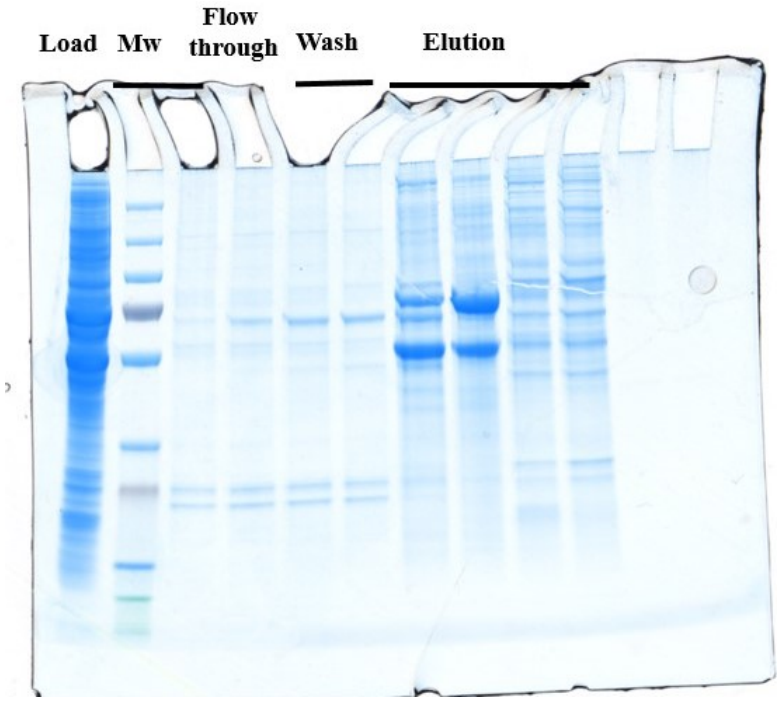


FIGURE 7.14 – **Gel SDS de la purification *POT1* à l'issue de l'étape de chromatographie par affinité sur colonne His-Select (2).** Gel SDS-Page gradient acrylamide 8%-16% révélé par coloration au bleu de Coomassie.

Étape 5 : Colonne Mono-Q

Enfin nous réalisons une étape de purification à l'aide d'une résine échangeuse d'anion (colonne Mono-Q). Pour ce faire, nous chargeons les fractions éluées sur une colonne Mono-Q préalablement équilibrée à l'aide du tampon Low Salt(Tris HCl pH 7.8 25mM, NaCl 150mM, DTT 1mM). Enfin nous effectuons un lavage de notre colonne à l'aide du même tampon Low Salt avant de réaliser l'élution à l'aide d'un gradient en Nacl de 150 mM à 300 mM.

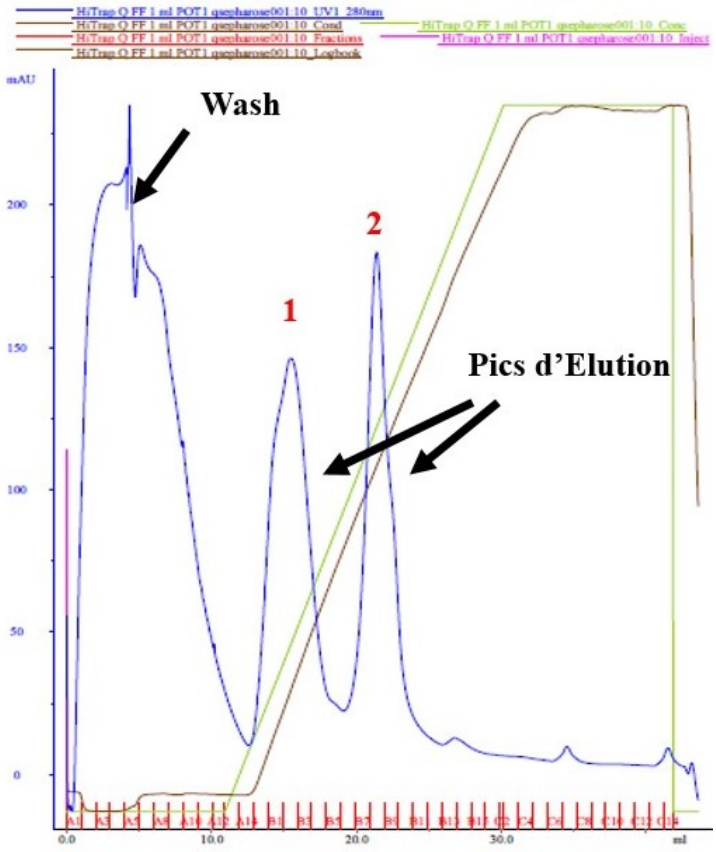


FIGURE 7.15 – **Chromatogramme de purification du complexe *POT1-TPP1* sur une colonne Mono-Q.** les courbes correspondent : en bleu à absorbance a 280 nm, en marron à la conductivité, en vert à la concentration en tampon, en orange la vitesse du flux.

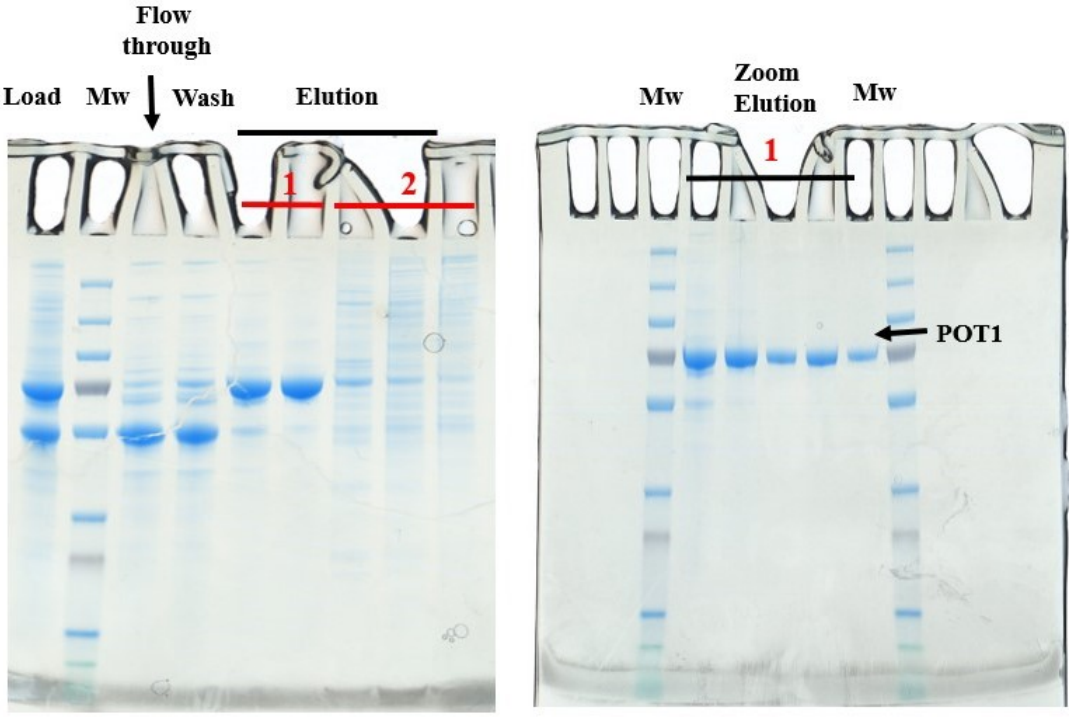


FIGURE 7.16 – **Gel SDS de la purification *POT1-TPP1* à l'issue de l'étape de chromatographie par affinité sur colonne Mono-Q.** Gel SDS-Page gradient acrylamide 8%-16% révélé par coloration au bleu de Coomassie.

Nous décidons de travailler avec les fractions éluées issues du pic 1 car elles sont dépourvues de contaminants.

Chapitre 8

Annexes Figures

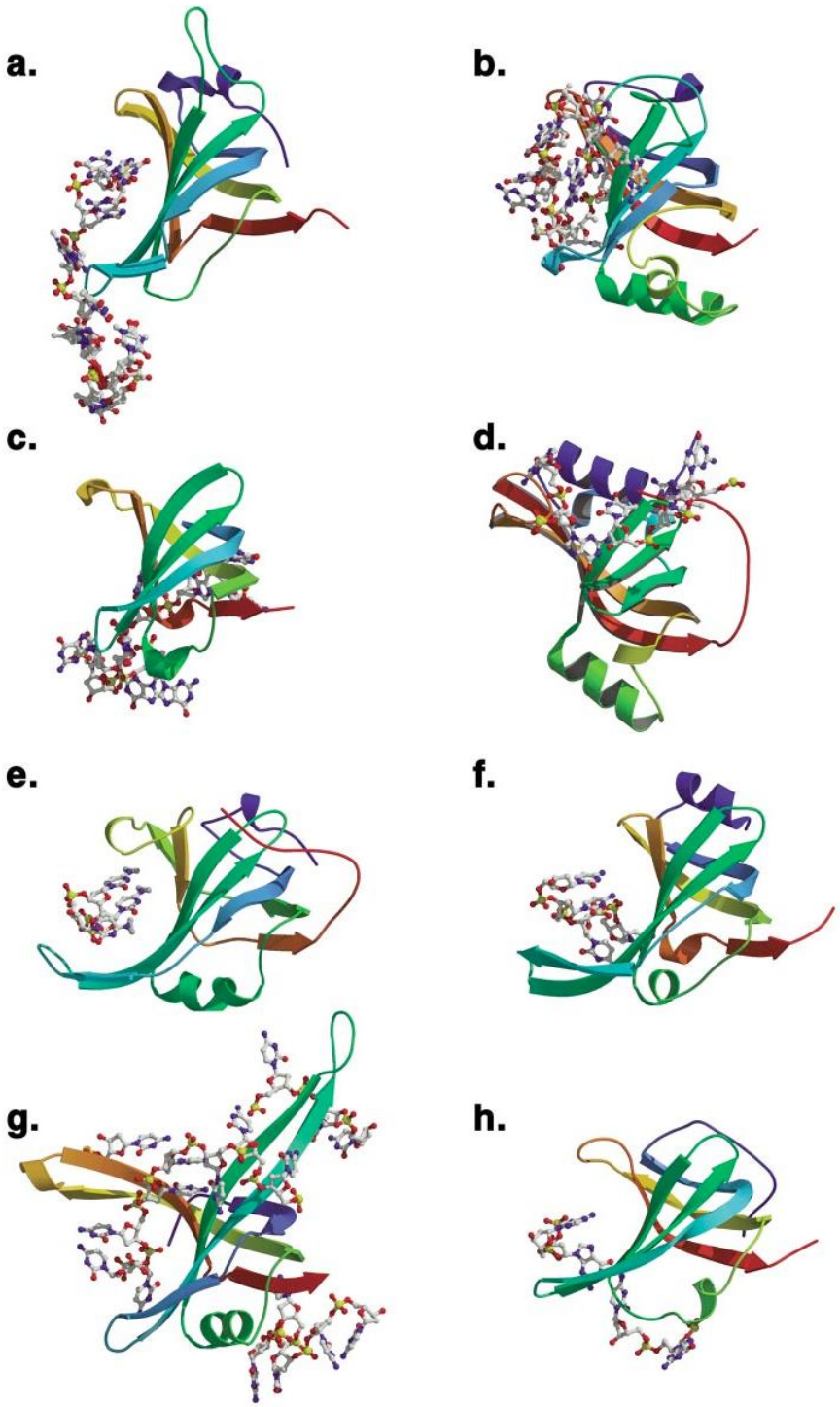


FIGURE 8.1 – **Comparaison des domaines *OB fold* en présence de leur ligand (Theobald et al., 2003)** (a) OnTEBP α1, (b) Cdc13, (c) OnTEBP α2, (d) OnTEBP β, (e) RPA-A, (f) RPA-B, (g) EcSSB, (h) RecG.

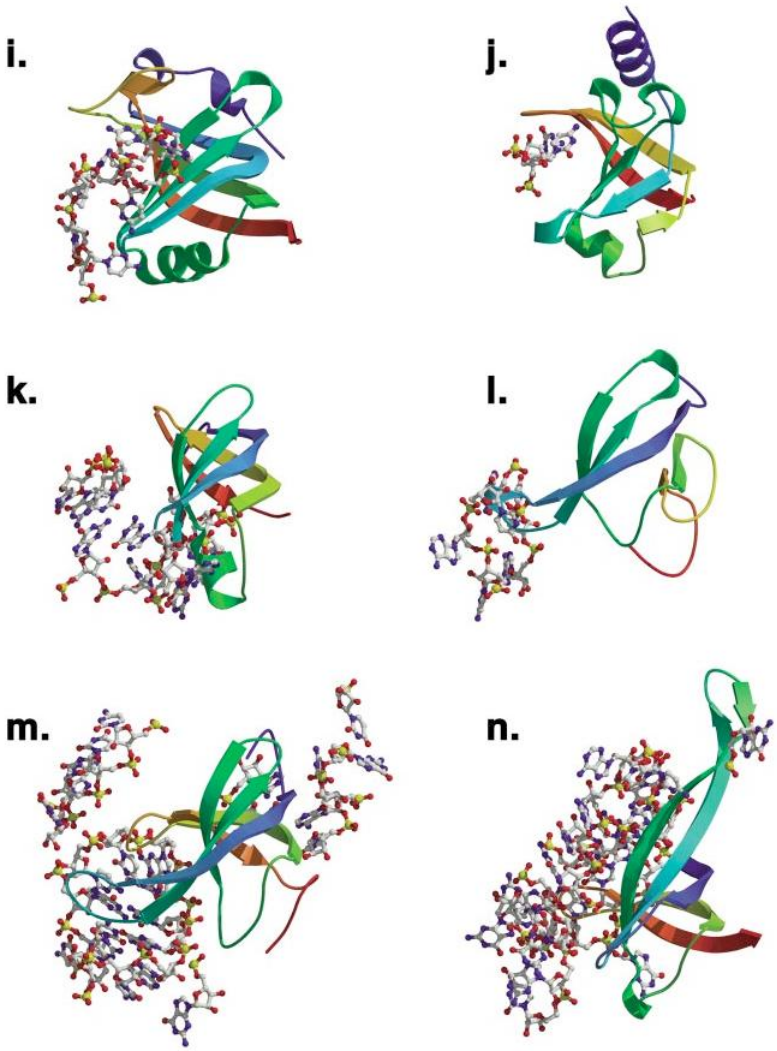


FIGURE 8.2 – **Comparaison des domaines *OB fold* en présence de leur ligand (suite) (Theobald et al., 2003)**(i) EcAspRS, (j) EcRho, (k) IF1, (l) L2, (m) S12, (n) S17

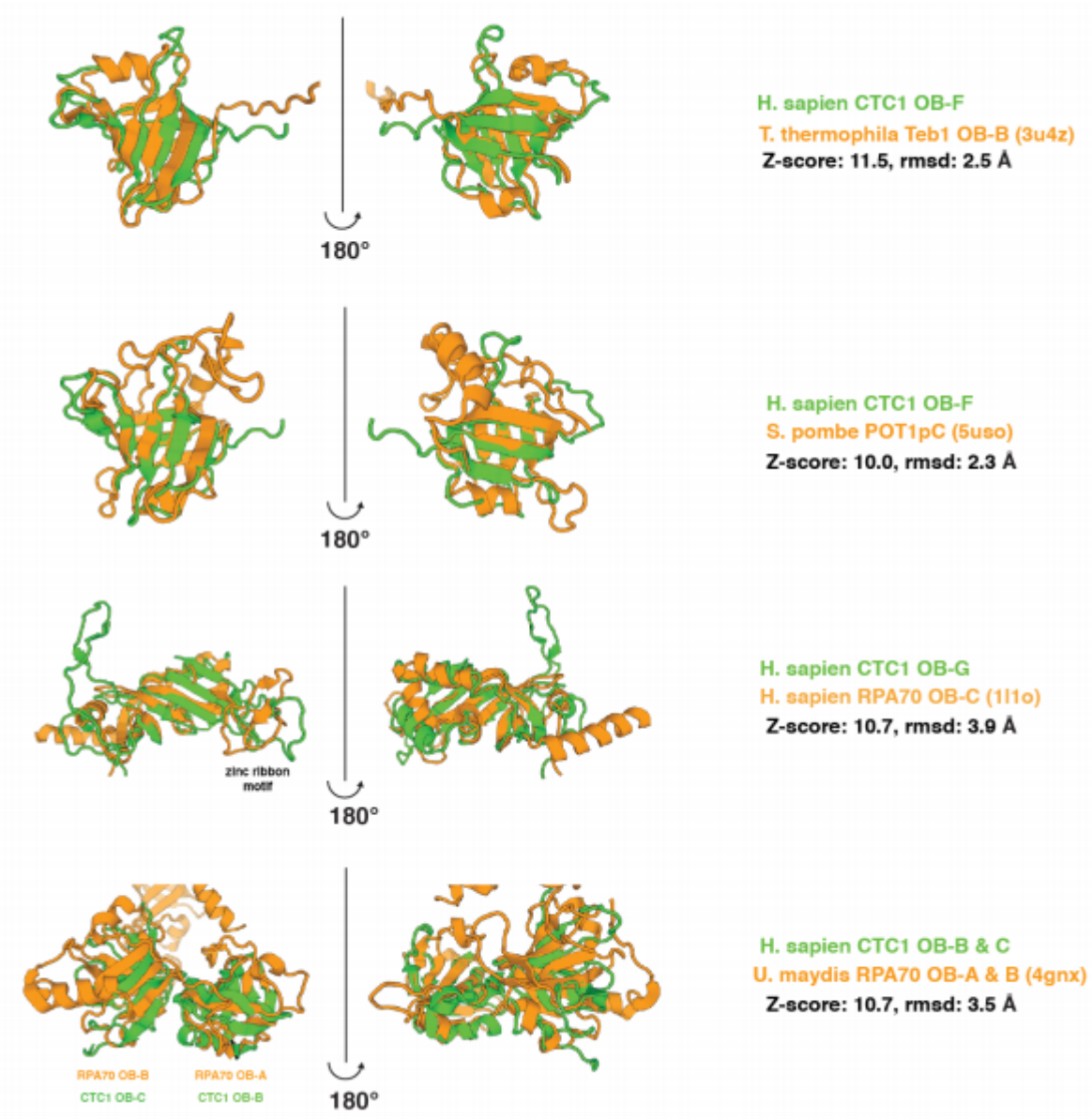


FIGURE 8.3 – **Homologie de structure des domaines *OB fold* de *CTC1*** (Lim et al., 2020)

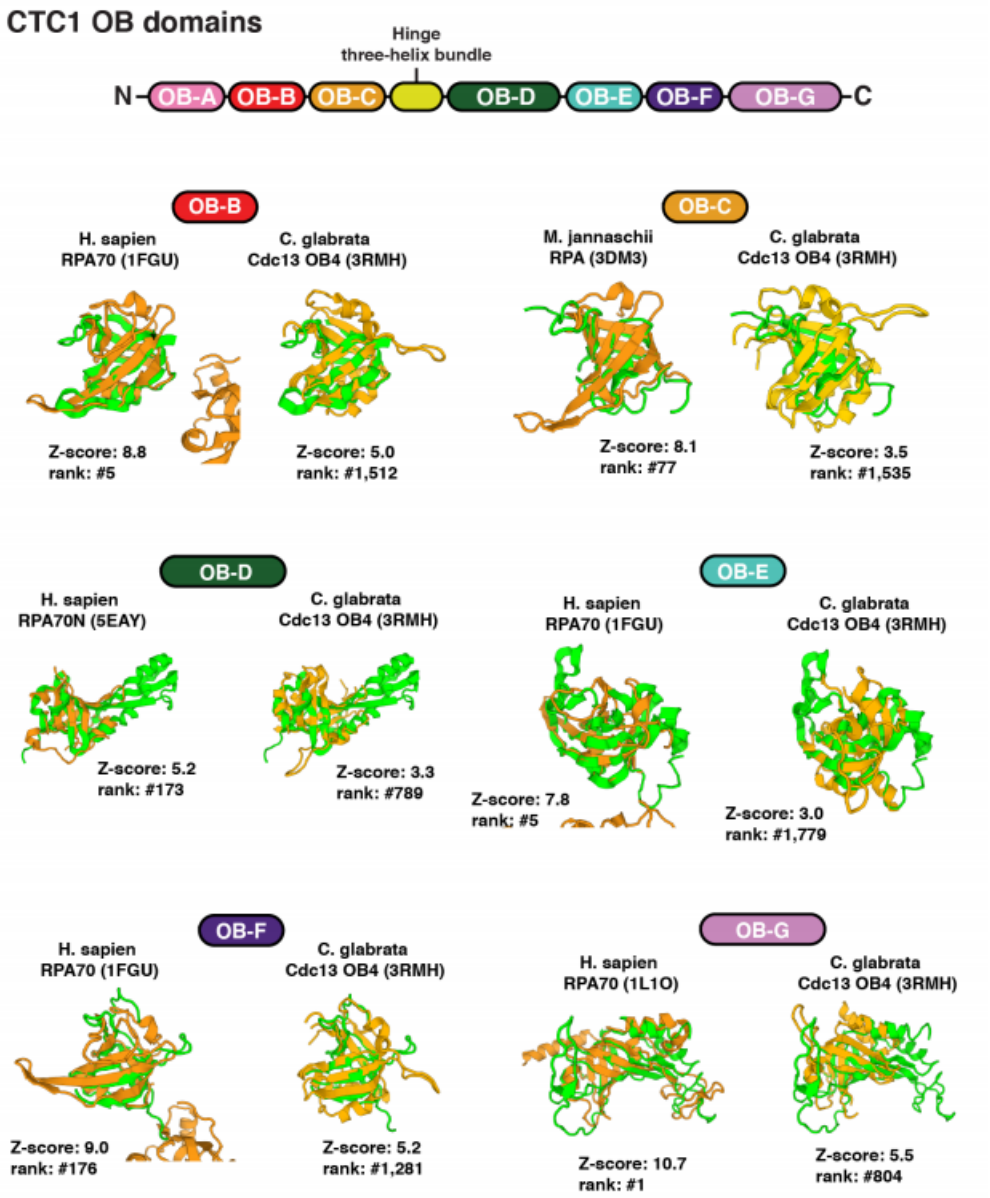
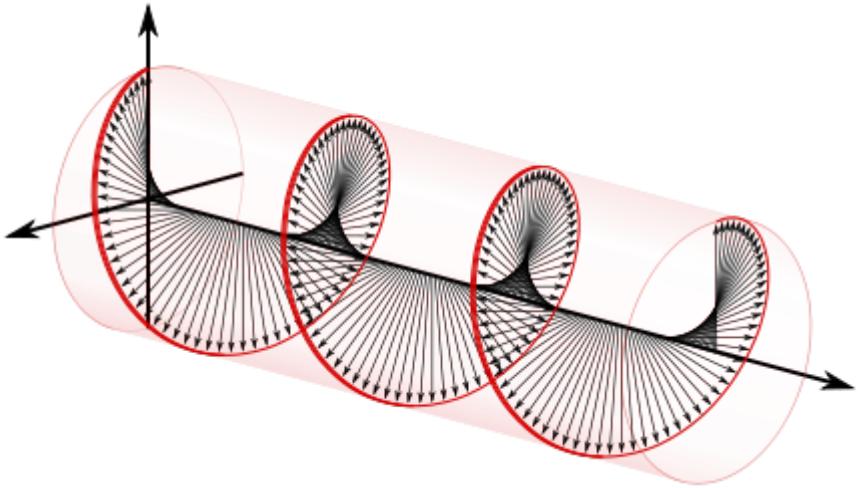
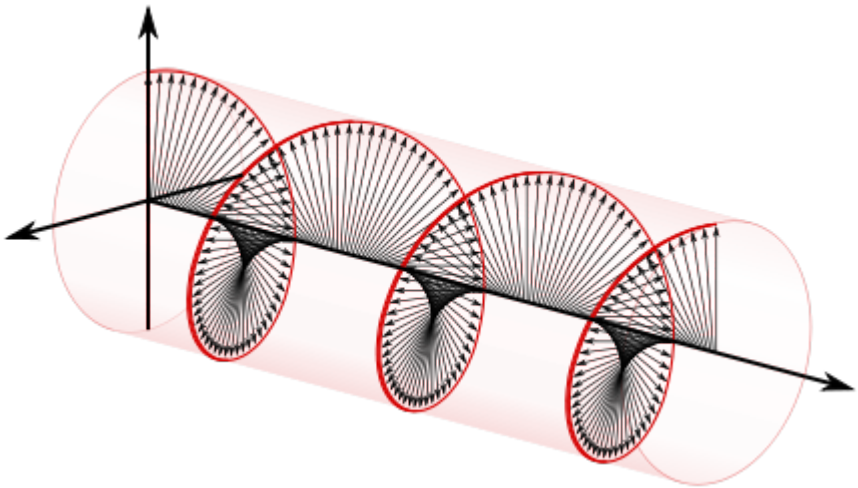


FIGURE 8.4 – **Comparaison d'homologie structurale des domaines *OB fold* de *CTC1* entre ceux de *RPA70* et *Cdc13*** Ce résultat est basé sur l'utilisation d'un *Z-score* calculé avec *DALI structural homology* (Lim et al., 2020)



A. Polarisation Circulaire Droite (PCD)



B. Polarisation Circulaire Gauche (PCG)

FIGURE 8.6 – **Schéma des sens de propagation de la polarisation circulaire.** ©Wikipédia

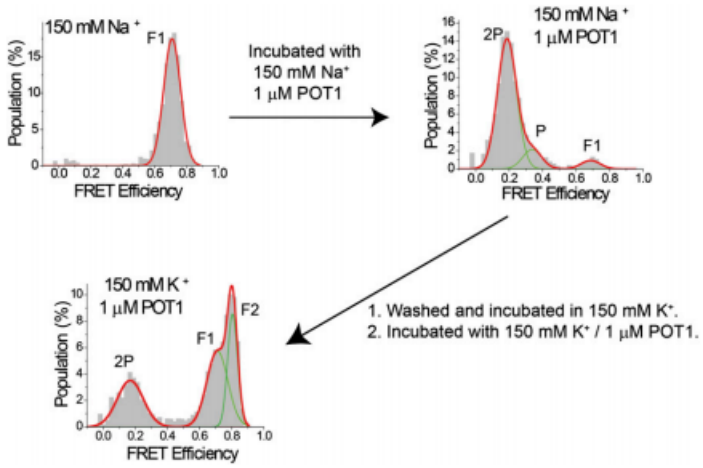


Fig. S6. Transitioning from the distribution in Na^+ to that in K^+. Because the folded GQ state is proposed to be more stable in 150 mM K^+ compared with 150 mM Na^+, we tested whether it is possible to recover the K^+ state distribution, shown in Fig. 1*C*, when the starting distribution is the Na^+ state distribution, shown in Fig. 1*B*. ssTEL4s were initially folded in 150 mM Na^+. Only the F1 folded peak was observed. One μM POT1 in 150 mM Na^+ was then introduced to the sample chamber. This resulted in a distribution that has P, 2P, and F1 states. The sample chamber was then washed multiple times to remove POT1 and incubated in 150 mM K^+ for half an hour before 1 μM POT1 in 150 mM K^+ was introduced. This resulted in a distribution where the P population is eliminated, the 2P population is significantly reduced, and both F1 and F2 conformations are recovered.

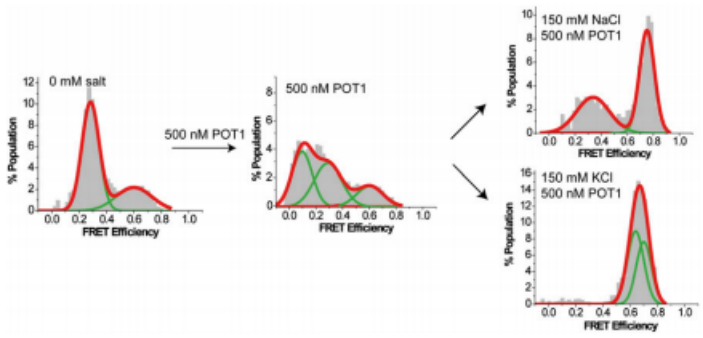


Fig. S7. In steady state, binding of one POT1 can stabilize unfolded ssTEL4 in 150 mM Na^+, but not in 150 mM K^+. The $ssTEL4^{1P}$ sequence is able to form a stable GQ at 150 mM salt (Na^+/K^+) but contains only one POT1 binding site, as opposed to two sites in ssTEL4. (*Left*) To test whether a single POT1 binding can maintain the POT1-bound unfolded state, $ssTEL4^{1P}$ was imaged in the absence of monovalent cation (<1 mM concentration), evidenced by two major peaks at lower FRET efficiencies ($E_F = 0.28 \pm 0.06$ and 0.60 ± 0.12). (*Center*) Addition of 500 nM POT1 (no salt) resulted in the formation of a POT1-bound unfolded population at $E_F = 0.10 \pm 0.07$, representing a single POT1-bound unfolded state for $ssTEL4^{1P}$. The FRET level of this state is significantly lower than the P peak ($E_F = 0.32$) observed in ssTEL4 in 150 mM Na^+. This difference in FRET efficiency is due to lower ionic strength in the $ssTEL4^{1P}$ case (no salt) compared with ssTEL4 (150 mM Na^+). With increasing ionic strength, screening length decreases, which gives rise to a greater persistence length for $ssTEL4^{1P}$ and hence a lower E_F (1). (*Right*) GQ formation was enforced by adding either 150 mM Na^+ or 150 mM K^+. POT1 concentration was maintained at 500 nM POT1 in both cases. In Na^+, a significant fraction of the DNA molecules remains in POT1-bound state with $E_F = 0.34 \pm 0.13$, which matches the P population ($E_F = 0.32$) for ssTEL4. In contrast, addition of 150 mM K^+ completely removes all of the unfolded and POT1-bound populations. The resulting FRET distribution is very similar to that observed for ssTEL4 at 150 mM K^+ in the absence of POT1.

FIGURE 8.5 – **Expérience de *single-molecule* suivant la restructuration d'un G4 en présence de *POT1* (Ray et al., 2014)**

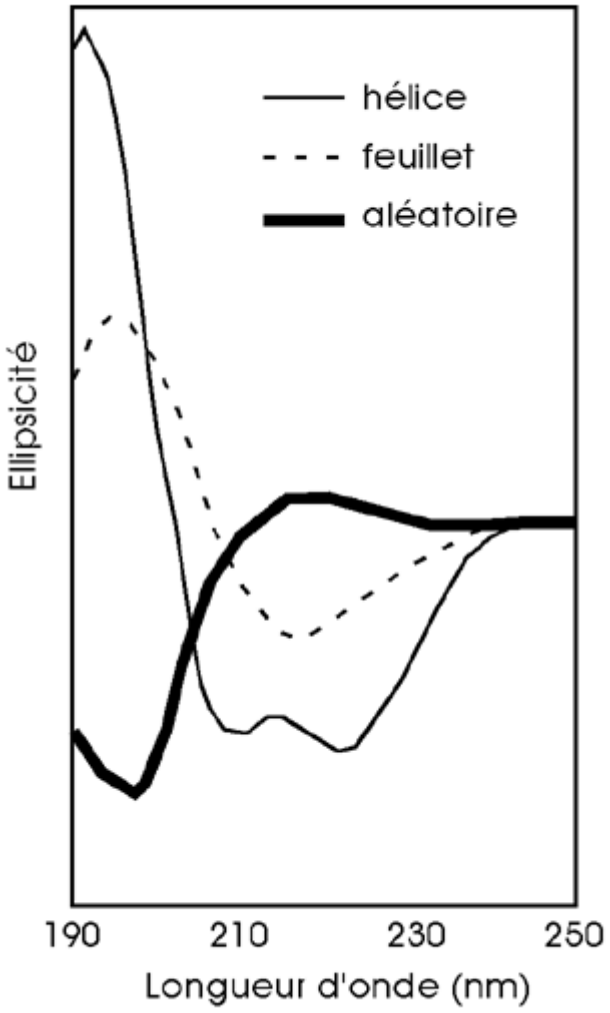


FIGURE 8.7 – **Spectre CD caractéristique d'une hélice α et d'un feuillet β** ©Université d'Angers

Résumé

Les télomères sont des séquences d'ADN (ADNtelo), associées à des protéines, localisées aux extrémités des chromosomes. L'ADNtelo est capable de former des structures très stables nommées G-quadruplexes (G4telo), et qui doivent être résolues pour le maintien des télomères. Il a été montré que POT1-TPP1 et RPA, protéines de liaison à l'ADN simple brin OB-fold (SSB) présentes aux télomères, ont la capacité de dérouler in vitro de simples G4telo. Toutefois aucune étude n'a été réalisée sur de longs ADNtelo formant des multimères de G4s en tandem (multiG4telo).
Mon travail de thèse a consisté à étudier par des approches biochimiques et biophysiques : les mécanismes d'interaction de ces SSB sur ces multiG4telo ; les cinétiques de formation des complexes SSB-ADNtelo ; l'enveloppe du complexe POT1-TPP1-ADNtelo. Les données montrent le mode de liaison et d'ouverture de multiG4telo varie selon la SSB, et que le complexe POT1-TPP1 forme avec l'ADNtelo un complexe dynamique en forme de fer à cheval.

Mots-clefs : Télomères, ADN télomérique, G-quadruplexe, SSB

Abstract

Telomeres are DNA sequences (teloDNA), associated with proteins, located at the ends of chromosomes. teloDNA can form very stable structures called G-quadruplexes (G4telo), which must be resolved for the maintenance of telomeres. POT1-TPP1 and RPA, OB-fold single-stranded DNA binding proteins (SSB) are present in telomeres, can unfold G4telo in vitro. However, no studies have been carried out on long teloDNA forming multimers of G4s in tandem (telomultiG4).
My thesis work consisted of studying by biochemical and biophysical approaches : the interaction mechanisms of these SSBs on this telomultiG4 ; the kinetics of formation of SSB-teloDNA complexes ; the envelope of the POT1-TPP1-teloDNA complex. The data show the mode of binding and opening of telomultiG4telo varies according to the SSB, and that the POT1-TPP1 complex forms with the teloDNA a dynamic complex in the shape of a horseshoe.

Keywords : Telomere, telomeric DNA, G-quadruplex, SSB