

Université Bordeaux Segalen

Année 2008-2011

Thèse n° 1888

THÈSE

pour le

DOCTORAT DE L'UNIVERSITÉ BORDEAUX 2

Mention : Sciences, Technologies, Santé

Option : Biochimie

Présentée et soutenue publiquement

Le 09 décembre 2011

Par Phong Lan Thao TRAN

Née le 07/07/1984 à Ho Chi Minh ville

**Quadruplexes de guanines:
formation, stabilité et interaction**

Membres du Jury

Dr. Geneviève Pratviel	CNRS, Toulouse	Rapporteur
Dr. Hervé Moine	IGBMC, ULP, Strasbourg	Rapporteur
Dr. Sylvain Ladame	Imperial College, Londres	Examineur
Dr. Frédéric Godde	CBMN, Université Bdx, Bordeaux	Examineur
Dr. Patrizia Alberti	MNHN, Paris	Directeur de thèse
Dr. Jean-Louis Mergny	INSERM U869, Université Bdx, Bordeaux	Directeur de thèse

Remerciements

Tout d'abord, je tiens à remercier personnellement les Docteurs Geneviève Pratviel et Hervé Moine, pour avoir accepté d'être rapporteurs de cette thèse et d'avoir consacré du temps à la lecture de ce manuscrit. Je remercie également les Docteurs Sylvain Ladame et Frédéric Godde, pour avoir gentiment accepté d'être membres du jury et d'évaluer mon travail.

Je remercie le Ministère de l'Enseignement Supérieur et de la Recherche, l'Ecole Doctorale Iviv de l'UPMC et l'Ecole Doctorale Science de la vie et de la santé de l'Université Bordeaux Segalen, pour avoir financé ces trois années de recherche. Je remercie vivement le Docteur Jean-Louis Mergny, pour avoir financé les derniers mois du projet, ce qui m'a permis de finir cette thèse dans de bonnes conditions.

Je tiens également à remercier le Docteur Carine Giovannangeli et le Professeur Jean-François Riou, pour m'avoir intégrée dans leur unité pendant la première moitié de ma thèse. Mes remerciements s'adressent également aux Docteurs Jean-Jacques Toulmé et Jean-Louis Mergny, pour m'avoir donné la chance de travailler dans leur unité à Bordeaux durant la deuxième moitié de ma thèse.

J'aimerais exprimer ma gratitude envers tous les collaborateurs qui ont participé, de près ou de loin, à ce travail de thèse, et en particulier à :

- Dr. Aldo Galeone et son équipe, ainsi que Dr. Bernard Rayner, qui nous ont fourni les oligonucléotides modifiés.
- Dr. Marie-Paule Teulade-Fichou et son équipe, qui nous ont fourni les ligands testés dans le criblage G4-FID.
- Pr. Atsushi Maruyama et Dr. Rui Moriyama, pour avoir partagé leur technique d'échange de brins.
- L'Université de Kyoto ("Global COE Program"), pour avoir financé mon séjour au Japon et à Jean-Louis Mergny, pour m'avoir financé le voyage jusqu'à Kyoto. J'aimerais également adresser un grand merci au Pr. Hiroshi Sugiyama et toute son équipe, pour m'avoir accueillie dans leur laboratoire et pour m'avoir enseigné l'ADN Origami. Cette expérience restera à jamais chère à ma mémoire.

Je tiens sincèrement à exprimer ma reconnaissance envers mes directeurs de thèse :

- Un grand merci à toi, Jean-Louis, pour m'avoir donné la chance de travailler dans ton laboratoire. Merci pour ta disponibilité, tes conseils et pour toutes les opportunités que tu m'as donné au cours de ces trois années. Merci très sincèrement de m'avoir fait « grandir » - aussi bien moralement que scientifiquement - et de m'avoir fait aimer mon travail. Tu as beaucoup contribué à ce que je suis devenue ("quelqu'un de bien, j'espère !!!"). Merci pour tout !!!
- Merci à Patrizia, pour m'avoir guidée dans mes premiers pas dans la science. Merci pour m'avoir donné envie de faire de la Recherche, même si ce n'est pas toujours rose, et pour m'avoir soutenue dans les moments difficiles. Je suis un peu comme ton bébé que tu vois grandir de jour en jour. Merci pour tout ce que tu as fait pour moi.

Ce travail de thèse n'aurait pu être réalisé dans d'aussi bonnes conditions sans le soutien, les conseils et l'attention que m'ont portés les membres de l'équipe :

- Je remercie les Docteurs Anne de Cian, Julien Gros et Laurent Lacroix pour leur disponibilité et leurs conseils.
- Merci au Docteur Sébastien Lyonnais, pour m'avoir encadrée pendant l'été 2007. Merci pour ta disponibilité, ta bonne humeur et ta décontraction. J'ai réellement beaucoup apprécié travailler avec toi.
- Je remercie vivement les Docteurs Anne Bourdoncle, Gilmar Salgado et Samir Amrane, pour votre disponibilité, votre soutien et votre gentillesse. Merci pour votre support et vos conseils au cours de la rédaction de ce manuscrit. Merci d'être là, tout simplement !!!
- Merci tout particulièrement à Aurore, Amandine, Gaëlle et Lionel, qui sont devenus bien plus que des collègues de travail. Vous êtes un peu ma deuxième famille, ici, à Bordeaux... Merci pour votre disponibilité et votre gentillesse. Merci d'être présents dans les moments difficiles ! Il est difficile d'exprimer les sentiments que je vous porte avec des mots ...
- Merci aussi à Daniel, Liliya et Nicole, pour votre passage dans le laboratoire. J'aimerais dire quelques mots en anglais pour vous remercier : " Thanks, guys, for your kindness and good mood ! I am so happy to have met you, we had nice discussions together... You will always be in my heart !!! ".
- Merci aux stagiaires avec qui j'ai eu le plaisir de travailler : Delphine Albrecht, Emilien Dubuc, Jack Nicoludis et Katrin Fischer ...

Je remercie sincèrement Daniel Da Silva pour avoir partagé le même bureau que moi pendant la rédaction de ce manuscrit. Merci pour ton aide et ton soutien dans les moments difficiles de ce travail. Merci aussi à Leyla pour ta bonne humeur et tes mots réconfortants.

J'aimerais également remercier toutes mes amies : Alexandra, Amal, Asta, Meriem, Mariam, Najah, ... pour m'avoir autant soutenue durant ces années. Merci, les filles, pour votre compréhension et votre bonne humeur. Merci pour les bons moments passés ensemble... comment dire ... tout ça ne pourrait pas s'oublier.

Enfin, merci à papa et maman, sans qui je n'en serais pas là où je suis aujourd'hui. Un grand merci à ma sœur qui m'a beaucoup aidée pour la finalisation de ce manuscrit. Merci à vous pour avoir été là et pour m'avoir réconfortée dans les moments difficiles. Merci pour tout l'amour que vous me portez, j'ai beaucoup de chance de vous avoir.

Ces deux pages ne sont certainement pas suffisantes pour exprimer toute ma gratitude envers les personnes qui m'ont aidée ces dernières années. Merci à tous et à bientôt !!!

*La vie est comme un instrument de musique;
il faut la tendre et la relâcher, pour la rendre agréable.
Demophile*

TABLE DES MATIERES

TABLE DES ILLUSTRATIONS.....	5
ABREVIATIONS.....	7
CHAPITRE I. INTRODUCTION GÉNÉRALE	10
1) Structures des Quadruplexes de Guanines	10
1.1 Structure de l'ADN : de la double-hélice aux G-quadruplexes.....	10
1.2 Caractéristiques des G-quadruplexes de guanines.....	12
1.2.1 Formation de Quartets.....	12
a) G-quartet classique.....	12
b) Quartets inhabituels.....	14
1.2.2 Structures en G-quadruplexes.....	16
a) Molécularité comprise entre 1 et 4.....	16
b) Configuration des G-quartets et orientation des brins	16
c) Configuration des boucles.....	18
1.2.3 Conditions de formation de G-quadruplexes	20
1.2.4 Méthodes d'étude des G-quadruplexes	20
2) Fonctions biologiques : arguments en faveur de l'existence de G4 <i>in vivo</i>	22
2.1 Analyses génomiques: localisation de G-quadruplexes potentiels	22
2.2 Régulation génique.....	23
2.3 Protéines reconnaissant les G4.....	26
2.4 Preuve directe de l'existence de G-quadruplexes <i>in vivo</i>	28
3) Applications pour les nanotechnologies et biotechnologies	30
3.1 Nanostructures d'ADN : "G-wires", "Frayed wires" ou ADN synaptiques.....	30
3.2 Formation de biocapteurs et de nanomachines avec l'ADN G-quadruplexe.....	32
3.3 Canaux ioniques à base de G-quadruplexes	34
3.4 Pontage de nanoparticules via la formation de G4	35
4) Objectifs de ce travail	35
CHAPITRE II. MATÉRIEL ET MÉTHODES.....	37
1) Produits	37
1.1 Les tampons	37
1.2 Les oligonucléotides	39
1.3 Les ligands	39
2) Spectroscopies.....	39
2.1 Absorbance.....	39
2.1.1 Matériel	39
2.1.2 Principe	40
2.1.3 Méthodes	41
a) Spectres d'absorption, détermination de concentrations et Spectres de Différence Thermique ou Isothermique	41
b) Dénaturation thermique de structures d'ADN	42
c) Association isotherme de quadruplexes tétramoléculaires.....	44
2.2 Dichroïsme Circulaire.....	45
2.2.1 Matériel	45
2.2.2 Principe	45
2.2.3 Méthode.....	46
a) Dichroïsme circulaire des acides nucléiques en particulier des G4	46
b) Mesures et traitement des données	47
2.3 Fluorescence	48
2.3.1 Matériel	48
2.3.2 Principe de G4-FID ("Fluorescent Intercalator Displacement")	48
2.3.3 Mesures et traitements des données.....	49
3) Analyse et purification des acides nucléiques.....	50
3.1 Radiomarquage des acides nucléiques.....	50

3.1.1 Marquage en position 5'	50
3.1.2 Marquage en position 3'	50
3.2 Electrophorèse sur gel de polyacrylamide.....	50
3.2.1 Conditions dénaturantes.....	50
3.2.2 Conditions natives	51
3.2.3 Modes de détection	52
a) Détection radioactive.....	52
b) Détection par ombrage de fluorescence ("UV-Shadow")	52
c) Détection par fluorescence.....	53
3.3 Purification d'oligonucléotides sur gel de polyacrylamide	54
4) ADN origami.....	55
4.1 Construction de l'ADN origami.....	55
4.2 Observation de l'ADN origami en AFM (Microscope à Force Atomique)	56
4.2.1 Matériel	56
4.2.2 Principe	57
 CHAPITRE III. STRUCTURES, STABILITÉ ET CINÉTIQUE DE G-QUADRUPLEXES	
TÉTRAMOLÉCULAIRES MODIFIÉS.....	59
1) Introduction	59
1.1 Stabilité et cinétiques de G-quadruplexes tétramoléculaires.....	59
1.2 L'effet des modifications chimiques sur la formation de G4 tétramoléculaires.....	62
1.2.1 Modification du squelette (sucre - phosphate)	62
a) Modification de l'orientation des brins.....	62
b) LNA-quadruplexes	63
c) PNA-quadruplexe	65
d) L - acides nucléiques	65
1.2.2 Modification des bases.....	66
2) Propriétés structurales et cinétiques des G-quadruplexes tétramoléculaires modifiés, 8-méthyleguanine et L-ADN en particulier	68
2.1 Article 1 : "Effects of 8-methylguanine on structure, stability and kinetics of formation of tetramolecular quadruplexes". Biochimie - 2011; 93(3): 399-408.....	68
2.2 Article 2 : "A mirror-image tetramolecular DNA quadruplex". Chemical Communications (Camb) - 2011; 47(19): 5437-9	69
3) Résultats et Discussion : Chapitre III.....	70
3.1 Effet de la 8-méthyleguanine sur la structure, stabilité et cinétique de formation de G4 tétramoléculaires :	70
3.1.1 L'insertion de 8-méthyleG aux extrémités favorise la formation de quartet en syn.	70
3.1.2 L'introduction de quartet en syn pourrait favoriser un signal CD de type hybride ou de type II.....	70
3.1.3 Une modification à l'extrémité 5' stabilise et augmente la vitesse de formation des G4 tétramoléculaires.	71
3.1.4 Certaines positions de modification favoriseraient la formation de plusieurs espèces en solution.	72
3.2 Sélection chirale pour les G4 tétramoléculaires de D- et L-ADN	72
 CHAPITRE IV. L'ADN ORIGAMI, NANOPLATEFORME POUR LA VISUALISATION D'ADN SYNAPTIQUES	
.....	73
1) Introduction	73
1.1 L'ADN origami.....	75
1.1.1 Principe de la construction de l'ADN origami et sa préparation.....	75
1.1.2 Structures d'ADN origami bi- et tri- dimensionnelles.....	79
a) Structures bidimensionnelles	79
b) Comment augmenter la surface d'ADN origami en 2D ?	79
c) Structures tridimensionnelles	80
1.1.3 Applications de l'ADN origami	81
a) Détection des molécules uniques	81
b) Organisation de matériaux.....	83
c) Système dynamique.....	85
d) ADN origami et G-quadruplexes.....	85
1.2 "ADN synaptique", G4 tétramoléculaire antiparallèle.....	87

2) Visualisation de l'ADN synaptique sur une nanostructure d'ADN.....	89
2.1 <i>Système d'étude et préparation des échantillons.....</i>	89
2.2 <i>Résultats et discussions.....</i>	91

CHAPITRE V: FORMATION DE G4 PAR DES SÉQUENCES BIOLOGIQUES, EN PARTICULIER LES SÉQUENCES TÉLOMÉRIQUES 95

1) Introduction: Rôle, Structure et Maintenance des télomères.....	95
1.1 <i>Structure du télomère et protection télomérique.....</i>	95
1.1.1 Séquences d'ADN télomériques.....	95
1.1.2 G-quadruplexes télomériques.....	97
1.1.3 T-loop.....	101
1.1.4 G-quadruplexes et T-loop.....	103
1.1.5 Complexe "Shelterin" ou protéines du "télosome".....	103
1.2 <i>Réplication télomérique et protéines impliquées.....</i>	107
1.2.1. Extension télomérique par des mécanismes de maintenance spécifiques.....	107
a) Télomérase.....	107
b) Mécanisme ALT ("Alternative Lengthening of Telomeres").....	108
1.2.2. Protéines permettant une réplication fidèle des télomères.....	109
1.3 <i>Dysfonctionnement télomérique et maladies humaines associées.....</i>	110
2) Article 3: "Stability of telomeric G-quadruplexes". Nucleic Acids Research - 2011; 39(8): 3282-94.	111
3) Résultats et Discussion : Chapitre V.....	112

CHAPITRE VI: DÉVELOPPEMENT DE G4-FID SUR PLAQUE POUR L'IDENTIFICATION DE LIGANDS SPÉCIFIQUES D'UNE TOPOLOGIE PARTICULIÈRE DE G-QUADRUPLEXES..... 114

1) Introduction: Études des ligands de G-quadruplexes.....	114
1.1 <i>Classement des ligands de G-quadruplexes.....</i>	115
1.1.1 Ligands basés sur des interactions d'empilement π - π	115
a) Ligands à chaînes latérales protonées <i>in situ</i>	116
b) Ligands N-méthylés.....	116
c) Ligands métallo-organiques.....	117
d) Ligands neutres ou négativement chargés.....	117
1.1.2 Ligands des boucles et des sillons de G-quadruplexe.....	118
1.2 <i>Techniques existantes pour l'étude des ligands.....</i>	118
1.2.1 Spectroscopie en UV-Visible.....	118
1.2.2 Dichroïsme circulaire.....	119
1.2.3 Spectroscopie RMN et cristallographie aux rayons X.....	119
1.2.4 Calorimétrie isothermale (ITC) et calorimétrie à balayage différentiel (DSC).....	120
1.2.5 Spectrométrie de masse à ionisation électrospray (ESI-MS).....	121
1.2.6 Résonance des plasmons de surface (SPR).....	121
1.2.7 Retard sur gel (EMSA).....	122
1.2.8 Dialyse à l'équilibre.....	123
1.2.9 Expérience de dénaturation suivi par Transfert d'Energie d'Excitation par Résonance (FRET).....	123
1.2.10 Test de criblage fluorescent basant sur la compétition duplexe-quadruplexe.....	124
1.2.11 Test de criblage combinant plusieurs molécules fluorescences.....	125
1.3 <i>Les raisons de cette étude.....</i>	126
2) Article 4: "Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures". Biochimie - 2011; 93(8): 1288-96.	127
3) Résultats et Discussion : Chapitre VI.....	128

CHAPITRE VII : DISCUSSION ET PERSPECTIVES..... 130

1) Effets des modifications sur la formation de G4 tétramoléculaires.....	130
1.1 <i>Mécanismes réactionnels envisagés pour la formation de G4 tétramoléculaires.....</i>	130
1.2 <i>Applications possibles de G4 modifiés, notamment 8-MeG et L-ADN.....</i>	132
1.2.1 Étude de ligands spécifiques de G4.....	132
1.2.2 Étude de spectres de dichroïsme circulaire (CD) de G-quadruplexes.....	133
1.2.3 Conception de sonde fluorescente interne.....	133

1.2.4 Application en nanotechnologie	134
1.2.5 Application en biotechnologie.....	136
2) Étude comparative de séquences télomériques et criblage de ligands sélectifs pour une conformation de G4 donnée.....	137
2.1 Difficultés de l'étude de G4 intramoléculaires	137
2.2 Existe-il des ligands capables de discriminer différentes conformations de G4?.....	138
2.3 Structures G-quadruplexes d'ADN et d'ARN	139
3) L'ADN origami, un système d'observation de molécule unique.....	142
3.1 Système de visualisation de structures de G4.....	142
3.2 Système d'étude des interactions protéine / ligand – G4.....	142
 ANNEXE 1 : SÉQUENCE DU PHAGE M13	 144
ANNEXE 2 : SÉQUENCES D'“AGRAPHES” UTILISÉS	145
LISTE DES PUBLICATIONS SCIENTIFIQUES CO-SIGNÉES	147
BIBLIOGRAPHIE	148

TABLE DES ILLUSTRATIONS

Table des Figures

Figure 1. Nucléobases et structures d'ADN	9
Figure 2. Guanine et G-quartet	11
Figure 3. Schématisation des homo-quartets et des différentes possibilités d'appariements relevées dans la littérature	13
Figure 4. Schématisation des quartets "mixtes" et des différentes possibilités d'appariements relevées dans la littérature	15
Figure 5. Polymorphisme des G-quadruplexes	17
Figure 6. Localisation de G4 potentiels et leurs fonctions biologiques	25
Figure 7. Existence de G-quadruplexes in vivo	27
Figure 8. Structures supramoléculaires formées par des oligonucléotides riches en G	29
Figure 9. Biocapteurs et nanomachines basés sur la formation de G-quadruplexes	31
Figure 10. Canaux ioniques à base de G-quadruplexes	33
Figure 11. Pontage de nanoparticules d'or via des G4 tétramoléculaires[122]	35
Figure 12. Schéma représentatif d'un faisceau lumineux traversant un milieu d'épaisseur l	40
Figure 13. Spectrophotomètre à double faisceau	41
Figure 14. Courbe de dénaturation thermique de Hum21 ($3 \mu\text{M}$) dans NaCl 100 mM,	43
Figure 15. Principe du dichroïsme circulaire	46
Figure 16. Représentation schématique de deux principales étapes du test G4-FID	49
Figure 17. Principe de la détection par "UV-Shadow"	53
Figure 18. Principe de l'expérience d'échange de brins	54
Figure 19. Principe de la microscopie à force atomique (AFM)	57
Figure 20. Stabilité et cinétique d'association des G4 tétramoléculaires [140, 141]	60
Figure 21. Analogues de guanines et d'autres modifications du squelette	64
Figure 22. Les bases de la nanotechnologie de l'ADN	74
Figure 23. Principe de l'ADN origami: de la conception à la structure[133, 166]	76
Figure 24. Augmentation de la taille des nanostructures par l'association de tuiles d'ADN origamis	78
Figure 25. Structures d'ADN origamis en 3D	78
Figure 26. Détection des molécules uniques sur surface d'ADN origami	82
Figure 27. Organisation des matériaux sur une structure d'ADN origami	82
Figure 28. Étude et construction des systèmes dynamiques sur une plate-forme d'ADN origami	84
Figure 29. Observation de la formation de G4 en temps réel sur surface d'ADN origami [197]	86
Figure 30. Propriétés de l'ADN synaptique [198, 199].	86
Figure 31. Système de visualisation d'ADN synaptique sur une nanostructure d'ADN	88
Figure 32. Images d'AFM des structures obtenues avec des duplexes courts	90
Figure 33. Images d'AFM des structures obtenues avec des duplexes longs	92
Figure 34. ADN et structures télomériques chez les eucaryotes	94
Figure 35. Structures G-quadruplexes intramoléculaires résolues à partir des motifs télomériques de quelques organismes	98
Figure 36. Enchaînement hypothétique de G-quadruplexes intramoléculaires à l'extrémité télomérique 3' sortante humaine	100
Figure 37. Complexes protéiques des télomères permettant leur protection	102
Figure 38. Problème de réplication télomérique	104
Figure 39. Mécanismes de maintenance télomérique	106
Figure 40. Modèle de progression de la fourche de réplication à l'extrémité des chromosomes de mammifères [282]	106
Figure 41. Schéma représentatif des différents sites d'interactions entre ligands et G-quadruplexe	115
Figure 42. Exemples de ligands de G-quadruplexes à chaînes latérales protonables in situ	116
Figure 43. Exemples de ligands de G-quadruplexes protonés par N-méthylation	116
Figure 44. Exemple de ligands de G-quadruplexes métallo-organiques	117

Figure 45. Exemples de ligands de G-quadruplexes neutres ou négativement chargés	117
Figure 46. Exemples de ligands de G-quadruplexes interagissant avec les boucles et les sillons des G-quadruplexes	118
Figure 47. Interactions ligand-G4 observées en spectroscopie UV-visible [327].	119
Figure 48. Dessins représentatifs des structures de complexes ligand-G4[327]	120
Figure 49. Schéma expliquant le principe de la résonance des plasmons de surface (SPR)	122
Figure 50. Schéma de la dénaturation de G-quadruplexes suivie par FRET	124
Figure 51. Schéma expliquant le principe du criblage fluorescent basé sur la compétition duplexe-quadruplexe [344].	125
Figure 52. Schéma réactionnel proposé pour la formation de G-quadruplexe tétramoléculaire (d(TGnT)) [355].	131
Figure 53. Signatures CD de G-quadruplexes	131
Figure 54. Schéma représentatif d'une sonde fluorescente interne [358]	134
Figure 55. Applications potentielles de L-ADN en nanotechnologie et en biotechnologie	135
Figure 56. Schéma représentatif des structures de G4 d'ARN et de G4 hybride ARN-ADN	140
Figure 57. ADN origami et système d'observation de molécule unique	141

Index des Tables

Table 1. Comparaison des paramètres structuraux d'un ADN en double-hélice et ceux d'un G-quadruplexe d'ADN	12
Table 2. Séquences formant des quartets inhabituels (homo-quartets ou quartets "mixtes")	15
Table 3. Liste des oligonucléotides utilisés dans ce manuscrit excepté ceux utilisés pour l'ADN origami (Annexes 1 et 2)	38
Table 4. Résumé des effets de différents paramètres sur la stabilité et la vitesse d'association des G4 tétramoléculaires [140].	61
Table 5. Les effets de l'inversion de polarité sur la structure, stabilité et cinétique de formation des G4 tétramoléculaires	63
Table 6. Liste des séquences d'ADN testées	89
Table 7. Tableau récapitulatif de l'ensemble des résultats obtenus	91
Table 8. Motifs télomériques de différents organismes	96
Table 9. Etudes structurales menées sur les séquences télomériques de différents organismes	98

ABBREVIATIONS

A, T, G, C et U :	Adénine, Thymine, Guanine, Cytosine et Uracile
ADN :	Acide DésoxyriboNucléique
AFM :	<i>“Atomic Force Microscopy”</i>
ARN :	Acide RiboNucléique
CD :	<i>“Circular Dichroism”</i>
G-quartet :	Quartet de Guanines
G4-FID :	<i>“G-quadruplex Fluorescent Intercalator Displacement”</i>
G4, G4-ADN et G-quadruplexe :	Quadruplexe de Guanines
IDS :	<i>“Isothermal Difference Spectra”</i>
liaison H :	liaison Hydrogène
M :	8-Méthyle-2'-déoxyguanosine
MET :	Microscopie Electronique en Transmission
nts :	nucléotides
PCR :	<i>“Polymerase Chain Reaction”</i>
PEG :	PolyEthylèneGlycol
PQS :	<i>“Putative Quadruplex Sequence”</i>
RMN :	Résonance Magnétique Nucléaire
TDS :	<i>“Thermal Difference Spectra”</i>
T _m :	<i>“Melting Temperature”</i>
UTR :	<i>“UnTranslated Region”</i>
UV :	Ultra-Violet

Introduction Générale

Introduction Générale

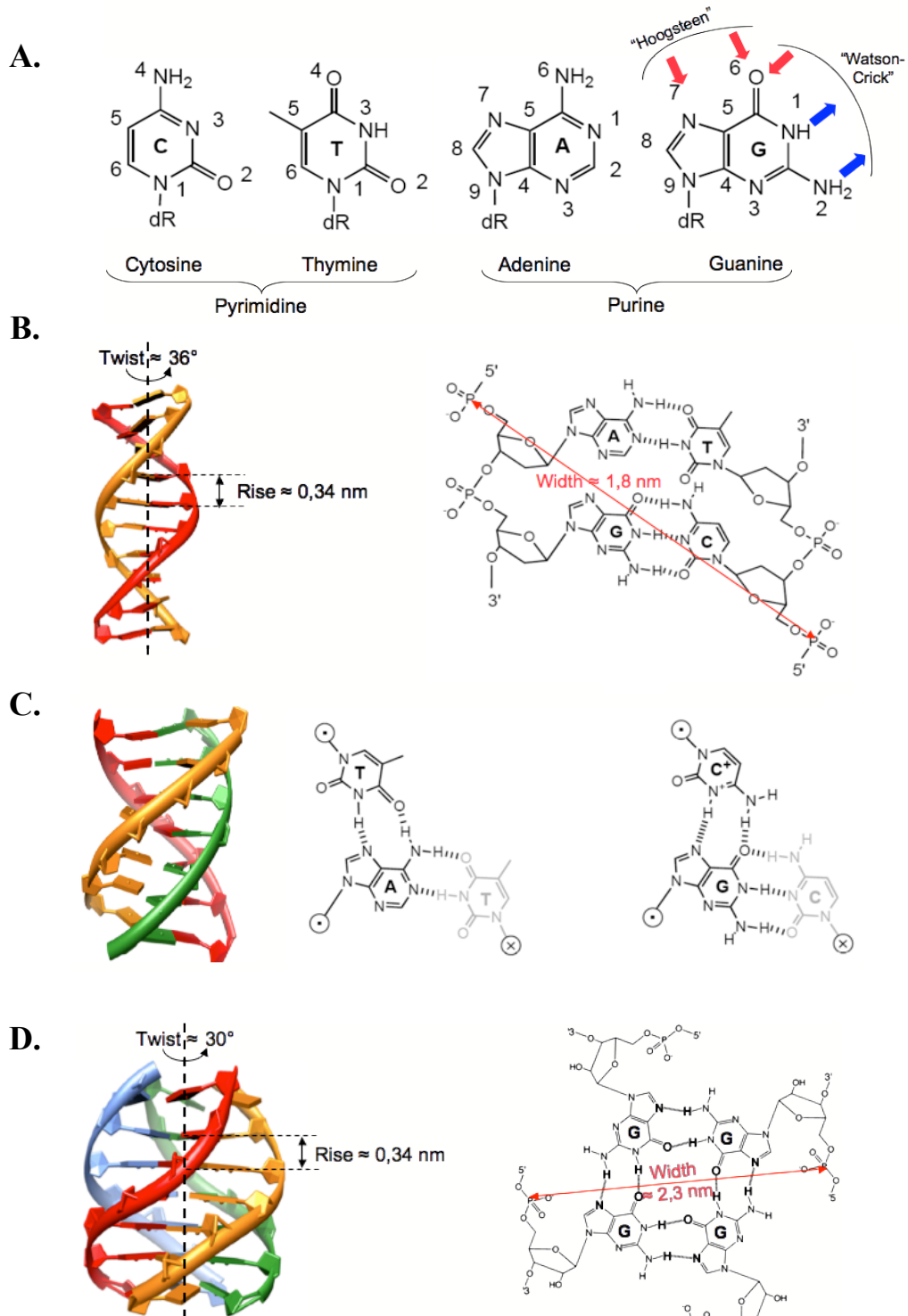


Figure 1. Nucléobases et structures d'ADN

(A) Formule chimique des 4 bases canoniques constituant l'ADN et les sites de liaison H (site donneur en bleu et site accepteur en rouge) impliquées dans des appariements de type "Watson-Crick" et "Hoogsteen". (B) Schéma de double-hélice d'ADN et appariement de type "Watson-Crick" entre A-T et G-C. (C) Schéma de triple-hélice d'ADN et exemple d'appariement de type "Hoogsteen" (en noir) entre A x T et G x C. (D) Schéma de G-quadruplexe d'ADN et de G-quartet formé par l'association de 4 guanines. Les schémas d'ADN double-hélice, triple-hélice et G-quadruplexe sont tirées d'un site web: <http://www.icmub.fr/185-personnel%3Fr%3D185&action%3Dview&id%3D15?r=185&action=view&id=239>

Chapitre I. Introduction Générale

1) Structures des Quadruplexes de Guanines

1.1 Structure de l'ADN : de la double-hélice aux G-quadruplexes

La structure en double-hélice de l'ADN a été révélée par James Watson et Francis Crick en 1953 [1, 2]. Elle correspond à l'appariement de 2 brins anti-parallèles par des liaisons hydrogènes (H) entre les bases A-T et G-C (Figure 1B). Chaque nucléobase possède de multiples sites accepteurs et de donneurs de liaisons hydrogènes (Figure 1A). La double-hélice est basée sur la formation de liaisons H des sites dites "Watson-Crick" des bases nucléotidiques. Cette structure hélicoïdale est stabilisée par d'importantes interactions d'empilement de type π - π entre les bases. L'ADN de forme B est la forme la plus fréquemment rencontrée. Cette structure forme une hélice droite et possède 10,5 paires de bases par tour (soit 21 nucléotides) soit une rotation moyenne d'environ 36° entre chaque sucre. Les bases sont en position *anti* et les sucres C2'-*endo* (Figure 2C). L'espace vertical ("rise") entre chaque paire de base de la structure est de 0,34 nm (soit 3,4 Angströms) (Table 1).

Quatre ans après la découverte de la structure d'ADN en double-hélice, G. Felsenfeld et A. Rich observaient pour la première fois la formation des triple-hélices d'acides nucléiques [3]. Il s'agit d'une structure à 3 brins formée par l'insertion d'un troisième brin dans le grand sillon de l'ADN en duplexe (Figure 1C) [4]. Le troisième brin interagit avec le brin riche en purines de la double-hélice via des appariements dits "Hoogsteen" (liaisons H sur le côté non "Watson-Crick" de la nucléobase) (Figure 1A).

En 1962, M. Gellert et ses collaborateurs ont élucidé le phénomène d'agrégation observé dans une solution d'acide guanylique [5]. La formation de gel est due à l'empilement de quatre guanines 5'-monophosphates associées entre elle. Cette auto-association de quatre guanines par des liaisons H de leurs sites "Watson-Crick" et "Hoogsteen" permet la formation d'une unité structurale dite quartet de guanines (ou G-quartet) (Figure 1D). L'empilement de ces G-quartets forme une structure non canonique en hélice droite appelée quadruplexe de guanines (ou G-quadruplexe, G4-ADN ou G4). Les paramètres structuraux d'un G-quadruplexe et d'un duplexe sont comparés dans la table 1. Ces structures en G4 sont l'objet de ce travail de thèse et seront détaillées ci-dessous.

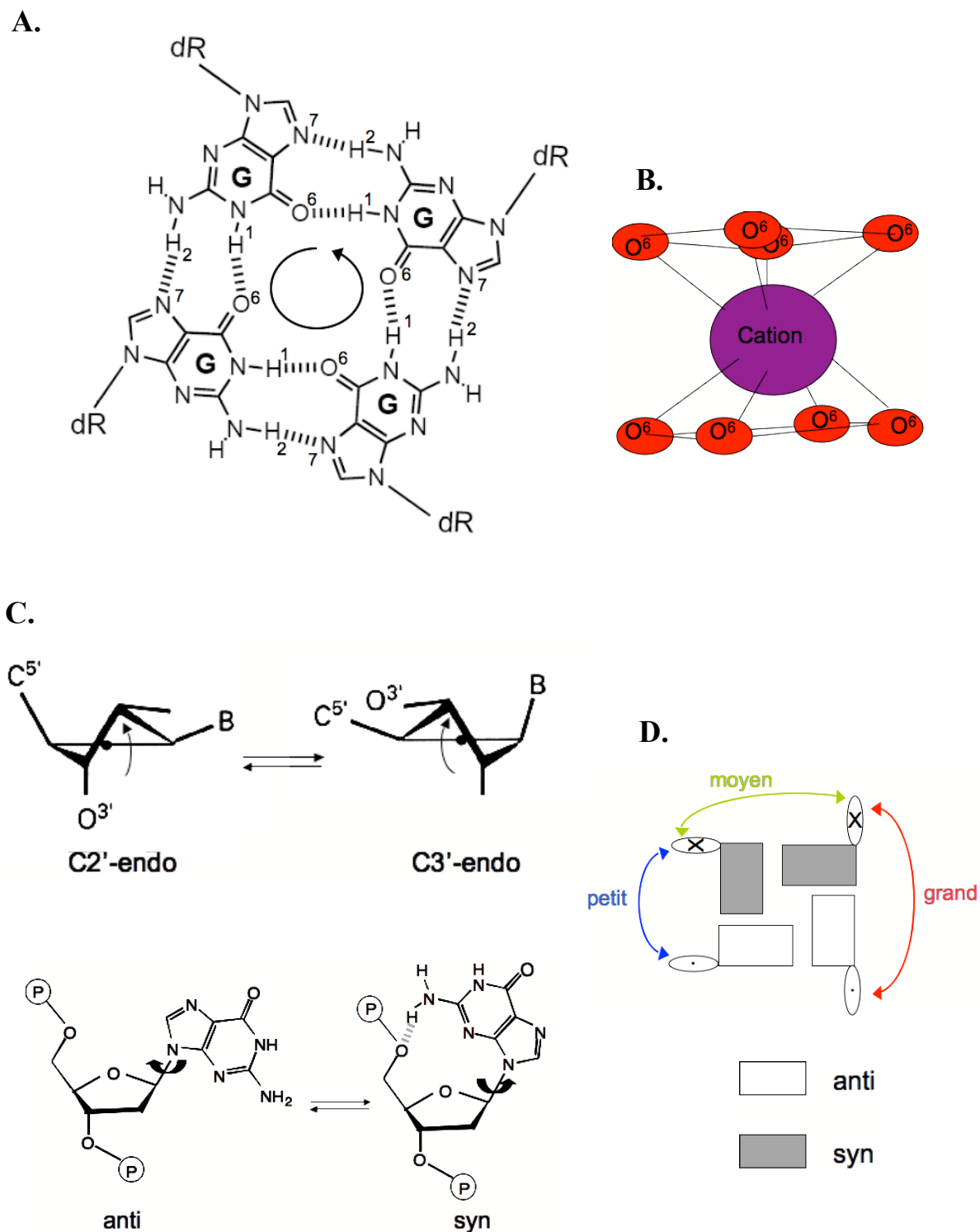


Figure 2. Guanine et G-quartet

(A) G-quartet classique et la rotation dans le sens anti-horaire du donneur N^1-H^1 vers l'accepteur $C^6=O^6$ (face “-”). (B) Cation métallique stabilisant 8 atomes O^6 de deux quartets consécutifs. (C) Conformations du sucre ($C2'$ - ou $C3'$ -endo) et des liaisons glycosidiques (anti ou syn). (D) Représentation de différentes tailles de sillons d'un G-quartet : grand, moyen et petit. La succession de différentes conformations de guanines (anti ou syn) au sein d'un G-quartet définit la dimension des sillons.

Type de structure	ADN forme B	G4-ADN
Hélice	Droite	Droite
Sucre	C2'-endo	C2'-endo
Largeur ("Width") (en nm)	1,8	2,3
Nombre de bases par tour d'hélice	10,5	12
Torsion par paire de base ("Twist") (en °)	~ 36	30
Distance axiale entre paires de bases consécutives ("Rise") (en nm)	0,34	0,34

Table 1. Comparaison des paramètres structuraux d'un ADN en double-hélice et ceux d'un G-quadruplexe d'ADN

1.2 Caractéristiques des G-quadruplexes de guanines

1.2.1 Formation de Quartets

a) G-quartet classique

Un G-quartet résulte de l'auto-association de quatre guanines coplanaires. Ces guanines sont reliées par deux réseaux de liaisons H : un premier réseau "interne" établi entre l'oxygène du groupement carbonyle en position 6 ($C^6=O^6$) et l'hydrogène de l'azote en 1 (N^1-H^1) de deux guanines voisines et un second réseau "externe" constitué de liaisons H via l'azote en position 7 (N^7) et l'hydrogène de l'amine en position 2 (N^2-H^2) (Figure 2A).

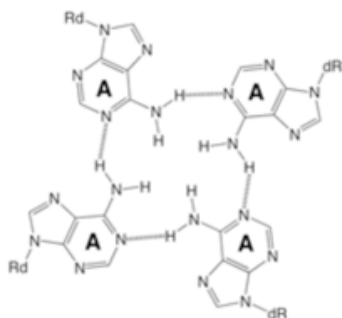
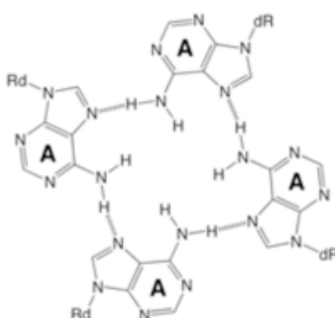
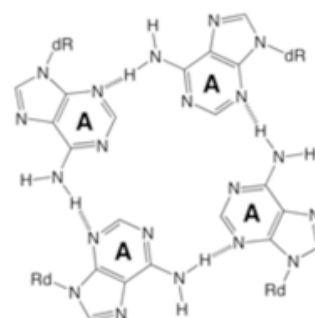
La polarité d'un G-quartet est déterminée par la direction de la liaison H du donneur N^1-H^1 à l'accepteur $C^6=O^6$. Les deux faces d'un G-quartet ne sont pas équivalentes: la face "+" (ou "tête/face") d'un G-quartet est définie par convention par une rotation dans le sens horaire du N^1-H^1 vers le carbonyle $C^6=O^6$ et l'autre face du quartet devient par opposition la face "-" (ou "queue/pile") [6] (Figure 2A).

Le sucre de nucléoside (ribose ou désoxyribose) peut adopter une multitude de conformations différentes dont deux les plus favorables : C2'-endo et C3'-endo (Figure 2C). La conformation C2'-endo est observée en général dans une hélice de type B d'ADN tandis que C3'-endo est trouvé dans l'ARN avec une hélice de type A. Dans une structure en G-quadruplexe, les sucres des guanosines adoptent généralement une conformation C2'-endo.

L'azote N^9 (pour une purine) porte la liaison glycosidique qui relie la base azotée au carbone C1' d'un sucre, ribose (pour l'ARN) ou désoxyribose (pour l'ADN). On peut observer deux orientations différentes *syn* et *anti* de la base azotée par rapport au sucre selon l'axe définit par la liaison glycosidique (Figure 2C). Dans une conformation *syn*, la base azotée et le sucre pointent du même côté tandis que la conformation *anti* définit une orientation opposée des deux composés. La conformation *anti* est privilégiée par des bases

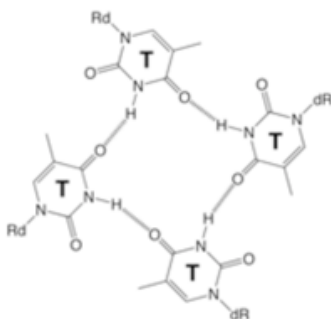
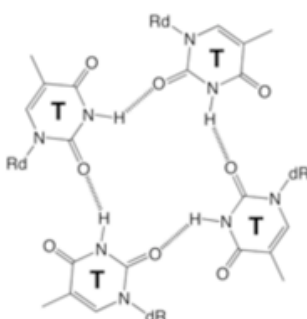
A.

A-quartet



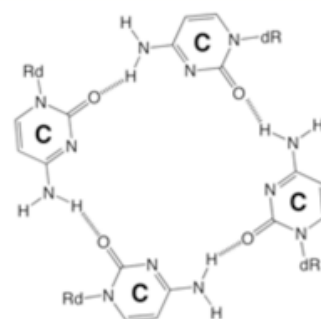
B.

T-quartet



C.

C-quartet



D.

U-quartet

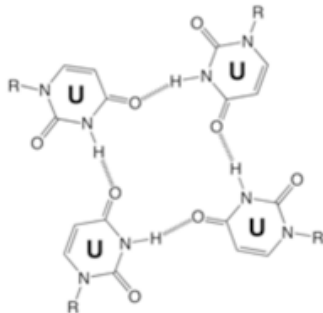
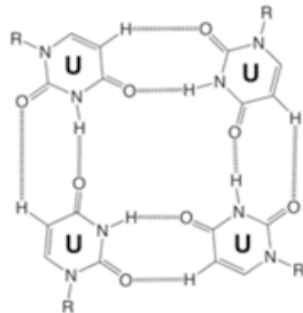


Figure 3. Schématisation des homo-quartets et des différentes possibilités d'appariements relevées dans la littérature

(A) Quartets d'adénines ; **(B)** Quartets de thymines ; **(C)** Quartet de cytosines ; **(D)** Quartets d'uraciles.

non modifiées quelque soit la conformation du sucre (C2'- ou C3'-*endo*) tandis que la position *syn* est plus défavorable en particulier dans le cas d'une conformation C3'-*endo* du sucre. Malgré cet encombrement stérique, la conformation *syn* peut exister dans le cas des guanines, en raison de l'interaction stabilisante entre le phosphate et le N²-H². L'orientation *syn* ou *anti* d'une guanine dans un G-quartet détermine la dimension du sillon entre deux guanines voisines (Figure 2D). Un G-quartet présente 4 sillons, ils peuvent être grands, moyens ou petits [7]. La combinaison de ces différentes conformations constitue la diversité structurale des G-quadruplexes.

L'empilement de G-quartets est stabilisé par la coordination de certains cations métalliques aux oxygènes des fonctions carbonyles des guanines (Figure 2B). Cette coordination limite la répulsion électrostatique des fonctions carbonyles entre-elles. En effet, il a été montré que la température de fusion de gels d'analogues de guanosine variait en fonction de la nature et de la concentration des cations utilisés, plus précisément en fonction de leurs rayons ioniques respectifs [8]. Cette dépendance en fonction de la nature et de la concentration en sels est une des caractéristiques des G-quadruplexes. Des cations monovalents (Li⁺, Cs⁺, Rb⁺, NH₄⁺, K⁺, Na⁺) ou divalents (Sr²⁺, Ca²⁺, Pb²⁺) peuvent être incorporés entre les G-quartets consécutifs et stabilisent plus ou moins la formation de G4 selon leur rayon ionique et leur coût énergétique de déshydratation [9, 10]. Voici le classement général des ions selon leur capacité d'induction et de stabilisation de structures de G4 : Sr²⁺ > K⁺ > NH₄⁺ > Ca²⁺ > Na⁺ ≥ Rb⁺ > Cs⁺ > Li⁺ [11-14].

Il est intéressant de remarquer que, sans être reliées par un squelette phosphaté, les guanosines sont aussi capables de former une structure hélicoïdale en G4 pouvant atteindre jusqu'à 30 nm de long [15]. Ceci souligne une forte coopérativité entre la formation de liaisons H, la coordination avec des cations et les interactions d'empilement π - π , essentielles aux assemblages moléculaires en G-quartets.

b) Quartets inhabituels

Des quartets peuvent être formés par d'autres bases canoniques des acides nucléiques. On trouve non seulement des homo-quartets (quartets de A, C, T ou U) mais également des quartets "hybrides" (ou "mixtes") (quartets G+C et quartets A+T). Les quartets correspondants et les séquences permettant leur formation ont été reportés dans la table 2 et les figures 3 (homo-quartets) et 4 (quartets "mixtes"). La formation de quartets inhabituels a été observée par RMN ou cristallographie dans des séquences riches en guanines comprenant plusieurs G-quartets classiques. La formation d'un quartet inhabituel peut entraîner une

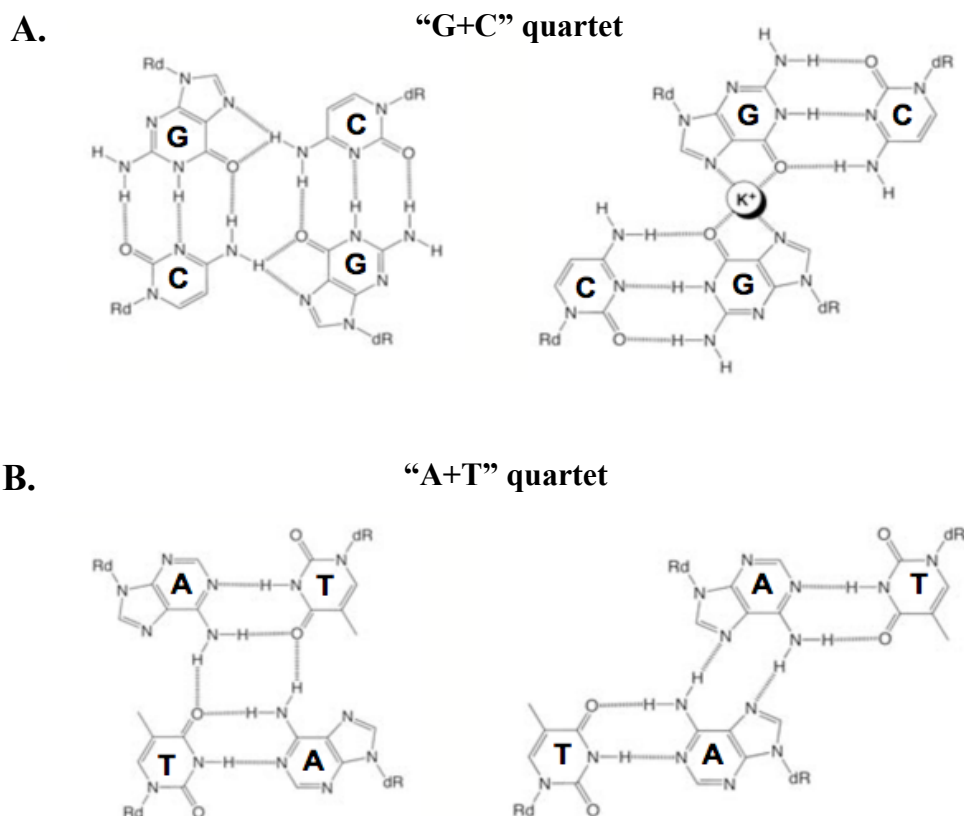


Figure 4. Schématisation des quartets “mixtes” et des différentes possibilités d’appariements relevées dans la littérature

(A) Quartets “mixtes” “G+C”; (B) Quartets “mixtes” “A+T”.

Quartets inhabituels	Séquences (de 5’ à 3’)	Condition de formation	Références
Homo-quartets	A-quartet d(AG ₃ T) d(^{Br} U)-r(GAG ₂ U) r(U)-d(^{Br} G)-r(AG ₂ U)	K ⁺ Na ⁺ Na ⁺	[16] [17] [17]
	T-quartet d(TG ₄ T) d(TG ₃ TG ₃ C)	Na ⁺ , K ⁺ et NH ₄ ⁺ K ⁺	[18] [19]
	C-quartet d(TG ₃ CG ₃ T)	Na ⁺	[20]
	U-quartet r(UG ₄ U) d(UG ₄ T) r(UG ₂ UGU)	K ⁺ et Sr ²⁺ Na ⁺ K ⁺	[21, 22] [18] [23]
	“G+C” quartet d(GCG ₂ X*G ₂ Y*) d(GCG ₂ T ₃ GCG ₂) d(G ₃ CT ₄ G ₃ C) d(GGGATC)	Na ⁺ Na ⁺ Na ⁺ , K ⁺ et NH ₄ ⁺ K ⁺	[24] [25] [26-28] [29]
quartets “mixtes”	“G+C” et “A+T” quartets d(GCG ₂ TTG ₂ AT) d(GAGCAG ₂ T)	Na ⁺ Na ⁺	[30] [31]

Table 2. Séquences formant des quartets inhabituels (homo-quartets ou quartets “mixtes”)

* X = T, U, TT, UC, TC ou A et Y = AT, AU, U ou A

modulation de la coordination des cations insérés entre les quartets, et notamment ralentir le mouvement des ions [18].

1.2.2 Structures en G-quadruplexes

Nous avons présenté précédemment le premier niveau architectural envisagé pour les G-quadruplexes, la formation d'un G-quartet. L'empilement de plusieurs G-quartets forme ce qu'on appelle le G-quadruplexe. Les structures en G-quadruplexe présentent un important polymorphisme. Leur diversité structurale peut être due à plusieurs paramètres (la molécularité, l'orientation des brins, la configuration des G-quartets et des boucles) qui seront abordés ci-dessous.

a) Molécularité comprise entre 1 et 4

Les G-quadruplexes peuvent être formés de quatre (tétramoléculaires), de deux (bimoléculaires) ou d'un seul brin(s) (mono- ou intramoléculaire) d'ADN (Figure 5B). Le repliement en G4 à partir de trois brins est également envisageable, mais à notre connaissance, cette stoechiométrie n'a jamais été rapportée, si l'on excepte l'observation d'espèces trimériques intermédiaires par spectrométrie de masse. Cette différence de molécularité représente le niveau le plus simple du polymorphisme des G4.

b) Configuration des G-quartets et orientation des brins

Chaque brin du G-quadruplexe a une direction définie par l'orientation 5' vers 3' du squelette phosphaté. Il existe quatre possibilités : (i) 4 brins orientés dans la même direction (G4 parallèle), (ii) 3 brins orientés dans une direction et le 4^{ème} dans la direction opposée (G4 "hybride" ou (3+1)), (iii) 2 brins voisins orientés dans une direction et les 2 autres dans la direction opposée (G4 antiparallèle) et (iv) 2 brins diamétralement opposés orientés dans une direction et les 2 autres dans la direction opposée (G4 antiparallèle) (Figure 5B). Notons au passage que le terme "G4 antiparallèle" regroupe en réalité deux types de structures de géométrie bien différente. La variabilité dans la polarité des brins est liée à l'enchaînement des conformations des liaisons glycosidiques *anti* ou *syn* des guanines au sein d'un quartet. Ainsi les quatre possibilités d'orientations de brins décrites précédemment sont associées à ces possibilités d'arrangement des 4 guanines d'un même G-quartet : (i) *anti-anti-anti-anti* ou *syn-syn-syn-syn*, (ii) *syn-anti-anti-anti* ou *anti-syn-syn-syn*, (iii) *syn-syn-anti-anti* et (iv) *syn-anti-syn-anti*.

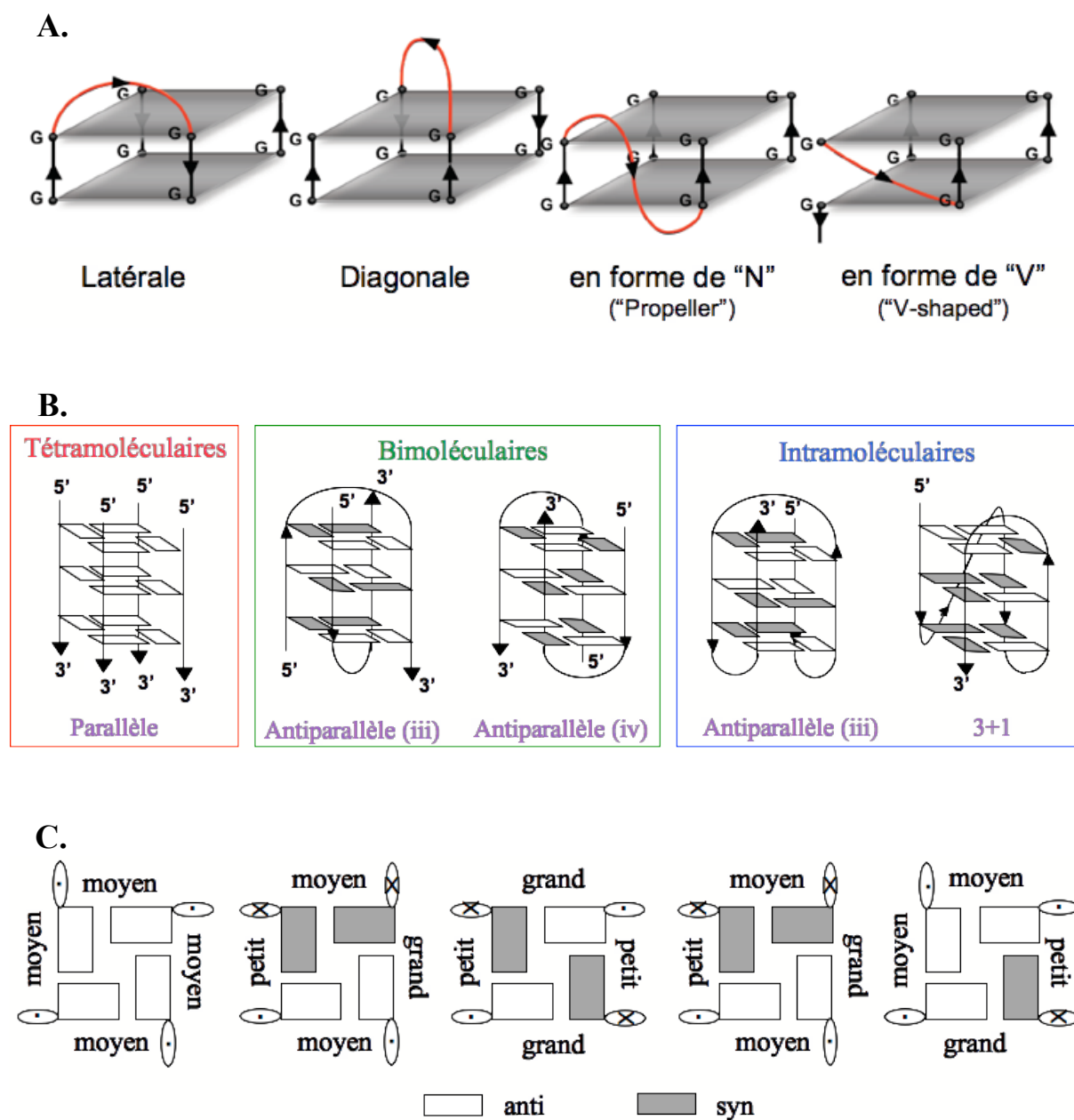


Figure 5. Polymorphisme des G-quadruplexes

(A) Différentes configurations de boucles: latérale, diagonale, en forme de "N" ou de "V".
 (B) Exemples de G-quadruplexes avec des molécularités (tétra-, bi- ou intramoléculaire), des orientations de brins (parallèle, antiparallèle ou 3+1), des arrangements de boucles et des quartets variés. (C) Structure des quartets présentés en (B) et leur taille de sillons.

Par ailleurs, nous avons mentionné précédemment que la direction des liaisons H au sein d'un quartet détermine sa polarité (cf. : §I.1.2.1 a). Elle peut être dans le sens horaire (face "+") ou anti-horaire (face "-"). Ceci est directement relié à la conformation des liaisons glycosidiques des guanines pour chaque type d'orientation de brins. L'empilement entre deux quartets adjacents de même polarité diffère de l'empilement de deux quartets de polarité opposée.

De plus, l'enchaînement des guanines en *syn* ou *anti* au sein d'un quartet entraîne la formation de sillons de tailles variables (Figure 5C).

c) Configuration des boucles

Les structures de G4 bi- ou intramoléculaires font apparaître des boucles. Ces boucles peuvent adopter quatre configurations différentes : (i) boucle diagonale reliant deux brins antiparallèles diagonalement opposés, (ii) boucle latérale connectant deux brins adjacents antiparallèles, (iii) boucle en forme de "N" ("propeller" ou "double chain reversal") reliant deux brins adjacents parallèles et (iv) boucle en forme de "V" ("V-shaped") (Figure 5A). La configuration des boucles est intimement liée aux contraintes d'orientation des brins d'un G4.

La configuration des boucles dépend énormément de leur taille et de leur séquence. En effet, une boucle en forme de "N" ne comportant aucun nucléotide peut relier 2 G-quartets successifs [32]. Une boucle en forme de "N" avec un ou plusieurs nucléotides peut relier 2 ou 3 quartets consécutifs. A partir de 3 nucléotides dans une boucle, on peut avoir différentes configurations dépendant de leur séquence.

Plusieurs équipes, dont la nôtre, se sont intéressées aux variations structurales de G4 intramoléculaire en fonction des longueurs [33-35] et/ou des séquences [36-39] de boucles.

Concernant la taille des boucles, les résultats publiés jusqu'à présent semblent indiquer que : (i) les boucles courtes (1 nucléotide) favorisent la formation de G4 parallèles stables ; (ii) les boucles longues (≥ 3 nucléotides) favorisent une conformation antiparallèle mais diminuent la stabilité du G4. En général, l'augmentation de la taille d'une boucle diminue la stabilité de la structure, chaque nucléotide ajouté entraîne une diminution d'environ 2°C du T_m dans K^+ . La tendance est moins claire en présence de Na^+ ; (iii) une séquence contenant une boucle très longue (≥ 9 nucléotides) peut également former des G4 avec une stabilité thermique acceptable en conditions physiologiques.

En général, les expériences sont réalisées en présence de K^+ et/ou de Na^+ . Les structures obtenues en présence de K^+ sont généralement plus stables que celles dans Na^+ . La nature des cations influence également la conformation des G4, le potassium semble favoriser la

formation de G4 parallèles. Même s'il semble peu probable que la nature des cations puisse influencer l'orientation finale des brins, il est cependant raisonnable de penser que la force ionique peut influencer les interactions secondaires trouvées éventuellement dans les boucles ou entre les boucles et le G4.

Enfin, l'effet de la séquence de boucles sur la stabilité des G4 a été mis en évidence. Rachwal *et al.* ont montré que dans une boucle d'une seule base, l'adénine donne un G4 moins stable qu'une pyrimidine (C ou T). Notre équipe a également observé des résultats similaires sur des boucles plus longues (3 nucléotides) [38]. Dans ce cas, l'adénine à la première position d'une boucle déstabilise considérablement la structure de G4 formée. La cytosine déstabilise légèrement le G4 quand elle est placée au centre d'une boucle.

Ces travaux ont permis également d'enrichir les données de l'algorithme développé par J. Huppert et ses collègues [40]. L'algorithme permet de prédire la formation et la stabilité d'un G4 *in vitro* à partir d'une séquence primaire. Mais, il n'y a, jusqu'à présent, aucune règle fiable permettant de prédire la conformation de G4 obtenue à partir d'une séquence donnée. Néanmoins, l'insertion d'une guanine modifiée (notamment, 8-bromoG [41] ou ribo-G) à des positions spécifiques permet de contrôler la topologie de la structure et donc du type de boucles [42]. L'étude des guanines modifiées sera abordée dans le chapitre III du manuscrit.

Par ailleurs, il est important de noter que toutes ces études utilisent le dichroïsme circulaire comme principale méthode d'analyse de l'orientation des brins au sein d'un G4. Cette technique est adaptée pour révéler la formation de G4 par une séquence riche en guanines, mais elle n'est pas fiable pour évaluer la polarité des brins au sein d'un G4. En effet, un spectre CD résulte de l'empilement de G-quartets successifs de différente polarité. Ce point sera discuté dans les chapitres II et III de ce manuscrit. Nous verrons dans le chapitre III qu'une même conformation de G4 parallèle à partir d'enchaînement de configurations de G-quartets différentes peut donner des spectres CD différents.

Nous pouvons également remarquer que dans certaines structures, des guanines peuvent se retrouver dans les boucles, par exemple le G4 formé par la séquence télomérique de *Tetrahymena thermophila* dans Na⁺[43] et la structure constituée de 2 quartets résolue pour la séquence télomérique humaine dans K⁺ [44]. À l'inverse, certains nucléotides autres que G peuvent former des quartets inhabituels, notamment un variant du motif télomérique humain d(GGGATC) [29].

1.2.3 Conditions de formation de *G*-quadruplexes

Nous avons mentionné précédemment que l'empilement des G-quartets nécessitait la présence de cations. La nature du cation présent dans la solution peut favoriser certains types de repliements et influencer la stabilité de la structure formée. En effet, les cations interagissent avec les G-quadruplexes de deux manières différentes : (i) par interaction avec les 8 atomes O⁶ de deux plateaux de G-quartets successifs (ions centraux) (Figure 2B) et (ii) par interaction avec les charges négatives du squelette phosphaté du quadruplexe (ions externes). Les ordres établis pour les cations monovalents ($K^+ > NH_4^+ > Rb^+ > Na^+ > Cs^+ > Li^+$) [13] et pour les cations bivalents ($Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$) [14] peuvent être corrélés au rayon ionique et au coût énergétique de déshydratation des cations. Les cations positionnés entre les quartets effectuent des échanges avec les ions dans la solution. Il a été montré que la demi-vie de l'interaction entre les cations et les quartets était généralement inférieure à la seconde, donc beaucoup plus courte que la demi-vie d'un G4 [45, 46]. Cela suggère que le mouvement des cations au sein du G4 ne nécessite pas son ouverture. Le G-quadruplexe pourrait donc être comparé à un canal ionique permettant le transport des ions. Cette caractéristique du G-quadruplexe pourrait être utilisée en nanotechnologie, notamment pour la fabrication de nanofils ("*G-wires*") pour l'électronique moléculaire et de nanopores, qui seront présentées un peu plus loin dans ce chapitre.

En absence de cations, les agents comme le polyéthylène glycol (PEG), peuvent aussi stabiliser la structure de G4 [47]. La présence de PEG influence la structure de G4 formée, et favorise la formation de G4 parallèle [48]. On peut s'interroger sur la pertinence d'une structure définie en solution diluée, plutôt que dans un milieu plus dense, étant donné le fort encombrement qui doit exister dans le noyau cellulaire (acides nucléiques, protéines, cofacteurs, ...). D'autres cosolutes peuvent stabiliser la structure en G4, notamment les polyamines (par exemple la spermidine) [49] et le triéthylène tétramine [50]. Il faut cependant noter que certains agents mimant l'encombrement moléculaire ("crowding agents") ont également un effet sur l'activité de l'eau, ce qui rend délicate la déconvolution de ces deux effets.

1.2.4 Méthodes d'étude des *G*-quadruplexes

De nombreuses techniques ont été utilisées pour identifier et caractériser les structures en G4. La cristallographie et la RMN sont des méthodes les plus puissantes pour cette étude structurale.

De nombreuses structures de G4 [51], incluant celles en présence de ligands ont été résolues par cristallographie. Cette technique a permis d'accéder à des données précieuses sur les distances et l'emplacement des ions au sein du G4. L'inconvénient de cette méthode est que la structure résolue étant à l'état solide (cristal) pourrait être différent de la structure adoptée en solution. En particulier pour les structures polymorphiques comme le G4, la conformation observée pourrait être la forme la plus facilement cristallisable et non celle majoritaire en solution.

La RMN constitue aujourd'hui une technique de référence pour identifier la structure d'une séquence donnée en solution. Cependant, en raison du polymorphisme des G4 et de l'augmentation de complexité en fonction du nombre de blocs de G consécutifs, il peut être nécessaire d'ajouter des bases aux extrémités de la séquence d'intérêt ou de modifier certaines bases (G => 8-bromoG ou inosine; T => U) "stratégiques" afin de favoriser une forme particulière et d'obtenir un spectre RMN exploitable [52, 53]. Ceci peut évidemment conduire à un biais sur l'identification de la structure réelle de la séquence native, puisque seules les formes exploitables sont caractérisées. De plus, la RMN nécessite une concentration relativement élevée des échantillons, ceci peut influencer la structure de G4 formée. En effet, certaines séquences peuvent montrer un changement de conformation en fonction de la concentration d'ADN utilisée [54].

En raison du débit limité et des contraintes de séquence inhérentes aux deux méthodes précédentes, d'autres techniques complémentaires comme les spectroscopies d'absorption et de fluorescence (FRET), le dichroïsme circulaire, la spectrométrie de masse à ionisation par "électrospray" (ESI-MS), la calorimétrie, la spectroscopie Raman et la mobilité électrophorétique permettent d'acquérir des informations sur la molécularité, la stabilité et le type de conformation adoptée. Certaines de ces méthodes (dénaturation thermique obtenue par spectroscopies d'absorption et de fluorescence, par dichroïsme circulaire ou par calorimétrie) permettent en plus d'accéder à des paramètres thermodynamiques et cinétiques essentiels à la caractérisation de ces structures, mais reposant souvent sur des hypothèses pas toujours validées.

D'autres techniques comme la sensibilité aux nucléases et les arrêts d'ADN polymérases permettent de déterminer la formation et l'emplacement de G4 sur une séquence longue. Les guanines impliquées dans les G-quartets peuvent être identifiées par des empreintes au diméthylsulfate (DMS). Si l'azote N⁷ d'une guanine participe à la formation d'un quartet, il est résistant à la méthylation par DMS et donc protégé du clivage. Les guanines non-clivées sont, par définition, impliquées dans la formation de G-quartets.

La propriété physique de ces structures peut être étudié par des pinces optiques (“laser tweezer”), méthode développée par l’équipe de H. Mao et S. Basu en 2009 [55]. Cette expérience de molécule unique permet de mesurer la force de dissociation de la structure.

Pour des applications en nanotechnologie, la formation de structures supramoléculaires de G4 comme les “G-wires” peut être observée au microscope à force atomique (AFM) ou au microscope électronique en transmission (MET).

2) Fonctions biologiques : arguments en faveur de l’existence de G4 in vivo

Depuis la fin des années 80, les structures en G-quadruplexes sont l’objet d’un intérêt tout particulier car elles pourraient avoir des rôles biologiques essentiels, que nous allons aborder ci-dessous.

2.1 Analyses génomiques: localisation de G-quadruplexes potentiels

Des études bioinformatiques effectuées sur le génome humain ont permis de mettre en évidence un grand nombre de séquences pouvant former des G-quadruplexes, et d’étudier leur distribution au sein du génome.

L’algorithme “*Quadspaser*” en recherchant les séquences du type $d(G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3})$ (où $N = A, T, C$ ou G) a permis d’identifier environ 376000 séquences susceptibles de former un G4 intramoléculaire (PQS pour “*Putative Quadruplex Sequences*” dans le génome humain [56]. Il est intéressant de noter que la distribution de ces séquences PQS n’est pas aléatoire dans le génome. Les PQS sont sous-représentés dans les régions codantes des gènes, mais sur-représentées dans des régions régulatrices comme les promoteurs (régions situées en amont du site d’initiation de la transcription (TSS)) (Figure 6A-B). Plus de 40% des promoteurs de gènes humains contiennent au moins un motif G4 [57]. Par ailleurs, la densité de PQS est également corrélée à la présence de sites hypersensibles à la nucléase (NHS). Ces sites sont des “fenêtres ouvertes” requises dans les promoteurs pour permettre à la machinerie de la transcription d’accéder aux séquences d’ADN. Certains gènes possèdent un nombre plus important de PQS dans leurs promoteurs, par exemple les proto-oncogènes [58]. À l’inverse, les gènes suppresseurs de tumeurs contiennent peu de PQS.

Néanmoins, l’ADN génomique est majoritairement sous forme d’une double-hélice, dans laquelle l’appariement Watson-Crick G-C, très stable, doit limiter le repliement en G4.

La formation de ces structures nécessite donc l'ouverture du duplexe, qui se produit au cours de la réplication (Figure 6B), de la transcription ou de la recombinaison. Les régions du génome sous forme simple-brin, telles que l'extrémité des télomères (Figure 6B) ou les ARN, sont énergétiquement plus favorables à la formation de ces structures. La présence de structures en G4 et leurs implications éventuelles au niveau des télomères ne seront pas abordées dans ce chapitre. Cet aspect sera traité dans le chapitre V avec les études de séquences télomériques. Les ARN, quant à eux, sont connus pour former des structures comme les triplexes ou les G-quadruplexes. Il a été reporté que plus de 3000 ARNm possèdent des motifs G4 dans leur région 5' non traduite (5'UTR pour "*UnTranslated Region*") (Figure 6A-B) [59].

Des études bioinformatiques récentes [60] ont observé un enrichissement de motifs G4 dans plusieurs régions régulatrices des gènes, y compris les sites de fixation de facteurs de transcription, des sites hypersensibles à la nucléase, des îlots CpG (au niveau du promoteur) et au niveau des insulateurs (régions permettant de limiter l'effet activateur ou répresseur de la transcription du gène adjacent). D'autres régions montrant un potentiel de formation de G4 sont les régions de commutation de classe des immunoglobulines, l'ADN ribosomique et les répétitions minisatellites.

La localisation de PQS dans des régions spécifiques du génome (ADN et ARN) suggère que les G-quadruplexes pourraient jouer des rôles dynamiques dans plusieurs voies cellulaires. Certaines de ces fonctions seront développées dans le paragraphe suivant.

2.2 Régulation génique

Comme nous l'avons mentionné précédemment, les G-quadruplexes sont susceptibles de se former dans de nombreuses régions régulatrices de la transcription (îlots CpG, sites hypersensibles à la nucléase, ...) et de la traduction (5'UTR, introns) du gène. Ceci implique que la présence de ces structures pourrait jouer un rôle dans la régulation génique.

En effet, il a été montré *in vitro* que la formation et la stabilisation de G4 au niveau du promoteur d'un proto-oncogène, comme *c-myc* [61, 62] ou *c-kit* [63], pourrait inhiber leur transcription. La surexpression de ces proto-oncogènes est souvent corrélée à la prolifération des cellules cancéreuses. Par conséquent la répression de la transcription de ces gènes par la formation du G4 aurait un effet antitumoral. Le développement de ligands permettant la stabilisation de ces structures constitue donc une nouvelle stratégie anticancéreuse. En 2009, l'équipe de L.H. Hurley a montré que le promoteur de hTERT contient une séquence riche en guanines susceptible de former une structure en G-quadruplexe. La formation de G4 à cet

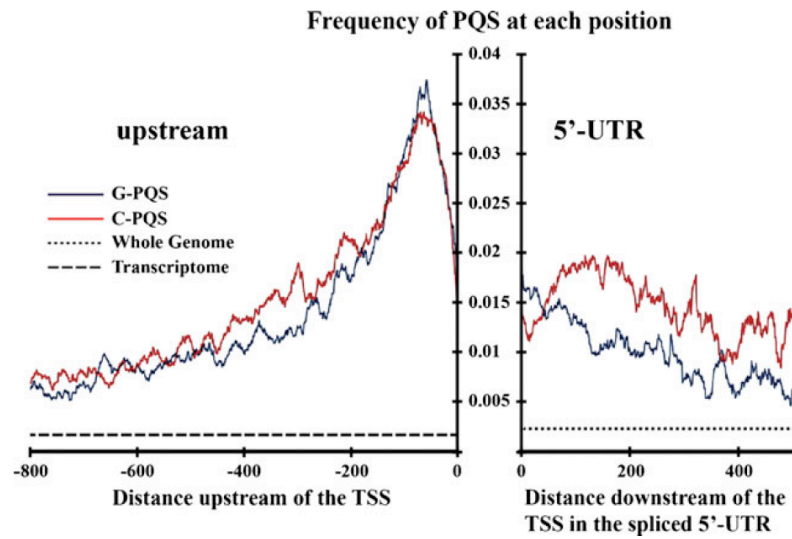
endroit pourrait masquer le site de fixation de Sp1 (pour “*specific protein-1*”) et inhiberait ainsi l’activité du promoteur de hTERT [64]. D’autres exemples de gènes ayant des motifs G4 dans leur promoteur sont : *VEGF*, *HIF1- α* , *Bcl-2*, *Ret* et *KRAS*. Bien que la fonction cellulaire de la formation de G4 dans certains de ces promoteurs ne soit pas clairement caractérisée, leur formation, leurs structures et leurs propriétés biophysiques *in vitro* sont bien connues (pour revue : [65]).

Par ailleurs, la présence de motifs G4 dans les minisatellites, connus comme des sites d’hypermutableté et d’instabilité génomique, pourrait jouer un rôle dans la régulation génique. En effet, la formation de G4 dans la région polymorphique associée au gène de l’insuline, ILPR (pour “*Insulin-Linked Polymorphic Region*”) pourrait, via des interaction avec la protéine insuline, moduler la transcription du gène [66, 67]. De plus, la stabilisation de ces structures dans les minisatellites humains (notamment, MS1 et CEB1) par certains ligands de G4 pourrait expliquer la forte instabilité génomique observée dans ces régions [68-70].

La formation de G-quadruplexes semble également avoir un rôle dans la régulation de la traduction des ARNm. La formation d’un G4 d’ARN intramoléculaire dans le 5’UTR du proto-oncogène NRAS a été démontrée par l’équipe de S. Balasubramanian [59]. Le motif G4 à ce niveau inhibe la traduction du gène en question. Par ailleurs, de nombreuses études suggèrent que des motifs G4 dans les introns pourraient jouer un rôle dans la régulation de l’épissage des ARNm [71, 72]. La présence de plusieurs PQS dans l’intron 6 du gène de hTERT, et l’effet d’un ligand de G4 sur l’augmentation de l’épissage de cet intron suggèrent un rôle potentiel de ces structures dans la maturation des ARN [72]. Nous avons également remarqué que la formation de G4 dans l’intron 3 de l’ARNm de TP53 (gène suppresseur de tumeurs) permettait la régulation de l’épissage de l’intron 2 du même gène [71]. Récemment, il a été montré que la formation de G4 dans la région 5’UTR de l’ARNm de TRF2 réprimait la traduction de celui-ci [73]. Les G-quadruplexes d’ARN possèdent une stabilité remarquable : ils pourraient donc jouer des rôles essentiels au niveau cellulaire, notamment dans la régulation de la maturation et/ou de la transcription des ARNm. Cependant, malgré leur stabilité, il est possible que d’autres types d’appariements (G-C et G-U) puissent empêcher l’association des guanines entre elles. De plus, l’interaction avec des protéines reconnaissant de régions consensus riches en G pourrait aussi entraver la formation de G4. L’étude de la formation de G4 dans un contexte cellulaire semble donc essentielle pour valider l’implication de ces structures dans de telles régulations.

Certaines de ces séquences PQS ont été étudiées et utilisées pour le criblage de ligands de G4 sur plaque présenté dans le chapitre VI.

A.



B.

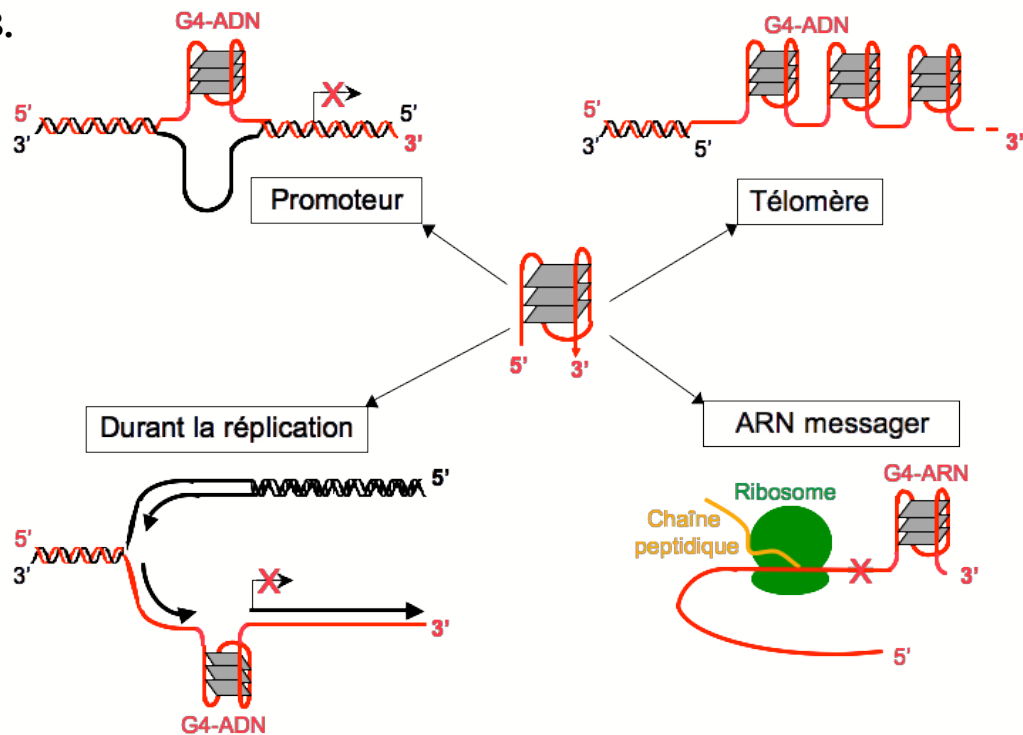


Figure 6. Localisation de G4 potentiels et leurs fonctions biologiques

(A) Forte densité de PQS ("Putative Quadruplex Sequences") en amont du site d'initiation de la transcription (TSS) de l'ADN (région promotrice) et au niveau du 5'UTR de l'ARNm [74].

(B) Implication possible de G4 in vivo, notamment au niveau du promoteur, du télomère, de l'ARNm ou durant la réplication.

2.3 Protéines reconnaissant les G4

L'existence de protéines naturelles interagissant spécifiquement avec les structures en G-quadruplexes constitue un argument indirect de l'existence des G4 *in vivo*. On peut classer ces protéines dans plusieurs groupes en se basant sur leurs fonctions distinctes.

Le premier groupe est constitué de protéines permettant la stabilisation de ces structures. On peut citer la **protéine de nucléocapside (NCp)** du VIH-1 (Virus de l'Immunodéficience Humaine de type 1) qui est impliquée dans la stabilisation du génome viral au cœur du virion et participe à la machinerie pré-intégrative du cycle viral [75]. On peut mentionner également la protéine **uqTBP25** (pour "*unimolecular quadruplex Telomere-Binding Protein 25*") qui est potentiellement impliquée dans les interactions avec l'ADN télomérique [76]. Un autre exemple intéressant dans ce groupe est la **thrombine humaine**. L'aptamère de la thrombine (court acide nucléique sélectionné contre la thrombine humaine par l'approche SELEX) formant un G4 intramoléculaire est capable de lier spécifiquement la **thrombine humaine** (en dehors de son site actif). L'interaction de l'aptamère avec sa cible permet de diminuer l'agrégation des plaquettes et d'empêcher l'interaction de la thrombine avec le fibrinogène, en lui conférant des propriétés anticoagulantes et anti-thrombotiques [81, 82]. Récemment, il a été montré que l'interaction entre la **protéine Huntingtin** et une séquence riche en G capable de former un G-quadruplexe inhibait l'agrégation de la protéine qui serait responsable de la maladie neurodégénérative de Huntington [77].

Les protéines induisant la formation de structures intra- et intermoléculaires de G4 sont classées dans le deuxième groupe. On peut citer trois exemples: **RAP1** (pour "*Repressor Activator Protein-1*") [78], **Hop1** ("*Human Osteogenic Protein-1*") [79] et la **topoisomérase I humaine**[80].

Des protéines permettant de dérouler (hélicases), déstabiliser ou cliver les structures de G4 constituent les deux autres groupes. Les hélicases ATP-dépendantes de type **RecQ**, en particulier les hélicases humaines **BLM** et **WRN**, peuvent résoudre des structures en G-quadruplexes très stables [83] et faciliter l'activité de l'ADN polymérase au travers de régions contenant des répétitions de trinuécléotides d(CGG) liées au syndrome de l'X fragile [84, 85]. La protéine hélicase **Pif1** de levure est impliqué dans la maintenance de la stabilité de minisatellites CEB1, insérés dans le génome de *S. cerevisiae*, en résolvant les G4 formés dans ces régions [69, 86, 87].

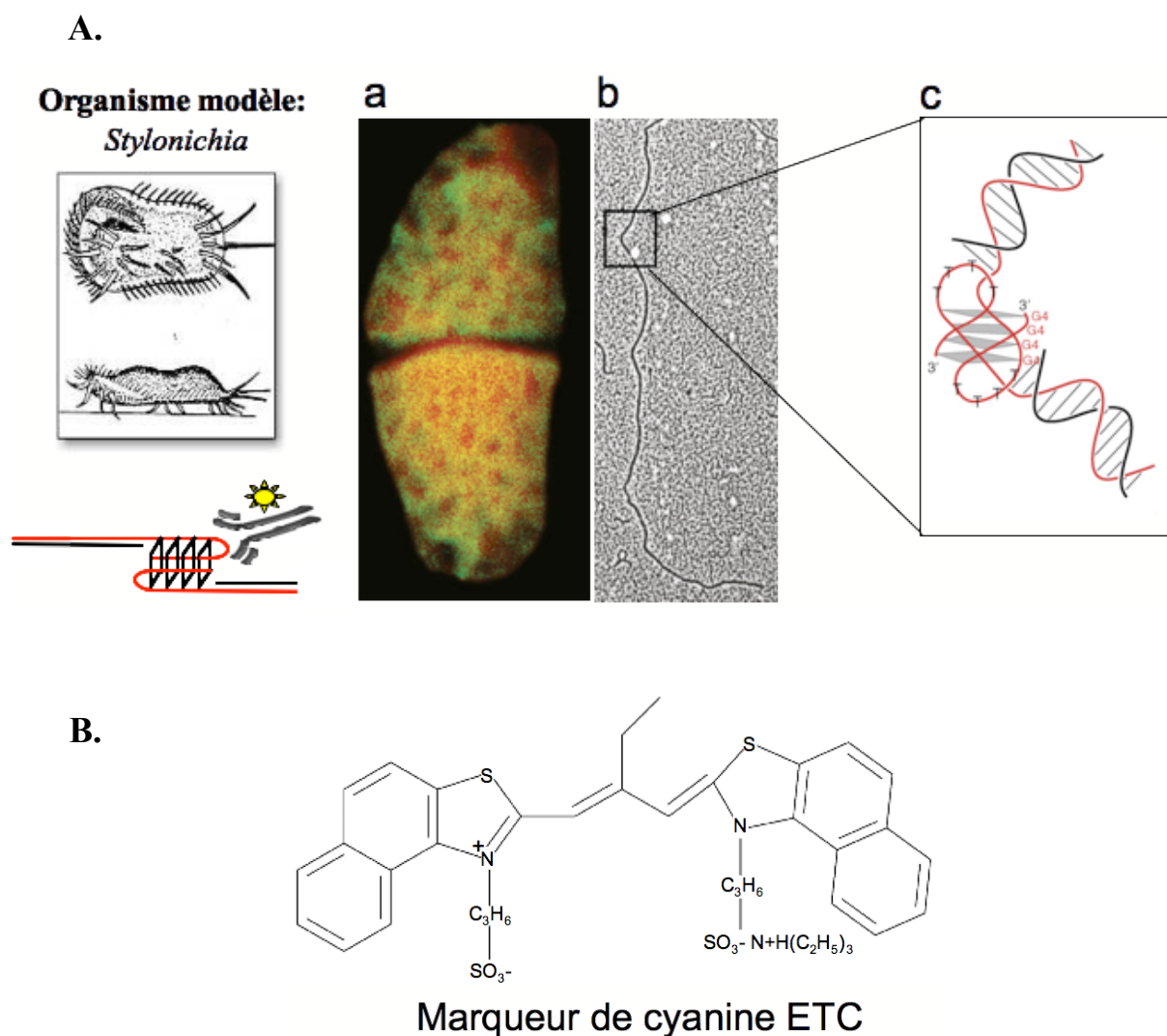


Figure 7. Existence de G-quadruplexes in vivo

(A) Preuve directe de l'existence de G-quadruplexes dans les macronoyaux des ciliés *Stylonia* [88]: (a) Image d'un macroyau de cilié observée au microscope confocal. L'image montre une superposition du marquage immunohistochimique (utilisation d'anticorps : Sty3 et Sty49) de G4 télomérique antiparallèle (en jaune) et celui de l'ADN (en rouge). Le marquage de G4 n'est pas observé au niveau de la bande de réplication. (b) Image observée au microscope électronique de l'ADN extrait du macroyau. Le carré indiquerait la position de formation de G4 reliant les deux chromosomes. (c) Structure de G4 formée entre deux télomères et confirmée par une analyse biochimique. (B) Formule chimique de la cyanine ETC, marqueur fluorescent de G4 [89].

Certaines protéines, en se fixant préférentiellement sur la forme simple-brin peuvent déstabiliser les quadruplexes. Les protéines **hnRNP** (pour “*heterogenous nuclear RiboNucleoProtein*”) ont été identifiées comme des protéines déstabilisant ces structures. Parmi ces protéines hnRNP, on peut citer **UP1** [90] (correspondant à l'extrémité N-terminale de hnRNP A1) qui est impliquée dans la maintenance télomérique. La protéine télomérique **hPOT1** (pour “*human Protection Of Telomeres I*”) [91] et la protéine **hRPA** (pour “*human Replication Protein A*”) [92] déstabilisent les G4 intramoléculaires formés par la séquence du télomère humain.

Les nucléases clivant les structures en G4 constituent le dernier groupe de ce classement. Nous pouvons citer deux nucléases, la **topoisomérase II** [93] et la **GQN1** [94] qui sont connues pour couper le G4 tétramoléculaire issu des régions de commutation de classe des immunoglobulines. La **GQN1** clive spécifiquement l'ADN G4, elle n'a pas d'effet sur l'ARN G4. La protéine **Kem1p**, une nucléase identifiée dans *S. cerevisiae*, peut cliver l'ADN G4 [95, 96] et serait impliquée dans la sénescence cellulaire et le raccourcissement des télomères [97].

2.4 Preuve directe de l'existence de G-quadruplexes *in vivo*

Les études sur les ciliés ont fourni la preuve la plus directe de l'existence de G-quadruplexes bimoléculaires au niveau des télomères. Les macronoyaux des ciliés, contenant des millions de minichromosomes, offrent ainsi un modèle privilégié pour l'étude des télomères. La formation *in vivo* de G-quadruplexes antiparallèles par les séquences télomériques de ciliés a été montrée par microscopie confocale, en utilisant des anticorps spécifiques de G4 [98] (Figure 7A). Il a également été démontré que ces structures étaient résolues au cours de la réplication, ceci est caractérisé par une absence de marquage de la bande de réplication des ciliés par un des anticorps.

Jusqu'à présent, l'existence *in vivo* de G4 au niveau des télomères d'organismes, autres que les ciliés, n'est pas démontrée ; le nombre moins important de télomères rend la révélation d'éventuelles G4 *in vivo* plus difficile. Néanmoins, le développement de molécules de G4 fluorescents, comme la cyanine ETC [89, 99] (Figure 7B) ou radioactifs (360A tritié) [378] pourrait permettre de détecter la formation de G4 dans les télomères humains. Cela constitue une nouvelle approche prometteuse pour l'identification de motifs ADN spécifiques.

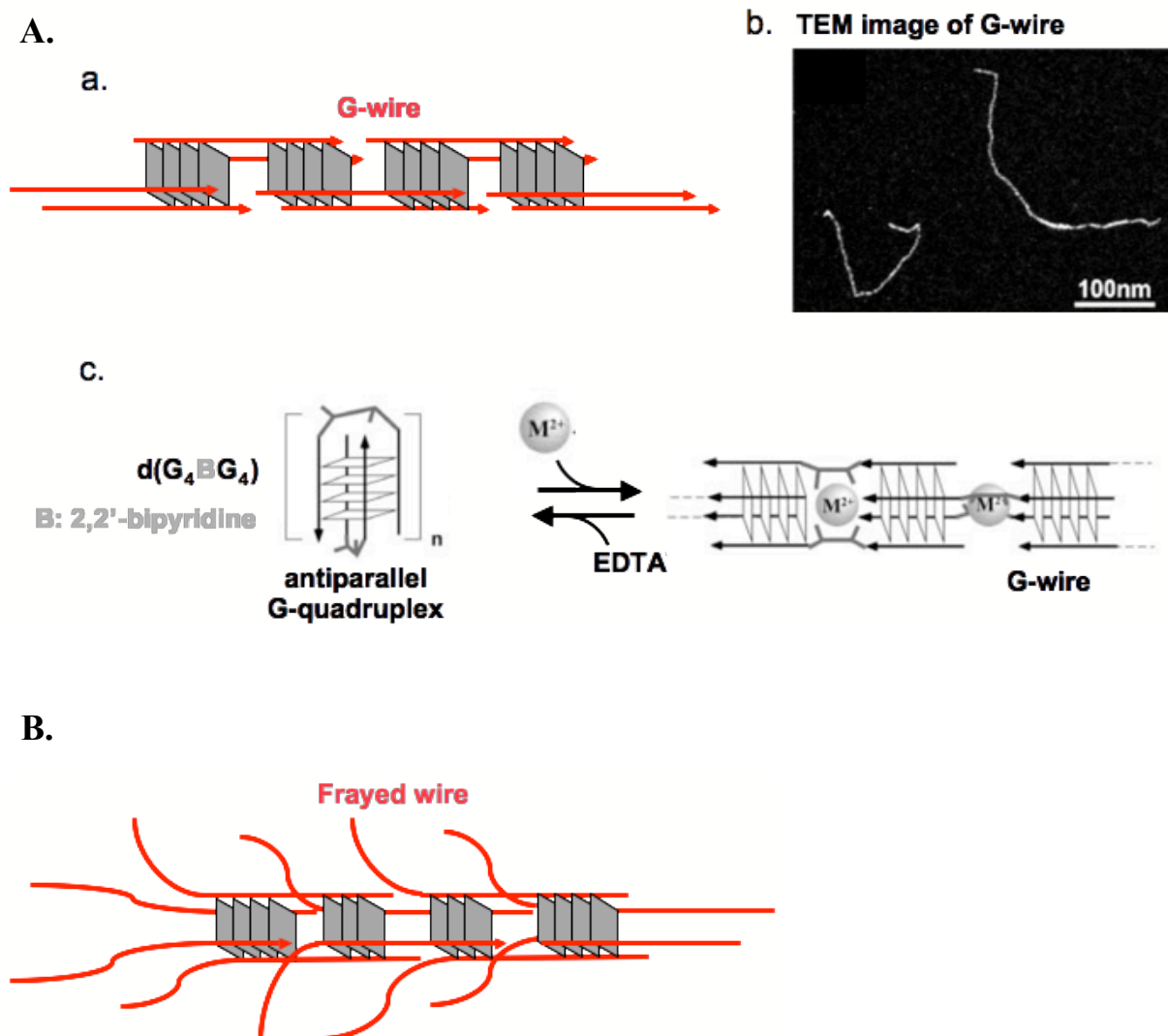


Figure 8. Structures supramoléculaires formées par des oligonucléotides riches en G

(A) G-wires : (a) Représentation schématique “G-wire” formé par des séquences $d(G_4T_2G_4)$ [100]. (b) Image au microscope électronique en transmission (MET) de “G-wires” formés par des poly-dG non-modifiés [101]. (c) Schématisation de “G-wire” modifié (insertion d’un espace éthylène glycol et une bipyridine centrale). (B) Représentation schématique de “Frayed wire” formé par des séquences $d(A_{15}G_{15})$ [102].

3) Applications pour les nanotechnologies et biotechnologies

Outre les implications biologiques de ces structures, les propriétés d'auto-assemblage des G-quadruplexes, leur stabilité et leur rigidité s'avèrent particulièrement intéressantes pour les nanotechnologies et biotechnologies. Nous présenterons ci-dessous quelques exemples d'applications potentielles concernant la fabrication de nanofils pour l'électronique moléculaire, de machines moléculaires, de nanopores ou de nanoparticules.

3.1 Nanostructures d'ADN : “*G-wires*”, “*Frayed wires*” ou ADN synaptiques

L'idée de former de longues structures interconnectées basées sur la formation de G-quadruplexes a été formulée dans les années 90 [103]. Les équipes de E. Henderson et R.D. Sheardy ont montré par gel d'électrophorèse qu'en présence de K^+ et de Mg^{2+} , la séquence télomérique du cilié d(G₄T₂G₄) était capable de former une structure supramoléculaire en fil, appelées “*G-wire*” [100, 104-106]. Ces polymères mesuraient entre 10 et 1000 nm de long et 2,5 nm de diamètre, en accord avec le diamètre d'un G₄. Une fois formés, ces polymères sont extrêmement stables et peuvent résister à des conditions dénaturantes classiques (40% formamide / 7 M urée, 100°C). Ces superstructures d'ADN parallèles sont formées de G-quadruplexes avec des appariements décalés : l'extrémité 5' d'un duplexe G-G est relié à l'extrémité 3' du duplexe G-G opposé (Figure 8A.a) [100]. Les *G-wires* peuvent également être obtenus par des séquences poly-d(G) comme montré sur la figure 8A.b [101].

N. Sugimoto et ses collaborateurs ont introduit un espaceur éthylène glycol et une bipyridine centrale entre les deux blocs de G, à la place des thymines [107, 108]. La bipyridine permettant la chélation de cation divalent, notamment Ni^{2+} , induit une transition du G₄ antiparallèle bimoléculaire et la formation de “*G-wires*” modifiés (Figure 8A.c). L'ajout d'EDTA chélate Ni^{2+} et permet une réversibilité du système. La taille de ces structures peut atteindre près de 200 nm. Néanmoins ces fils ne sont pas très rigides, ceci est dû à un manque de rigidité des espaceurs éthylène glycols introduits.

Les structures en G-quadruplexes, et en particulier les “*G-wires*” constituent des approches séduisantes pour la construction de fils moléculaires pour des applications en microélectronique. En effet, la bonne conductivité supposée de ces structures pourrait résulter d'un transport électronique créé par la coordination de cations centraux avec les fonctions carbonyles des G-quartets. Il a été montré par microscopie à force électrostatique que des ADN G₄ formés par des poly-dG en présence de K^+ étaient polarisables, tandis que des ADN

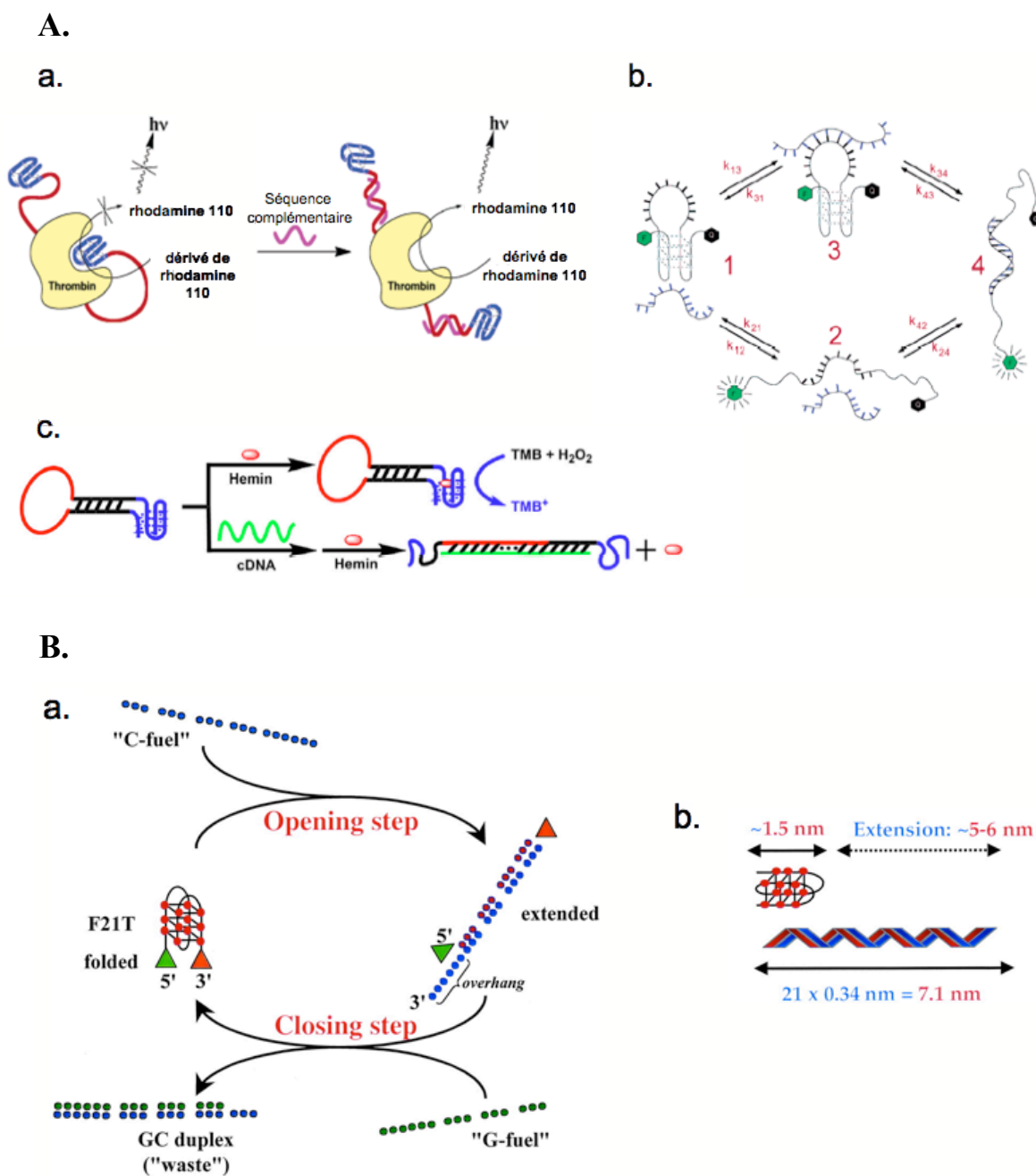


Figure 9. Biocapteurs et nanomachines basés sur la formation de G-quadruplexes

(A) Biocapteurs de séquence d'ADN: (a) détection de séquence d'ADN par l'activité catalytique du complexe aptamère-thrombine [110] ; (b) balise moléculaire (MB) fluorescente basée sur l'ouverture et la fermeture de G4 [111] ; (c) balise moléculaire catalytique (MCB) basée sur un désoxyribozyme structuré en G4 [112]. En bleu, nous avons la partie permettant la formation de G4. En noir et en rouge, ce sont la tige et la boucle de la balise. En absence de la séquence complémentaire (cADN), l'hémine se lie à la structure en G4 et permet la réaction d'oxydation de TMB (pour "3,3',5,5'-TetraMethylbenzidine Sulfate"). La présence de cADN empêche la formation de G4. L'hémine ne peut plus fixer la structure et ne permet donc pas la réaction d'oxydation. (B) Schéma d'une machine moléculaire basée sur la formation et la dissociation d'un G4: (a) cycle d'ouverture et de fermeture de la machine ; (b) mouvement d'extension généré par l'ouverture du G4 [113].

en double-hélice ne le sont pas [109]. Ceci peut s'expliquer par une plus grande rigidité et une plus grande résistance aux forces de surfaces des structures de G4. L'ADN G4 pourrait donc être un meilleur conducteur et candidat que l'ADN duplexe pour la fabrication de fils moléculaires pour la microélectronique. Cependant les "*G-wires*" sont des structures dont la longueur, l'orientation et la molécularité sont relativement difficiles à contrôler, leur production et leur utilisation peuvent s'avérer plus complexes que celle d'un duplexe classique d'ADN. Pour cela, A.B. Kotlyar et ses collaborateurs ont biotinylé les séquences poly-dG afin de réunir les 4 brins par interaction avidine-biotine, permettant ainsi d'augmenter la chance de former des structures tétramoléculaires [114]. Cette nouvelle approche permet donc de contrôler la molécularité des structures formées.

MacGregor et ses collègues ont identifié des structures allongées formées par les séquences $d(A_{15}G_{15})$ et $d(T_{15}G_{15})$ en présence de K^+ [102]. Ils ont appelé ces structures "*Frayed wires*" pour différencier des "*G-wires*" (Figure 8B). Bien qu'ils soient moins étudiés que les "*G-wires*", les "*Frayed wires*" sont tout aussi intéressants. Ces structures sont plus stables en Na^+ qu'en K^+ ou en NH_4^+ [102]. Les queues poly-dT ou poly-dA permettent d'attacher des objets, tels que des enzymes ou d'entrevoir un moyen de connecter ces structures entre elles.

L'ADN synaptique n'est pas abordé dans ce paragraphe, puisque cette structure est utilisée dans un des projets de cette thèse et est présentée dans l'introduction du chapitre IV.

3.2 Formation de biocapteurs et de nanomachines avec l'ADN G-quadruplexe

Les aptamères d'ADN ont non seulement des potentiels thérapeutiques, mais peuvent également être utilisés pour la chimie bioanalytique. Notamment l'ADN G4 formé par la séquence TBA $d(G_2T_2G_2TGTG_2T_2G_2)$ (aptamère de la thrombine) est très utilisé pour le développement de biocapteurs. La détection de séquence d'ADN par les interactions TBA-thrombine et l'utilisation de "balise catalytique" a été exploitée par Willner et ses collègues (Figure 9A.a) [110, 115]. La protéine thrombine est liée covalamment à l'oligonucléotide contenant la séquence TBA. En absence de la séquence complémentaire, la séquence TBA forme un G-quadruplexe qui va bloquer le site actif de l'enzyme et empêcher ainsi la réaction catalytique. La présence de la séquence complémentaire entraîne la dissociation du complexe thrombine-aptamère, ceci permet l'hydrolyse d'un peptide lié à la rhodamine 110. Cette réaction libère un fluorophore, la rhodamine 110 qui est quantifiable par spectroscopie de fluorescence.

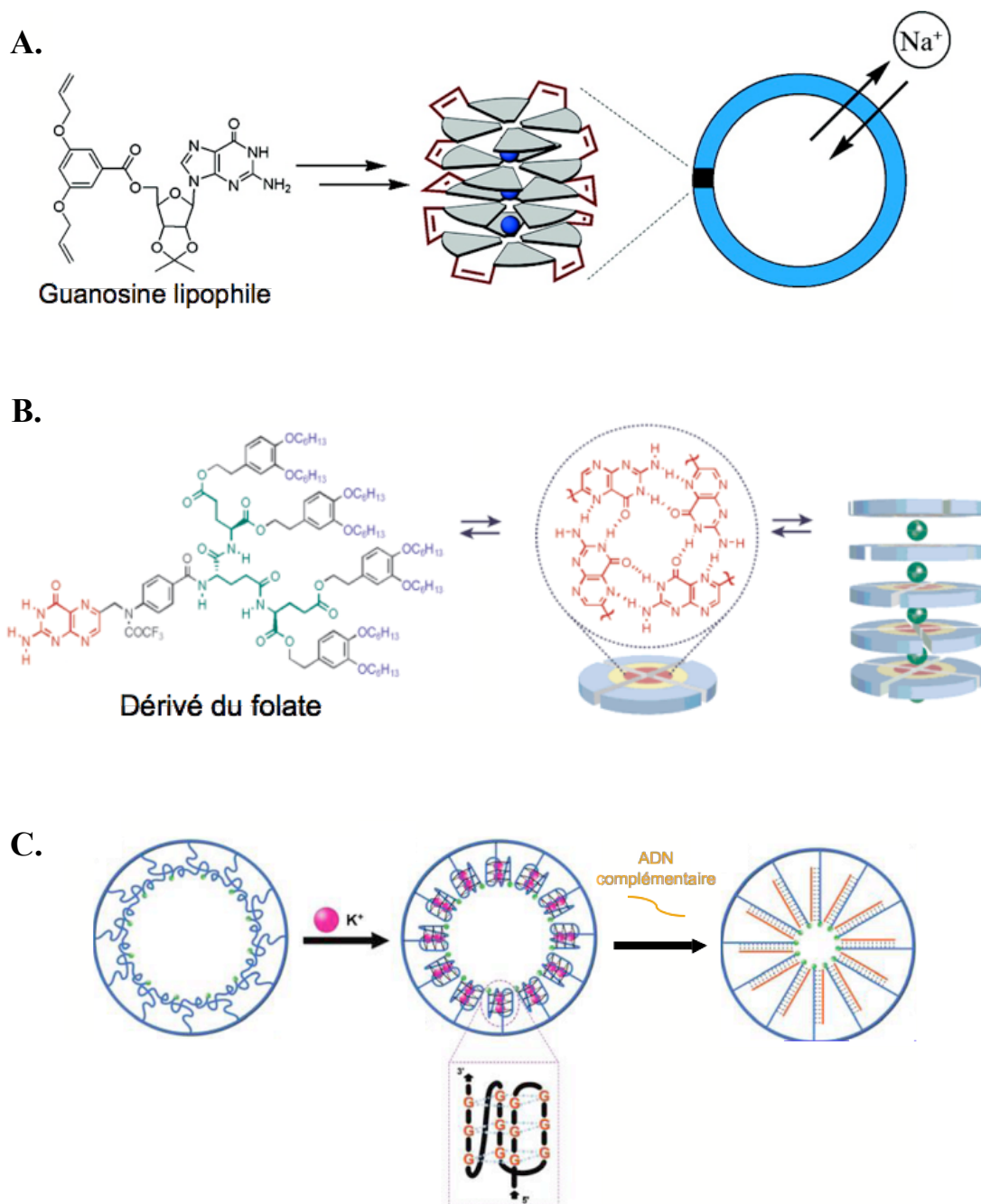


Figure 10. Canaux ioniques à base de G-quadruplexes

(A) Canal ionique formé par auto-assemblage de guanines lipophiles [116]. (B) Exemple de canal ionique formé à partir de dérivés du folate [117]. Les longues chaînes carbonées de l'acide folique permet de rigidifier l'empilement des quartets. (C) Immobilisation de l'ADN G4 sur la surface interne du nanopore [118]. Le diamètre du pore est contrôlé par l'ajout de K^+ et de séquences d'ADN complémentaires.

Une balise moléculaire (MB) fluorescente basée sur la structure de G4 a été mise en place par notre équipe en 2006 [111]. Le système consiste en un G4 avec une boucle centrale longue (12 à 21 nucléotides). En présence de la séquence complémentaire, la boucle centrale forme un duplexe avec la séquence, ce qui entraîne l'ouverture du G-quadruplexe. Cette ouverture de G4 est détectée par une augmentation de l'émission de fluorescence qui est due à l'éloignement entre le fluorophore et le quencher (Figure 9A.b).

Récemment, l'équipe de E. Wang a construit des portes logiques en utilisant une balise moléculaire catalytique (MCB) basée sur un désoxyribozyme structuré en G4 [112]. La formation et la dissociation du G4 sont liées à une réaction chimique qui permet de changer la couleur de la solution, visible à l'œil nu et mesurable par spectroscopie d'absorption UV-vis (Figure 9A.c).

L'ouverture et la fermeture d'un G4 intramoléculaire formé par la séquence télomérique humaine ont également permis de créer une machine moléculaire [113, 119]. L'ajout d'un brin complémentaire ("*C-fuel*") de la séquence télomérique induit un mouvement d'extension causé par l'ouverture de la structure de G4. À l'inverse, l'ajout d'un oligonucléotide ("*G-fuel*") formant un duplexe plus stable avec "*C-fuel*" permet le repliement de la séquence télomérique en G4, conduisant à un mouvement de contraction. Ce mouvement d'extension/contraction est, en théorie, sur une distance de 5,5 nm : de 1,5 nm pour le G4 à 7 nm pour un duplexe (Figure 9B).

3.3 Canaux ioniques à base de G-quadruplexes

Nous avons mentionné précédemment que la dynamique des cations au sein d'un G4 faisait de lui un canal ionique potentiel. En effet, Jeffery Davies et ses collaborateurs ont construit des canaux ioniques artificiels à partir d'empilement de G-quartets modifiés lipophiles [116]. Les nucléosides modifiés en 5' par un groupement (3,5-bis(allyloxy)benzoyl)-2',3'-isopropylidene) permettent de former des liaisons covalentes entre les G-quartets (réaction de métathèse oléfinique) et donc de rigidifier la structure formée (Figure 10A). Cette structure de G4 lipophile permet un transport efficace de cations, ici Na^+ , au travers d'une bicouche lipidique mimant une membrane. La rigidification de l'empilement peut également être obtenu par association de longues chaînes carbonées qui contraignent l'ensemble (Figure 10B) [117, 120].

Par ailleurs, X. Hou et ses collègues ont construit un nanopore dont la surface interne est recouverte de séquences d'ADN pouvant former des G4 [118]. Le diamètre du pore est influencé par la présence de K^+ . L'ajout de K^+ permet un mouvement de contraction dû à la

formation de G4 et à l'inverse, l'ajout d'une séquence complémentaire induit la formation de duplexe et donc un mouvement d'extension qui réduit le diamètre du pore (Figure 10C).

3.4 Pontage de nanoparticules via la formation de G4

L'équipe de A. Kotlyar a montré que l'insertion de G-quadruplexes tétramoléculaires parallèles contenant des résidus phosphorothioates à chaque extrémité permettait l'assemblage de nanoparticules d'argent [121] ou d'or [122] et pouvait entraîner la formation de structures en forme de fleur. Ces nanoparticules reliées par des structures en G4 permettent un assemblage stable sans entraîner d'agrégation même à haute concentration en sel (0,2 M) (Figure 11).

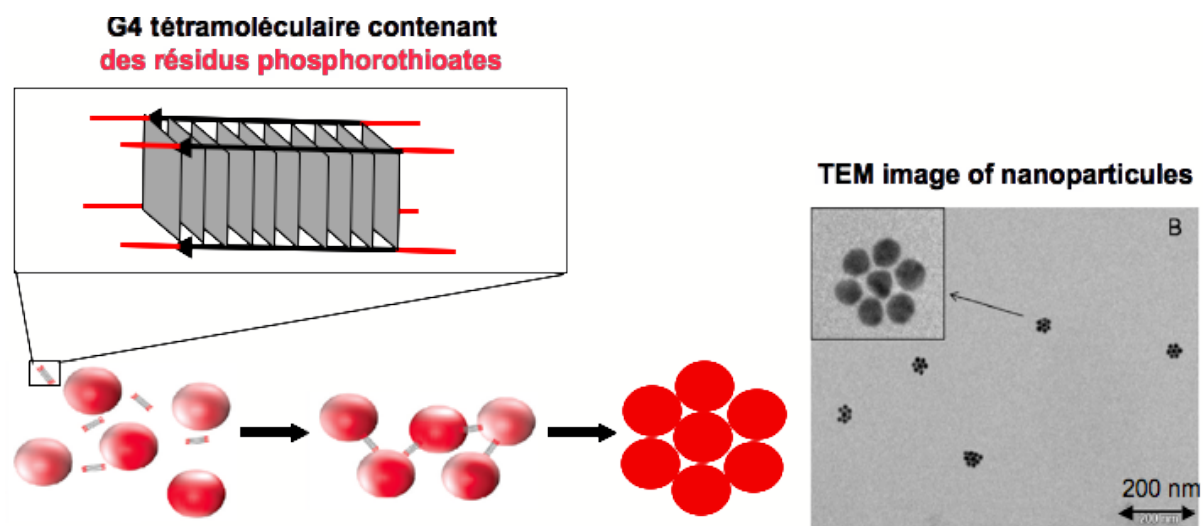


Figure 11. Pontage de nanoparticules d'or via des G4 tétramoléculaires [122]

À droite: Schématisation de nanoparticules d'or (rond rouge) reliées par des structures de G4 tétramoléculaires parallèles (en noir) contenant des résidus phosphorothioates (en rouge). Les résidus phosphorothioates permettent l'ancrage de la structure dans les nanoparticules. *À gauche:* image de nanoparticules d'or en forme de fleurs observée au MET.

4) Objectifs de ce travail

Comme nous l'avons montré dans les paragraphes précédents, les structures en G-quadruplexes sont susceptibles de se former dans le génome humain et d'avoir un rôle essentiel au niveau cellulaire ainsi que des applications en nanotechnologies. Il est donc important de bien comprendre la formation de ces structures. De nombreuses études structurales ont été réalisées ces dernières années ; cependant, compte tenu du polymorphisme structural des G4, en particulier des G4 intramoléculaires, la tâche s'avère difficile et est donc toujours d'actualité. En effet, il n'existe, jusqu'à présent, aucune règle qui puisse prédire la

conformation de G4 à partir d'une séquence primaire. De plus, pour des applications en nanotechnologies, il est nécessaire de pouvoir contrôler et anticiper la formation de ces structures. Tout cela nous a amené à réaliser ces travaux de thèse.

Dans la première partie des résultats (Chapitre III), nous avons étudié la structure, la stabilité et la cinétique d'association de G-quadruplexes tétramoléculaires modifiés, en particulier la 8-méthyleguanine et la guanine en série L ("*levogyre*"). Ces résultats ont contribué à la description des mécanismes d'association de ces espèces qui est encore limitée. Les connaissances acquises ont été utilisées pour les applications en nanotechnologie, notamment pour la formation de l'ADN synaptique sur un support d'ADN origami. Cet aspect sera présenté dans le chapitre IV.

Nous nous sommes ensuite intéressés à l'étude structurale des séquences biologiques. Dans le chapitre V, nous avons fait une étude comparative de la structure et de la stabilité des séquences télomériques de différents organismes allant des protozoaires aux vertébrés en passant par les levures, les plantes et les invertébrés. Cette étude a permis de comparer les structures et la stabilité de ces différentes séquences télomériques et a participé à l'enrichissement des données nécessaires au développement d'un algorithme prédisant la stabilité des G4.

La diversité des conformations adoptées par une séquence riche en G et la présence de ces séquences dans nombreuses régions du génome rendent difficile le développement de ligands de G4 efficaces en thérapie. En effet, le grand défi d'aujourd'hui est de trouver des molécules qui peuvent cibler une région spécifique du génome. Nous avons décrit dans le dernier chapitre des résultats (chapitre VI), le développement d'une méthode de criblage rapide de l'interaction de ligands avec différentes conformations de G4. Le développement de G4-FID sur plaque constitue une méthode simple, rapide et peu coûteuse pour l'identification de ligands spécifiques d'une conformation particulière de G4. Nous discuterons également dans ce chapitre les avantages et les limites de la technique.

Matériel et Méthodes

Chapitre II. Matériel et Méthodes

1) Produits

1.1 Les tampons

Pour les expériences de spectroscopie, nous utilisons généralement des tampons qui n'absorbent pas ou peu dans le domaine de longueurs d'onde étudiées, entre 200 nm et 500 nm, et qui n'ont pas de propriétés de fluorescence notables. D'autre part, les expériences de dénaturation thermique nécessitent des tampons dont le pH varie peu avec la température. Le tampon composé par le couple acide faible/base faible, acide cacodylique/cacodylate ($pK_a = 6,27$ à 25°C) possède ces trois propriétés, avec un pouvoir tamponnant suffisant entre pH 5,0 et 7,4. Il constitue donc le tampon de choix pour les études spectroscopiques à pH physiologique en absorbance, dichroïsme circulaire et fluorescence.

Le tampon que nous appelons dans ce manuscrit cacodylate de lithium, utilisé à 10 mM et pH 7,2, est obtenu par basification d'une solution d'acide cacodylique ($(\text{CH}_3)_2\text{AsO}_2\text{H}$) avec une solution d'hydroxyde de lithium (LiOH). La quantité d'ions Li^+ ajoutés au cours de cette opération est sensiblement égale à la quantité de cacodylate, 8,9 mM pour une solution à 10 mM en cacodylate à pH 7,2. L'ion Li^+ n'a pas ou peu d'effet stabilisant sur les G-quadruplexes, certainement en raison de son plus faible rayon ionique par rapport à la cavité formée par un quartet de guanines. L'utilisation de ce contre-ion nous permet donc de négliger son effet sur la stabilisation de G-quadruplexes et d'étudier l'effet d'autres ions, comme le sodium et le potassium, sur la stabilité et conformation de ces structures, sans changer de tampon. Des expériences réalisées préalablement à mon arrivée dans l'équipe avaient montré que l'addition de quantités croissantes de lithium ne modifiait pas ou très peu le T_m d'un G-quadruplexe dans une solution contenant du sodium ou du potassium. L'utilisation de ce tampon a donc remplacé progressivement celle du sel commercial de cacodylate de *sodium* qui introduisait systématiquement 10 mM d'ions Na^+ , pouvant ainsi jouer un rôle important dans la stabilité des structures.

M: 8-méthyle-2'-déoxyguanosine; *FAM*: 6-carboxyfluorescéine; *en rouge*: les guanines pouvant être impliquées dans la formation des G-quartets.

1.2 Les oligonucléotides

Les oligonucléotides utilisés sont synthétisés soit par la société *Eurogentec* (Liège, Belgique), soit par nos collaborateurs (équipe d'Aldo Galeone, Université de Naples, Italie et Bernard Rayner, INSERM U869, Université Bordeaux Segalen, France). Toutes les synthèses sont réceptionnées après déprotection sous forme lyophilisée, et sont resuspendues dans de l'eau bidistillée sauf les ADN modifiés et les séquences d'agrafes utilisées pour l'ADN origami qui sont directement livrés en solution.

Une fois resuspendus, leur concentration est déterminée par spectroscopie d'absorption à 260 nm (cf. : §II.2.1.3 a) en utilisant les coefficients d'extinction molaire fournis par le fabricant et en utilisant la loi de Beer-Lambert (voir page suivante).

Tous les stocks d'oligonucléotides sont ensuite conservés à -20°C dans de l'eau bidistillée pour éviter l'introduction d'ions non désirés, pouvant ainsi influencer la stabilité et la cinétique de formation de certaines structures. La qualité (pureté) des oligonucléotides est estimée pour certaines synthèses par électrophorèse en conditions dénaturantes.

1.3 Les ligands

Les molécules chimiques testées et caractérisées pour leur capacité à lier des structures particulières d'ADN ont des origines diverses. PIPER est commercialisé par Calbiochem et BRACO19 a été synthétisé dans le cadre de l'ACI « Médicaments et Cibles thérapeutiques » par la société Sanofi-Aventis. Les autres molécules testées ont été synthétisées par les chimistes du groupe de Marie-Paule Teulade-Fichou à l'Institut Curie. Toutes les molécules sont stockées à -20°C dans 100% DMSO.

2) Spectroscopies

2.1 Absorbance

2.1.1 Matériel

Les expériences de spectroscopie d'absorption ont été réalisées sur des spectrophotomètres Uvikon 940 (Kontron), Uvikon XL et XS (Secomam) présentant un passeur de 2 x 6 positions permettant l'analyse de 9 échantillons, 2 références et une sonde température qui peut être thermostaté par cryostat externe. Pour les mesures d'absorbance dans l'UV à l'aide de ces spectrophotomètres, des cuves en quartz (Hellma, France) de différents parcours optiques ($l = 1$ et $0,2$ cm) et contenance (600 μ l ou 300 μ l respectivement)

peuvent être utilisées. L'utilisation d'un Nanodrop ND-1000 a également permis ponctuellement la détermination rapide et économique de concentrations d'oligonucléotides pour des applications en biologie moléculaire ou en nanotechnologie (ADN origami).

2.1.2 Principe

En fonction de sa longueur d'onde, une radiation UV-visible (200 - 700 nm) traversant une solution est susceptible d'être absorbée par celle-ci (Figure 12). L'absorption d'une telle radiation correspond à une transition de la molécule absorbante de son état électronique fondamental à un état électronique excité. À une longueur d'onde λ donnée, l'absorbance A est définie par la loi de Beer-Lambert:

$$(1) \quad A(\lambda) = \log(I_0 / I) = \epsilon(\lambda).l.C$$

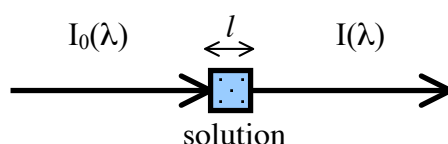


Figure 12. Schéma représentatif d'un faisceau lumineux traversant un milieu d'épaisseur l

où I_0 est l'intensité de la lumière incidente, I est l'intensité de la lumière sortante ($I \leq I_0$), ϵ est le coefficient d'extinction molaire de la molécule absorbante (en $L.mol^{-1}.cm^{-1}$), l est la longueur du trajet optique de la solution traversée (en cm) et C est la concentration molaire de la molécule absorbante (en $mol.L^{-1}$).

L'absorbance des acides nucléiques entre 200 et 300 nm est exclusivement due aux systèmes insaturés conjugués et aux fonctions carbonyles présents dans les bases puriques et pyrimidiques. En effet, la contribution du squelette phosphaté n'est mesurable qu'à des longueurs d'onde inférieures à 190 nm. Le maximum d'absorption dépend donc de la séquence, de la structure et de la composition en bases ; il est compris entre 254 nm pour des séquences riches en purines (A et G) et 270 nm pour des séquences riches en pyrimidines (C, T et U).

Les mesures de Spectres de Différence Thermique (TDS) ou Isotherme (IDS) (cf. : §II.2.1.3 a) et de dénaturation thermique (cf. : §II.2.1.3 b) ont été réalisées avec un spectrophotomètre à double faisceau, qui permet la soustraction, à l'acquisition, de l'absorbance éventuelle du solvant ou tampon de référence (Figure 13). Nous utilisons généralement de l'eau bidistillée en référence, l'absorbance des tampons utilisés étant négligeable.

$$(2) \quad A_{\text{soluté}}(\lambda) = \log(I_{\text{solution}} / I_{\text{solvant}}) = \varepsilon(\lambda) l C$$

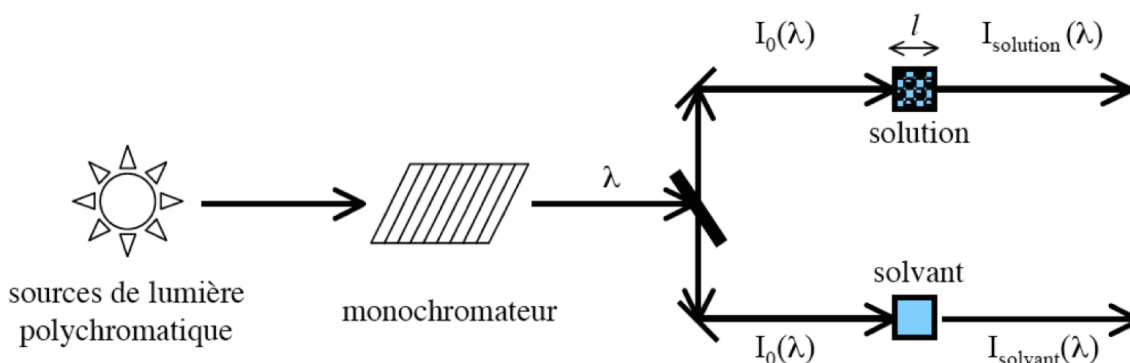


Figure 13. Spectrophotomètre à double faisceau

2.1.3 Méthodes

a) Spectres d'absorption, détermination de concentrations et Spectres de Différence Thermique ou Isothermique

Comme mentionné précédemment, l'absorption d'un acide nucléique dépend de sa séquence, de sa structure et de sa composition en bases. Pour certaines structures d'ADN, une importante variation d'absorbance peut en effet être observée pour des formes structurées et déstructurées. Les coefficients d'extinction molaire étant calculés à 260 nm pour des oligonucléotides non structurés, les concentrations doivent en général être déduites des spectres enregistrés à haute température, lorsque les oligonucléotides sont dénaturés. Néanmoins, pour certaines structures où la variation d'absorbance à 260 nm est faible, c'est généralement le cas des G-quadruplexes [123], l'approximation de concentration réalisée à partir des spectres à basse température est acceptable.

Ces observations ont permis de montrer que la courbe, obtenue par soustraction du spectre d'absorption de la forme structurée à celui de la forme non-structurée, nommée TDS ou IDS pour "*Thermal ou Isothermal Difference Spectra*", pouvait révéler la présence de certaines structures d'acides nucléiques (ADN B, ADN Z, duplexes parallèles, triplexes, G-quadruplexes ou motif i) [124]. Le TDS est défini comme la différence entre le spectre à haute température et celui à basse température, tandis que l'IDS résulte de la soustraction des spectres enregistrés au début d'une cinétique d'association à ceux obtenus à la fin d'une cinétique d'association.

Le TDS permet également de choisir les longueurs d'onde appropriées pour suivre les changements conformationnels induits par une variation de température. L'absorbance à une

longueur d'onde adaptée en fonction de la température donne une courbe dite de "dénaturation thermique" (paragraphe suivant).

b) Dénaturation thermique de structures d'ADN

Les expériences de dénaturation thermique suivies par spectroscopie d'absorption permettent d'étudier la stabilité d'une structure en fonction de la température. En fonction de sa structure, la dénaturation thermique d'un ADN structuré peut s'accompagner d'un hyperchromisme (augmentation de l'absorbance) ou d'un hypochromisme (diminution de l'absorbance) à une longueur d'onde choisie. Les spectrophotomètres sont reliés à des bains-maries extérieurs qui permettent la régulation thermique des échantillons, à des gradients de température bien définis, généralement compris entre 0,2 et 1°C.min⁻¹. Les programmes "MacroTm" sur Uvikon 940 et "ThermAlys" sur Uvikon XL et XS, permettent l'enregistrement à intervalles réguliers de l'absorbance des échantillons (9 au maximum) et à 1-5 longueurs d'ondes différentes (335, 295, 273, 260 et 240 nm en général). La température est mesurée tout au long de l'expérience par une sonde immergée dans une cuve remplie d'eau (Uvikon 940) ou par une sonde placée dans une cuve solide (Uvikon XL et XS). Les absorbances à une longueur d'onde de référence où les acides nucléiques n'absorbent pas (405 ou 335 nm) sont soustraites ensuite aux absorbances mesurées aux autres longueurs d'onde.

Les courbes de dénaturation thermiques obtenues présentent généralement une transition qui se déroule sur une gamme de température plus ou moins large selon l'échantillon considéré. Ce type de transition traduit le passage de la forme associée ou repliée de la molécule (à basse température) vers sa forme dissociée ou dépliée (à haute température). À partir de ces courbes sigmoïdales, nous pouvons déterminer la température de demi-transition T_m , qui est définie comme la température à laquelle 50% des structures ont été dénaturées.

Comment déterminons nous un T_m (Figure 14)? Le T_m correspond à l'intersection entre la courbe de dénaturation thermique et la médiane entre les deux lignes de bases à basse (forme structurée) et à haute température (forme dénaturée). Le choix des lignes de bases est parfois arbitraire, mais nous avons vérifié qu'il n'affecte que modérément la valeur du T_m ($\pm 1^\circ\text{C}$). On notera que les lignes de base à basse $L_1(T)$ et haute température $L_0(T)$ correspondent à la variation d'absorbance des états structurés et non-structurés respectivement. La fraction d'ADN structurée à une température T donnée $\theta(T)$ peut alors

s'exprimer en fonction de la courbe de dénaturation thermique obtenue, c'est à dire de l'absorbance en fonction de la température ($A(T)$) :

$$(3) \quad \theta(T) = [L_0(T) - A(T)] / [L_0(T) - L_1(T)]$$

le T_m étant définie comme la température pour laquelle $\theta = 0,5$ ($\theta = 1$ quand $T \ll T_m$ et $\theta = 0$ quand $T \gg T_m$)

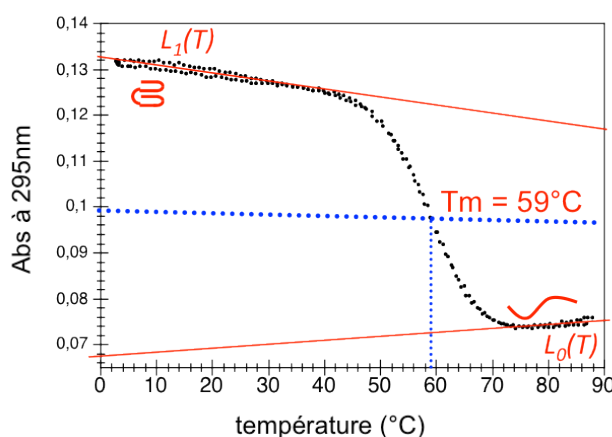


Figure 14. Courbe de dénaturation thermique de Hum21 (3 μ M) dans NaCl 100 mM, suivie à 295 nm

$L_1(T)$ est la ligne de base extrapolée de l'absorbance à basse température (séquence structurée); $L_0(T)$ est celle extrapolée à haute température (séquence dénaturée).

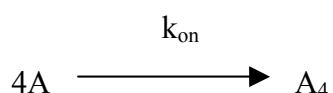
Des courbes de dénaturation et de renaturation superposables indiquent que les valeurs d'absorbance enregistrées à chaque température correspondent à des états d'équilibre thermodynamique. Il est alors possible, moyennant quelques hypothèses, de déterminer des paramètres thermodynamiques des courbes de dénaturation thermique [125] en exprimant les constantes d'association et de dissociation en fonction de θ . Dans le cas contraire, un phénomène d'hystérésis est observé : les courbes de montée et de descente en température ne sont pas superposables. Ce phénomène est lié à des cinétiques de dénaturation et/ou de renaturation lentes par rapport au gradient de température. C'est le cas des G-quadruplexes tétramoléculaires que nous avons étudiés avec des séquences modifiées (cf. : §III) et en présence ou non de ligands des G4 (cf. : §VI). Dans ce cas précis, il est nécessaire de préformer les structures à haute concentration ($\sim 300 \mu$ M) avant de pouvoir étudier leur stabilité thermique. Aux concentrations utilisées (quelques μ M), la réassociation des brins d'ADN en G-quadruplexes tétramoléculaires est négligeable dans la gamme de température où ces structures se dissocient, par conséquent la transition observée ne reflète pas l'équilibre thermodynamique. Il est donc préférable de parler de demi-dissociation ($T_{1/2}$) plutôt que de demi-transition (T_m) dans ce cas là.

Inversement, si l'on se place à plus basse température et à une concentration suffisante, nous pouvons suivre par spectroscopie d'absorption la cinétique d'association isotherme de G-quadruplexes tétramoléculaires (cf. : §II.2.1.3 c).

c) Association isotherme de quadruplexes tétramoléculaires

La mesure de la cinétique d'association d'un G4 tétramoléculaire est décrite ci-dessous. La cuve, contenant l'ADN préalablement dénaturé, est ensuite placée dans le spectrophotomètre thermostaté à froid (4°C) et restera à cette température pendant toute l'expérience. À basse température, l'association des G4 tétramoléculaires est favorisée. En utilisant les mêmes programmes que pour les expériences de dénaturation thermique, il est possible de mesurer l'absorbance, aux longueurs d'onde caractéristiques de la structure étudiée (240 et 295 nm pour les G-quadruplexes), en fonction du temps. La formation d'un G-quadruplexe est accompagnée d'une variation d'absorbance caractérisée par une augmentation de l'absorbance à 295 nm et une diminution de l'absorbance à 240 nm. Les courbes expérimentales sont traitées ensuite par KaleidaGraph et ajustées avec une courbe théorique qui dépend de l'ordre de la réaction, et permettent de déduire la constante de vitesse d'association k_{on} des ADN simple-brins en G4 tétramoléculaire.

L'établissement de l'équation permettant d'ajuster les points expérimentaux enregistrés à 240 et 295 nm est possible en considérant qu'à basse température, l'association d'un G4 tétramoléculaire est un phénomène quasi-irréversible (la demi-vie du complexe formé étant suffisamment longue pour pouvoir négliger la réassociation). Cette réaction unidirectionnelle peut être écrite sous forme :



Où A correspond à la forme simple-brin, A_4 à la forme G4 et k_{on} à la constante d'association. La vitesse de la réaction d'association est alors définie par l'équation :

$$V_{on} = k_{on} [A]^n = -d[A] / dt$$

Où V_{on} est la vitesse d'association, n l'ordre de la réaction, $[A]$ la concentration en simple-brin et t le temps. La variation de la fraction simple-brin est définie au cours du temps par :

$$\alpha(t) = [A](t) / C_0$$

Où α représente cette fraction non-appariée, et C_0 la concentration initiale en simple brin. De ce fait, l'équation de vitesse peut être exprimée sous forme d'une équation différentielle :

$$d\alpha(t) / dt = - k_{on} C_0^{(n-1)} dt$$

Dont la résolution donne comme solution l'évolution de α au cours du temps :

$$\alpha(t) = {}^{1-n}\sqrt{(1 + C_0^{(n-1)} (n-1) k_{on} t)}$$

Dans toutes nos analyses, n est arbitrairement fixé à 4, et k_{on} est exprimé en $[M^{-3}.s^{-1}]$. L'ordre de la réaction dépend en réalité des conditions expérimentales et serait plutôt compris entre 3 et 4. En pratique, les points expérimentaux obtenus à 240 et 295 nm lors d'une expérience d'association isotherme d'un G4 tétramoléculaire sont ajustés par l'équation décrivant la variation de α au cours du temps, en considérant que $n = 4$ et que k_{on} est inconnu.

2.2 Dichroïsme Circulaire

2.2.1 Matériel

Les expériences de dichroïsme circulaire (CD) ont été réalisées sur des spectropolarimètres monofaisceau J-810 ou J-815 (Jasco). Tous les deux sont équipés d'une lampe Xénon, purgés par l'azote afin de limiter l'oxydation des miroirs et thermostatés par effet Peltier. Le Jasco J-810 possède un passeur de 6 échantillons tandis que le Jasco-815 ne permet de mesurer qu'un échantillon à chaque fois. Les mesures sont réalisées en général dans des cuvettes en quartz (Hellma, France) de 1 cm ou 0,2 cm de parcours optique.

2.2.2 Principe

Les molécules chirales présentent une activité optique, c'est-à-dire qu'elle peuvent changer les propriétés de polarisation de la radiation qui les traverse. En particulier elle peuvent absorber différemment de la radiation polarisée circulairement gauche et de la radiation polarisée circulairement droite ; cette propriété est dite dichroïsme circulaire (CD). Pour une longueur d'onde donnée, le signal CD est défini comme la différence entre l'absorbance de la lumière polarisée circulairement à gauche (A_G) et celle de la lumière polarisée circulairement à droite (A_D) (Figure 15).

$$(4) \quad CD = A_G - A_D$$

Nous pouvons mettre en relation le signal CD (en unité d'absorbance) avec la

concentration C de l'échantillon et le parcours optique l par une loi analogue à celle de Beer-Lambert pour l'absorbance.

$$(5) \quad CD = \Delta A = \Delta(\epsilon l C) = (\epsilon_G - \epsilon_D) l C$$

La conversion du signal CD en ellipticité θ , selon la formule (6), autorise son expression en millidegrés. Cette relation entre CD et ellipticité n'est valide que pour des θ petits (de quelques dizaines de mdeg).

$$(6) \quad \theta (mdeg) = 32980 \times CD$$

La forme d'un spectre CD donne, entre autres, des informations concernant la structure secondaire des macromolécules, comme des acides nucléiques ou des protéines.

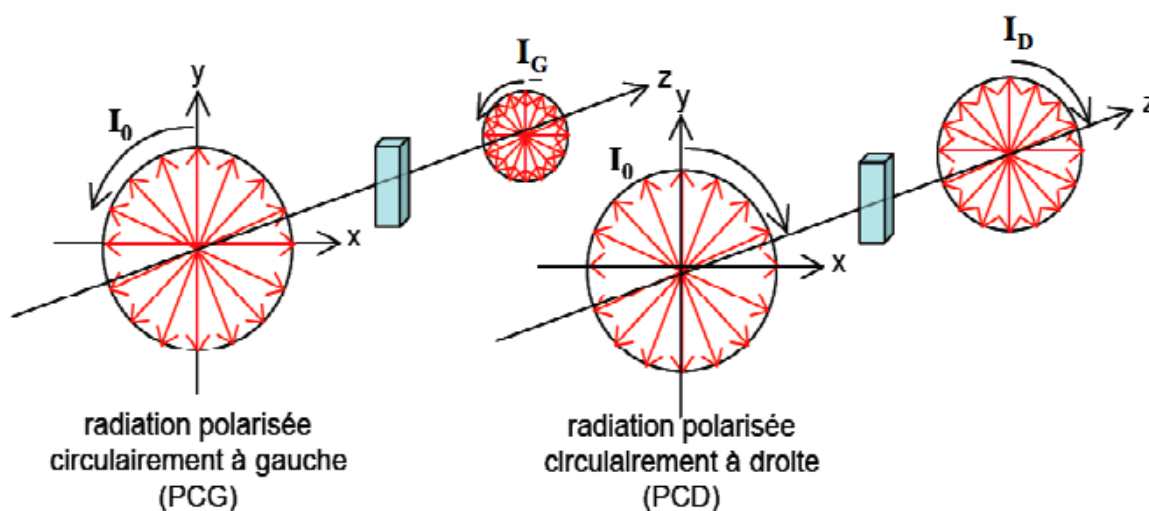


Figure 15. Principe du dichroïsme circulaire

Une radiation électromagnétique qui se propage le long de la direction z est dite polarisée circulairement à gauche (PCG) quand son vecteur champ électrique tourne dans le plan (x, y) dans le sens inverse des aiguilles d'une montre. Quand une radiation PCG d'intensité I_0 est absorbée par une solution, la radiation sortante sera encore PCG, mais son intensité I_G sera plus faible ; et idem pour une radiation PCD (polarisée circulairement à droite). Le signal CD est définie comme $CD = A_G - A_D = \log(I_0/I_G) - \log(I_0/I_D)$. La composition des deux radiations sortantes donne une radiation polarisée elliptiquement ; l'ellipticité θ est définie comme l'arc qui a comme tangente le rapport entre l'axe mineur et l'axe majeur de l'ellipse.

2.2.3 Méthode

a) Dichroïsme circulaire des acides nucléiques en particulier des G4

Les spectres CD des acides nucléiques résultent des transitions dues aux couplages entre les bases, lorsqu'elles sont empilées dans une structure chirale, par exemple hélicoïdale. Le signal CD reflète ainsi l'orientation des bases entre elles. En cela le dichroïsme circulaire s'avère être un outil précieux pour l'étude conformationnelle de structures d'acides nucléiques et en particulier des structures hautement polymorphiques comme les G4. Selon

l'analyse publié par Gray *et al.* en 2008 [126] et l'article publié par Masiero *et al.* en 2010 [127], la polarité des G-quartets successifs empilés dans une structure G4 influence le signal CD. Nous allons détailler l'interprétation des spectres CD des G4 et en particulier des G4 tétramoléculaires dans les chapitres III et VII de ce manuscrit.

Le polymorphisme des G4 qui conduit au mélange de diverses conformations G4 en solution pour certaines séquences, explique en partie la difficulté d'interprétation des spectres CD.

Comme pour les mesures d'absorbance, l'acquisition du signal CD en fonction de la température permet de tracer des courbes de dénaturation thermique. La forme structurée, à basse température (4°C ou 25°C) possède des bandes CD, positives ou négatives, relativement importantes dû à l'empilement et à l'appariement des bases entre-elles, alors qu'à haute température, le couplage entre bases (et donc le signal CD) de la forme déstructurée est fortement réduit. Nous observons donc en général une transition d'un signal positif ou négatif, selon la longueur d'onde observée, vers un signal proche de 0. Le traitement des courbes de dénaturation thermique obtenues par dichroïsme circulaire est effectué de la même manière que décrite précédemment (cf. : §II.2.1.3 b).

b) Mesures et traitement des données

L'acquisition des spectres de CD est réalisée à l'aide du logiciel "Spectra Manager" développé par Jasco. Le balayage est généralement effectué de 335 à 220 nm à une vitesse de 200 à 500 nm.s⁻¹, avec un pas et une largeur de bande à 1 nm et un temps de réponse de 1 s. Le signal CD est généralement faible et les spectres obtenus sont plus bruités que des spectres d'absorbance ; il est donc préférable d'accumuler plusieurs spectres (2 à 3 accumulations en moyenne) pour améliorer le rapport signal/bruit.

Pour toute mesure réalisée, nous avons enregistré une ligne de base correspondant à la cuve remplie de tampon. Étant donné que la qualité des cuves peut influencer la qualité des spectres, il est donc préférable d'utiliser des cuves neuves pour les mesures de CD.

La soustraction de la ligne de base est effectuée manuellement après exportation des données brutes sous KaleidaGraph. Les mesures enregistrées en ".txt" sont regroupées sous un même tableau grâce au logiciel "DataHelper".

Pour les expériences de dénaturation thermique, l'échantillon est thermostaté par un effet Peltier piloté directement par le logiciel "Spectra Manager". Pour les expériences de dénaturation avec les G4 tétramoléculaires modifiés présentées dans l'**article 1** de ce

manuscrit, nous avons utilisé un gradient de température de $0,2^{\circ}\text{C}.\text{min}^{-1}$ et enregistré à des longueurs d'onde de 335, 240, 260 et 295 nm.

2.3 Fluorescence

2.3.1 Matériel

Les spectres de fluorescence ont été effectués sur un spectrofluorimètre FluoroMax-3 (Jobin Yvon), à lampe de Xénon (150 W), équipé d'un passeur de 4 échantillons et relié à un cryostat permettant de contrôler la température. Les expériences sont réalisées dans des cuvettes possédant 4 faces polies, de parcours optiques de 10 mm, et de contenance 3 mL.

Pour augmenter le débit, les expériences de FID (*"Fluorescence Intercalator Displacement"*) sont réalisées dans des plaques de 96 puits, la fluorescence est détectée avec une machine Mx3005P (Stratagène) (conçue pour la PCR quantitative et ici utilisée simplement comme lecteur de fluorescence de plaques). Elle est munie d'un système de filtres permettant la détection de la fluorescence du Thiazole Orange utilisé dans ce test (filtre d'excitation à 492 nm ± 5 nm, filtre d'émission à 516 nm ± 5 nm avec un gain de 8X). Les expériences sont enregistrées à 25°C .

2.3.2 Principe de G4-FID (*"Fluorescent Intercalator Displacement"*)

Le principe de FID est basé sur le déplacement du thiazole orange (une molécule fluorescence) de la structure étudiée (quadruplexes ou duplexes), par des ligands de G4. La fluorescence du Thiazole Orange (TO) est exaltée (maximum d'émission à 525 nm) lorsqu'il est lié à l'ADN (quadruplexes ou duplexes). De ce fait le déplacement de la sonde par fixation du ligand sur l'ADN entraîne une diminution de la fluorescence du TO (Figure 16). Ce test permet de classer les ligands potentiels en fonction de leur affinité pour les G-quadruplexes et leur sélectivité pour ces derniers par rapport aux duplexes. Cette expérience est simple et peu coûteuse puisqu'elle ne demande pas l'apport d'oligonucléotides modifiés ou d'équipement spécifique. Elle est également rapide en termes de temps de lecture et d'interprétation et permet d'analyser une quantité importante de molécules. Cependant, cette méthode représente certaines limites qui sont détaillées dans le chapitre VI.

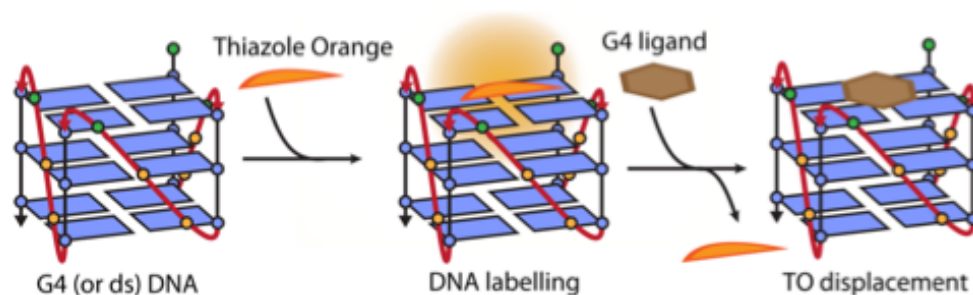


Figure 16. Représentation schématique de deux principales étapes du test G4-FID

(a) marquage de la matrice d'ADN (quadruplexes ou duplexes) par le Thiazole Orange (TO) et (b) déplacement de la sonde fluorescente de la matrice d'ADN par des ligands. (Ces deux étapes ne sont pas forcément découplées ; il est possible d'ajouter le ligand et le TO simultanément)

2.3.3 Mesures et traitements des données

Les expériences sont réalisées sur des plaques 96 puits, dans un tampon 10 mM cacodylate de lithium (pH 7,2) et 100 mM KCl. Le thiazole orange (1 μ M) est incubé avec les différentes structures d'oligonucléotides préformées (0,5 μ M en brin pour les G4 intramoléculaires, 2 μ M en brins pour les G4 tétramoléculaires, 1 μ M pour les duplexes) en présence ou non de ligands (2,5 μ M). Le déplacement du TO est évalué en mesurant sa fluorescence sur la machine Mx3005P (Stratagène).

Les intensités de fluorescence sont ensuite traitées par Excel. Le bruit de fond (F_b) est soustrait de l'intensité de fluorescence initiale (FI et FI_0) :

$$(7) \quad FA = FI - F_b \text{ ou } FA_0 = FI_0 - F_b$$

Où F_b (bruit de fond) est la moyenne de fluorescences du tampon, du TO non lié à l'ADN et/ou du ligand, FI_0 est l'intensité de fluorescence du TO lorsqu'il est lié à l'ADN, FI est l'intensité de fluorescence après l'ajout du ligand, FA et FA_0 sont les intensités de fluorescence du TO après soustraction du bruit de fond.

Nous pouvons ensuite calculer le pourcentage de déplacement du TO en utilisant la formule ci-dessous :

$$(8) \quad \text{Déplacement du TO (\%)} = 100 - ((FA/FA_0) \times 100)$$

3) Analyse et purification des acides nucléiques

3.1 Radiomarquage des acides nucléiques

3.1.1 Marquage en position 5'

Les oligonucléotides possédant une extrémité 5'-OH libre sont marqués par incorporation d'un phosphate radioactif. La réaction s'effectue à 37°C pendant 30 à 40 minutes : 20 pmoles d'oligonucléotides sont incubées en présence de 10 unités de T4 polynucléotide kinase (T4 PNK, New England Biolabs (NEB)) et de 10 μCi de [$\gamma^{32}\text{P}$ -ATP] (3000 Ci.mmol⁻¹, PerkinElmer) dans 70 mM Tris-HCl, 10 mM MgCl₂, 5 mM Dithiothréitol (DTT), pH 7,6 (tampon T4 PNK, NEB).

Dans le cas d'oligonucléotides suffisamment longs (≥ 10 nucléotides), le [$\gamma^{32}\text{P}$ -ATP] non incorporé et l'enzyme sont éliminés après passage sur colonne d'exclusion (Biospin SP6, Biorad) pendant 4 minutes à 1000g. Dans le cas d'oligonucléotides courts (< 10 nucléotides), la purification s'effectue sur un gel de polyacrylamide 20% en conditions dénaturantes (TBE 1X, Urée 7M) (cf. : §II.3.2.1).

3.1.2 Marquage en position 3'

Les oligonucléotides (20 pmol) avec une extrémité 3'-OH sont marqués par 20 unités de Terminal Transférase (TdT, New England Biolabs) dans le tampon suivant: 20 mM Tris-acétate, 50 mM acétate de potassium, 10 mM acétate de magnésium, 1 mM Dithiothréitol (DTT), pH 7,9 (tampon NEB 4), en présence de 250 μM de CoCl₂ et de 10 μCi de [$\alpha^{32}\text{P}$ -dATP] ou [$\alpha^{32}\text{P}$ -dNTP] (3000 Ci.mmol⁻¹) selon l'application. L'incubation a lieu à 37°C pendant 30 à 40 minutes. Les nucléotides radiomarqués non incorporés ainsi que l'enzyme sont éliminés après passage sur une colonne d'exclusion 4 minutes à 1000 g (Biospin SP6 Biorad).

3.2 Electrophorèse sur gel de polyacrylamide

3.2.1 Conditions dénaturantes

Les oligonucléotides peuvent être analysés par radiomarquage en 3' ou 5' au ^{32}P (cf. : §II.3.1) ou par « UV-shadow » sur gel de polyacrylamide 15 ou 20 % (acrylamide : bisacrylamide (19 :1)) en condition dénaturantes (urée 7M, TBE 1X) dans un tampon de migration TBE 1X. Le gel d'acrylamide est polymérisé par 0,1% d'ammonium persulfate (APS) et 0,1% de TEMED (N,N,N',N'-Tetraméthylènediamine). Une migration du gel (7 Watts.cm⁻²) est effectuée avant le dépôt des échantillons afin d'atteindre une température de

50 à 60°C. Les échantillons sont dilués dans un tampon de charge dénaturant contenant 95% de formamide déionisée (25% de formamide final) et pouvant contenir éventuellement des marqueurs colorés (xylène-cyanol et bleu de bromophénol). Cependant, l'ajout des bleus directement dans les échantillons n'est pas recommandé dans le cas d'une détection par "UV-shadow" (cf. : §II.3.2.3 b), ceux-ci pouvant masquer la bande d'ADN à analyser. Les échantillons sont dénaturés 2 à 3 minutes à 90°C dans un bain à sec. Ils sont ensuite centrifugés rapidement avant d'être déposés sur gel : environ 30 à 60 μ M d'ADN (selon la taille des oligonucléotides) par puits pour une visualisation par "UV-shadow" et 0,5 à 1 pmol pour une détection après radiomarquage classique (0,5 μ Ci/pmol d'ADN).

3.2.2 Conditions natives

L'électrophorèse de gels de polyacrylamide en conditions natives permet d'observer d'éventuelles structures d'acide nucléiques, et les éventuelles interactions entre elles ou avec d'autres agents du milieu. Des conditions de température et de sels contrôlées permettent de conserver certaines structures intra- ou intermoléculaires ou certaines interactions avec des ligands ou des protéines. Ces espèces migrent dans le gel en fonction de leur taille et de leur charge apparentes.

Pour la visualisation de G-quadruplexes tétramoléculaires, nous avons utilisé des gels de polyacrylamide (acrylamide : bisacrylamide (19 : 1)) à 20% contenant 10 à 20 mM de NaCl ou de KCl et du TBE 1X. Les oligonucléotides sont préformés à 4°C pendant 48h dans un tampon 10 mM cacodylate de lithium pH 7,2 et 110 mM de NaCl ou de KCl. Nous ajoutons ensuite 10% de sucrose pour permettre de charger les échantillons sur gel. Les gels sont migrés dans du TBE 1X supplémenté par 10 à 20 mM en NaCl ou KCl, en chambre froide ou thermostaté par cryostat. La migration est effectuée à 3-4W/gel, soit environ 120V, pour une température des gels de 15°C en fin de migration. Ces conditions permettent de préserver les structures formées à basse température et de minimiser la dénaturation au cours de la migration. Ceci est particulièrement important dans le cas des G4 intramoléculaires où la stabilité thermique est souvent plus faible que des G4 tétramoléculaires.

Pour l'observation de G-quadruplexes intramoléculaires, nous avons utilisé de 12 à 20% de polyacrylamide (acrylamide : bisacrylamide (19 : 1)) selon la taille des oligonucléotides testés. Le gel contient 10 à 20 mM de NaCl ou de KCl et du TBE 1X. Les oligonucléotides sont préalablement chauffés à 90°C puis refroidi lentement à la température ambiante ($\approx$ 25°C) ou à 4°C à une vitesse d'environ 1°C.min⁻¹. La migration s'effectue dans un tampon

TBE 1X supplémenté par 10 à 20 mM NaCl ou KCl, en chambre froide ou thermostaté par cryostat. La migration est effectuée à 3-4W/gel.

3.2.3 Modes de détection

a) Détection radioactive

Pour une détection radioactive, les gels doivent être fins (0,4 mm d'épaisseur) pour être ensuite séchés sur du papier Whatman 3MM. Une fois sec, le gel est posé sur un écran radiosensible à base de cristaux de BaFBr : Eu²⁺ dans une matrice organique (Molecular Dynamics ou Amersham Biosciences). Ces écrans permettent la conversion des électrons émis par le rayonnement β^- du ³²P en photons : la radiation induit l'oxydation de Eu²⁺ en Eu³⁺ ainsi que la réduction de BaFBr en BaFBr⁻ ; les complexes BaFBr⁻ absorbent autour de 600 nm ; l'absorption d'une telle radiation libère des électrons qui réduisent l'Eu³⁺ à un état excité de l'Eu²⁺ ; le retour à l'état fondamental de ce dernier se fait par l'émission d'un photon à 390 nm. Après l'exposition au gel, l'écran est ensuite numérisé à l'aide d'un PhosphorImager (Molecular Dynamics) ou d'un Typhoon (Amersham Biosciences). L'instrument détecte l'émission des photons à 390 nm après une excitation à 633 nm par un laser. L'image obtenue est analysée à l'aide du logiciel ImageQuant 5.2 (Molecular Dynamics).

b) Détection par ombrage de fluorescence ("UV-Shadow")

Dans le cas d'une visualisation de l'ADN par "UV-Shadow", le gel est transféré sur un film plastique puis placé sur une plaque de silice fluorescente (pour chromatographie en couche mince). La plaque de silice en-dessous du gel devient fluorescente quand elle est excitée aux UV (254 nm). Les acides nucléiques présents dans le gel absorbant eux-aussi les UV, ils sont révélés par l'ombre qu'ils produisent sur la plaque de silice lorsque nous les éclairons aux UV en épifluorescence (Figure 17). Les images sont générées et analysées directement avec une GBox (Syngene, Synoptics) ou du système "Bio-imaging" (MF-ChemiBIS 3.2 – ETNA). Cette technique est simple, elle ne nécessite aucun pré- ou post-marquage des matériels, mais est beaucoup moins sensible que le radiomarquage (cf. : §II.3.2.3 a). Par ailleurs, une analyse quantitative des bandes reste délicate, du fait d'un éclairage peu homogène et d'un "bruit de fond" important.

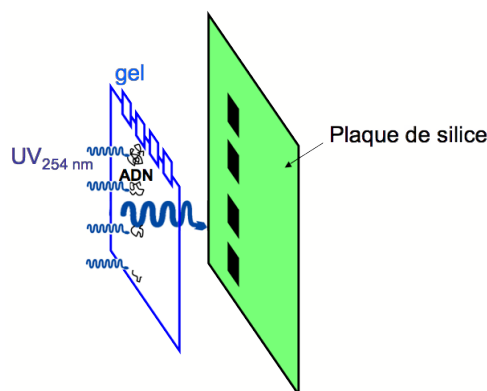


Figure 17. Principe de la détection par “UV-Shadow”

c) Détection par fluorescence

Dans le cas d’acides nucléiques conjugués à des molécules fluorescentes, il est possible de les révéler en utilisant les propriétés d’excitation et d’émission des fluorophores. Pour des oligonucléotides non marqués, il est également possible de “colorer” le gel après migration, avec un bain de tampon contenant une molécule qui fluoresce davantage lorsqu’elle est liée à l’ADN (par exemple le bromure d’éthidium ou le SyBR green, gold ou safe). La visualisation peut alors se faire par transillumination dans l’UV (cas du bromure d’éthidium) et capture par caméra numérique (MF-ChemiBIS 3.2 – ETNA) ou à l’aide d’un Typhoon (Amersham Biosciences).

- Gels révélés par “post-marquage”

Après migration, le gel peut être coloré par un bain contenant des différents colorants de l’ADN tels que le bromure d’éthidium (BET), le SyBR (green, gold ou safe), le “Stains-All” ou encore le thiazole orange.

La détection par fluorescence de colorants de l’ADN a l’avantage d’être plus simple et plus rapide que celle par radioactivité. Les inconvénients de cette méthode de révélation est qu’elle nécessite plus de matériel que la radioactivité et qu’il est très difficile de s’assurer que l’intensité des bandes reflète bien l’abondance relative des espèces. En effet, ces colorants interagissent différemment avec les structures d’acides nucléiques, ce qui revient à dire que les différentes structures ne sont pas colorées de manière similaire. En particulier, nous n’avons pas trouvé, pour l’instant, de colorant permettant de révéler de façon identique de courtes séquences d’ADN du type d(TG_nT) (n = 3-5), sous leur forme simple-brin ou G-quadruplexe tétramoléculaire. Le signal obtenu pour le simple-brin était souvent inexistant, ce qui rend délicate l’utilisation de cette méthode pour la quantification.

- Réaction d'échange de brins observée en électrophorèse

Cette réaction d'échange de brins est facilitée par un copolymère cationique [128-131] (CCC, pour “*cationic comb-type copolymer*”) synthétisé et mise en place par l'équipe du Pr. Atsushi Maruyama à Fukuoka (Japon). Le copolymère possède des propriétés de chaperone d'acides nucléiques. Il réduit la barrière énergétique nécessaire à la dissociation et à l'association des brins impliqués dans une structure et permet donc de promouvoir l'échange de brins entre l'ADN fluorescent et la structure préformée (Figure 18).

Dans le cas de cette étude, le brin TG₄T couplé à la fluorescéine en position 3' (0,5 µM en brin) est incubé en présence d'excès de G4 tétramoléculaires préformés (20 µM) et avec ou sans 1 mM de CCC. La réaction est effectuée à température ambiante dans un tampon composé de 10 mM cacodylate de lithium (pH 7,2) et de 100 mM NaCl. Avant le dépôt des échantillons, nous ajoutons 12% de glycérol ou sucrose et 5% de sulfate de polyvinyle (PVS) qui permet de dissocier le CCC de l'ADN [132]. Pour la visualisation de cette réaction, nous avons utilisé des gels de polyacrylamide (acrylamide : bisacrylamide (19 : 1)) à 20% contenant 10 à 20 mM de NaCl et du TBE 1X. La migration des échantillons est effectuée dans un tampon de migration (TBE 1X supplémenté de 10 à 20 mM en NaCl), à température ambiante et à 3-4 W par gel. Le gel est ensuite révélé par fluorescence à 520 nm ($\lambda_{\text{exc}} = 480$ nm) en utilisant une caméra numérique (ETNA – MF-ChemiBIS 3.2).

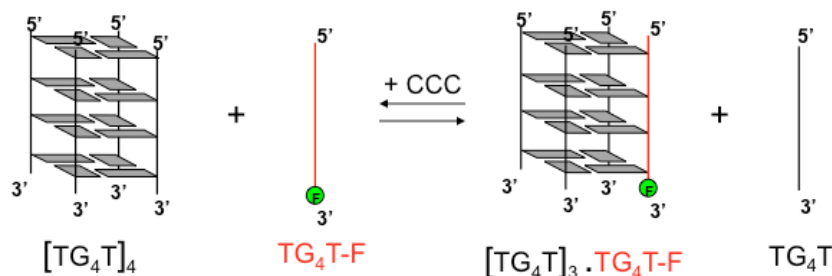


Figure 18. Principe de l'expérience d'échange de brins

Le CCC (“*cationic comb-type copolymer*”) permet au brin fluorescence (TG₄T-F) de s'incorporer dans le G4 tétramoléculaire préformé. Cette réaction peut être observée par électrophorèse.

3.3 Purification d'oligonucléotides sur gel de polyacrylamide

Les oligonucléotides peuvent être purifiés par migration sur gel de polyacrylamide, en les détectant par “UV-shadow” ou radiomarquage. Les oligonucléotides radiomarqués ou non sont déposés, selon leur taille, sur un gel de polyacrylamide de 12 à 20 % (acrylamide : bisacrylamide (19 : 1)) en conditions dénaturantes (urée 7M, TBE 1X). Après migration, une bande majoritaire d'ADN est révélée, par “UV-shadow” dans le cas d'oligonucléotides non

radiomarqués ou par exposition courte contre un écran radiosensible, numérisation et impression en taille réelle de l'image générée dans le cas d'un radiomarquage.

La bande correspondant à l'ADN désiré est ensuite découpée, broyée et éluée dans 500 μL d'eau bidistillée. Pour les oligonucléotides résistants comme l'ADN, le broyat est chauffé à trois reprises pendant 5 min à 50°C. Pour les oligonucléotides plus fragiles comme l'ARN, l'éluât est mis sous agitation à 4°C sur la nuit. L'échantillon est ensuite déposé sur une colonne de purification Spin-X 0,22 μm (Costar) et centrifugé pendant 1 minute. Le surnageant récupéré après centrifugation est mis en présence de 0,3 M d'acétate de sodium ou 2,5 M d'acétate d'ammonium. Les oligonucléotides sont ensuite précipités par au moins 3-4 volumes d'éthanol, incubés 30 minutes à -80°C et centrifugés 30 minutes à 12000 rpm à 4°C. Le culot récupéré est mis en présence de 200 μL d'éthanol 85% et est centrifugé 15 minutes à 12000 rpm à 4°C. Une fois l'éthanol éliminé, le précipité est resuspendu dans de l'eau bidistillée et la concentration est déterminée par spectroscopie comme décrit précédemment (cf. : §II.2.1.3 a). Les oligonucléotides resuspendus sont conservés à -20°C. L'utilisation d'acétate d'ammonium pour la précipitation d'acides nucléiques est particulièrement intéressante pour l'étude des G4, car ces ions sont volatiles, ils peuvent être éliminés par une ou deux étapes de lyophilisation supplémentaires, après reprise dans l'eau bidistillée. La précipitation à l'acétate d'ammonium reste cependant généralement moins efficace que la précipitation à l'acétate de sodium.

4) ADN origami

4.1 Construction de l'ADN origami

Il est possible de concevoir des édifices supramoléculaires auto-organisés complexes à partir d'ADN. La nanotechnologie des acides nucléiques consiste à utiliser l'ADN non plus comme le code génétique de la cellule vivante, mais comme un matériau de construction permettant d'élaborer des objets de quelques nanomètres. En utilisant les propriétés d'auto-assemblage de l'ADN, l'ADN origami développé par Paul W.K. Rothmund [133], en 2006, permet de créer arbitrairement des formes bi- [133, 134] ou tridimensionnelle [135].

Comment construisons-nous une structure d'ADN origami ? L'ADN origami est composé de 2 éléments: un long brin d'échafaudage ("scaffold strand" en anglais) et d'une multiplicité de brins courts et généralement synthétiques, appelés "agrafes" ("staple strands" en anglais). Le brin d'échafaudage utilisé provient généralement de la séquence circulaire du

génom viral du phage M13 (M13mp18, New England Biolabs (NEB)) dont la séquence est connue (7,249 nucléotides) (annexe 1). Les petites séquences d’“agrafes” permettent de plier le brin d’échafaudage et de lui donner une forme désirée. Ainsi, en changeant juste les “agrafes”, nous pouvons obtenir toutes les formes bi-dimensionnelles désirées, comme par exemple un carré, un rectangle, un triangle, une étoile à cinq branches, un “smiley” [133], un dauphin [134], ou une carte de l’Amérique. Il est également possible de réaliser des structures tridimensionnelles comme une boîte à ADN [135]. Il est désormais possible de créer des ADN origamis en utilisant des logiciels destinés à cela. Le logiciel utilisé dans cette étude se nomme “caDNAno”, il peut être téléchargé gratuitement sur Internet (<http://cadnano.org/>).

Comment fonctionne un tel programme ? Nous commençons par définir le schéma de la forme que nous souhaitons obtenir. Nous décrivons ensuite les liens possibles entre les différents points de l’échafaudage, posons les “agrafes” sur le brin principal, et si nous ne nous trompons pas le programme va alors générer la liste des séquences des “agrafes” nécessaires pour la construction de l’ADN origami. Les séquences générées sont ensuite synthétisées par la société *Eurogentec* (Liège, Belgique) et sont délivrées directement en solution.

Pour obtenir la structure de l’ADN origami en solution, nous mélangeons tout simplement l’ADN circulaire (10 nM) avec les séquences d’“agrafes” (50 nM) dans un tampon contenant 20 mM Tris-HCl, 10 mM MgCl₂, 1mM EDTA (Acide Ethylène Diamine Tétracétique) à pH 7,6. À l’aide d’un thermocycleur (Mastercycler Pro, Eppendorf), nous chauffons le mélange (25 µL) à 85°C pendant 2 minutes, puis le refroidissons à 15°C avec un gradient de température de 1°C.min⁻¹. Les “agrafes” non liées aux brins d’échafaudage sont éliminées après passage sur colonne de Sephacryl[™] S-300 (éliminant les séquences de taille inférieure à 118 nucléotides) pendant 2 minutes à 750 g (MicroSpin[™] S-300, GE Healthcare). La solution obtenue est ensuite visualisée en AFM (Microscope à Force Atomique) sur une surface de mica.

4.2 Observation de l’ADN origami en AFM (Microscope à Force Atomique)

4.2.1 Matériel

Les images AFM sont obtenues en utilisant un système “fast-scanning AFM” (Nano Live Vision, RIBM Co. Ltd., Tsukuba, Japon). L’AFM contient un levier en nitrure de silicium (fréquence de résonance = 1-2 MHz, “spring constant” = 0,1-0,3 N/m, “EBD Tip radius” < 15 nm, Olympus BL-AC10EGS-A2). L’échantillon (2 µL) est déposé sur une plaque de mica fraîchement clivée (φ 1.5 mm, RIBM Co. Ltd., Tsukuba, Japon), laissé 5

minutes à température ambiante, et lavé ensuite plusieurs fois avec le même tampon (20 mM Tris-HCl, 10 mM MgCl_2 , 1mM EDTA et pH 7,6). Nous utilisons ici, le mode “tapping” pour balayer la surface de l'échantillon.

4.2.2 Principe

Le microscope à force atomique (AFM) est un microscope à sonde locale qui permet d'observer la topographie de la surface d'un échantillon. Le principe se base sur les interactions (attraction-répulsion, selon la distance) entre les atomes de surface de l'échantillon et une pointe montée sur un levier flexible et réfléchissant (Figure 19). La pointe balaye la surface à étudier, la flexion du levier donne une mesure de la force d'interaction entre la surface et la pointe. La flexion du levier est mesurée par l'utilisation d'un rayon laser, qui se réfléchit sur le levier : la position du rayon réfléchi, qui dépend de la flexion du levier, est détectée par un quadrant de photodiodes. Un ordinateur élabore les données pour déterminer la hauteur de l'échantillon et reconstituer l'image de la surface.

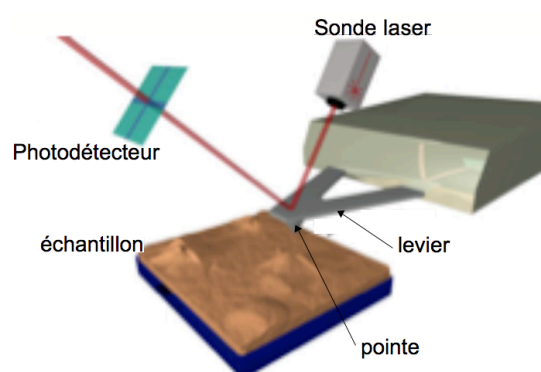


Figure 19. Principe de la microscopie à force atomique (AFM)

Il existe plusieurs modes d'utilisation de l'AFM :

Le mode contact : Ce mode consiste à utiliser les forces de répulsion. La pointe appuie sur la surface, elle est donc repoussée, et le levier est dévié. La rétroaction s'effectue sur la mesure de la direction de la déviation.

Le mode dynamique : Le mode modulation d'amplitude (mode “tapping”) est le plus utilisé et consiste à faire vibrer le levier à sa fréquence propre de résonance, avec une certaine amplitude. Quand la pointe interagit avec la surface, l'amplitude décroît parce que la fréquence de résonance change. La rétroaction se fait sur l'amplitude d'oscillation du levier.

La résolution des images correspond essentiellement à la dimension du sommet de la pointe. Nous utilisons pour cette étude le mode “tapping”, appelé aussi contact intermittent. Cela permet l’utilisation de pointes très fines (de l’ordre d’une dizaine de nm) et les pointes s’usent moins vite.

Les images d’AFM obtenues sont ensuite traitées par le logiciel “AFM scan system” et enregistrées sous BitMap (“.bmp”).

Résultats

Chapitre III. Structures, stabilité et cinétique de G-quadruplexes tétramoléculaires modifiés

1) Introduction

Cette dernière décennie, plusieurs équipes ont étudié les cinétiques d'association et de dissociation de G-quadruplexes. Les séquences formant des G4 tétramoléculaires sont utilisées principalement pour ces études.

En effet, les cinétiques d'association et de dissociation des G4 intramoléculaires sont trop rapides pour être déterminées en cinétique classique (cf. : §II.2.1.3 c). Les cinétiques de ces structures peuvent être estimées indirectement par l'ajout du brin complémentaire [111, 119, 136]. Dans ce cas, nous devons supposer que le brin complémentaire ne participe pas lui-même au processus de dissociation du G4, mais piège la forme simple-brin en s'y associant. La formation d'un quadruplex peut être suivie en cinétique rapide après addition d'un cation comme le potassium. Néanmoins, le polymorphisme des G4 intra- et bi-moléculaires rend les études plus compliquées [137].

Nous avons donc choisi d'utiliser un modèle plus facilement exploitable pour l'étude présentée dans ce chapitre, les G4 tétramoléculaires. Deux avantages majeurs expliquent ce choix: i) la quasi absence de polymorphisme (même si des formes octamériques ont pu être observées [138], et si les tétrades de T coiffant les quartets de guanines pourraient adopter plusieurs géométries [18]) ; ii) la relative lenteur du phénomène d'association, qui en plus est modulable par la concentration en brin, autorise le suivi par des techniques classiques.

1.1 Stabilité et cinétiques de G-quadruplexes tétramoléculaires

Les G-quadruplexes tétramoléculaires que nous étudions sont formés par de courtes séquences d'ADN, du type d(TG_nT). Tous les brins de ces G4 sont parallèles, toutes les guanines sont orientées en *anti*, tous les sucres sont configurés en C2'-*endo* et tous les sillons sont identiques [139]. La présence de T aux extrémités permet d'éviter la formation de structures où les brins sont décalés l'un par rapport à l'autre. Ces structures sont particulièrement inertes du point de vue cinétique. Leur formation est lente (Figures 20.A et B), mais une fois formées, elles peuvent être extrêmement stables. L'utilisation d'hydroxyde de lithium est parfois nécessaire pour la dissociation du complexe.

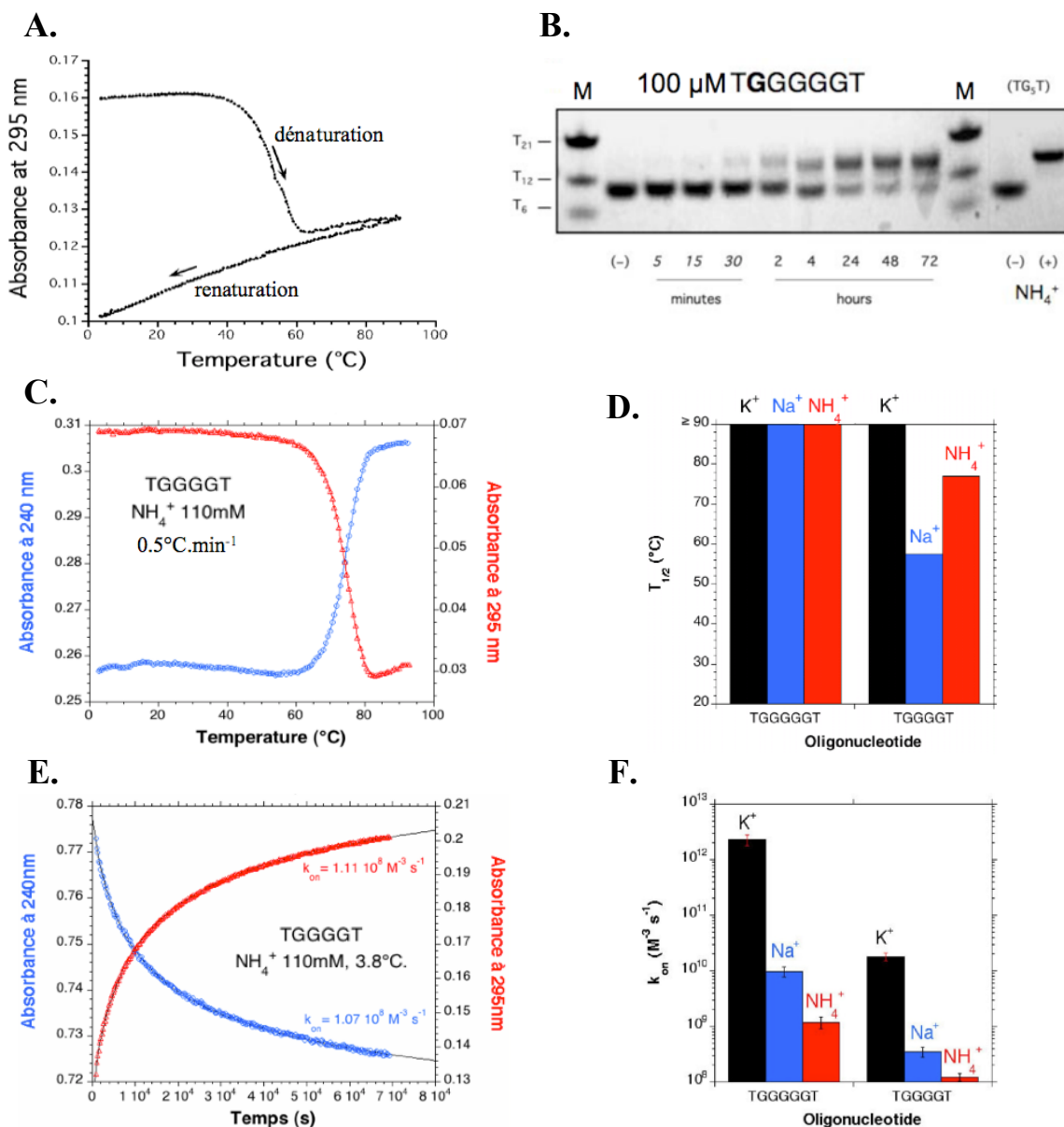


Figure 20. Stabilité et cinétique d'association des G4 tétramoléculaires [140, 141]

(A) Courbes de dénaturation et de renaturation de 20 µM d(TG₄T) dans un tampon de 10 mM cacodylate de sodium (à pH 7,0) et de 0,1 M Na⁺. L'absorbance est enregistrée à 295 nm à un gradient de température de ~0,18°C.min⁻¹. Une fois dissocié le G4 tétramoléculaire n'est pas reformé dans le temps de l'expérience. (B) Cinétique d'association de d(TG₅T) (100 µM en brins) suivie par gel d'électrophorèse en conditions natives. Les échantillons sont incubés dans 0,11 M NH₄⁺, à 4°C pendant une durée variable, de 5 minutes à 72 heures. M = Marqueurs simple-brins ; (-) = condition défavorable à la formation de G4 (sans ion et traitement à l'hydroxyde de lithium) ; (+) = condition favorable à la formation de G4 (présence de ion NH₄⁺). (C) Dissociation du G4 d(TG₄T) préformé en cacodylate de lithium 10mM, NH₄⁺ 110mM suivie à 240 nm (bleu) et 295 nm (rouge) selon un gradient de température de 0,5°C.min⁻¹. (D) Températures de demi-dissociation (T_{1/2}) obtenues pour les G4 d(TG₅T) et d(TG₄T) en K⁺ (noir), Na⁺ (bleu) et NH₄⁺ (rouge). (E) Association à basse température (3,8°C) de la séquence d(TG₄T) en cacodylate de lithium 10mM, NH₄⁺ 110mM, suivie à 240nm (bleu) et 295nm (rouge) en fonction du temps. (F) Les constantes de vitesse d'association (k_{on}) ont été calculées à partir de l'ajustement des points expérimentaux (lignes noires sur la courbe (E)) en utilisant l'équation développée dans le Matériel et Méthodes (cf. §II.2.1.3 c) de ce manuscrit ; k_{on} mesurés pour d(TG₅T) et d(TG₄T) en K⁺ (noir), Na⁺ (bleu) et NH₄⁺ (rouge).

Comme nous l'avons mentionné dans le chapitre précédent, le repliement ou la dénaturation des G4 tétramoléculaires s'accompagne de variations d'absorbance à des longueurs d'ondes caractéristiques qui peuvent être suivies en fonction de la température ou du temps. Nous pouvons ainsi déterminer la stabilité thermique ($T_{1/2}$ = température de demi-dissociation - qui dépend du gradient de température puisque la dénaturation de ces structures n'est généralement pas réversible cinétiquement) et la cinétique d'association (k_{on} = constante d'association) des structures par spectroscopie d'absorption (cf : §II.2.1). Plusieurs paramètres peuvent influencer la stabilité et la cinétique des G4 tétramoléculaires; ils sont regroupés dans la table 4.

Paramètres	Association	Stabilité ($T_{1/2}$)
$Na^+ \rightarrow K^+$	+	+
Augmenter la concentration en cations monovalents	+	≈ 0
Augmenter la concentration en oligonucléotide	+	$\approx 0^a$
Augmenter le nombre de Guanines consécutives	+	+
Ajouter des nucléotides non-G aux extrémités 5' ou 3'	-	+/- ^b
Diminuer la température	+	nd
Présence de copolymère cationique	+	-

Table 4. Résumé des effets de différents paramètres sur la stabilité et la vitesse d'association des G4 tétramoléculaires [140].

(-) = diminution ; (+) = augmentation ; (0) = pas d'effet ; (nd) = non-déterminé ; (^a) $T_{1/2}$ ne varie pas en fonction de la concentration en oligonucléotides (pour concentration $\leq 200 \mu M$) ; (^b) En général, l'ajout de nucléotides aux extrémités stabilise la structure mais les effets observés dépend beaucoup de la longueur de l'insertion, des bases (dT, dC ou dA) et de l'extrémité auquel les nucléotides sont ajoutés (5' ou 3').

La stabilité et la cinétique d'association de ces structures dépendent fortement de la nature des cations monovalents utilisés: $K^+ > Na^+$ (Figures 20C-F). La vitesse d'association varie en fonction des concentrations en oligonucléotides et en cation monovalent : l'augmentation des concentrations (oligonucléotides ou cation) induit une augmentation de la vitesse de formation (Table 4). La séquence peut également influencer les valeurs de $T_{1/2}$ et de k_{on} . Ces derniers varient en fonction du nombre de guanines consécutives (Figures 20D-F et Table 4) : l'augmentation du nombre de guanines entraîne une augmentation de la stabilité et de la cinétique d'association de la structure. Les bases flanquantes aux extrémités 5' et 3' du bloc de guanine influencent à la fois l'association et la dissociation, mais pas nécessairement dans le même sens : l'ajout de bases aux extrémités (par exemple d(TTG₄TT)) ralentit la formation de G4 tétramoléculaire mais augmente la stabilité de la structure.

L'équipe de Janez Plavec a montré par RMN, que la thymine (pour l'ADN) ou l'uracile (pour l'ARN) de l'extrémité 5' des séquences d(TG₄T) et d(UG₄T) (cf. : § I.1.2.1 b) est capable de former des T- ou U-quartets qui ralentissent le mouvement des cations [18]. La formation de ces quartets inhabituels (T- et U-quartets) pourrait participer à la stabilisation de la structure. Par ailleurs, la constante k_{on} dépend également de la température ; en effet la cinétique d'association est accélérée à basse température.

Enfin, la formation de G4 peut être modulée par la présence d'un cosolute, tels que des ligands [142], le polyéthylène glycol (PEG) [47, 48], les polyamines (spermine, spermidine ou triéthylène tétramine) [140] ou, comme montré plus récemment un copolymère cationique [132]. Ce copolymère (en anglais "*cationic comb-type copolymer*") augmente la cinétique d'association mais aussi de dissociation du G4 tétramoléculaire étudié d(TG₄T), sans pour autant altérer l'effet des cations sur sa formation. Il présente donc des propriétés de chaperone d'acides nucléiques qui favorise la réaction d'échange de brins ; nous l'avons utilisé dans l'article 2 de ce manuscrit (cf. : §II.2.2).

1.2 L'effet des modifications chimiques sur la formation de G4 tétramoléculaires

Nous avons évoqué en introduction générale (cf. : §I.1.2.1 b) la possibilité de formation de quartets autres que les G-quartets (A-, T-, C-, U- quartets et quartets mixtes "G+C" ou "A+T") par des bases dites "naturelles". Nous allons introduire ci-dessous les impacts des différentes modifications chimiques sur la stabilité thermique et les cinétiques de formation des G4 tétramoléculaires : en particulier, les effets de la 8-méthylguanine et du L-ADN qui font l'objet de 2 articles présentés dans ce chapitre.

1.2.1 Modification du squelette (sucre - phosphate)

a) Modification de l'orientation des brins

En 2009, Veronica Esposito et ses collègues ont montré par RMN et dichroïsme circulaire (CD), l'influence de l'inversion de polarité 3'-3' ou 5'-5' sur la structure et la stabilité de G4 tétramoléculaires notamment la séquence d(TG₃T) [143] (Table 5). Nous avons étudié l'effet de ces modifications sur la cinétique d'association de ces complexes (résultats non publiés) (Table 5).

Cette inversion de polarité n'empêche pas la formation de G4, mais elle peut introduire des quartets composés de G en conformation *syn*, ce qui se traduit en général par un changement du spectre de CD. Dans l'ensemble, ces modifications n'améliorent ni la stabilité ni la cinétique de formation du G4 tétramoléculaire, sauf pour la séquence avec l'inversion de

polarité 5'-5' entre la 2^{ème} et la 3^{ème} G (VI = 3'TG^{5'}-5'GGT^{3'}). Cette modification introduit des G en *syn* sur le deuxième quartet et présente un spectre CD de type II (ou “antiparallèle”) (cf. : §VII.1.3.2). La structure formée est très stable : $T_{1/2} = 70^\circ\text{C}$ au lieu de $47,5^\circ\text{C}$ pour le G4 non modifié. La modification en question augmente la cinétique d'association par 2 par rapport à la structure canonique.

Séquences	$T_{1/2}$ (°C)	k_{on} ($\text{M}^{-3} \cdot \text{s}^{-1}$)	Empilement quartets	Spectre CD
5'TGGGT ^{3'} (ODN)	47.5	8.3×10^7	a-a-a	I
5'dS ^{3'} -3'GGGT ^{5'} (I)	42	2.7×10^7	a-a-a	I
5'T ^{3'} -3'GGGT ^{5'} (II)	47	9.1×10^7	a-a-a	I
5'TG ^{3'} -3'GGT ^{5'} (III)	51	t.l.	a-a-a	II
3'dS ^{5'} -5'GGGT ^{3'} (IV)	41	t.l.	a-a-a	I
3'T ^{5'} -5'GGGT ^{3'} (V)	47.5	9.3×10^7	s-a-a	h
3'TG ^{5'} -5'GGT ^{3'} (VI)	70	1.9×10^8	a-s-a	II

Table 5. Les effets de l'inversion de polarité sur la structure, stabilité et cinétique de formation des G4 tétramoléculaires

(t.l.) = trop lent ; (a) = anti ; (s) = syn ; (I) = spectre CD de type I (1 pic < 0 vers 240 nm et 1 pic > 0 vers 260 nm) ; (II) = spectre CD de type II (1 pic < 0 vers 260 nm et 2 pics > 0 vers 245 et 295 nm) ; (h) = spectre CD de type hybride (2 pics > 0 vers 260 et 295 nm) ; (dS) = “dSpacer” (nucléoside ne possédant pas de nucléobase).

b) LNA-quadruplexes

L'acide nucléique bloqué (LNA pour “Locked Nucleic Acid”) est un analogue contenant un pont méthylène entre l'hydroxyle en position 2 et l'atome de carbone 4 du sucre (Figure 21). Ce pont bloque le ribose en C3'-*endo* (observé dans l'ARN) et défavorise la formation de guanine en conformation *syn*. Ceci est probablement dû à la gêne stérique entre le H3' du sucre et le N³ de la nucléobase [144]. La première structure de LNA en G4 tétramoléculaire a été obtenue par RMN en 2004 [145]. Tous les sucres de la séquence d(TG₃T) ont été modifiés. Cependant la modification ne change pas énormément la structure formée : la séquence de LNA forme bien un G4 tétramoléculaire parallèle avec une hélicité droite, et toutes les guanines des quartets sont en position *anti*. Les G4 formés par les séquences de LNA ont une stabilité comparable à celle de l'ARN (r(UG₃U)), ils sont donc nettement plus stables que les G4 formés par les brins d'ADN. L'étude des cinétiques d'association et de dissociation de TG₃T modifié est réalisée, en 2006 [146]. Il a été montré que cette modification des sucres du G4 accélère son association et ralentit sa dissociation en comparant au G4 canonique. De plus, ils ont trouvé que l'association du G4-LNA n'est pas d'ordre ≈ 4 comme celui du G4-ADN, mais est seulement d'ordre 2.

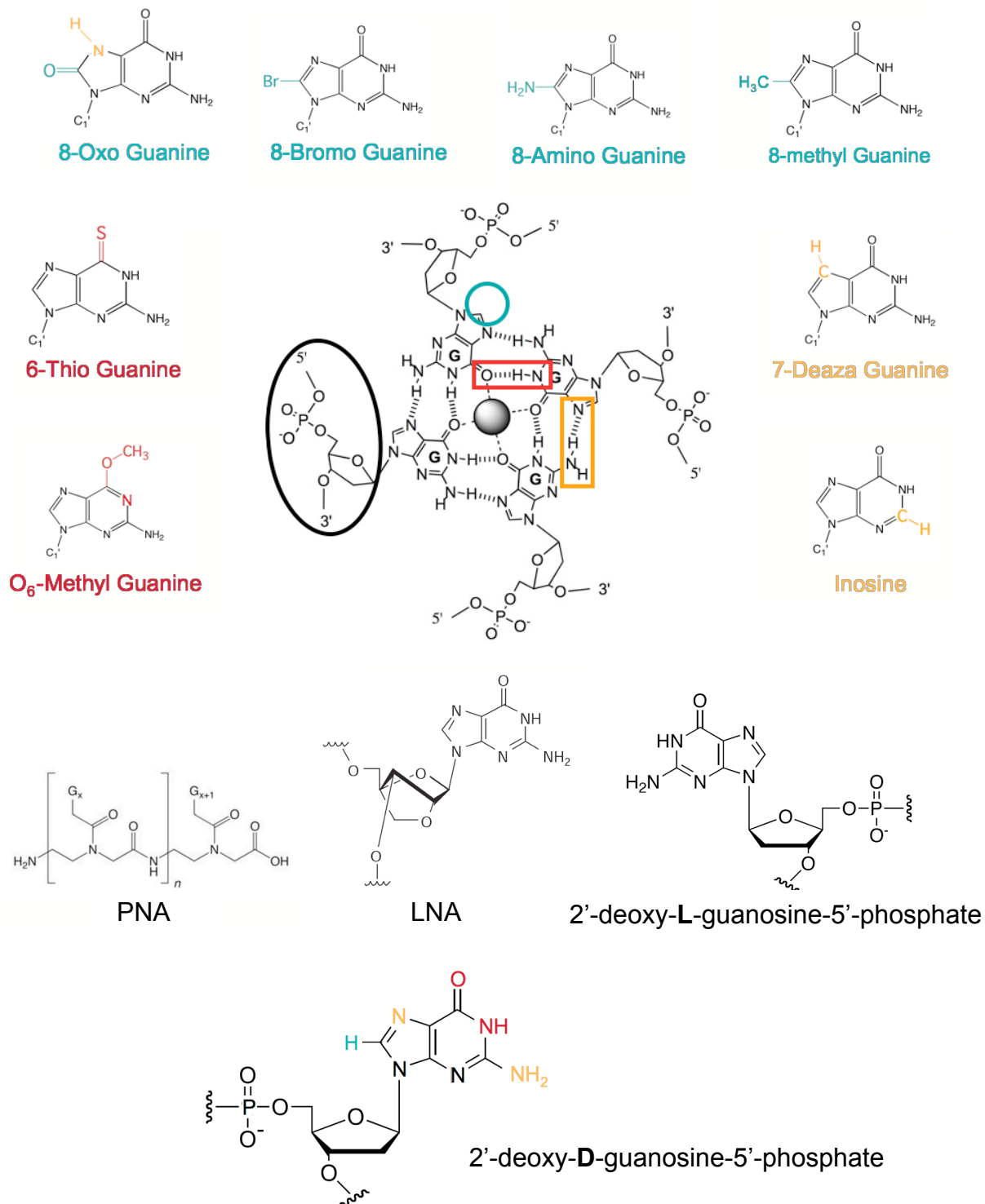


Figure 21. Analogues de guanines et d'autres modifications du squelette

En vert: H⁸ de guanine remplacé par d'autres substituants plus encombrants et polarisés. En rouge: modifications perturbant le réseau interne de liaisons H. En orange: modifications perturbant le réseau externe de liaisons H. En noir: modifications du squelette de l'ADN (PNA, LNA et L-ADN).

c) PNA-quadruplexe

L'acide nucléique peptide (PNA pour "*Peptide Nucleic Acid*") est un analogue d'acide nucléique où la liaison phosphodiester est remplacée par une liaison peptidique (Figure 21). Le squelette du PNA ne contient pas de groupement phosphate, et n'est donc pas chargé. La formation de G4 composé entièrement de PNA est possible, d'après l'article publié par l'équipe de Shankar Balasubramanian en 2004 [147]. Cependant, le G4-PNA formé (pTG₃) est beaucoup moins stable que le G4-ADN et, contrairement au G4 canonique, il est peu stabilisé, voire déstabilisé, par les cations (Na⁺ ou K⁺). Du fait de l'absence de charge dans le squelette PNA, l'augmentation de la force ionique ne stabilise pas les structures PNA en solution basées sur des interactions non covalentes. Malgré l'achiralité du squelette PNA, il semblerait possible de comparer les spectres de dichroïsme circulaire obtenus par des séquences PNA et ADN. Le G4-PNA présente un spectre de CD de type II ("antiparallèle") au lieu de type I ("parallèle") pour le G4-ADN. Comme nous l'avons évoqué dans le chapitre VII (cf. : §VII.1.3.2), l'interprétation des spectres CD est plutôt délicate. Nous ne pouvons pas affirmer que le spectre CD de type II est dû à la formation d'un G4 antiparallèle. Cela dit, l'arrangement antiparallèle du G4-PNA alignerait les extrémités N- et C-terminales (chargées positives et négatives respectivement) des brins de PNA, ce qui serait plus favorable qu'un arrangement parallèle.

Nous devons signaler qu'au départ, les séquences PNA ont été utilisées pour l'invasion de G4 en ADN [148, 149] ou en ARN [150]. Cette invasion de la séquence PNA pourrait participer à la formation et à la stabilisation du G4 [151]. En effet, le "G4 hybride" composé d'ADN et de PNA est plus stable que le G4-ADN. Récemment, il a été montré que la vitesse de formation de cet hybride ADN-PNA pourrait varier selon la conformation du G4 [152] : les structures G4-ADN parallèles seront plus accessibles pour le PNA ce qui permettrait une hybridation du PNA plus rapide pour ces cibles.

d) L - acides nucléiques

Les acides nucléiques sont des molécules chirales. Ils possèdent deux formes énantiomères : la forme "naturelle" D (pour "*dextrogyre*") qui compose l'ADN ou l'ARN canonique et la forme L (pour "*levogyre*") qui est l'image de l'ADN ou l'ARN naturel dans un miroir (Figure 21). Les premiers nucléotides en séries L- sont synthétisés par Holy et ses collègues, en 1969 [153]. Les L - acides nucléiques sont connus pour leur résistance à certaines nucléases, permettant à un "*spiegelmer*" ("*Spiegel*" signifie miroir en allemand)

d'entrer en essai clinique [154] et à de nombreuses applications en nanotechnologie et en biotechnologie qui seront détaillées dans le chapitre VII (cf. : §VII.1.3.4 et §VII. 1.3.5).

L'ADN "*levogyre*" (L-ADN) forme des duplexes avec les mêmes propriétés physiques (solubilité et stabilité) que l'ADN "*dextrogyre*" (D-ADN) sauf pour la chiralité, entraînant la formation d'hélices gauches [155]. Cependant, le duplexe hétérochiral résultant de l'association de deux brins L et D n'est pas stable [156].

À ce jour, très peu de recherches s'intéressent à la formation de G4 par les L-ADN. Seule H. Urata et ses collègues ont affirmé en 2002 que la séquence L-d(GTG₂TG₃TG₃T) est capable de former des G-quadruplexes avec une chiralité inversée par rapport à l'ADN non modifié [157]. Cette séquence est l'énantiomère de l'analogue de Zintevir et présente une structure et une stabilité thermique équivalentes à son homologue. Le Zintevir est un ADN 17 mer capable de former un G4 et présente une forte activité anti-VIH-1. Il a été montré, en 2004, que le L-ADN montrait aussi une activité anti-VIH-1 comparable [158]. Cela nous a emmené à étudier la formation de G4 par L-ADN en utilisant la séquence d(TG₄T) comme modèle d'étude et des méthodes telles que la spectroscopie d'absorption (détermination des $T_{1/2}$ et des k_{on}), le dichroïsme circulaire (CD) et la mobilité électrophorétique. Ce travail est présenté dans l'**article 2**. Récemment, l'équipe d'Aldo Galeone a étudié l'effet de l'insertion de L-nucléotides sur la structure et la stabilité du G4, en utilisant également le modèle d(TG₄T) [159]. Les G4 chimériques (mariant D- et L- nucléotides) forment également des G4 tétramoléculaires stables, ce qui a été montré par spectres CD et RMN. Néanmoins la position de l'insertion peut influencer la chiralité et la stabilité du G4. L'extrémité 3' et le centre du bloc de guanines sembleraient plus importants que l'extrémité 5' pour la détermination de son hélicité.

1.2.2 Modification des bases

L'incorporation des bases modifiées dans les oligonucléotides peut entraîner un changement de structure et/ou un changement des propriétés physiques et biologiques des structures d'ADN résultantes. Ces dernières années, des équipes se sont attachées à la synthèse d'analogues de guanines. Ces modifications peuvent être classées en fonction de leurs impacts sur la reconnaissance entre guanines voisines dans un G-quartet, via des liaisons H (Figure 21) [141, 160]. Les modifications dont le nom est marqué en rouge (6-thioguanine, O⁶-méthyleguanine) perturbent le réseau interne de liaisons H, celles dont le nom est écrit en orange (inosine, 7-déazaguanine) perturbent le réseau externe de liaisons H, tandis que les modifications présentées en bleu-vert (8-oxoguanine, 8-bromoguanine, 8-aminoguanine [141,

160] et 8-méthyleguanine [161]) remplacent le H en position 8 de la guanine par d'autres substituants plus encombrants et polarisés.

Notre équipe s'intéresse également au sujet. En 2007, Julien Gros a réalisé une analyse des propriétés physiques de séquences modifiées (12 types de modifications différents) par rapport à leurs homologues, d(TG₄T) et d(TG₅T) [141]. Il démontre que les effets de l'incorporation de la base modifiée dépendent fortement du type et de la position de la modification. La formation d'un G4 tétramoléculaire comportant un quartet modifié est possible bien que la présence d'une base modifiée entraîne le plus souvent une double pénalité, pour l'association et la dissociation du complexe. Néanmoins, certaines modifications échappent à la règle : notamment la 8-bromoguanine et la 8-aminoguanine, insérées aux extrémités 5' des séquences d(TG₄T) et d(TG₅T), permettent d'augmenter la vitesse d'association et le temps de demi-vie des G4 modifiés. En 2005, Virgilio et ses collègues ont également montré que l'insertion de la 8-méthyleguanine peut stabiliser la structure [161].

Certains analogues de guanines, comme la 8-bromoguanine [162] et la 8-méthyleguanine [161] favorisent la conformation *syn*. Cependant, la modification 8-bromoguanine n'est certainement pas la plus appropriée pour une étude détaillée des structures. En effet, le signal du proton H⁸ n'est plus observable en RMN dans ce cas. Pour étudier la conformation *syn* ou *anti* d'une guanine, il serait préférable d'utiliser la 8-méthyleguanine. Nous reportons, dans l'**article 1**, l'effet de cette modification sur la structure, la stabilité et la cinétique de formation des G4 tétramoléculaires, en utilisant la RMN, la spectroscopie d'absorption, le dichroïsme circulaire et la mobilité électrophorétique.

2) Propriétés structurales et cinétiques des G-quadruplexes tétramoléculaires modifiés, 8-méthyleguanine et L-ADN en particulier

2.1 Article 1 : “Effects of 8-methylguanine on structure, stability and kinetics of formation of tetramolecular quadruplexes”. Biochimie - 2011; 93(3): 399-408

Tran, P.L.T., Virgilio, A. & Esposito, V., Citarella, G., Mergny, J.L. and Galeone, A.

Résumé en français :

Les G-quadruplexes tétramoléculaires résultent de l'association de quatre brins riches en guanines. La modification du squelette ou des bases guanines de l'oligonucléotide pourrait améliorer la stabilité. Dans cet objectif, la position 8 de la guanosine est particulièrement appropriée pour l'introduction des modifications puisqu'elle est située dans les sillons de la structure G4. Les modifications à cette position ne devraient pas perturber l'assemblage structural qui modifierait les sites de Watson-Crick et de Hoogsteen. Dans cette étude, nous analysons l'effet d'un résidu 8-méthyle-2'-déoxyguanosine (M) sur la structure et la stabilité de G-quadruplexes tétramoléculaires parallèles. Dans certains cas, la présence de cette modification entraîne la formation d'un quadruplexe tétramoléculaire inhabituel contenant des quartets appelés tout-*syn* (4 guanines du quartet sont en conformations *syn*). De plus, le nucléoside modifié (M) à l'extrémité 5' de la séquence accélère 15 fois ou plus la formation du quadruplex comparé à l'oligonucléotide non modifié. Cela rend cet analogue de guanine attrayant dans le contexte des G4 tétramoléculaires parallèles.

Article 1



Research paper

Effects of 8-methylguanine on structure, stability and kinetics of formation of tetramolecular quadruplexes

Phong Lan Thao Tran^{a,c,1}, Antonella Virgilio^{b,1}, Veronica Esposito^b, Giuseppe Citarella^b, Jean-Louis Mergny^{a,c,**}, Aldo Galeone^{b,*}

^aINSERM, U869, European Institute of Chemistry and Biology, Bordeaux University, 2 rue Robert Escarpit, Pessac F-33607, France

^bDipartimento di Chimica delle Sostanze Naturali, Università degli Studi di Napoli Federico II, via D. Montesano, 49, I-80131 Napoli, Italy

^cINSERM, U565, Acides Nucléiques: Dynamique, Ciblage et Fonctions Biologiques, Muséum National d'Histoire Naturelle (MNHN) USM503, CNRS, UMR5153, Département de "Régulations, Développement et Diversité Moléculaire", 43 rue Cuvier, CP26, Paris Cedex 5, F-75231 France

ARTICLE INFO

Article history:

Received 12 October 2010

Accepted 19 October 2010

Available online 27 October 2010

Keywords:

Parallel G-quadruplex

8-Methyl-2'-deoxyguanosine

Glycosidic conformation

Abbreviations:

TDS

thermal difference spectra

IDS

isothermal difference spectra

CD

circular dichroism

ABSTRACT

Tetramolecular G-quadruplexes result from the association of four guanine-rich strands. Modification of the backbone strand or the guanine bases of the oligonucleotide may improve stability or introduce new functionalities. In this regard, the 8 position of a guanosine is particularly suitable for introduction of modifications since as it is positioned in the groove of the quadruplex structure. Modifications at this position should not interfere with structural assembly as would changes at Watson–Crick and Hoogsteen sites. In this study, we investigated the effect of an 8-methyl-2'-deoxyguanosine residue (M) on the structure and stability of tetramolecular parallel G-quadruplexes. In some cases, the presence of this residue resulted in the formation of unusual quadruplex structures containing all-syn tetrads. Furthermore, the modified nucleoside M at the 5'-end of the sequence accelerated quadruplex formation by 15-fold or more relative to the unmodified oligonucleotide, which makes this nucleobase an attractive replacement for guanine in the context of tetramolecular parallel quadruplexes.

© 2010 Elsevier Masson SAS. All rights reserved.

1. Introduction

G-quadruplexes are a polymorphic class of higher-order nucleic acid structures in which the structural unit is formed by a planar arrangement of four guanines, known as *G-quartets* or *G-tetrads*. A vertical stacking of several G-quartets and the presence of monovalent cations provide these structures with remarkable stabilities. G-quadruplexes may find application in fields of molecular biology, genetics, pharmaceuticals and nanotechnology [1].

Modifications to the bases or the backbone of oligonucleotides that form quadruplexes are performed with the aims of improving

stability or providing G-quadruplexes with unique properties. Chemical modification of strands can also reveal clues to features of these higher-order structures. A number of modifications to the base moiety and/or the sugar-phosphate backbone are known [2,3]. One of the simplest ways to prepare a modified quadruplex is to introduce a guanine base analogue into the sequence. The effects of incorporation of a number of base analogues on tetramolecular parallel quadruplex (G4) formation have been reported. This research suggests that most quartets formed by oligonucleotides with modified bases do not improve the structural stability and exist only thanks to the docking platform provided by the neighbouring G-quartets: As stated before, "guanines are a quartet's best friend" [3].

Noteworthy exceptions are guanine analogues modified at the 8 position. This position is positioned in the groove of the quadruplex structure, and modifications at this site do not hamper Watson–Crick and Hoogsteen pairing. 8-Bromo-guanine and 8-amino-guanine are particularly interesting since these modifications have been shown to accelerate quadruplex formation [3,4]. Incorporation of 8-amino-guanines can lead to a significant quadruplex polymorphism [4]. However, for these analogues, the absence of non-exchangeable

* Corresponding author. Dipartimento di Chimica delle Sostanze Naturali, Università degli Studi di Napoli Federico II, via D. Montesano, 49, I-80131 Napoli, Italy. Tel.: +39 081678542; fax: +39 081678552.

** Corresponding author. INSERM, U869, European Institute of Chemistry and Biology, Bordeaux University, 2 rue Robert Escarpit, Pessac F-33607, France. Tel.: +33 5 4000 30 22; fax: +33-5 57 571 015.

E-mail addresses: jean-louis.mergny@inserm.fr (J.-L. Mergny), galeone@unina.it (A. Galeone).

¹ These authors contributed equally to this work.

protons at position 8 represents a major drawback in view of a detailed structural investigation. For example, the direct proof of a *syn* or *anti* glycosidic conformation can be easily afforded by using NMR and, particularly, by comparing the NOEs involving the 8 proton and the H1' or the H2' sugar protons. A methyl group, by virtue of its protons, is more useful for structural investigations, particularly taking into account its steric size comparable to that of the bromine atom, that makes it similarly able to promote the *syn* glycosidic conformation (Fig. 1). Furthermore, the presence of non-exchangeable protons at position 8 may provide structural information in addition to glycosidic angle due to NOE contacts.

The 8-methyl-2'-deoxyguanosine residue favours the *syn* glycosidic conformation as already proven both in Z-DNA [5–7] and in antiparallel quadruplex structures. For example, the substitution of one or more *syn* 2'-deoxyguanosines by 8-methyl-2'-deoxyguanosine residues improves the thrombin inhibitory activity of the quadruplex forming aptamer TBA [8,9]. Similarly, the 8-methyl-2'-deoxyguanosine stabilizes the quadruplex structure proposed for the 5'-end of the retinoblastoma susceptibility gene, when it replaces a residue that adopts the *syn* conformation [10]. On the other hand, the same substitution for G-residues that adopt the *anti* conformation usually produces a lower biological activity or stability [8,9]. However, the effects of this substitution on the parallel quadruplex structure [d(TG₃T)]₄, in which all G-residues are known to adopt an *anti* conformation, are sequence-dependent [11]. The substitution of the first G-residue, quite unexpectedly, results in a parallel quadruplex containing an unusual all-*syn* G-tetrad characterized by an apparent melting temperature higher than its natural counterpart. In contrast, the substitution of the second G-residue did not change the glycosidic preference compared to the natural quadruplex [d(TG₃T)]₄, notwithstanding the presence of the methyl group usually prone to induce a *syn* conformation.

In order to investigate the relationship among the glycosidic angle preference, the thermal stability, the kinetic properties and the modified position in the sequence, six oligodeoxynucleotides (ODNs) (I–VI, Table 1) containing an 8-methyl-2'-deoxyguanosine (M) have been investigated by several techniques. In a previous report [11], we investigated the quadruplex structures formed by the mono-substituted TG₃T analogues (ODNs I–II) by NMR, CD and molecular mechanics techniques. In this work we have synthesized and analyzed the quadruplex structures adopted by the mono-substituted TG₄T analogues (ODNs III–VI) by the same techniques. Furthermore, a kinetic analysis has been performed for all modified ODNs.

2. Materials and methods

2.1. Oligonucleotides synthesis and purification

Unmodified oligonucleotides TG₃T and TG₄T were purchased by Eurogentec (Seraing, Belgium). Their concentrations were estimated using molar extinction coefficients provided by the

Table 1

Sequence of the oligonucleotides investigated, apparent melting temperatures ($T_{1/2}$) and association kinetic constants (k_{on}). M = 8-methyl-2'-deoxyguanosine. See text for details.

Name	Sequence	$T_{1/2}$ (°C) at λ_{max} (CD) (70 mM K ⁺ , 100 μ M s.s. ODN)	$T_{1/2}$ (°C) at λ_{max} (CD) (70 mM Na ⁺ , 100 μ M s.s. ODN)	k_{on} (4 °C) M ⁻³ s ⁻¹
TG ₃ T	TGGGT	45 ^a	^c	1.26×10^8
I	TMGGT	66 ^a	^c	1.88×10^9
II	TGMGT	52 ^a	^c	2.49×10^8
TG ₄ T	TGGGGT	^b	65	2.15×10^9
III	TMGGGT	^b	75	3.15×10^{10}
IV	TGMGGT	^b	65	5.41×10^9
V	TGGMGT	^b	47	1.46×10^9
VI	TGGGMT	79	32	2.13×10^8

^a See ref. [11].

^b Not determined (too stable).

^c Not determined.

manufacturer (47,700 and 57,800 M⁻¹ cm⁻¹, respectively). The modified oligonucleotides I–VI were synthesized on a Millipore Cyclone Plus DNA synthesizer using solid phase β -cyanoethyl phosphoramidite chemistry at 15- μ mol scale. The synthesis of the suitably protected 8-methyl-2'-deoxyguanosine-3'-phosphoramidite was performed following the synthetic strategy described by Khoda et al. [12]. The contribution of the methyl group to the molar extinction coefficient was considered negligible. Oligomers were detached from the support and deprotected by treatment with concentrated aqueous ammonia at 55 °C for 12 h. The combined filtrates and washings were concentrated under reduced pressure, redissolved in H₂O and analysed and purified by high-performance liquid chromatography (HPLC) on a Nucleogel SAX column (Macherey–Nagel, 1000-8/46). Buffer A was 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN, and buffer B was 1 M KCl, 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN. A linear gradient from 0 to 100% B for 30 min and flow rate 1 ml/min were used. Oligomers were desalted using Sep-pak cartridges (C-18). The isolated oligomers proved to be >98% pure by NMR.

2.2. Non-denaturing gel electrophoresis

Non-denaturing gel electrophoresis allows separation of single-stranded oligonucleotides from tetramolecular G-quadruplex structures. Samples were loaded on a 20% polyacrylamide (acrylamide/bis-acrylamide 19:1) gel containing TBE and KCl at 10 mM. Electrophoresis was performed at 4 W/gel to reach a temperature close to 19 °C (migration in the cold room) or 35 °C (electrophoresis performed at room temperature). To achieve complete quadruplex formation, samples (indicated by +) were incubated during 48 h, at 4 °C and at high strand concentration (300 μ M), in 100 mM KCl. In parallel, samples (indicated by –) were incubated in 40 mM LiOH at 37 °C during 15 min and were neutralized by 40 mM HCl. Lithium cacodylate buffer was prepared by mixing cacodylic acid with LiOH. Bands were revealed by UV shadowing (90 μ M of oligonucleotide) using a UV light source (254 nm) and a digital camera. This method does not require any labeling of any kind and relies solely on the absorbance of the nucleic acid in the far UV region (254 nm). We compared migrations of both unmodified and modified sequences with or without cations.

2.3. Circular dichroism (CD) and absorbance differential spectra (TDS and IDS)

Our reference conditions for this study were 10 mM Lithium cacodylate pH 7.2 supplemented with 100 mM KCl. CD spectra were

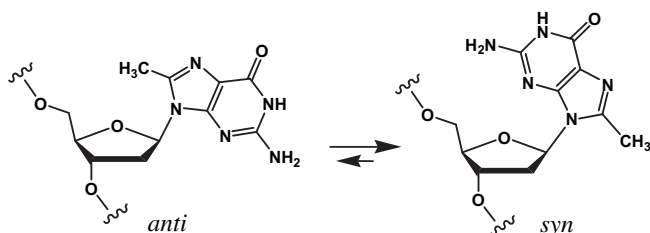


Fig. 1. The *syn/anti* equilibrium of a 8-methyl-2'-deoxyguanosine residue.

recorded on a JASCO-810 spectropolarimeter using a 0.2- to 1 cm path length quartz cuvettes as previously described [4]. Thermal difference spectra (TDS) were obtained by calculating the difference between the absorbance spectra recorded above and below the observed transition [13]. Isothermal difference spectra (IDS) were obtained by calculating the difference between the absorbance spectra of the folded and unfolded forms of a sample (after and before an isothermal kinetics experiment).

2.4. Isothermal kinetics

These experiments were performed by starting from isolated strands and comparing association of both modified and unmodified sequences by UV-isothermal experiments at 6.6 °C in K⁺ conditions as previously described [3]. To study the impact of M at a given position, data was fit using a model previously published [3,4] and association rate constants (k_{on}) were calculated. The corresponding order of the association reaction was assumed to be $n = 4$, as in previous studies. Strand concentrations between 10 and 100 μ M were tested.

2.5. Melting experiments

CD samples of the quadruplexes **III**–**VI** and their natural counterpart [d(TGGGGT)]₄ were prepared at a concentration of 1×10^{-4} M, by using the buffer solution used for NMR experiments (10 mM KH₂PO₄, 70 mM KCl, 0.2 mM EDTA, pH 7.0), and the corresponding Na⁺ buffer (10 mM NaH₂PO₄, 70 mM NaCl, 0.2 mM EDTA, pH 7.0). CD melting curves were registered on a Jasco 715 CD spectrophotometer in a 0.1 cm pathlength cuvette using a thermoelectrically-controlled cell holder (Jasco PTC-348). CD melting curves were determined as a function of temperature from 20 to 90 °C at the maximum effect Cotton wavelengths for all quadruplexes with a scan rate of 10 °C h⁻¹.

2.6. Nuclear magnetic resonance

NMR samples were prepared at a concentration of ~5 mM in 0.6 mL (H₂O/D₂O 9:1 v/v) buffer solution containing 10 mM KH₂PO₄/K₂HPO₄, 70 mM KCl and 0.2 mM EDTA (pH 7.0). All the samples were heated for 5–10 min at 80 °C and slowly cooled (10–12 h) to room temperature. The solutions were equilibrated for several weeks at 4 °C and ¹H NMR spectra were recorded using pulsed-field gradient WATERGATE [15] for H₂O suppression. The annealing process was assumed to be complete when ¹H NMR spectra were superimposeable on changing time. For D₂O experiments, the H₂O was replaced with D₂O by drying down the sample, lyophilization and redissolution in D₂O alone. NMR spectra were recorded with a Varian Unity INOVA 700 MHz spectrometer. ¹H chemical shifts were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). Phase sensitive NOESY spectra [16] were recorded with mixing times of 180 ms ($T = 25$ °C). Pulsed-field gradient WATERGATE was used for NOESY spectra in H₂O with 200 ms mixing times. TOCSY spectra [17] with mixing times of 120 ms were recorded with D₂O solutions. NOESY and TOCSY were recorded using a TPPI [18] procedure for quadrature detection. In all 2D experiments, the time domain data consisted of 2048 complex points in t₂ and 400–512 fids in t₁ dimension. The relaxation delay was kept at 1.2 s for all experiments.

2.7. Molecular modelling

The main conformational features of the quadruplexes **III**, **IV** and **VI** were explored by means of a molecular modelling study. The AMBER force field using AMBER 99 parameter set was used [19].

The initial coordinates for the starting model of the quadruplex [d(TGGGGT)]₄ were taken from the NMR solution structure of the quadruplex [d(TTGGGGT)]₄ (Protein Data Bank entry number 139D), with one of the four available structures chosen randomly. The initial [d(TGGGGT)]₄ G-quadruplex model was built by deleting the first thymidine residue in each of the four d(TTGGGGT) strands. The complete structures of quadruplexes were then built using the Biopolymer building tool of Discover by deleting, one at a time, 2'-deoxyguanosines G2 (**III**), G3 (**IV**) and G5 (**VI**) and replacing them with an 8-methyl-2'-deoxyguanosine residue for each strand. As for NMR results, the modified residues were arranged in the *syn* conformations for G2 of **III** and G5 of **VI** and in the *anti* conformation for G3 of **IV**. The calculations were performed using a distance-dependent macroscopic dielectric constant of 4 ϵ_r and an infinite cut-off for non-bonded interactions to partially compensate for the lack of solvent used [20]. Using the steepest descent followed by quasi-Newton–Raphson method (VA09A), the conformational energy of each complex was minimized until convergence to an RMS gradient of 0.1 kcal/mol Å was reached. Illustrations of structures were generated using the INSIGHT II program, version 2005 (Accelrys, San Diego, CA, USA). All the calculations were performed on a PC running Linux ES 2.6.9.

3. Results and discussion

3.1. Biophysical analysis

In order to demonstrate the formation of tetramolecular quadruplex structures, we first used classical biochemical and biophysical methods. A simple assay usually used for this aim is the migration analysis on a non-denaturing polyacrylamide gel (PAGE). As shown in Fig. 2, incubation of oligonucleotides **I**–**VI**, as well as their unmodified counterparts TG₃T and TG₄T, in the presence of potassium ions, led to the appearance of a retarded band in all samples, as compared to a single band in the single-stranded controls. This retarded band was preferentially stained with different dyes, such as SYBR Gold, suggesting G4 formation (data not shown). However, it should be noted that this difference in mobility depended on the gel conditions: at lower acrylamide content (12% rather than 15%) an abrogation or a reversion of the migration behaviour was observed (i.e., in 12% acrylamide, the G4 migrates *faster* than the single strands!). Although these data are interesting, their analysis using Ferguson plots in order to extrapolate mobility to 0% acrylamide is beyond the scope of this work. These data emphasize the complexity of the electrophoresis process in quadruplex structures analysis and suggest a complex relationship between molecular shape, size and charge. The intricate dependence of quadruplex motility on molecular properties is also corroborated by the relative positions of the G4 bands: Whereas the positions of the single strands are relatively homogeneous (compare “–” bands in Fig. 2), positions of the quadruplexes greatly differ (“+” bands). In general, modified quadruplexes migrated faster than their natural counterparts. Conversion to the associated species was nearly complete in all cases (little or no material remained at the original position). In general, a main band was found for most of the quadruplexes, although minor amounts of different species were detected for some oligonucleotides in the gel electrophoresed at 19 °C (Fig. 2A). A complete conversion to a single band was observed for all modified oligonucleotides when the gel was run at approximately 35 °C (Fig. 2B).

Formation of quadruplexes was further confirmed by classical spectroscopic methods. Fig. 3A shows the isothermal absorbance difference spectra (IDS) of the modified quadruplex structures and their natural counterparts. Although differences in intensities were found, the shape of the IDS profiles were relatively similar and very

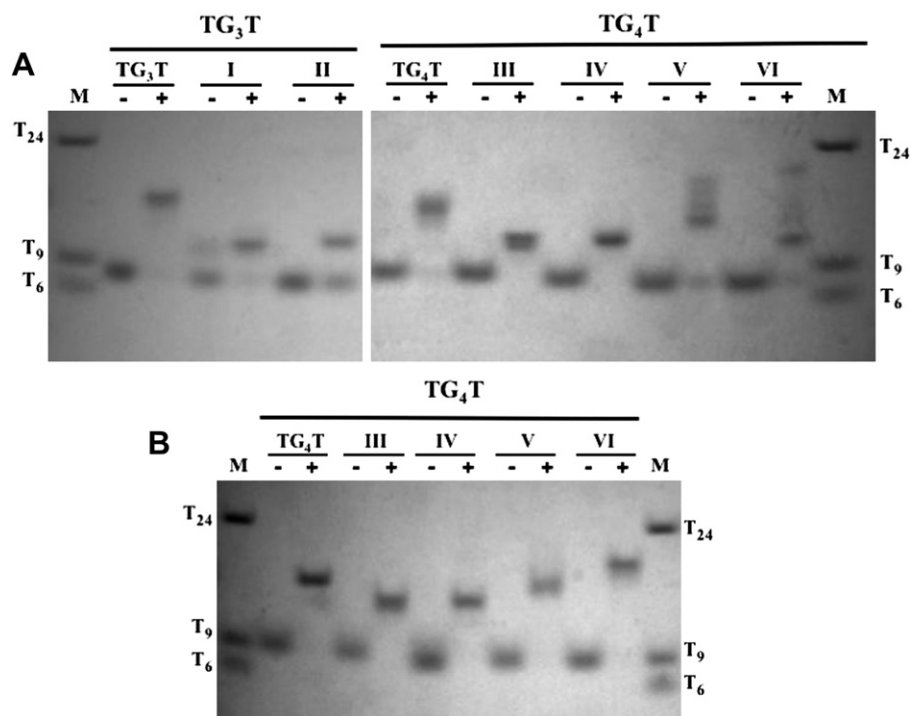


Fig. 2. Behaviour of the G-quadruplex forming oligonucleotides on a non-denaturing gel. Samples were loaded on non-denaturing gels, and gels were run at 19 °C (A) and 35 °C (B). Oligonucleotides were revealed by UV shadowing. Each oligonucleotide was loaded at 90 μ M strand concentration. Lanes “–” correspond to the sequence pre-treated with LiOH (40 mM, 15' at 37 °C) followed by neutralization by 40 mM HCl and immediate loading on a gel. Lanes “+” correspond to the sequence incubated in 10 mM lithium cacodylate 100 mM KCl for 48 h at 4 °C M corresponds to the migration markers (dT₂₄, dT₉ and dT₆).

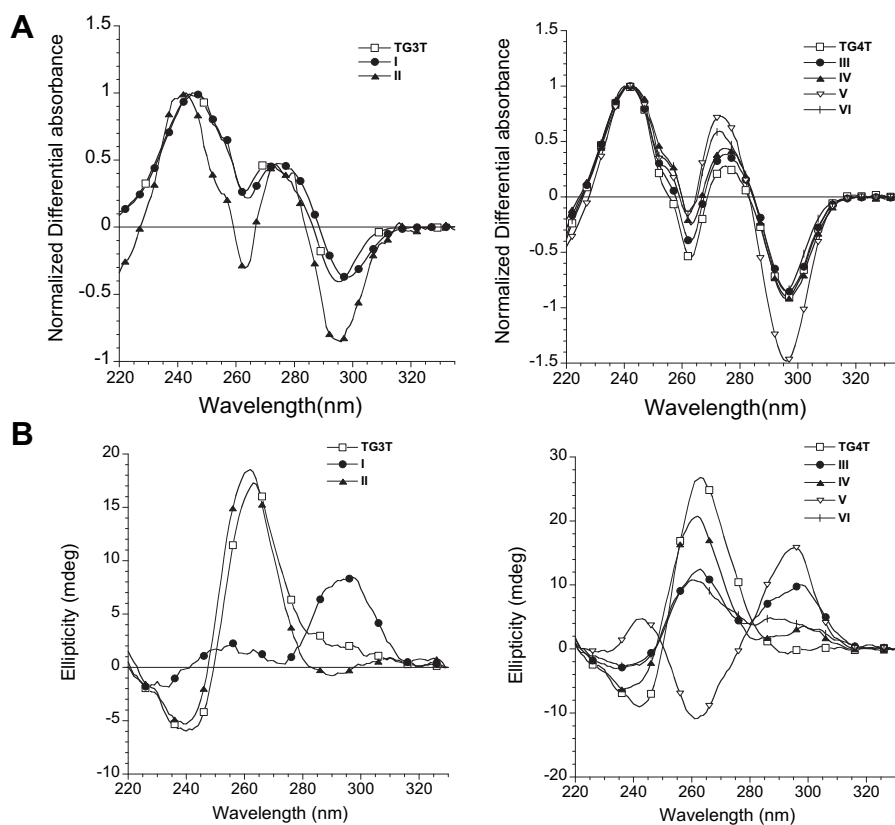


Fig. 3. Circular dichroism (CD) and isothermal difference spectra (IDS). (A) Isothermal difference spectra resulting from the difference between the absorbance recorded at 4 °C before and after annealing (for 24 h) in 100 mM KCl containing 10 mM lithium cacodylate at pH 7.2. Data were normalized ($IDS_{norm} = IDS/\max(IDS)$) over the 220–335 nm wavelength range. (B) CD spectra recorded at 4 °C (in 100 mM K⁺). Oligonucleotides were prepared at a concentration 10 μ M. Spectra were recorded one month after the annealing.

close to those previously reported for tetramolecular complexes [3,14,21]. These profiles confirm that long incubations of the single strands at low temperature in the presence of potassium ions lead to the slow conversion of single strands to quadruplexes.

In contrast, analysis of CD spectra revealed differences undetectable by the IDS profiles (Fig. 3B). It is rather instructive to note that quadruplex structures with the same maxima on IDS spectra had completely different CD spectra. In Fig. 3B the CD spectra (4 °C, 100 mM KCl and 10 μ M oligonucleotide strand) of the modified ODNs I–VI and their natural counterparts are shown. ODNs II and IV showed Type I (previously referred to as “parallel”) CD spectra resembling the spectra of their natural counterparts TG₃T and TG₄T with a maximum around 265 nm and a minimum around 240 nm. In contrast, CD spectra of I and V showed Type II “antiparallel” profiles with two maxima around 245–250 and 295 nm and minima around 265 nm, whereas CD spectra of III and VI showed two maxima around 260 and 290–295 nm. The interpretation of these CD spectra and the attribution of strand polarity (“parallel” or “antiparallel”) remains complicated, although the occurrence of a maximum around 290–295 nm generally suggests the presence of G residues in a *syn* glycosidic conformation [22]. It is noteworthy that for ODNs IV and VI, CD profiles strongly depended on incubation time after annealing. Fig. S1 shows CD spectra of ODNs IV and VI recorded after a relatively short incubation time (48 h) and after a long incubation time (more than one month). In the case of IV, the maximum around 265 increased, while maximum around 290–295 decreased with incubation time. In contrast, ODN VI

showed the opposite behaviour. These data suggest that after annealing two or more quadruplex species could be involved in an equilibrium that, very slowly, leads to the formation of a major species.

In order to evaluate the thermal stability of the quadruplex structures, we performed CD melting experiments (Fig. S2); the results are summarized in Table 1 and Fig. 4C. Oligonucleotides III, IV and V formed remarkably heat-resistant structures in a buffer containing 70 mM K⁺, as did their natural counterpart TG₄T, with a significant quadruplex fraction remaining even at 90 °C. On the other hand, in these conditions, VI had an apparent melting temperature ($T_{1/2}$) of 79 °C, thus suggesting that the introduction of M at the 3'-end of the G-run resulted in decreased thermal stability as compared to TG₄T. In order to better understand the dependence of thermal stability on the position of M in the sequence, CD melting experiments in 70 mM NaCl solution were performed. The CD profiles in Na⁺ solution (data not shown) were very similar to those in K⁺ solution, indicating no significant structural differences between quadruplexes formed in sodium and potassium solutions. On the other hand, the trend of the apparent melting temperatures (Table 1) clearly shows that thermal stability decreases as the position of M is moved toward the 3'-end, and III (the modified quadruplex in which M is adjacent to the 5'-end) melted at a higher $T_{1/2}$ than TG₄T. As previously reported [11], $T_{1/2}$ values for the modified TG₃T structures show a similar trend (Fig. 4C).

Finally we performed a series of isothermal association experiments to compare the association kinetics of the different

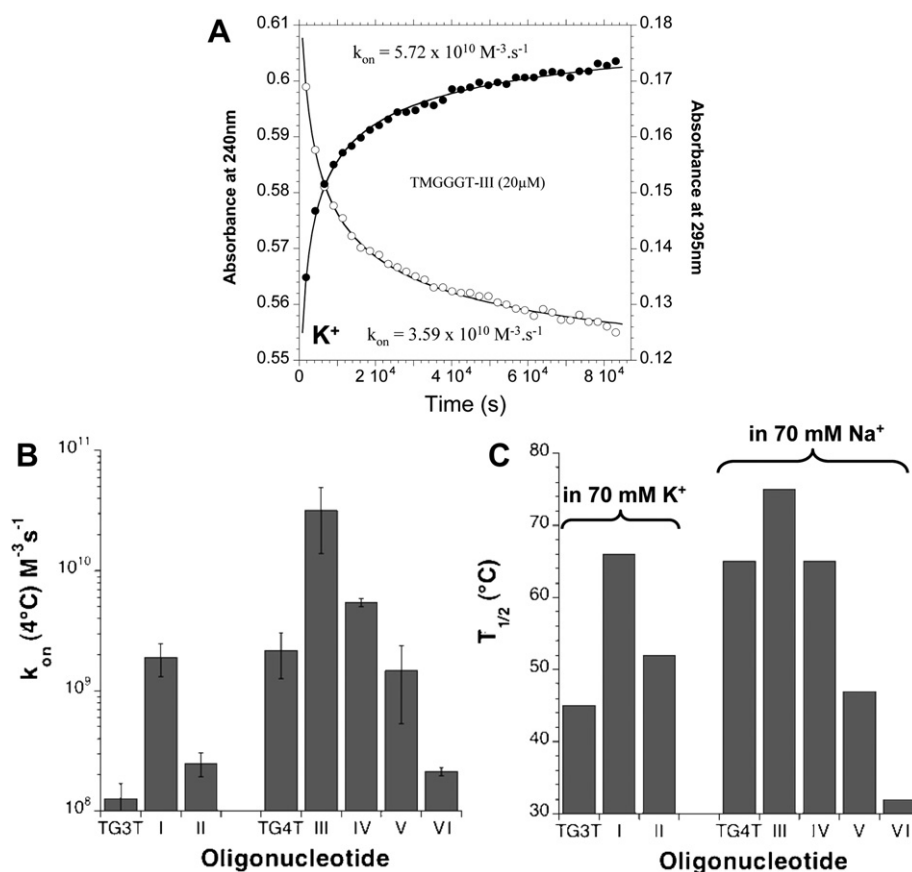


Fig. 4. Association kinetics. (A) Representative example of an isothermal renaturation experiment. Quadruplex III was formed at 20 μ M strand concentration, at 6.6 °C, in 100 mM KCl containing 10 mM lithium cacodylate (pH7.2). Raw absorbance was recorded simultaneously at two wavelengths (240 nm: open circles $\circ$ and 295 nm: filled circles $\bullet$). The fitted curves (black full lines) are nearly indistinguishable from the experimental data. In this example, fitted k_{on} values are provided for each curve (240 and 295 nm). All the values of k_{on} analysed in this paper come from fitted curves at 295 nm which are more reliable. (B) Summary of the observed values of k_{on} . (C) Summary of the apparent melting temperature values determined by CD. Note that ionic conditions are different between TG₃T and TG₄T due to major differences in stabilities.

sequences. An example of such an experiment is presented in Fig. 4A and results are summarized in Fig. 4B. A striking position-dependent effect was found: M was extremely favourable when inserted at the 5' position, leading to 15- (for **III**) to 16-fold (for **I**) faster association kinetics as compared to their unmodified counterparts. This effect was inverted when M was at the opposite end of the strand: **VI** associated 10 times slower than TG₄T. Internal positions had an intermediate behaviour: Insertion of M at the second position (**II** and **IV**) had a small beneficial effect (≈ 2 fold) on rate, whereas the association rate for **V** was not significantly different from that of TG₄T. From these experiments, we draw two conclusions: 1) When located at the 5'-end, M leads to a significant increase in the association rate (compare TG₃T with **I** and TG₄T with **III**) and 2) M affects the association rate in a position-dependent manner as modification of the 3' most guanine lead to a lower association rate than the unmodified oligonucleotide (compare TG₄T with **VI**).

3.2. Nuclear magnetic resonance and molecular modelling

As a long incubation time is required to obtain formation of a major structure in solution, NMR studies were performed several weeks after sample preparation. The CD profiles of ODNs **III–VI** obtained from the NMR samples (70 mM KCl, 100 μ M oligonucleotide strand) (Fig. S3) strictly resembled the CD profiles obtained in 100 mM KCl (Fig. 3B).

One of the distinctive features of structures containing G-tetrads is the appearance of imino proton resonances in the region between 10.5 and 12.0 ppm in ¹H NMR spectra [23]. Examination of this region is commonly used to assess whether the oligonucleotide

adopts a unique structure and to provide insight into its symmetry. Indeed, the simple appearance of ¹H NMR spectra of ODNs **III** and **IV**, with four main signals clearly evident in this region, indicates that, under the conditions utilized, both the modified oligomers form a single, well-defined hydrogen-bonded conformation consistent with highly symmetric G-quadruplex structures containing four G-tetrads (Fig. 5). The ¹H spectrum of **VI**, instead, had only two signals between 10.4 and 11.0 ppm, each of which, however, were attributable to two protons (Fig. 5).

Since all ODNs evaluated contain four G-residues in their sequences, the quadruplex structures formed by **III**, **IV** and **VI**, possess a four-fold symmetry (C₄) with all strands equivalent and parallel to each other. This symmetry was corroborated by the presence of five main singlets in the aromatic region between 7.0 and 8.0 ppm: three belonging to the guanine H8 and two to the thymine H6 protons. Furthermore, three methyl resonances in the range between 1.3 and 1.6 ppm (attributed to the two T-CH₃) and between 2.2 and 2.4 ppm (attributed to the M-CH₃) were observed for all three samples. The ¹H spectrum of **V** was more complicated than that of the other ODNs. Indeed, the imino proton region was more crowded and the number of signals suggests the presence of several types of quadruplex structures. The coexistence of multiple species prevented us from performing a resonance assignment and a structural study of this modified oligonucleotide.

The NOESY and TOCSY spectra of **III**, **IV** (500 MHz) and **VI** (700 MHz) at 25 °C showed well-dispersed cross peaks and consequently, both exchangeable and non-exchangeable protons could be nearly completely assigned following the standard procedures [24] (Table 2). As reported for other parallel quadruplex structures [25–27], the observed NOEs among G-H8 and T-H6 and

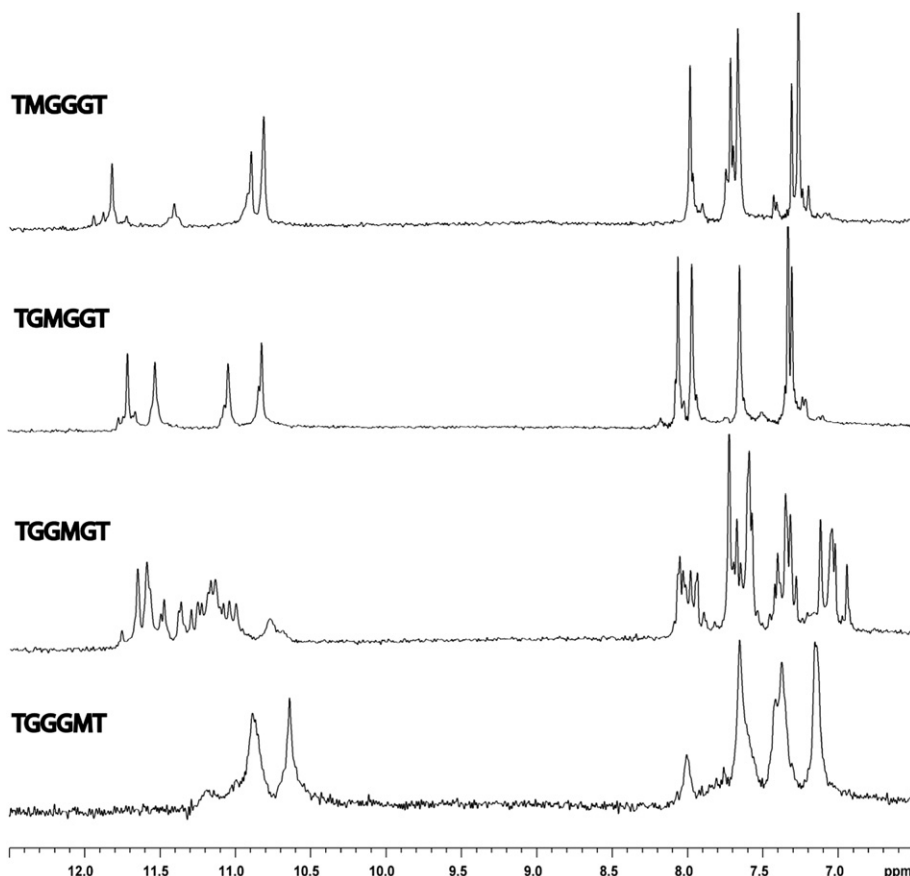


Fig. 5. ¹H NMR spectrum imino-aromatic regions of the mono-substituted TG₄T analogues **III–VI**.

Table 2

Proton chemical shifts for quadruplexes formed by ODNs **III**, **IV** (500 MHz) and **VI** (700 MHz) in 10 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 70 mM KCl and 0.2 mM EDTA (pH 7.0, $T = 25^\circ\text{C}$). N.D. = not determined.

	H8/H6	H1'	H2'/H2''	H3'	H4'	H5'/H5''	CH ₃	NH
III d(T ₁ M ₂ G ₃ G ₄ G ₅ T ₆)								
T ₁	7.37	5.99	2.08/2.47	4.69	3.90	N.D.	1.60	—
M ₂	—	5.93	2.92	4.91	4.29	4.18/3.79	2.29	11.90
G ₃	8.04	5.80	2.64/2.71	5.06	4.47	4.28	—	10.91
G ₄	7.77	6.09	2.69	5.02	4.55	4.33/4.24	—	10.95
G ₅	7.72	6.26	2.52/2.69	4.92	4.53	4.20/4.06	—	10.90
T ₆	7.32	6.05	2.17	4.46	4.05	N.D.	1.62	—
IV d(T ₁ G ₂ M ₃ G ₄ G ₅ T ₆)								
T ₁	7.29	5.87	2.07/2.26	4.61	4.00	3.92/3.61	1.36	—
G ₂	8.05	6.16	2.89/3.04	4.98	4.38	4.02/3.89	—	11.70
M ₃	—	6.07	2.59/2.98	5.05	4.22	4.18	2.29	11.52
G ₄	7.96	5.96	2.61/2.77	5.07	4.51	4.28/4.23	—	11.05
G ₅	7.64	6.29	2.53/2.68	4.85	4.27	4.19/4.07	—	10.84
T ₆	7.32	6.05	2.16	4.45	4.05	N.D.	1.60	—
VI d(T ₁ G ₂ G ₃ G ₄ M ₅ T ₆)								
T ₁	7.15	5.93	2.15/2.45	4.47	N.D.	N.D.	1.47	—
G ₂	7.68	5.99	2.53/2.68	5.02	4.49	4.23/3.37	—	10.67
G ₃	7.42	5.99	2.51	5.01	N.D.	3.98	—	10.89
G ₄	7.39	6.24	2.50/2.79	5.02	4.49	4.24/3.99	—	10.89
M ₅	—	5.99	2.46/3.36	4.95	4.46	4.17/4.00	2.20	10.70
T ₆	7.13	5.92	2.03/2.15	4.47	4.16	3.91	1.45	—

their own H1', H2' and H2'' ribose protons and the H1', H2' and H2'' protons on the 5'-side, suggest that all three quadruplexes assume a right-handed helical winding. As for the glycosidic torsion angles in **IV**, very weak NOEs between G-H8/M-CH₃8 and ribose H1' and strong NOEs between G-H8/M-CH₃8 and ribose H2' indicate that all residues (including the modified bases) possess an *anti* glycosidic conformation [28] (Fig. 6B).

Conversely, in the quadruplexes **III** and **VI** all the canonical guanine and thymine residues assume an *anti* conformation with the exception of the modified nucleotides (M), which adopt a *syn* glycosidic conformation as judged by the intense NOEs between the methyl group in 8-position and the H1' sugar proton and the weaker crosspeaks between methyl and H2' [29] (Fig. 6A and C). Because of the *syn* M nucleosides, the protons of methyl group in 8-position are >6 Å away from the sugar protons on the neighbouring 5' residue [30] and the sequential connectivity path through the strand is broken at 5'-T₁M₂-3' level for **IV** and at 5'-G₄M₅-3' level for **VI**. Furthermore, in the H₂O NOESY spectrum of **IV**, we observed sequential imino–imino NOEs arising from intra-strand contacts between the all-*anti* M-tetrad and both the overlying and underlying ones. Moreover, inter-strand NOEs were observed between the methyl group of the M residue and the NH proton of the modified base on the adjacent strand; inter-strand NOE contacts involving the H8 and NH protons are observed within the unmodified tetrads. This suggests that M residues are not randomly oriented but are in mutual close proximity to each other (data not shown).

In Fig. 7 schematic representations of the quadruplexes formed by ODNs **III**, **IV** and **VI** are shown. Using NMR data we built molecular models of quadruplex structures **III**, **IV** and **VI** (Fig. 8) as described in the experimental section. Apart from the presence of an additional G-tetrad, the model obtained for ODN **III** strictly resembled the previously reported model of **I** [11]. As expected, the structure shows a right-handed helical backbone geometry in which the strands are equivalent to each other. The modified residues assume *syn* glycosidic conformations without causing any distortions of the backbone, and the all-*syn* tetrad is planar. This results in good stacking between the five-membered rings of the M bases and the five-membered rings of the guanines beneath it. The model of quadruplex **IV** shows that all purines adopt an *anti* glycosidic conformation as shown by NMR. In this case, the

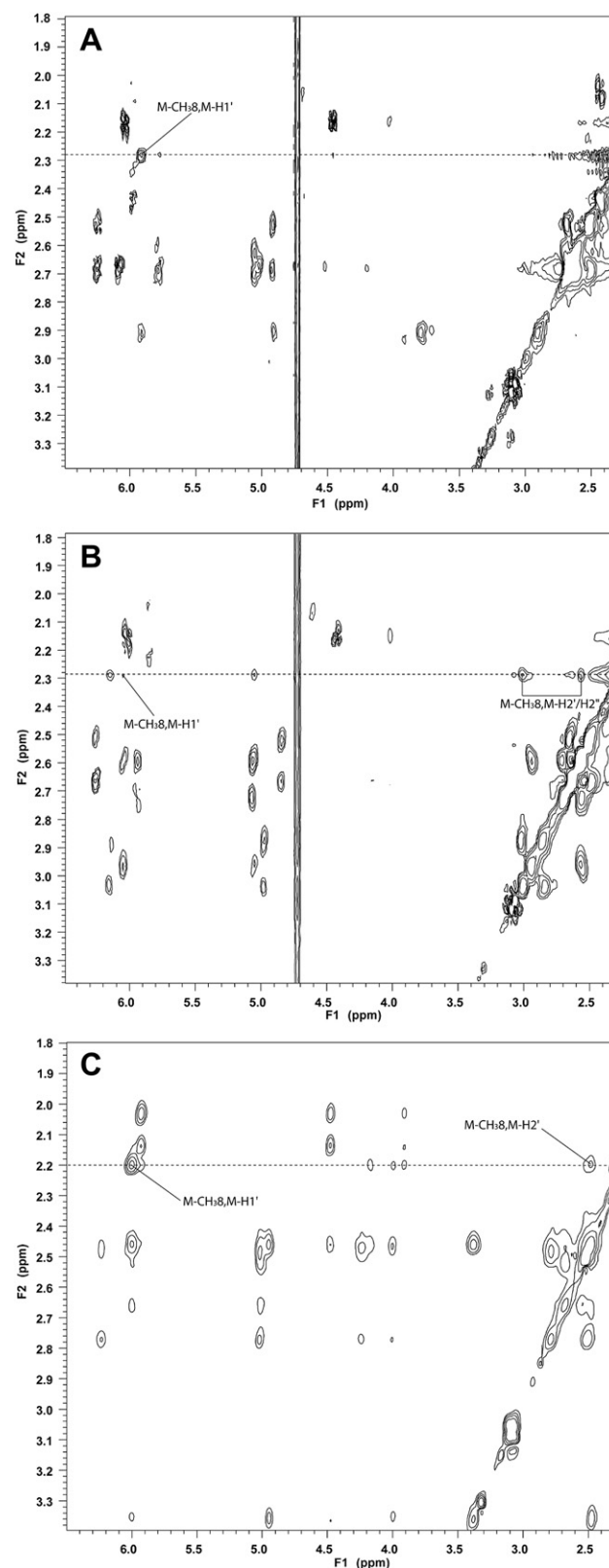


Fig. 6. Expanded NOESY spectra for quadruplex structures formed by **III** (A), **IV** (B) and **VI** (C) (500 MHz for **III** and **IV** and 700 MHz for **VI**, $T = 25^\circ\text{C}$; strands concentrations ~ 5 mM; solution: 10 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 70 mM KCl and 0.2 mM EDTA in D_2O , total volume = 0.6 ml; mixing time = 200 ms) correlating base M CH₃-8 protons (depicted by horizontal dashed lines) and sugar protons H1' and H2'/H2''.

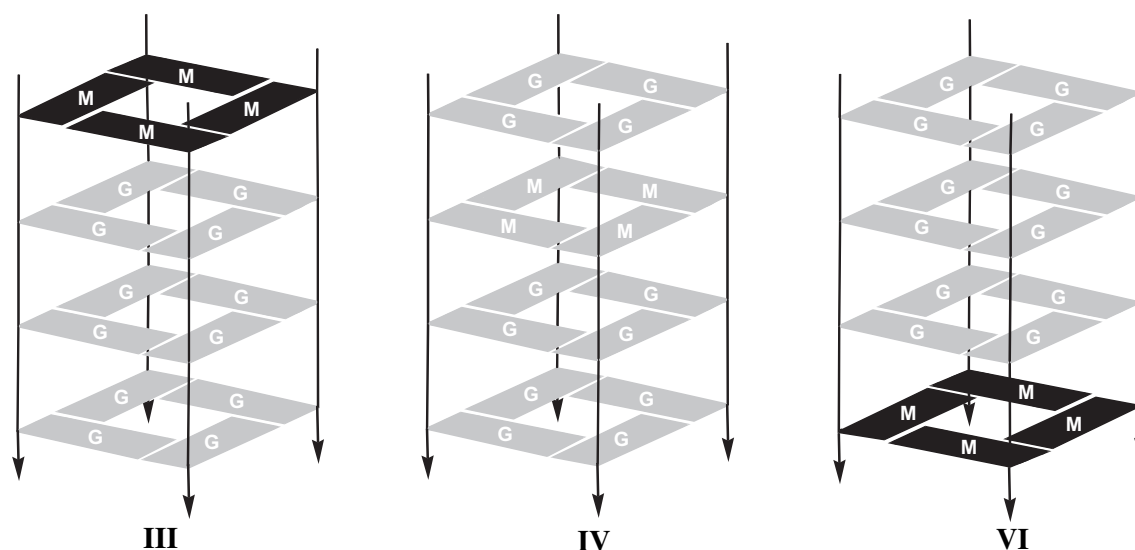


Fig. 7. Schematic representation of the quadruplexes formed by ODNs **III**, **IV** and **VI**. M = 8-methyl-2'-dG. *Anti* and *syn* residues are in grey and black, respectively. T residues have been omitted for clarity.

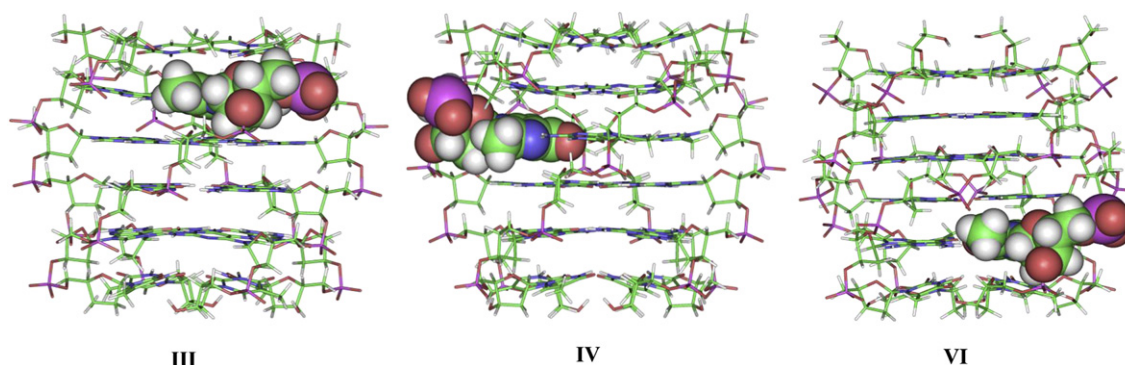


Fig. 8. Molecular models of the quadruplexes formed by ODNs **III**, **IV** and **VI**. The structures are oriented with the 5' end upward. Heavy atoms are shown with different colours (carbons, green; nitrogens, blue; oxygens, red; hydrogens, white). A 8-methyl-2'-dG residue per structure is reported in CPK.

structure is characterized by a lack of steric interactions. Unlike quadruplexes **III** and **IV**, in the model of quadruplex **VI** the M residues in the all-*syn* tetrad have slightly distorted methyl groups and the tetrad is not completely planar. This is consistent with the apparent melting temperatures in sodium buffer that clearly indicate that the quadruplex **VI** is less stable than the other quadruplex structures studied here.

4. Conclusions

In the present study, we analyzed the effects of 8-methyl-2'-deoxyguanine (M) on quadruplex structure, stability and formation kinetics. The influence of M on the molecular properties of the parallel quadruplex structures was sequence-dependent. Although the apparent melting temperatures cannot be considered an accurate evaluation of the thermal stability due to the slow kinetics of formation and dissociation of these structures, it provides insight into the behaviours of the modified ODNs (Table 1). The $T_{1/2}$ s for ODNs **I–VI** when compared with those of their natural counterparts, suggest that M significantly stabilizes the parallel quadruplex structures particularly when it is located at the 5'-end of the G-tract. This 5'-end effect was also observed when the association constants (k_{on}) were compared; association rate

increased as M approached the 5'-end (Table 1). This is consistent with results reported in studies of other modified parallel quadruplexes TRGGT ($R = 8$ -amino-2'-deoxyguanine or 8-bromo-2'-deoxyguanine) [3,4], although, in these cases, the improvement in k_{on} observed was higher than that we observed for quadruplex **III**.

NMR was used to confirm the ability of the 8-methyl group to induce the *syn* glycosidic conformation in the context of the parallel quadruplex. This effect was also sequence-dependent. In fact, the replacement of the first dG of the sequences TG₃T and TG₄T (**I** and **III**) by an M residue resulted in unusual all-*syn* tetrads. Unexpectedly, the replacement of the second dG with M (**II** and **IV**) resulted in quadruplex structures in which all purines adopted *anti* glycosidic conformations. It is noteworthy that these quadruplexes had similar thermal stabilities to the unmodified quadruplexes. As shown in a previous study [11], the substitution of the third dG by M in TG₃T, resulted in a relatively unstable quadruplex structure. This datum was confirmed to some extent in sequence TG₄T in which the replacement at the third dG (**V**) led to the formation of several types of quadruplex structures less stable than their unmodified counterpart. Finally, the introduction of M in the final position of the sequence TG₄T (**VI**) resulted in a quadruplex characterized by an all-*syn* G-tetrad with a $T_{1/2}$ significantly lower than those of quadruplex structures **III**, **IV** and **V**. Generally, NMR data

suggest that, in the modified quadruplexes investigated, the formation of an all-*syn* tetrad is particularly preferred when M is introduced at one end of the sequence.

The CD spectra of the modified ODNs showed several interesting features. The CD spectra of quadruplexes **II** and **IV** closely resemble those of their natural counterparts; this was not unexpected as, based on NMR data, the structures share the same features (parallel strand orientation and only all-*anti* G-tetrads). In contrast, spectra of quadruplexes **III** and **VI** show two positive bands around 260 and 295 nm. The band at 260 nm has been proposed to originate from the stacking involving two G-tetrads with the same hydrogen bonds polarities [31] or “head-to-tail” stacking [22]. The band at 295 nm may be due to stacking of two G-tetrads with opposite hydrogen bonds polarities [31] or “head-to-head” and “tail-to-tail” stacking [22]. Consistently with the NMR results, both these stacking types are present in **III** and **VI**.

NMR and kinetic studies on ODN **III**, in which an M residue replaces the first dG at the 5′-end, corroborated the hypothesis that a *syn* dG at the 5′-end of the strand is involved in the nucleation process that begins complex formation [3]. Furthermore, $T_{1/2}$ measurements strongly suggest that the 5′ *syn* dG influences quadruplex stability, probably due to favourable stacking between the all-*syn* tetrad and the adjacent all-*anti* tetrad, as suggested by molecular models.

The results described here supply further data regarding the relationship between the glycosidic conformation of the G bases and the thermal stability and the formation kinetics of quadruplexes. Furthermore, the ability of residues of 8-methyl-2′-deoxyguanosine to stabilize and accelerate formation of parallel quadruplexes, when present in specific positions, may be of interest in the area of anti-HIV aptamers such as the phosphorothioate $[d(T_2G_4T_2)]_4$ [32,33] and $[d(GAGGT)]_4$ analogues [34–36]. In fact, in order to improve the properties of an aptamer, 8-methyl-2′-deoxyguanosine could be introduced post-SELEX. Furthermore, the introduction of methyl groups into the grooves, which may establish hydrophobic interactions, may improve the affinity and/or the specificity towards the target molecule as introduction of the modified base at certain positions does not alter the quadruplex structure and preserves thermal stability of the unmodified complex.

Acknowledgements

This work was supported by Italian M.U.R.S.T. (P.R.I.N. 2007), the INSERM, the Fondation pour la Recherche Médicale (FRM), the University of Bordeaux 2, the agence Nationale de la recherche (ANR grants G4-Toolbox & QuantADN) and Région Aquitaine grants. We thank P. Alberti (MNHN, Paris) and L. Lacroix (Toulouse) for helpful discussions. The authors are grateful to “Centro di Servizio Interdipartimentale di Analisi Strumentale” (C.S.I.A.S.) for supplying NMR facilities. The authors are also grateful to Pasquale Paciello, Marilena Simonetti, Maria Fusco and Luisa Cuorvo for their collaboration.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.biochi.2010.10.011.

References

- [1] S. Neidle, S. Balasubramanian, Quadruplex Nucleic Acids. RSC Publishing, London, 2006.
- [2] S. Saha, J. Cai, D. Eiler, A.D. Hamilton, Programming the formation of DNA and PNA quadruplexes by π - π -stacking interactions, Chem. Commun. 46 (2010) 1685–1687 For a recent example see:.
- [3] J. Gros, F. Rosu, S. Amrane, A. De Cian, V. Gabelica, L. Lacroix, J.-L. Mergny, Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes, Nucleic Acids Res. 35 (2007) 3064–3075.
- [4] J. Gros, A. Avino, J. Lopez de la Osa, C. Gonzalez, L. Lacroix, A. Perez, M. Orozco, R. Eritja, J.-L. Mergny, 8-Amino guanine accelerates tetramolecular G-quadruplex formation, Chem. Commun. (2008) 2926–2928.
- [5] H. Sugiyama, K. Kawai, A. Matsunaga, K. Fujimoto, I. Saito, H. Robinson, A.H.-J. Wang, Synthesis, structure and thermodynamic properties of 8-methylguanine-containing oligonucleotides: Z-DNA under physiological salt conditions, Nucleic Acids Res. 24 (1996) 1272–1278.
- [6] I. Cherrak, O. Mauffret, F. Santamaria, A. Hocquet, M. Ghomi, B. Rayner, S. Fermandjian, L-nucleotides and 8-methylguanine of $d(C^{18}G_2C_3G_4C_5G_6C_7G_8C_9G_{10})_2$ act cooperatively to promote a left-handed helix under physiological salt conditions, Nucleic Acids Res. 23 (2003) 6986–6995.
- [7] Y. Xu, R. Ikeda, H. Sugiyama, 8-Methylguanosine: a powerful Z-DNA stabilizer, J. Am. Chem. Soc. 125 (2003) 13519–13524.
- [8] G.-X. He, S.H. Krawczyk, S. Swaminathan, R.G. Shea, J.P. Dougherty, T. Terhorst, V.S. Law, L.C. Griffin, S. Coutre, N. Bischofberger, N2- and C8-Substituted oligodeoxynucleotides with enhanced thrombin inhibitory activity in vitro and in vivo, J. Med. Chem. 41 (1998) 2234–2242.
- [9] S.H. Krawczyk, N. Bischofberger, L.C. Griffin, V.S. Law, R.G. Shea, S. Swaminathan, Structure-activity study of oligodeoxynucleotides which inhibit thrombin, Nucleosides Nucleotides 14 (1995) 1109–1116.
- [10] Y. Xu, H. Sugiyama, Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb), Nucleic Acids Res. 34 (2006) 949–954.
- [11] A. Virgilio, V. Esposito, A. Randazzo, L. Mayol, A. Galeone, 8-Methyl-2′-deoxyguanosine incorporation into parallel DNA quadruplex structures, Nucleic Acids Res. 33 (2005) 6188–6195.
- [12] K. Kohda, H. Tsunomoto, Y. Minoura, K. Tanabe, S. Shibutani, Synthesis, mis-coding specificity, and thermodynamic stability of oligodeoxynucleotide containing 8-methyl-2′-deoxyguanosine, Chem. Res. Toxicol. 9 (1996) 1278–1284.
- [13] J.-L. Mergny, J. Li, L. Lacroix, S. Amrane, J.B. Chaires, Thermal difference spectra: a specific signature for nucleic acid structures, Nucleic Acids Res. 33 (2005) e138/1–e138/6.
- [14] J.-L. Mergny, A. De Cian, A. Ghelab, B. Saccà, L. Lacroix, Kinetics of tetramolecular quadruplex, Nucleic Acids Res. 33 (2005) 81–94.
- [15] M. Piotto, V. Saudek, V.J. Sklenar, Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, J. Biomol. NMR 2 (1992) 661–665.
- [16] J. Jeener, B. Meier, H.P. Bachmann, R.R. Ernst, Investigation of exchange processes by two-dimensional NMR spectroscopy, J. Chem. Phys. 71 (1979) 4546–4553.
- [17] L. Braunschweiler, R.R. Ernst, Coherence transfer by isotropic mixing: application to proton correlation spectroscopy, J. Magn. Reson. 53 (1983) 521–528.
- [18] D. Marion, K. Wuthrich, Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of proton-proton spin-spin coupling constants in proteins, Biochem. Biophys. Res. Commun. 113 (1983) 967–974.
- [19] W.D. Cornell, P. Cieplack, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
- [20] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P.J. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, J. Am. Chem. Soc. 106 (1984) 765–784.
- [21] J.-L. Mergny, A.T. Phan, L. Lacroix, Following G-quartet formation by UV-spectroscopy, FEBS Lett. 435 (1998) 74–78.
- [22] S. Masiero, R. Trotta, S. Pieraccini, S. De Tito, R. Perone, A. Randazzo, G.P. Spada, A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures, Org. Biomol. Chem. 8 (2010) 2683–2692.
- [23] J. Feigon, K.M. Koshlap, F.W. Smith, ^1H NMR spectroscopy of DNA triplexes and quadruplexes, Meth. Enzymol. 261 (1995) 225–255.
- [24] K. Wüthrich, NMR of Proteins and Nucleic Acids. Wiley, New York, 1986.
- [25] Y. Wang, D.J. Patel, Guanine residues in $d(TG_4T)$ and $d(T_2G_4)$ form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution, Biochemistry 31 (1992) 8112–8119.
- [26] Y. Wang, D.J. Patel, Solution structure of a parallel-stranded G-quadruplex DNA, J. Mol. Biol. 234 (1993) 1171–1183.
- [27] F. Aboul-ela, A.I.H. Murchie, D.G. Norman, D.M.J. Lilley, Solution structure of a parallel-stranded tetraplex formed by $d(TG_4T)$ in the presence of sodium ions by nuclear magnetic resonance spectroscopy, J. Mol. Biol. 234 (1994) 458–471.
- [28] P.K. Patel, R.V. Hosur, NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats, Nucleic Acids Res. 27 (1999) 2457–2464.
- [29] Y. Wang, D.J. Patel, Solution structure of the human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex, Structure 1 (1993) 263–282.
- [30] F.W. Smith, J. Feigon, Strand orientation in the DNA quadruplex formed from the *Oxytricha* telomere repeat oligonucleotide $d(G_4T_4G_4)$ in solution, Biochemistry 32 (1993) 8683–8692.
- [31] D.M. Gray, J.-D. Wen, C.W. Gray, R. Repges, C. Repges, G. Raabe, J. Fleischhauer, Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities, Chirality 20 (2008) 431–440.

- [32] J.R. Wyatt, T.A. Vickers, J.L. Roberson, R.W. Buckheit Jr., T. Klimkait, E. DeBaets, P.W. Davis, B. Rayner, J.L. Imbach, D.J. Ecker, Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion, *Proc. Nat. Acad. Sci. USA* 91 (1994) 1356–1360.
- [33] D.J. Ecker, J.R. Wyatt, T.A. Vickers, R. Buckheit, J. Roberson, J.-L. Imbach, Novel guanosine quartet structure binds to the HIV envelope and inhibits envelope mediated cell fusion, *Nucleosides Nucleotides* 14 (1995) 1117–1127.
- [34] M. Koizumi, K. Akahori, T. Ohmine, S. Tsutsumi, J. Sone, T. Kosaka, M. Kaneko, S. Kimura, K. Shimada, Biologically active oligodeoxyribonucleotides. Part 12: N²-Methylation of 2'-deoxyguanosines enhances stability of parallel G-quadruplex and anti-HIV-1 activity, *Bioorg. Med. Chem. Lett.* 10 (2000) 2213–2216 and references cited therein.
- [35] J. D'Onofrio, L. Petraccone, E. Erra, L. Martino, G. Di Fabio, L. De Napoli, C. Giancola, D. Montesarchio, 5'-Modified G-quadruplex forming oligonucleotides endowed with anti-HIV activity: synthesis and biophysical properties, *Bioconjug. Chem.* 18 (2007) 1194–1204.
- [36] J. D'Onofrio, L. Petraccone, L. Martino, G. Di Fabio, A. Iadonisi, J. Balzarini, C. Giancola, D. Montesarchio, Synthesis, biophysical characterization, and anti-hiv activity of glyco-conjugated G-quadruplex-forming oligonucleotides, *Bioconjug. Chem.* 19 (2008) 607–616.

SUPPLEMENTARY MATERIAL

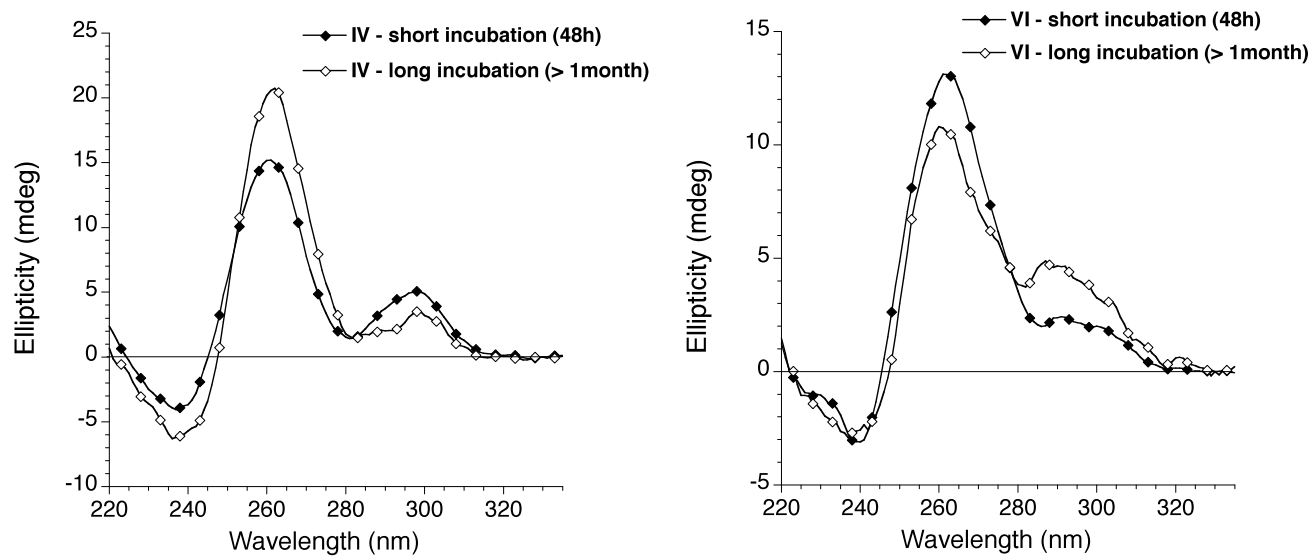


Figure S1. CD spectra of ODNs **IV** (left) and **VI** (right) (4°C in 100 mM K⁺) obtained 48 h and more than 1 month after annealing.

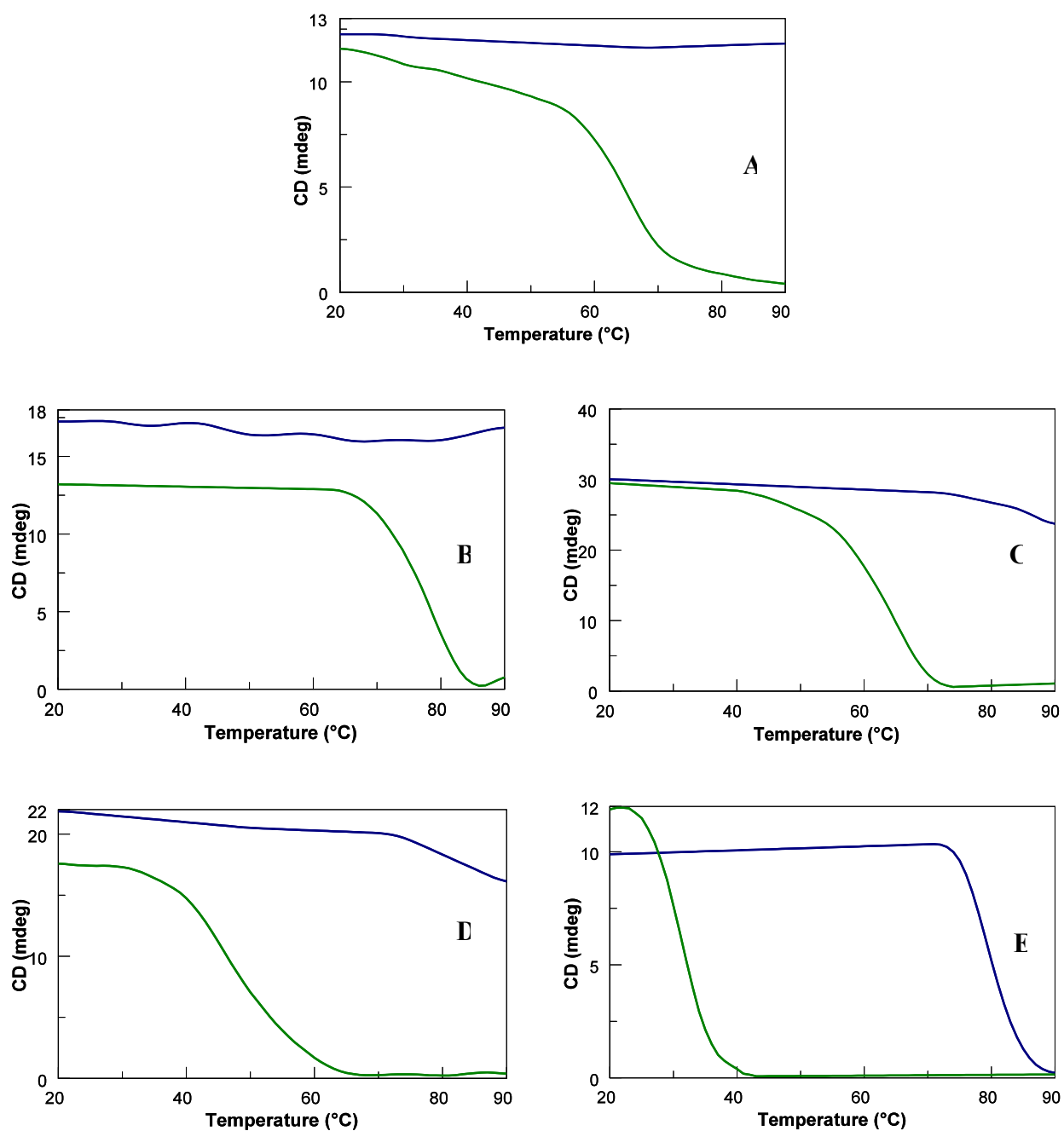


Fig. S2. CD melting of [d(TGGGGT)]₄ (A), **III** (B), **IV** (C), **V** (D), **VI** (E) in 70 mM K⁺ (blue) and 70 mM Na⁺ (green) buffer solution (wavelength = λ_{max} ; sample concentration: 1×10^{-4} M; scan rate: 10°C h^{-1}).

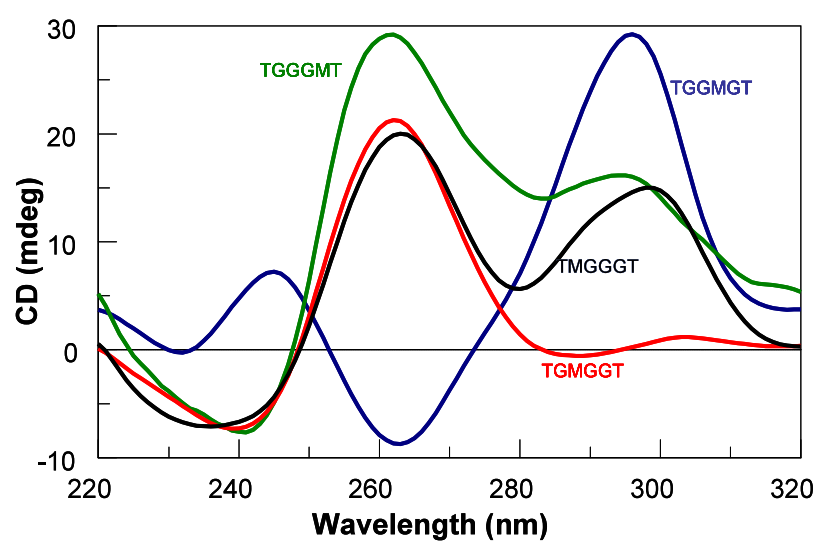


Fig. S3. CD spectra of ODNs **III** (black), **IV** (red), **V** (blue), **VI** (green) at 20° C, obtained from the NMR samples (70 mM KCl, 100 μ M oligo).

2.2 Article 2 : “A mirror-image tetramolecular DNA quadruplex”. Chemical Communications (Camb) - 2011; 47(19): 5437-9

Tran, P.L.T., Moriyama, R., Maruyama, A., Rayner, B. and Mergny, J.L.

Résumé en français :

L'ADN de série L-, l'image miroir de l'ADN naturel forme des structures de chiralité inversée. Nous démontrons ici qu'une courte séquence L-ADN riche en guanines forme un G4 tétramoléculaire avec les mêmes propriétés que son homologue D-ADN, à l'exception de son spectre de dichroïsme circulaire, qui est inversé. Les brins L- et D- s'auto-excluent quand ils sont mélangés ensemble, montrant que l'auto-assemblage des différents brins en parallèle est contrôlé.

Article 2

Cite this: *Chem. Commun.*, 2011, **47**, 5437–5439

www.rsc.org/chemcomm

COMMUNICATION

A mirror-image tetramolecular DNA quadruplex†

Phong Lan Thao Tran,^{ab} Rui Moriyama,^c Atsushi Maruyama,^c Bernard Rayner^{ab} and Jean-Louis Mergny^{*ab}

Received 4th March 2011, Accepted 25th March 2011

DOI: 10.1039/c1cc11293g

L-DNA, the mirror image of natural DNA forms structures of opposite chirality. We demonstrate here that a short guanine rich L-DNA strand forms a tetramolecular quadruplex with the same properties as a D-DNA strand of identical sequence, besides an inverted circular dichroism spectra. L- and D-strands self exclude when mixed together, showing that the controlled parallel self-assembly of different G-rich strands can be obtained through L-DNA use.

L-Nucleic acids, the mirror images of natural D-DNA and RNA, have found a number of interesting applications in biotechnology (as aptamers,^{1,2} PCR³ or microarray⁴ probes, agents for enantiomeric separation⁵ and molecular beacons)⁶ and nanotechnology.⁷ One of the advantages of L-nucleic acids is their nuclease resistance, allowing one *spiegelmer* (from “*Spiegel*” meaning mirror in German) to enter clinical trials.⁸

As a perfect mirror image of D-DNA, L-DNA forms duplexes with identical physical characteristics (solubility, stability) except for chirality, leading to left-handed double-helices.⁹ We investigate here their potential to form G-quadruplexes,¹⁰ using the well known tetramolecular [TG₄T]₄ complex as a model.¹¹ 2'-Deoxy-L-guanosine and L-thymidine were prepared according to slightly modified procedures.^{12,13} N-2-Isobutyryl-2'-deoxy-L-guanosine and L-thymidine were dimethoxytritylated and further converted to the corresponding β-cyanoethylphosphoramidites according to standard procedures. L-Oligonucleotides were assembled on a DNA synthesizer, deprotected and purified by reverse-phase HPLC (see full details in ESI†).

We expected that a short L-TG₄T strand should form a stable tetramolecular quadruplex with identical properties as its well-known regular DNA counterpart but opposite CD spectra. Fig. 1 summarizes the evidence for G4 formation obtained for L-TG₄T. L- and D-strands form complexes of identical retarded mobility when incubated under conditions favoring G4 formation. Isothermal difference spectra (IDS) were obtained by calculating the difference between the absorbance spectra of the folded and unfolded forms of a sample

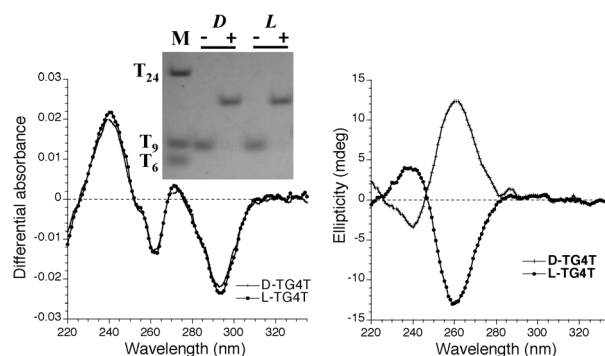


Fig. 1 Evidence for G4 formation by L- and D-TG₄T. Left: IDS spectra of the D- and L-strands in 0.5 M KCl. Gel: migration of D- and L-strands in 20% polyacrylamide under conditions favorable (+) or not (–) to G4 formation (60 μM strand concentration). “M” lanes correspond to oligo dT_n migration markers. Right: CD spectra of the L- and D- quadruplexes in 0.5 M KCl.

(after and before an isothermal kinetics experiment);¹⁴ they are nearly superimposable (left) while circular dichroism (CD) spectra are inverted (right), as expected for a left-handed L-DNA quadruplex.

We then compared the thermal stability and kinetics of the two quadruplexes. First, the apparent melting temperatures¹¹ of the preformed L- and D-DNA quadruplexes are very close (within experimental error; Fig. 2, left, and Table 1). As previously observed, thermal denaturation is quasi irreversible for both samples in the micromolar strand concentration range—no reassociation is observed upon cooling.¹¹ Second, the kinetics of association are identical, within experimental error. These experiments were performed by starting from isolated strands and comparing the association of L- and D-TG₄T and the L + D mixture by UV-isothermal experiments as previously described.^{11,15} One observes a time-dependent increase in absorbance at 295 nm that reflects the association of the different strands. These profiles may be fitted as previously described¹¹ and lead to similar *k*_{on} values (Table 1).

L-DNA cannot form stable duplexes with D-DNA.^{12,16} We therefore decided to investigate if the two enantiomers are compatible within an intermolecular quadruplex or if chiral selection occurs.¹⁷ Three independent experiments suggest that L/D hybrid structures are at most marginal:

(i) Kinetics analysis of G4 formation with D + L mixtures supports chiral selection: when equimolar concentrations

^a Université de Bordeaux, Laboratoire ARNA, F-33000 Bordeaux, France. E-mail: jean-louis.mergny@inserm.fr

^b INSERM, U869, Laboratoire ARNA, IECB, F-33600 Pessac, France

^c Institute for Materials Chemistry and Engineering, Kyushu University, Fukuoka, Japan

† Electronic supplementary information (ESI) available: Experimental procedures, CCC formulae, representative kinetic and gel experiments. See DOI: 10.1039/c1cc11293g

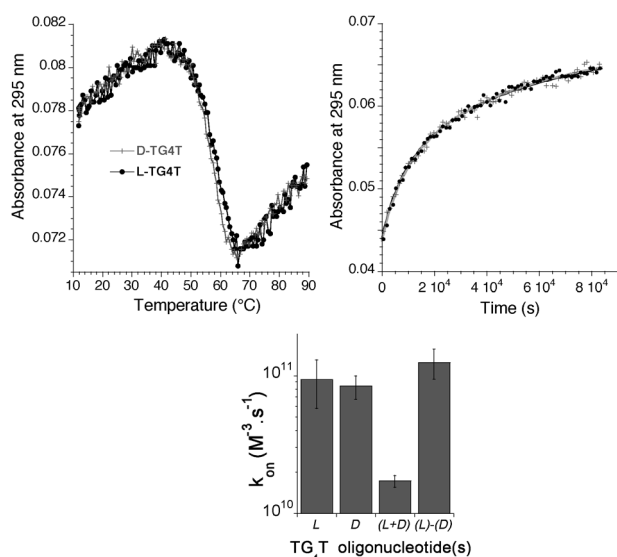


Fig. 2 Stability and kinetics of G4 formed by L- and D-TG₄T. Left: UV melting profiles at 295 nm in 0.1 M NaCl (0.3 °C min⁻¹; note that the apparent *T*_m depends on the scan rate as dissociation is nearly irreversible at this strand concentration). Right: isothermal association at 6 °C in 0.5 M KCl. Bottom: *k*_{on} values calculated for the L- and D-oligonucleotides as well as the L + D mixture using two different models (assuming the concentrations are additive (“L + D” case) or not (“(L) – (D)”)).

Table 1 Comparison of L- and D-TG₄T. Apparent melting temperatures are determined at 295 nm in 0.1 M or 0.5 M NaCl (no melting occurs in KCl) with a temperature gradient of 0.3 and 1 °C min⁻¹, respectively. *k*_{on} values were obtained in 0.5 M KCl from isothermal association experiments. Apparent thermal temperatures are determined with 1–2 °C accuracy

Oligonucleotide	<i>T</i> _{1/2} /°C 0.1 M Na ⁺	<i>T</i> _{1/2} /°C 0.5 M Na ⁺	<i>k</i> _{on} /M ⁻³ s ⁻¹
L-	58	67.5	(9.4 ± 3.6) × 10 ¹⁰
D-	57	67	(8.4 ± 1.6) × 10 ¹⁰

of D and L strands are present, the association proceeds with kinetics that suggest that concentrations are not additive. As the rate of this reaction is strongly dependent on strand concentration^{11,18} (with an experimentally determined experimental order of 3 to 4) this suggests that both strands associate independently from each other (Fig. 2, bottom).

(ii) When mixing two strands of different lengths, such as TG₄T and T₄G₄T, one expects to see 5 bands for a tetramolecular complex, corresponding to the different possibilities (four short strands, three short, one long, *etc.*)¹⁹ (Fig. 3, left). This is indeed the case when mixing two D-strands, while a very different situation (compare lanes 4 and 5) is observed for L-TG₄T + D-T₄G₄T, in which two dominant bands are obtained with migrations corresponding to the “pure” species [L-TG₄T]₄ and [D-T₄G₄T]₄.

(iii) Finally, using a fluorescent L-TG₄T strand, we demonstrate that a cationic comb-type copolymer (CCC) acting as a nucleic acid chaperone^{20,21} promotes its incorporation by strand exchange into a preformed [L-TG₄T]₄ quadruplex, but not into [D-TG₄T]₄ (Fig. 3, right and Fig. 4). A reciprocal experiment with a fluorescent D-TG₄T-fluorescein strand gives

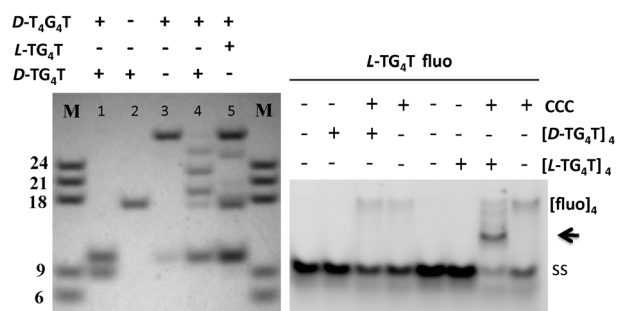


Fig. 3 Enantiomeric exclusion in quadruplex assembly. Left: self exclusion of L- and D-strands into quadruplexes. “M” lanes correspond to oligo dT_n migration markers (length in nucleotides provided on the left). The gel is revealed by UV shadow. Lane 1: conditions unfavourable for G4 formation, only single strands are seen. Lanes 2–5: preincubation in KCl favoring G4 formation (90 μM for L- and D-TG₄T; 180 μM for D-T₄G₄T). Right: using a CCC (a cationic comb-type copolymer),²² we demonstrate that a fluorescent L-DNA strand (0.5 μM strand concentration) may be “inserted” into a preformed tetramolecular L-quadruplex (arrow) but not D-quadruplex (both at 20 μM strand concentration). The gel is revealed by fluorescence (*i.e.*, only L-TG₄T fluo is seen, either single-stranded (ss) or incorporated into G4).

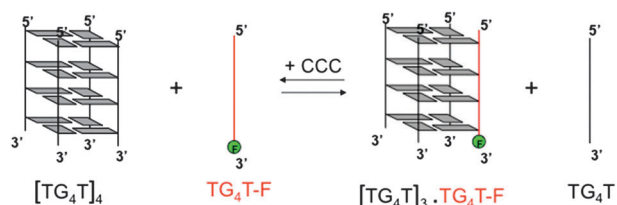


Fig. 4 Principle of the strand exchange experiment.^{21,22} In the presence of CCC, one can incorporate the fluorescent strand into a preformed tetramolecular quadruplex. This may be evidenced by electrophoresis. One may compare the strand exchange efficiency with L- or D-fluorescent strands, to be incorporated in a L- or D-preformed quadruplex.

a “mirror” result (incorporation into a preformed D-quadruplex, but not L-quadruplex; not shown).

Altogether, our results demonstrate that one can obtain left-handed L-DNA tetramolecular quadruplexes and that chiral selection may be achieved. Several applications may be proposed for L-DNA quadruplexes:

(i) As a number of aptamers adopt quadruplex folds,²³ it should be interesting to develop G4-based spiegelmers, providing extra resistance in a biological environment.

(ii) L- vs. D-Quadruplexes may be tested for quadruplex ligand affinity. A number of G4 ligands such as telomestatin possess chiral centers, potentially allowing chiral discrimination,²⁴ as found for duplex binders:²⁵ comparing their affinities for both quadruplexes might give clues on the binding mode, as little discrimination should be expected for terminal stacking while groove binding and interactions with the loops may differ.

(iii) Chiral selection is an advantage in nanotechnology as L- and D- quadruplexes may be assembled simultaneously with little or no “crosstalk”. This is in contrast with unnatural nucleic acids such as LNA,²⁶ PNA²⁷ and 2′O methyl RNAs¹¹ which have been reported to form G-quadruplexes but may form hybrid structures such as 2PNA–2DNA.

Funding was provided by INSERM, Université de Bordeaux, CNRS-PIR, *Fondation pour la Recherche Médicale*, *Région Aquitaine*, Japan Science and Technology Agency (JST) and ANR-09-Blanc-0355. J.L.M. would also like to thank N. Sharp and W. Bell (Massive Dynamic, Boston, MA, USA) for alternative funding.

Notes and references

- 1 S. Leva, A. Lichte, J. Burmeister, P. Muhn, B. Jahnke, D. Fesser, J. Erfurth, P. Burgstaller and S. Klussmann, *Chem. Biol.*, 2002, **9**, 351–359.
- 2 B. Wlotzka, S. Leva, B. Eschgfäller, J. Burmeister, F. Kleinjung, C. Kaduk, P. Muhn, H. Hess-Stumpp and S. Klussmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 8898–8902.
- 3 G. Hayashi, M. Hagihara, A. Kobori and K. Nakatani, *ChemBioChem*, 2007, **8**, 169–171.
- 4 N. C. Hauser, R. Martinez, A. Jacob, S. Rupp, J. D. Hoheisel and S. Matysiak, *Nucleic Acids Res.*, 2006, **34**, 5101–5111.
- 5 C. André, A. Berthelot, M. Thomassin and Y. C. Guillaume, *Electrophoresis*, 2006, **27**, 3254–3262.
- 6 Y. Kim, C. J. Yang and W. Tan, *Nucleic Acids Res.*, 2007, **35**, 7279–7287.
- 7 C. Lin, Y. Ke, Z. Li, J. H. Wang, Y. Liu and H. Yan, *Nano Lett.*, 2009, **9**, 433–436.
- 8 S. G. Sayyed, H. Hagele, O. P. Kulkarni, K. Endlich, S. Segerer, D. Eulberg, S. Klussmann and H. J. Anders, *Diabetologia*, 2009, **52**, 2445–2454.
- 9 H. Urata, E. Shinohara, K. Ogura, Y. Ueda and M. Akagi, *J. Am. Chem. Soc.*, 1991, **113**, 8174–8175.
- 10 S. Neidle and S. Balasubramanian, *Quadruplex Nucleic Acids*, RSC Biomolecular Sciences, Cambridge, 2006.
- 11 J. L. Mergny, A. De Cian, A. Ghelab, B. Saccà and L. Lacroix, *Nucleic Acids Res.*, 2005, **33**, 81–94.
- 12 A. Garbesi, M. L. Capobianco, F. P. Colonna, L. Tondelli, F. Arcamone, G. Manzini, C. W. Hilbers, J. M. E. Aelen and M. J. J. Blommers, *Nucleic Acids Res.*, 1993, **21**, 4159–4165 (correction published p. 5286).
- 13 H. Urata, T. Kumashiro, T. Otake, T. Kawahata and M. Akagi, *Nucleic Acids Res. Suppl.*, 2002, **2**, 163–164.
- 14 J. L. Mergny, J. Li, L. Lacroix, S. Amrane and J. B. Chaires, *Nucleic Acids Res.*, 2005, **33**, e138.
- 15 J. Gros, A. Avino, J. Lopez de la Osa, C. Gonzalez, L. Lacroix, A. Perez, M. Orozco, R. Eritja and J. L. Mergny, *Chem. Commun.*, 2008, 2926–2928.
- 16 S. Fujimori, K. Shudo and Y. Hashimoto, *J. Am. Chem. Soc.*, 1990, **112**, 7436–7438.
- 17 M. Froeyen, F. Morvan, J. J. Vasseur, P. Nielsen, A. Van Aerschot, H. Rosemeyer and P. Herdewijn, *Chem. Biodiversity*, 2007, **4**, 803–817.
- 18 C. Bardin and J. L. Leroy, *Nucleic Acids Res.*, 2007, **36**, 477–488.
- 19 D. Sen and W. Gilbert, *Nature*, 1988, **334**, 364–366.
- 20 R. Moriyama, N. Shimada, A. Kano and A. Maruyama, *Nucleic Acids Symp. Ser.*, 2008, **52**, 227–228.
- 21 R. Moriyama, N. Shimada, A. Kano and A. Maruyama, *Nucleic Acids Symp. Ser.*, 2009, **53**, 61–62.
- 22 R. Moriyama, N. Shimada, A. Kano and A. Maruyama, *Biomaterials*, 2011, **32**, 2351–2358.
- 23 B. Gatto, M. Palumbo and C. Sissi, *Curr. Med. Chem.*, 2009, **16**, 1248–1265.
- 24 T. Doi, K. Shibata, M. Yoshida, M. Takagi, M. Tera, K. Nagasawa, K. Shin-ya and T. Takahashi, *Org. Biomol. Chem.*, 2011, **9**, 387–393.
- 25 C. Dose, D. Ho, H. E. Gaub, P. B. Dervan and C. H. Albrecht, *Angew. Chem., Int. Ed.*, 2007, **46**, 8384–8387.
- 26 A. Randazzo, V. Esposito, O. Ohlenschläger, R. Ramachandran and L. Mayol, *Nucleic Acids Res.*, 2004, **32**, 3083–3092.
- 27 B. Datta, C. Schmitt and B. A. Armitage, *J. Am. Chem. Soc.*, 2003, **125**, 4111–4118.

A mirror-image tetramolecular DNA quadruplex

Phong Lan Thao Tran,^{a,b} Rui Moriyama,^c Atsushi Maruyama,^c Bernard Rayner,^{a,b} and Jean-Louis Mergny^{a,b}

Supporting Information

Experimental section

Oligonucleotide synthesis and purification

Unmodified oligonucleotides *D*-TG₄T, *D*-T₄G₄T and *D*-TG₄T with fluorescein at the 3' end were purchased from Eurogentec (Seraing, Belgium). Their concentrations were estimated using molar extinction coefficients provided by the manufacturer (57800 M⁻¹ cm⁻¹, 82100 M⁻¹ cm⁻¹ and 78800 M⁻¹ cm⁻¹ at 260 nm, respectively)¹.

2'-Deoxy-*L*-guanosine and *L*-thymidine were prepared according to slightly modified procedures². N-2-isobutyryl-2'-deoxy-*L*-guanosine and *L*-thymidine were dimethoxytritylated and further converted to the corresponding β -cyanoethylphosphoramidites according to standard procedures. *L*-oligonucleotides were synthesized on a Model 381A (Applied Biosystems Inc.) DNA synthesizer (1 μ mol scale) using β -*L*-Thymidine 3'-lcaa CPG (ChemGenes) solid support for *L*-TG₄T and 3'-(6-FAM) CPG (Link Technologies) solid support for *L*-TG₄T fluo. Oligonucleotides were cleaved from the support and deprotected by treatment with 30% aqueous ammonia for 5 hr at 55°C. Combined filtrate and washings were concentrated under reduced pressure, redissolved in water and the crude oligonucleotides were purified by reverse-phase HPLC on a Nucleodur 100-5 C18 ec column (Macherey Nagel) using a linear gradient of acetonitrile (5 to 22 % in 20 min) in 0.1 M triethylammonium acetate pH 7 at 1 ml/min flow rate. Fractions were collected and evaporated under reduced pressure. ms analyses were carried out on a MALDI-Tof-Tof mass spectrometer (Ultraflex, Brucker) operated in the reflectron mode. The matrix used was a 4:1 mixture of 2,4,6-trihydroxyacetophenone (10 mg/mL) in ethanol and 100 mM aqueous ammonium citrate. Samples (0.8 μ l of a 1:1 aqueous solution of oligonucleotide and matrix) were spotted on a stainless steel target and air-dried before analysis. For each sample, a MALDI mass spectrum was acquired by accumulating the ion signals from 100 UV laser shots with a constant laser power. *L*-TG₄T [M-H]⁻: calc. 1862.3, found 1861.2; 3'-(6-FAM)-*L*-TG₄T [M-H]⁻: calc. 2431.3, found 2430.6.

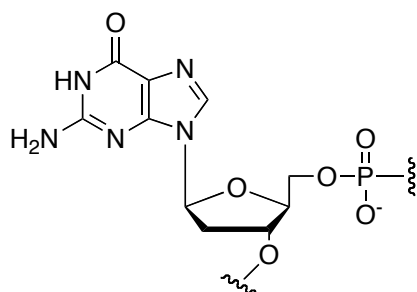


Figure S1. Structure of 2'-deoxy-*L*-guanosine -5'-phosphate as a constituent of *L*-DNA.

Cationic comb-type Copolymer (CCC).

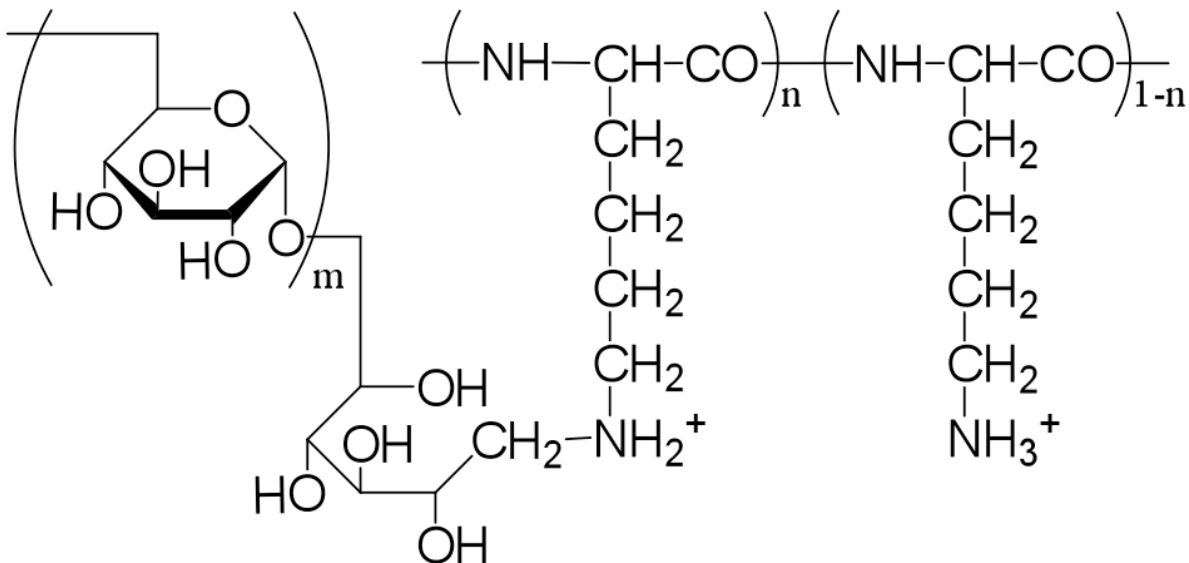


Figure S2. Structural formula of Cationic Comb-type Copolymer (CCC), poly(L-lysine)-*graft*-dextran³. MW of PLL backbone: 5.3×10^4 , MW of dextran grafts: 8.4×10^3 , grafting degree of Dex: 16 mol % ($n = 0.16$).

Absorbance differential spectra (TDS and IDS) and circular dichroism (CD)

Our reference conditions for this study were 10mM Lithium cacodylate pH 7.2 supplemented with 0.5M KCl. Oligonucleotides were prepared at $8 \mu\text{M}$ strand concentration. Thermal difference spectra (TDS) were obtained by calculating the difference between the absorbance spectra recorded above and below the observed transition (data not shown). Isothermal difference spectra (IDS) were obtained by calculating the difference between the absorbance spectra of the folded and unfolded forms of a sample (after and before an isothermal kinetics experiment) (**Figure 1 left**). CD spectra were recorded on a JASCO-810 spectropolarimeter using a 1-cm path length quartz cuvettes as previously described⁴.

UV melting experiments

Melting experiments were conducted as described earlier⁵. Denaturation was followed by recording the absorbance at 240, 260, 273 and 295 nm. Only apparent melting temperatures determined at 295 nm are shown in **Figure 2** and **Table 1**. The experiments were performed in 10 mM Lithium cacodylate pH 7.2 supplemented with 0.1 or 0.5 M NaCl and with $10 \mu\text{M}$ or $18 \mu\text{M}$ strand concentration respectively. For the melting experiments in 0.1M NaCl, samples were pre-folded at 4°C for 48 hours and were heated from 10°C to 96°C , at a rate of $0.3^\circ\text{C}/\text{min}$. For melting experiments in 0.5M NaCl, samples were also pre-folded at 4°C for 48 hours and were heated from 10°C to 96°C , at a rate of $1^\circ\text{C}/\text{min}$.

Isothermal kinetics

These experiments were performed by starting from isolated strands and comparing association of *L*- and *D*- TG4T and *L+D* mixture by UV-isothermal experiments at 6°C in 10 mM Lithium cacodylate and 0.5 M KCl conditions as previously described^{6,7}. Data was fitted using a model

previously published⁶ and association rate constants (k_{on}) were calculated (*Table 1*). The corresponding order of the association reaction was assumed to be $n = 4$, as in previous studies. The sequences were tested at 8 μ M strand concentration. Examples of isothermal kinetics are provided in the figure below:

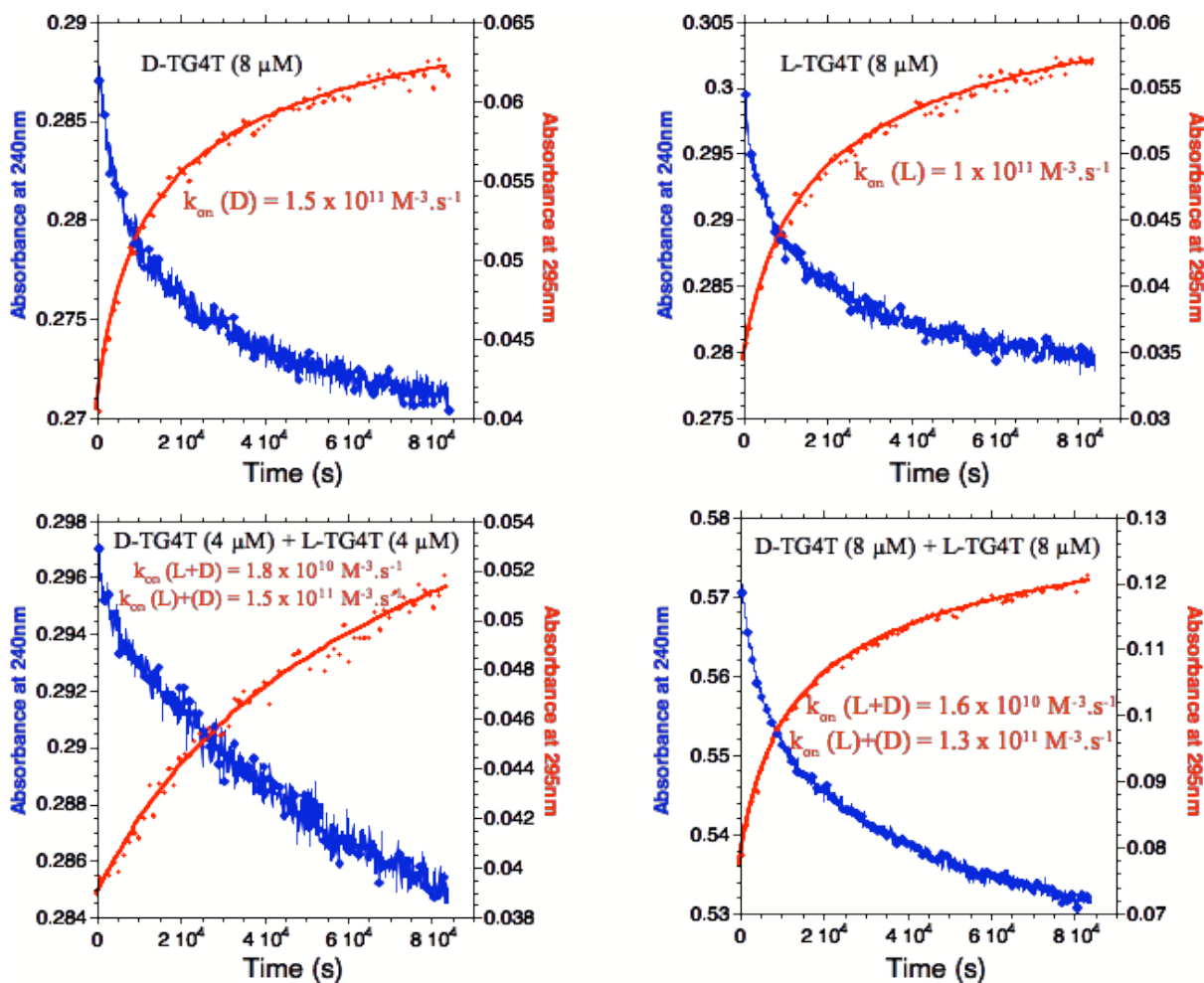


Figure S3. Example of isothermal experiments recorded at 240 nm (blue) and 295 nm (red). Top: left: D-TG₄T (8 μ M) right: L-TG₄T (8 μ M).

Bottom: Both L- and D- strands are present, each at 4 μ M (left: total concentration 8 μ M) or 8 μ M (right: total concentration 16 μ M). These samples indicate that the reaction proceeds at a much slower rate when having 4 μ M of each strand, instead of 8 μ M of one. In order to observe comparable rates, one must raise the concentration of each strand to 8 μ M, as for the experiments with L- or D- alone.

The curves at 295 nm (red dots) were fitted using Kaleidagraph software with a mathematical function (red curve) leaving k_{on} , starting absorbance and final absorbance as floating parameters. The values provided in the main table are average $\pm$ SD of 3-5 experiments. Two fits are performed: one assumes that concentrations are additive while the other treat each enantiomer independently.

Non-denaturing gel electrophoresis

Non-denaturing gel electrophoresis allows separation of single-stranded oligonucleotides from tetramolecular G-quadruplex structures. Samples were loaded on a 20% polyacrylamide (acrylamide/bis-acrylamide 19:1) gel containing TBE 1X and KCl at 10 mM. Electrophoresis was performed at 4 W/gel to reach a temperature close to 35°C (electrophoresis performed at room temperature). To achieve complete quadruplex formation, samples (indicated by +) were incubated during 48 hours, at 4°C and at high strand concentration (200 μ M), in 0.1M KCl and 10 mM LiCaco (pH 7.2). In parallel, samples (indicated by -) were diluted in water and heated 2 minutes at 90°C just before loading. Lithium cacodylate buffer was prepared by mixing cacodylic acid with LiOH. 12% glycerol was added just before loading. Bands were revealed by UV-shadow (60 μ M strand concentration) using a UV light source (254 nm) and an ETNA-MS-ChemiBis digital camera (FSVT). This method does not require any labeling of any kind and relies solely on the absorbance of the nucleic acid in the far UV region (around 260 nm). We compared migrations of *L*- and *D*- sequences with or without cations.

L- or *D*-TG₄T and *D*-T₄G₄T were mixed in the same buffer (10mM LiCaco and 0.1M KCl) and incubated during 5 days at 4°C. (in **Figure 3 left**, lane 4: 90 μ M *D*-T₄G₄T + 60 μ M *D*-TG₄T and lane 5: 180 μ M *D*-T₄G₄T + 90 μ M *D*-TG₄T is necessary to observe quadruplex formation). Oligothymidylate markers (dT₆, dT₉, dT₁₈, dT₂₁ or dT₂₄) were also loaded on the gel. One should note that the migration of the dT_n oligonucleotides does not necessarily correspond to single strands⁸: these oligonucleotides were chosen here to provide an internal migration standard, not to identify intramolecular or higher-order structures.

Besides the gel shown in **Figure 3 left**, we present here another independent experiment:

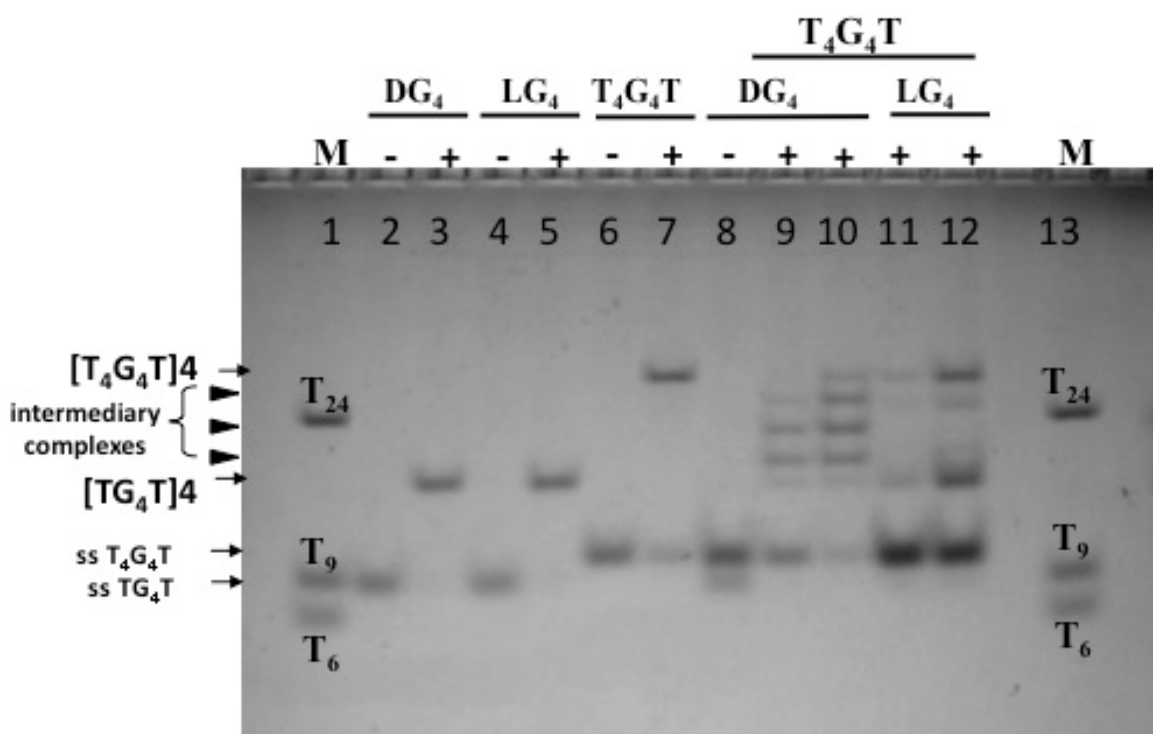


Figure S4. An independent gel experiment (the 5 leftmost lanes are shown in Figure 1). M corresponds to size markers. (+): conditions favorable to G4 formation; (-): conditions unfavorable to G4 formation (no cation and preheating). The gel was revealed by UV-shadow. Lanes 2-7: only one oligonucleotide is present, either *D*- or *L*-TG₄T or *D*-T₄G₄T. The difference in migration between the single strands is relatively modest (compare lane 6 with lanes 2 or 4) whereas the difference between the tetramolecular quadruplex is significant (compare lane 7 with lanes 3 or 5). Lanes 8-12: two oligonucleotides of different lengths are mixed. Lanes 9 and 10: the two *D*-strands are incubated together at two different concentrations. Besides a fraction of *D*-T₄G₄T that remain single-stranded, 5 bands are seen, corresponding to the "pure" *D*-[T₄G₄T]₄ and *D*-[TG₄T]₄ quadruplexes (slowest and fastest bands, respectively) and three intermediate complexes, in agreement with the formation of the tetramer. Lanes 11-12: when the *D*-T₄G₄T strand is mixed with *L*-TG₄T, the predominant species correspond to the "pure" *D*-[T₄G₄T]₄ and *L*-[TG₄T]₄ quadruplexes. A significant amount of *D*-T₄G₄T remains single-stranded. Lane 12: a minor band of intermediate mobility suggests that enantiomeric exclusion is not absolute.

Strand exchange reaction of intermolecular quadruplex by PAGE

L- or *D*-TG₄T with fluorescein at the 3' end (0.5 μM strand concentration) were incubated with an excess amount of preformed intermolecular DNA quadruplex, 20 μM TG₄T, at room temperature, with and without 1 mM of cationic comb-type copolymer (CCC). The reaction was done in 10 mM LiCaco and 0.1 M NaCl. 12% glycerol and 5% PVS were added just before loading. Samples were loaded on a 20% polyacrylamide (acrylamide/bis-acrylamide 19:1) gel containing TBE 1X and NaCl at 10mM. Electrophoresis was performed at 4W/gel to reach a temperature close to 35°C (electrophoresis performed at room temperature). The gels were revealed by recording fluorescence at 520 nm ($\lambda_{\text{exc}} = 480 \text{ nm}$).

The cationic comb copolymer (CCC) increased both the association and the dissociation rate of intermolecular quadruplex⁹. The copolymer acted as a nucleic acids chaperone that reduces the energy barrier need for both the dissociation and association folding¹⁰ and enables the strand exchange reaction between the fluorescent strand and the preformed unlabeled quadruplex

Supplementary references:

- (1) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059-1077.
- (2) Garbesi, A.; Capobianco, M. L.; Colonna, F. P.; Tondelli, L.; Arcamone, F.; Manzini, G.; Hilbers, C. W.; Aelen, J. M. E.; Blommers, M. J. J. *Nucleic Acids Res.* **1993**, *21*, 4159-4165 (correction published p5286); Urata, H.; Kumashiro, T.; Otake, T.; Kawahata, T.; Akagi, M. *Nucleic Acids Res Suppl* **2002**, 163-164.
- (3) Maruyama, A.; Katoh, M.; Ishihara, T.; Akaike, T. *Bioconj. Chem.* **1997**, *8*, 3-6; Maruyama, A.; Watanabe, H.; Ferdous, A.; Katoh, M.; Ishihara, T.; Akaike, T. *Bioconj. Chem.* **1998**, *9*, 292-299.
- (4) Guédin, A.; Alberti, P.; Mergny, J. L. *Nucleic Acids Res.* **2009**, *37*, 5559-5567.
- (5) Mergny, J. L.; Phan, A. T.; Lacroix, L. *Febs Lett.* **1998**, *435*, 74-78; Mergny, J. L.; Lacroix, L. *Curr. Protocols Nucl. Acids Chem.* **2009**, *17*, 17.11.11-17.11.15.
- (6) Mergny, J. L.; De Cian, A.; Ghelab, A.; Saccà, B.; Lacroix, L. *Nucleic Acids Res.* **2005**, *33*, 81-94.
- (7) Gros, J.; Avino, A.; Lopez de la Osa, J.; Gonzalez, C.; Lacroix, L.; Perez, A.; Orozco, M.; Eritja, R.; Mergny, J. L. *Chem. Comm.* **2008**, *2008*, 2926-2928.

- (8) Kejnovska, I.; Kypr, J.; Vorlickova, M. *Biochem. Biophys. Res. Commun.* **2007**, 353, 776-779.
- (9) Moriyama, R.; Shimada, N.; Kano, A.; Maruyama, A. *Nucleic Acids Symp. Ser.* **2008**, 227-228.
- (10) Moriyama, R.; Shimada, N.; Kano, A.; Maruyama, A. *Nucleic Acids Symp Ser (Oxf)* **2009**, 61-62.

3) *Résultats et Discussion : Chapitre III*

Nous avons présenté dans ce chapitre les résultats obtenus à partir de mesures des stabilités et des cinétiques d'association de G4 tétramoléculaires modifiés : G4 comportant une insertion de 8-méthyleguanine (M) à différentes positions de la séquence d(TG₄T) et G4 formés de L-ADN.

3.1 Effet de la 8-méthyleguanine sur la structure, stabilité et cinétique de formation de G4 tétramoléculaires

L'effet de la 8-méthyleguanine sur la structure, la stabilité et la cinétique de formation de G4 tétramoléculaires dépend de la position de la modification au sein de la séquence d(TG₃T) ou d(TG₄T).

3.1.1 *L'insertion de 8-méthyleG aux extrémités favorise la formation de quartet en syn*

Les résultats de RMN semblent montrer que toutes les séquences étudiées forment des G4 tétramoléculaires. Certaines positions de modification semblent favoriser un changement de conformation des guanines en *syn*, en particulier l'insertion de la 8-méthyleguanine aux extrémités (d(TMGGT), d(TMGGGT) et d(TGGGMT)). La modification de la deuxième guanine (d(TGMGT) [161] et d(TGMGGT)) ne change pas la conformation des guanines, elles restent en conformation *anti* comme la structure formée par les séquences non modifiées.

3.1.2 *L'introduction de quartet en syn pourrait favoriser un signal CD de type hybride ou de type II*

Ces séquences modifiées donnent des spectres de CD différents en fonction également de la position de l'insertion de guanine modifiée. Les séquences modifiées peuvent présenter un spectre CD de type I (pic positif à 260 nm et pic négatif à 240 nm), hybride (2 pics positifs à 260 et 295 nm) ou de type II (pic positif à 295 nm et pic négatif à 260 nm).

Les résultats de CD présentés dans ce chapitre confirment une fois de plus les résultats de calculs de spectres CD de deux G-quartets empilés, effectués par D.M. Gray *et al.* en 2008 [126] et plus récemment par S. Masiero *et al.* [127]. Ces calculs montrent que les spectres CD dépendent de l'orientation relative des deux G-quartets. Les deux faces d'un G-quartet ne sont pas équivalentes (cf. §I.1.2.1 a): la face “+” (ou “tête/face”) d'un G-quartet est définie, par convention, comme la face où les liaisons hydrogène du donneur N-H vers l'accepteur carbonyle C=O tournent dans les sens horaire ; l'autre face du quartet est définie

comme la face “-” (ou “queue/pile”) [6]. Si les deux quartets successifs ont les mêmes polarités (les calculs sont faits sur des G-quartets non reliés par un squelette phosphaté), le signal CD observé est de type I. Par contre, s'ils ont des polarités inversées, le signal CD est de type II. Les résultats de ces calculs permettent d'avancer des hypothèses pour l'interprétation des spectres CD obtenus expérimentalement.

Ainsi, la modification de la deuxième guanine (d(TG**M**GT) et d(TG**M**GGT)) montrant seulement des G-quartets en *anti* (d'après les résultats de RMN) présente un signal CD de type I similaire à leurs homologues. L'introduction des quartets tout-*syn*, dans le cas des insertions de 8-méthyleguanine aux extrémités, entraîne un signal CD de type hybride (d(TMGGGT) et d(TGGGMT)) ou de type II (d(TMGGT)).

La structure de la séquence modifiée d(TGG**M**GT) n'a pas pu être déterminée à cause de la présence de différentes structures G4 en solution qui rend l'analyse du spectre RMN plus difficile. Ce résultat est cohérent avec celui de la mobilité électrophorétique du fait de la présence de plusieurs bandes. Cette modification sur l'avant-dernière guanine de l'extrémité 3' de d(TG₄T) favorise un spectre CD de type II. Cette modification induirait donc la formation d'au moins un quartet en *syn*. En effet, à part le quartet contenant des guanines modifiées qui pourrait être en *syn*, nous ne pouvons pas exclure que d'autres quartets non modifiés basculeraient également en *syn*.

3.1.3 Une modification à l'extrémité 5' stabilise et augmente la vitesse de formation des G4 tétramoléculaires

Seules les séquences modifiées à l'extrémité 5' (d(TMGGGT) et d(TMGGT)) permettent la formation de G4 tétramoléculaires plus stables et sont caractérisés par une vitesse d'association supérieure aux G4 formés par les séquences non modifiées. Ce résultat est cohérent avec les résultats trouvés dans d'autres études de G4 tétramoléculaires modifiés d(TRGGGT) (où **R** = 8-amino-2'-déoxyguanine ou 8-bromo-2'-déoxyguanine) [379]. Plus la modification est proche de l'extrémité 3', plus nous réduisons la stabilité et la cinétique d'association du quadruplexe correspondant. Ce résultat suggère que la guanine en *syn* à l'extrémité 5' participerait au processus de nucléation du complexe [141]. La formation d'un quartet tout-*syn* favoriserait l'empilement du quartet tout-*anti* sous-jacent, résultat supporté par les modèles moléculaires.

3.1.4 Certaines positions de modification favoriseraient la formation de plusieurs espèces en solution.

Les G4 modifiés migrent, en général, plus vite que leurs homologues sur gels natifs de 20% polyacrylamide. Il est aussi intéressant de remarquer par gel d'électrophorèse qu'à basse température (19°C) certaines séquences modifiées (d(TGGMGT) et d(TGGGMT)) forment un mélange de plusieurs espèces en solution qui sont moins stables que la séquence non-modifiée. Ces structures, à plus haute température (à 35°C), semblent se dissocier pour ne former qu'une seule structure majoritaire. Ce résultat est confirmé par les spectres de RMN (résultats non présentés).

3.2 Sélection chirale pour les G4 tétramoléculaires de D- et L-ADN

La séquence L-d(TG4T) forme sans surprise un G4 tétramoléculaire de stabilité et de vitesse d'association équivalente à son énantiomère dextrogyre (D-ADN). Comme attendu, le G4 tétramoléculaire formé par la séquence L-ADN présente un spectre CD inversé par rapport à celui de son homologue, ce qui confirme une chiralité inversée des deux énantiomères. Le L-ADN forme donc un G4 tétramoléculaire parallèle avec une hélicité gauche au lieu de droite.

Les études de cinétique d'association, de mobilité électrophorétique et d'échange de brins par électrophorèse ont montré une exclusion chirale quand nous mélangeons les deux brins D- et L-ADN en solution. En effet, nous observons la formation de G4 comportant uniquement D- ou L-ADN mais jamais un mélange de deux énantiomères dans le même G4 tétramoléculaire. Cependant la formation de G4 chimériques composés de D- et L-nucléotides sur le même brin est possible [159].

Nous pourrions nous servir des connaissances acquises sur les G4 tétramoléculaires parallèles et sur les séquences modifiées pour des applications en nanotechnologie. Une de ces applications sera présentée dans le chapitre suivant. Nous avons utilisé ici l'ADN origami comme support pour visualiser la formation de G-quadruplexes "synaptiques".

Chapitre IV. L'ADN origami, nanoplateforme pour la visualisation d'ADN synaptiques

1) Introduction

En 1982, Nadrian C. Seeman propose d'utiliser les propriétés d'auto assemblage de l'ADN pour créer des structures artificielles. La suggestion de considérer des acides nucléiques non plus comme des molécules génétiques mais comme un matériau de construction des nanostructures est complètement révolutionnaire et fait naître ainsi la nanotechnologie de l'ADN.

La nanotechnologie de l'ADN est fondée sur le principe que l'ADN simple-brin s'auto assemble en double-hélice par complémentarité des bases. Ces interactions intermoléculaires facilement prédictibles permettent aux chercheurs de "programmer" l'association de molécules d'ADN individuelles. Cependant, l'ADN double-brin est linéaire et ne formerait donc que des assemblages unidimensionnels. Pour créer des structures complexes à deux ou trois dimensions, il est nécessaire de créer des jonctions d'ADN ramifié (Figure 22A). Ces jonctions ramifiées à 4 bras contenant des extensions simple-brins cohésives ("*sticky-ends*" en anglais) s'associent par complémentarité en une structure plus complexe. La flexibilité de l'ADN avec une simple jonction ne facilite cependant pas la création de structures plus évoluées. L'apparition des motifs à croisements multiples ("*crossover*" en anglais) offre la rigidité nécessaire pour l'assemblage d'objets plus larges. Les trois principaux motifs de croisements utilisés dans la nanotechnologie de l'ADN sont montrés dans la figure 22B. Le motif à double croisement (DX pour "*double crossover*") résultant de deux échanges entre les brins de polarités opposées est le motif fondamental dans la nanotechnologie de l'ADN, en particulier dans le domaine de l'ADN origami. Presque tous les motifs développés dans le domaine sont des dérivés du motif DX. Le motif à jointure paranémique (PX pour "*paranemic crossover*") résulte des échanges de brins d'ADN de même polarité à tous les sites possibles entre deux doubles hélices adjacentes. Le dernier motif JX₂ est un variant du motif PX avec deux croisements en moins au milieu.

Paul W. Rothmund, inventeur de l'ADN origami, a classé les approches existantes pour la construction de nanostructures d'ADN en 3 catégories (Figure 22C) : (1) "*multi-stranded design*" qui est composé entièrement d'oligonucléotides courts, (2) "*single-stranded*

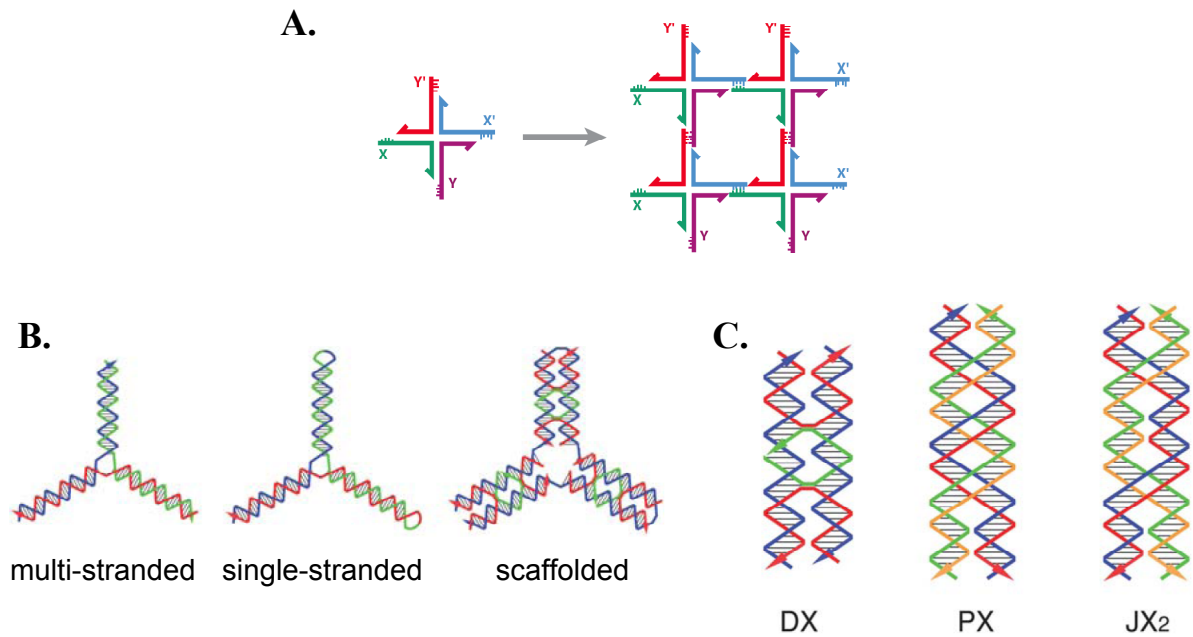


Figure 22. Les bases de la nanotechnologie de l'ADN

(A) Auto-assemblage des ADN ramifiés [164] : à gauche: un ADN ramifié à 4 bras contenant des extensions simples brins ("sticky-ends" en anglais); à droite: association des jonctions ramifiées par complémentarité en une structure plus complexe. (B) Trois principaux motifs de croisements utilisés dans la nanotechnologie de l'ADN. (C) Trois approches pour la construction de nanostructures d'ADN [165].

design” composé d’un “brin d’échafaudage” long (“*scaffold strand*” en anglais) avec peu ou pas de “brin assistant” (“*helper strand*” en anglais), et (3) “*scaffolded design*” composé d’un “brin d’échafaudage” long et de multiples “brins assistants” courts. L’approche “*multi-stranded*” est utilisée pour la construction des structures conventionnelles basées sur l’assemblage des tuiles d’ADN. Les deux autres approches, “*single-stranded*” et “*scaffolded designs*” sont utilisées pour l’ADN origami. L’octaèdre créé par W. Shih et ses collègues en 2004 [163] est un parfait exemple de l’utilisation de “*single-stranded design*”. Toutefois, la plupart des études d’ADN origami emploient l’approche “*scaffolded*”.

1.1 L’ADN origami

1.1.1 Principe de la construction de l’ADN origami et sa préparation

Un des progrès les plus marquants de la nanotechnologie de l’ADN a été développé par Paul W. Rothmund en 2006 [133]. Il décrivait dans son article une nouvelle approche pour la construction des nanostructures d’ADN en utilisant une multiplicité de brins courts et synthétiques (plus de 200 séquences), appelés “agrafes” (“*staple strands*” en anglais) pour diriger le repliement d’un brin long et circulaire, appelé “échafaudage” (“*scaffold strand*” en anglais) en une structure arbitraire. Comme nous l’avons décrit dans le chapitre “Matériels et Méthodes” (cf. : §II.4), le brin d’échafaudage provient généralement de la séquence circulaire du génome viral du phage M13 (7249 nucléotides) dont la séquence est connue (annexe 1).

La conception de l’ADN origami peut être divisé en 5 étapes (Figure 23A). Les deux premières étapes de la conception consistent à déterminer le motif de repliement du brin d’échafaudage et l’emplacement des croisements (“*crossover*”). Le diamètre d’une hélice d’ADN canonique est de 2 nm et un tour d’hélice mesure 3,4 nm, soit 10,5 nucléotides. Dans une structure d’ADN origami, un tour d’hélice mesure de 3 à 3,5 nm en longueur et occupe 3,5 nm en largeur ; il est composé de 10,7 nucléotides (plus large qu’une hélice d’ADN canonique). Des répulsions électrostatiques entre les différentes hélices sont probablement à l’origine de ce phénomène. Par conséquent, les 7249 nucléotides du brin d’échafaudage peuvent couvrir environ 8500 nm² quand il est complètement hybridé avec les séquences d’agrafes. Les trois dernières étapes permettent d’obtenir une géométrie des brins d’agrafes et de générer leurs séquences. Le repliement du brin d’échafaudage est fixé à l’aide des brins d’“agrafes”. Ces séquences d’“agrafes” sont reliées à trois hélices adjacentes formant une forme géométrique en S ou en Z. La longueur de ces brins est de 32 nucléotides pour des

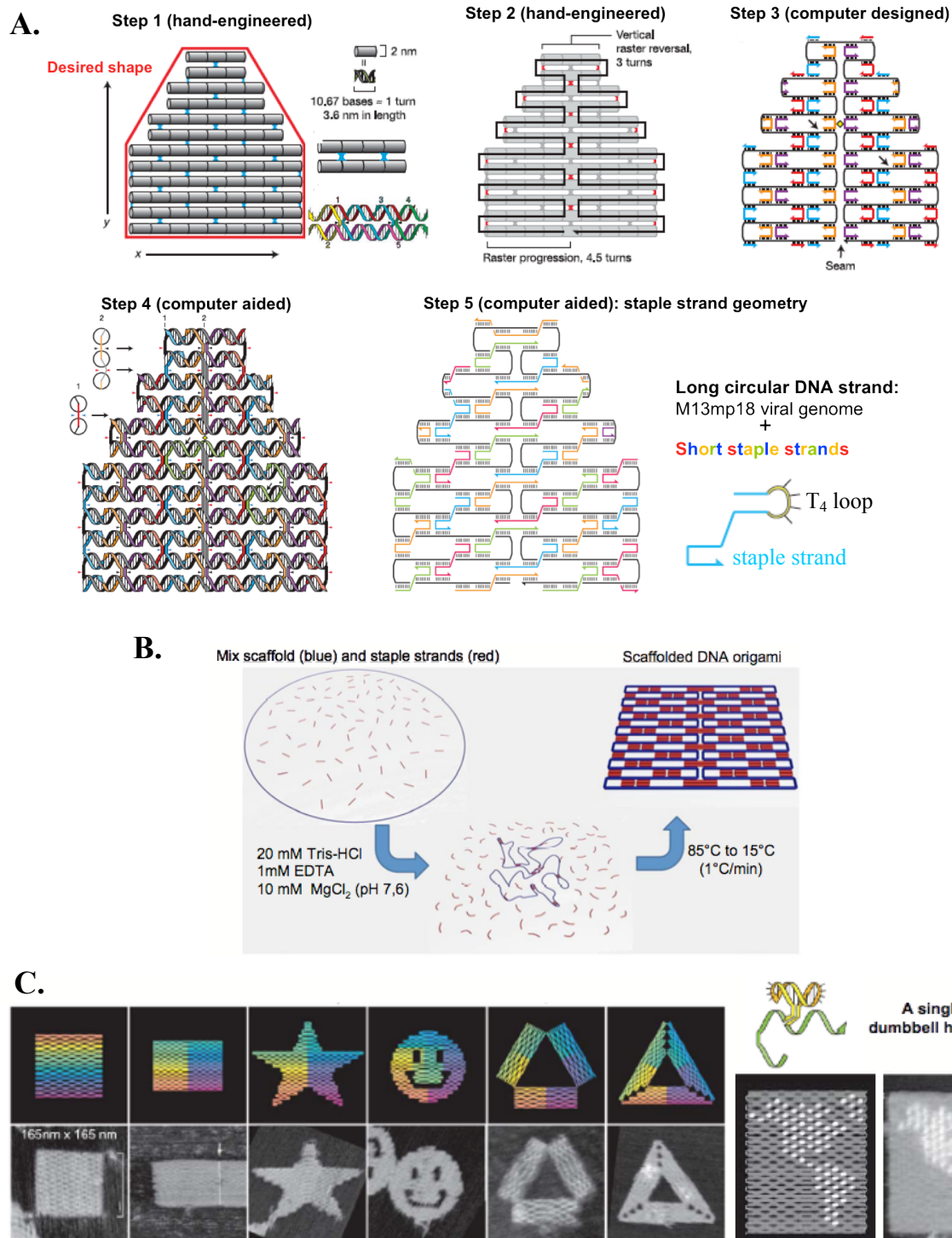


Figure 23. Principe de l'ADN origami : de la conception à la structure [133, 166]

(A) Cinq étapes de la conception de l'ADN origami et schéma d'une séquence d'"agrafe" se situant au bord de la structure avec une boucle de 4 thymines. (B) Formation d'ADN origami rectangulaire à 2D en mélangeant le brin d'échafaudage (en bleu) et les séquences d'"agrafes" (en rouge). (C) Structures d'ADN origamis bidimensionnelles observées en AFM : de gauche à droite: un carré, un rectangle, une étoile, un "smiley", un triangle et une carte de l'hémisphère ouest obtenue par insertion d'une tige boucle appelé "dumbbell hairpin" dans certaines séquences d'"agrafes".

croisements tous les 1,5 tours (soit 16 nucléotides) et de 52 nucléotides pour des croisements tous les 2,5 tours. L'utilisation des logiciels comme caDNAno permet de générer automatiquement les séquences d'agrafes [167]. Enfin, pour éviter l'agrégation des différents motifs d'ADN origami, nous insérons manuellement une boucle de 4 thymines dans les séquences d'"agrafes" se situant aux bords de la structure (Figure 23A).

Une fois les séquences d'"agrafes" synthétisées et préparées, la structure de l'origami est obtenue tout simplement en mélangeant toutes les séquences ("agrafes" et "échafaudage") dans un tampon contenant 20 mM Tris-HCl (pH 7,6), 1mM EDTA et 10 mM $MgCl_2$ (Figure 23B). Le mélange est chauffé à 85°C pendant 5 min afin de dénaturer les structures secondaires présentes dans l'ADN. La solution est ensuite refroidie lentement (à 1°C.min⁻¹) jusqu'à 15°C en utilisant un thermocycleur. La structure de l'ADN origami obtenue est visualisée en AFM sur une surface de mica fraîchement découpée. Le tampon utilisé pour la formation de la structure est aussi employé pour la visualisation en AFM. Le magnésium (Mg^{2+}) stabilise la structure de l'ADN origami en neutralisant les charges négatives des liaisons phosphodiester et permet la fixation de la structure obtenue sur la surface de mica.

L'ADN origami présente de nombreux avantages par rapport à la méthode d'assemblage décrite par N. Seeman. Cette nouvelle approche est d'abord très facile et rapide à réaliser. La méthode ne nécessite pas un contrôle précis de la stœchiométrie, elle ne nécessite pas l'utilisation d'oligonucléotides purifiés et réduit donc le temps nécessaire à la synthèse. En effet, les séquences d'"agrafes" sont supposées être uniques et ces brins ne s'hybrident donc pas entre eux. Leurs concentrations relatives sont, par conséquent, non essentielles à la formation de la structure. En général, nous mettons en excès les séquences d'"agrafes" par rapport au brin d'échafaudage (2 à 10 équivalents de séquences d'"agrafes" pour un équivalent de brin d'échafaudage). De plus, l'invasion des brins d'"agrafes" déplace les structures secondaires non-désirées du brin d'échafaudage et permet un repliement correct en éliminant la fixation des "agrafes" tronquées ou incorrectes. Les objets obtenus par cette nouvelle stratégie présentent peu de défauts et le rendement est plutôt bon (>70% de structures bien formées [133]) par rapport à la méthode classique (~34% de structures bien formées [165]).

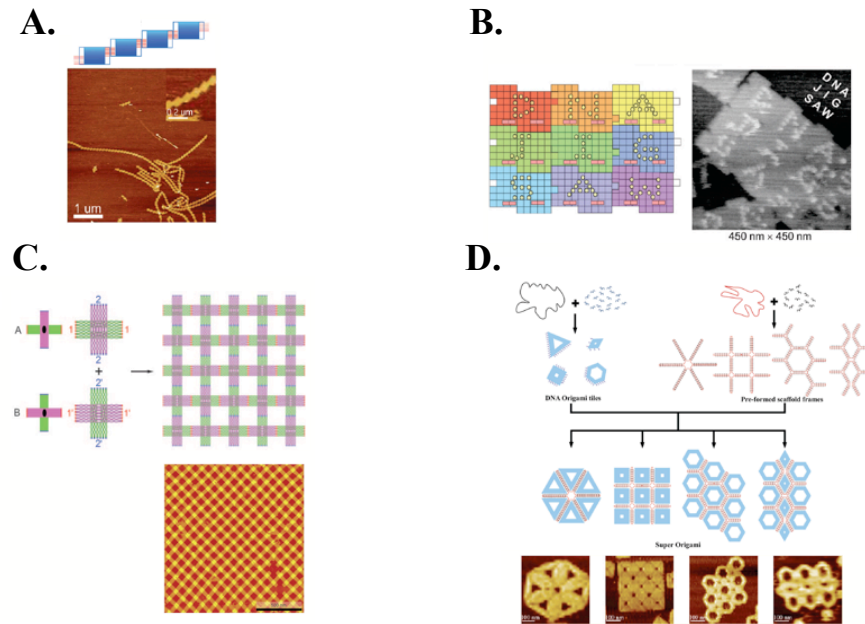


Figure 24. Augmentation de la taille des nanostructures par l'association de tuiles d'ADN origamis (A) Schéma et image d'AFM d'un ruban d'ADN formé par l'association linéaire d'origami [168]. (B) Dessin et image d'AFM d'un puzzle "DNA JIG SAW" basé sur la complémentarité des formes et sur l'hybridation par des extensions simples brins ("sticky-ends") des tuiles d'origami [169]. (C) Schéma illustratif et image d'AFM d'un treillage en deux dimensions à base de tuiles d'origami [170]. (D) Schéma d'illustration de "superorigami" et images d'AFM des superstructures obtenues [171].

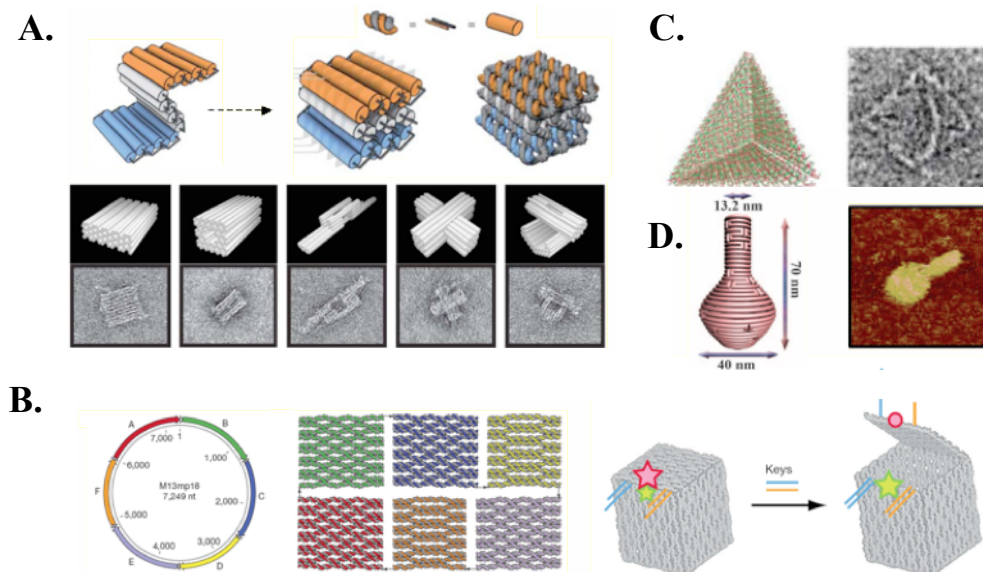


Figure 25. Structures d'ADN origamis en 3D

(A) Construction des objets tridimensionnels basés sur l'empilement des couches d'hélices formant une structure en nid d'abeille [172] : en haut : schéma illustratif de la formation des nids d'abeille par empilement des couches d'hélices (la double-hélice présentée par un cylindre); en bas : objets obtenus visualisés au microscope électronique en transmission (MET). (B) Construction d'une boîte d'ADN origami [135] : à gauche : division de la séquence circulaire du phage M13 (7249 nucléotides) en six domaines et les six faces correspondantes de la boîte; à droite : ouverture de boîte contrôlée par un système à double-serrures. (C) Construction d'un tétraèdre à partir des faces d'ADN origami triangulaires [173] : à gauche : dessin schématique du tétraèdre avec les faces fermées; à droite : image de MET correspondante. (D) Formation d'un nano-flacon [174] : à gauche : représentation schématique du nano-flacon avec indication des dimensions; à droite : image d'AFM correspondante.

1.1.2 Structures d'ADN origami bi- et tri- dimensionnelles

a) Structures bidimensionnelles

Différentes nanostructures bidimensionnelles ont été montrées dans l'article de Paul W. Rothmund publié en 2006 dans *Nature* [133]. En changeant simplement les séquences d'"agrafes", nous pouvons obtenir toutes les formes possibles et imaginables par exemple un carré, un rectangle, une étoile, un "smiley", des triangles, un dauphin, etc... (Figure 23C). Les séquences d'"agrafes" peuvent aussi être modifiées pour créer des motifs à la surface de la structure. Ainsi, en insérant une tige boucle appelée "*dumbbell hairpin*" dans certaines séquences d'"agrafes", nous induisons un relief local qui se traduit par un pixel plus clair sur les images d'AFM. La carte de l'hémisphère ouest montrée dans la figure 23C est obtenue par cette méthode. Néanmoins la méthode de l'ADN origami était limitée, à l'époque, par la taille du brin d'échafaudage, et le fait que les "agrafes" soient uniques ne permet pas d'augmenter arbitrairement la taille des objets.

b) Comment augmenter la surface d'ADN origami en 2D ?

Nous avons mentionné que la longueur du génome M13mp18 est de 7249 nucléotides couvrant en présence d'"agrafes" une surface de 8500 nm². Cette surface est suffisante pour répartir des nano-objets et observer leurs fonctions, mais elle est trop petite pour construire des nano-machines plus complexes notamment des circuits logiques. Deux méthodes ont été développées pour résoudre le problème lié à la taille du brin d'échafaudage. Pour la première approche, Högberg et ses collègues [175] ont produit en une seule hybridation deux structures d'origami différentes à partir d'une double-hélice comme brins d'échafaudages (7560 pb). Pour cela, l'ADN d'échafaudage double-brin est d'abord dénaturé à haute température supplémenté de formamide afin d'obtenir la séparation des deux brins complémentaires. Une fois dénaturé, le mélange (2 brins d'échafaudages + les "agrafes") est refroidi rapidement à la température ambiante afin de promouvoir l'hybridation des brins et de piéger les complexes "échafaudage-agrafes". La formation des structures est achevée en éliminant progressivement le formamide par dialyse [176]. La deuxième méthode a été présentée par Pound et ses collègues en 2009 [177]. Cette méthode consiste à utiliser des amorces biotinylées pour une amplification par PCR suivie d'une séparation de deux brins d'ADN via des billes magnétiques recouvertes de streptavidine, ce qui permet la production de brins d'échafaudages de longueur variable. Cependant, comme l'ADN d'échafaudage est allongé, le nombre des séquences d'"agrafes" nécessaires au repliement de la structure doit également être augmenté, ce qui peut créer de nombreux mésappariements. En outre, des brins

d'échafaudage trop longs peuvent entraîner des cassures de l'ADN et un repliement partiel des structures.

Un autre moyen pour augmenter la taille des structures d'ADN origami consiste à associer les tuiles d'origamis individuelles. De nombreuses équipes de recherche ont étudié cette question. Li et ses collègues présentent un élargissement linéaire de l'ADN origami en utilisant des sites de connexions procurés par les séquences d'"agrafes" (Figure 24A) [168]. La même équipe montre, en 2010, l'étendue des structures d'origami dans deux dimensions en utilisant plus de complexes d'"agrafes" [178]. Dans la même année, l'équipe de Sugiyama a employé une stratégie de "puzzle Jigsaw" qui est basée sur la complémentarité des formes et sur l'hybridation par des extensions simple-brin ("*sticky-ends*") pour élargir la structure bidimensionnelle (Figure 24B) [169, 179]. Récemment, l'équipe de Seeman présente un treillage en deux dimensions à base de tuiles d'ADN origami (Figure 24C) [170]. Et en dernier lieu, Yan et son équipe utilisent une nouvelle stratégie appelée "*superorigami*" ou "*origami of origami*" (Figure 24D) [171]. La méthode consiste à utiliser deux structures d'origami à partir de deux brins d'échafaudages différents. Le brin M13 (en noir) en présence des séquences d'"agrafes" forme des tuiles d'ADN origamis individuelles contenant chacune des extensions simple-brin ("*sticky-ends*"). Ces derniers permettent de relier les tuiles d'origami à l'ossature formée par le brin d'échafaudage PhiX174.

c) Structures tridimensionnelles

Nous avons montré dans le paragraphe précédent que cette technique permettait d'obtenir des nanostructures bidimensionnelles (2D) de tailles importantes. Rien n'empêche d'utiliser l'ADN origami pour former des structures tridimensionnelles (3D). En 2010, l'équipe de Shih a présenté la construction d'objets en 3D basée sur l'empilement de couches d'hélices permettant la formation de structure en nid d'abeille (Figure 25A) [172]. Le repliement de structures aussi compactes nécessite un refroidissement très lent (>174 heures) et une concentration de Mg^{2+} optimale pour la réaction.

Cependant la plupart des structures octaèdres en 3D utilisent l'ADN pour former les différentes faces en laissant le contenu vide. Une structure 3D creuse peut-être utilisée comme un cargo pour le transport des molécules pharmaceutiques. Le premier exemple d'une telle structure 3D est une "nanoboîte" créée par Andersen et ses collègues en 2009 [135]. Ils ont divisé le brin d'échafaudage (7249 nucléotides) en six domaines qui correspondent aux six faces d'un cube (Figure 25B). La formation de la boîte est contrôlée par la tension des brins liant les deux faces. La particularité de cette construction est qu'elle possède un système à

double-serrures permettant de contrôler l'ouverture de la boîte. Avec une autre stratégie, l'équipe de M. Komiyama a également rapporté la construction de boîte d'ADN en 3D [380]. D'autres structures 3D similaires ont été présentées par les équipes de H. Yan et de H. Sugiyama en 2009 (Figure 25C)[173, 180]. Plus récemment, l'équipe de H. Yan a montré la formation d'objet sphérique, un nano-flacon (Figure 25D) [174].

1.1.3 Applications de l'ADN origami

La structure d'ADN origami présente une surface plane permettant d'attacher différentes molécules ou matériels (des protéines, des nanotubes de carbone, des nanoparticules, des structures d'ADN, ...).

a) Détection des molécules uniques

Une des premières applications a été développée par Ke et ses collègues en 2008 [181]. Dans cette étude, ils utilisent l'origami rectangulaire comme une puce à ADN pour détecter l'expression des gènes. Des sondes d'oligonucléotides simple-brin ont été introduites dans certaines séquences d'"agrafes" situées à des positions précises de la surface de l'origami. Ces sondes sont complémentaires aux séquences d'ARN cibles. La capture des ARN est révélée par un contraste plus clair en AFM (Figure 26A). Bien que cette méthode ne puisse rivaliser avec la puce à ADN traditionnelle qui analyse, en une fois, des milliers de cibles, elle permet néanmoins la détection d'un faible niveau d'expression de la cible.

En 2010, Voigt et ses collègues ont démontré que des réactions chimiques de molécules uniques pouvaient être visualisées sur une surface d'ADN origami [182]. Plusieurs groupements fonctionnels sont disposés sur l'ADN origami afin d'observer leurs liaisons avec la streptavidine et d'étudier leurs clivages par différentes réactions (Figure 26B).

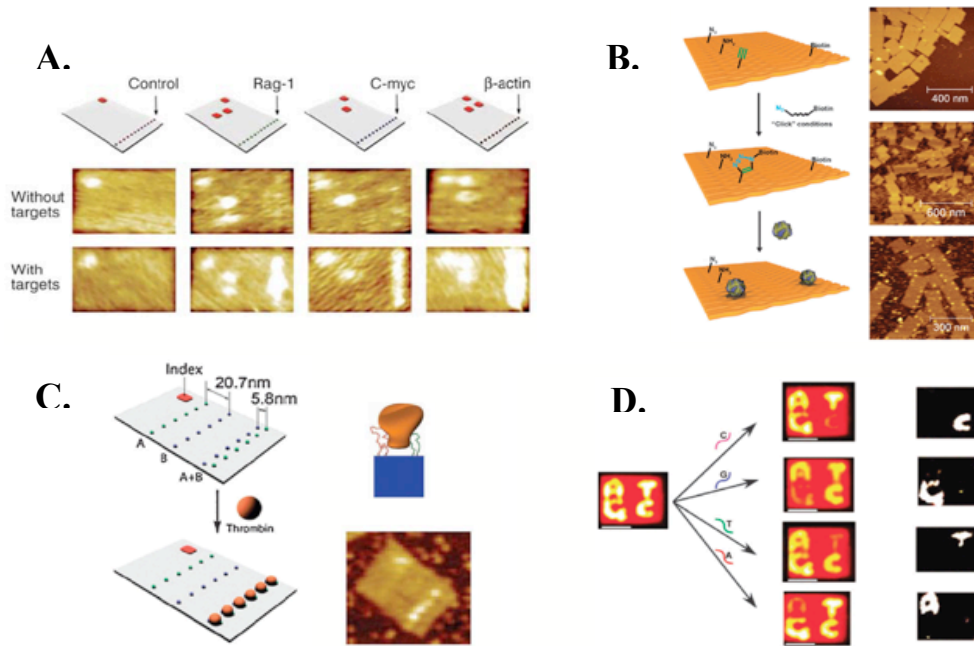


Figure 26. Détection des molécules uniques sur surface d'ADN origami

(A) Détection de l'expression génique en AFM sur une puce d'ADN à base d'origami [181]. (B) Visualisation de réaction chimique de molécule unique en AFM sur une surface d'ADN origami [182]. (C) Liaison bivalente thrombine-aptamères observée en AFM sur une nanostructure d'ADN [183]. (D) Détection des SNP sur une puce moléculaire à base d'ADN origami et visualisation en AFM [184].

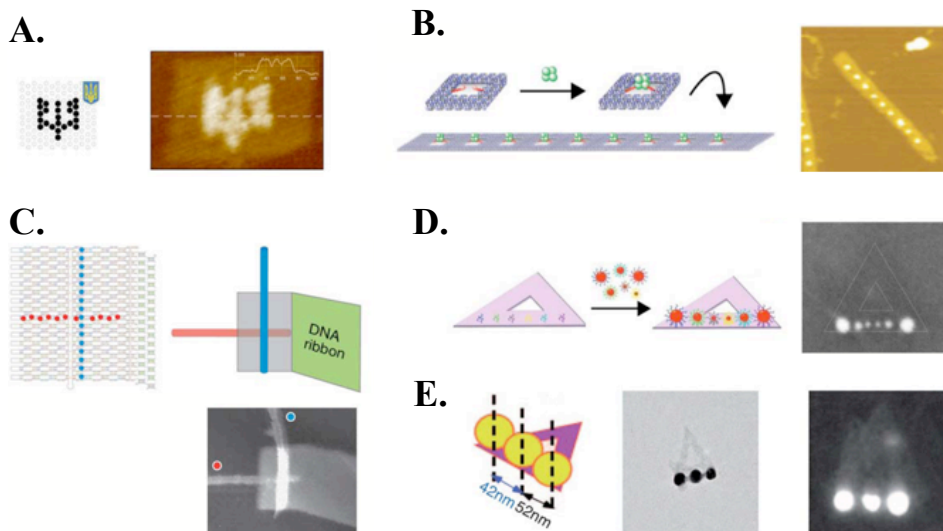


Figure 27. Organisation des matériaux sur une structure d'ADN origami

(A) Assemblage arbitraire des streptavidines sur une surface d'ADN rectangulaire [185]. (B) Insertion progressive et contrôlée des streptavidines sur des sites multiples [186]. A gauche: représentation schématique de la capture de streptavidine suivi par un assemblage multiple; à droite: assemblage de la chaîne protéique observé en AFM. (C) Organisation de deux populations de nanotubes de carbone sur une structure origami rectangulaire [187]. Schéma d'illustration de l'organisation des tubes et l'image d'AFM correspondante. (D) Organisation de six nanoparticules d'or différentes sur un ADN origami triangulaire [188]. Schéma d'illustration et image d'AFM correspondante. (E) Disposition de trois nanoparticules d'argent sur un ADN origami triangulaire [189]. De gauche à droite: schéma d'illustration de l'organisation des particules image de MET correspondante après une coloration négative et visualisation de l'échantillon sans coloration au microscope à balayage (SEM).

L'étude des molécules uniques sur une surface d'ADN origami va au-delà des réactions chimiques. Rinker et ses collègues ont montré que la liaison bivalente thrombine-aptamères dépendait énormément de la distance entre les deux aptamères [183]. Les aptamères sont introduits dans des "agrafes" et répartis sur une surface rectangulaire formant des rangées de distances variables (Figure 26C). Cette étude permet de déterminer la distance optimale pour une liaison bivalente thrombine-aptamères. Récemment, l'équipe de N. Seeman a introduit une puce moléculaire à base d'ADN origami pour détecter les SNP ("*Single Nucleotide Polymorphism*" en anglais) (Figure 26D) [184].

b) Organisation de matériaux

L'attache de nanomatériaux sur une structure d'ADN origami a été largement étudiée ces dernières années. Kuzyk et ses collègues ont décrit un système d'assemblage de protéines sur la surface d'ADN origami en 2D [185]. L'étude consistait à utiliser une surface ADN rectangulaire pour observer l'assemblage de streptavidines (Figure 27A). Plusieurs séquences d'"agrafes" contenant une biotine sont réparties arbitrairement sur la surface. La fixation des streptavidines sur les "agrafes" modifiées est visualisable par AFM. Numajiri et ses collègues ont développé un autre système permettant d'observer l'assemblage de streptavidines (Figure 27B) [186, 190]. Ils ont introduit sur les côtés du puits, deux crochets contenant chacun une biotine. Ce système permet de contrôler la fixation et le déplacement de la protéine du puits. L'équipe a montré également une insertion progressive et contrôlée des streptavidines sur des sites multiples.

Mises à part les protéines, l'ADN origami est une surface idéale pour l'organisation d'autres types de molécules, notamment des nanotubes de carbone. Maune et ses collègues ont développé une stratégie pour organiser deux populations de nanotubes de carbone sur une structure origami rectangulaire (Figure 27C) [187]. Cette étude a démontré le pouvoir d'organisation d'une structure d'ADN origami, qui pourrait être utilisée pour la synthèse des circuits électroniques.

L'assemblage des nanoparticules de métal sur la surface d'ADN origami a été également étudié. En 2010, Ding et ses collègues ont utilisé une structure triangulaire pour étudier l'organisation d'une chaîne de nanoparticules d'or (Figure 27D) [188]. Dans la même année, Pal et son équipe ont regardé l'organisation des nanoparticules d'argent sur une surface similaire (Figure 27E) [189].

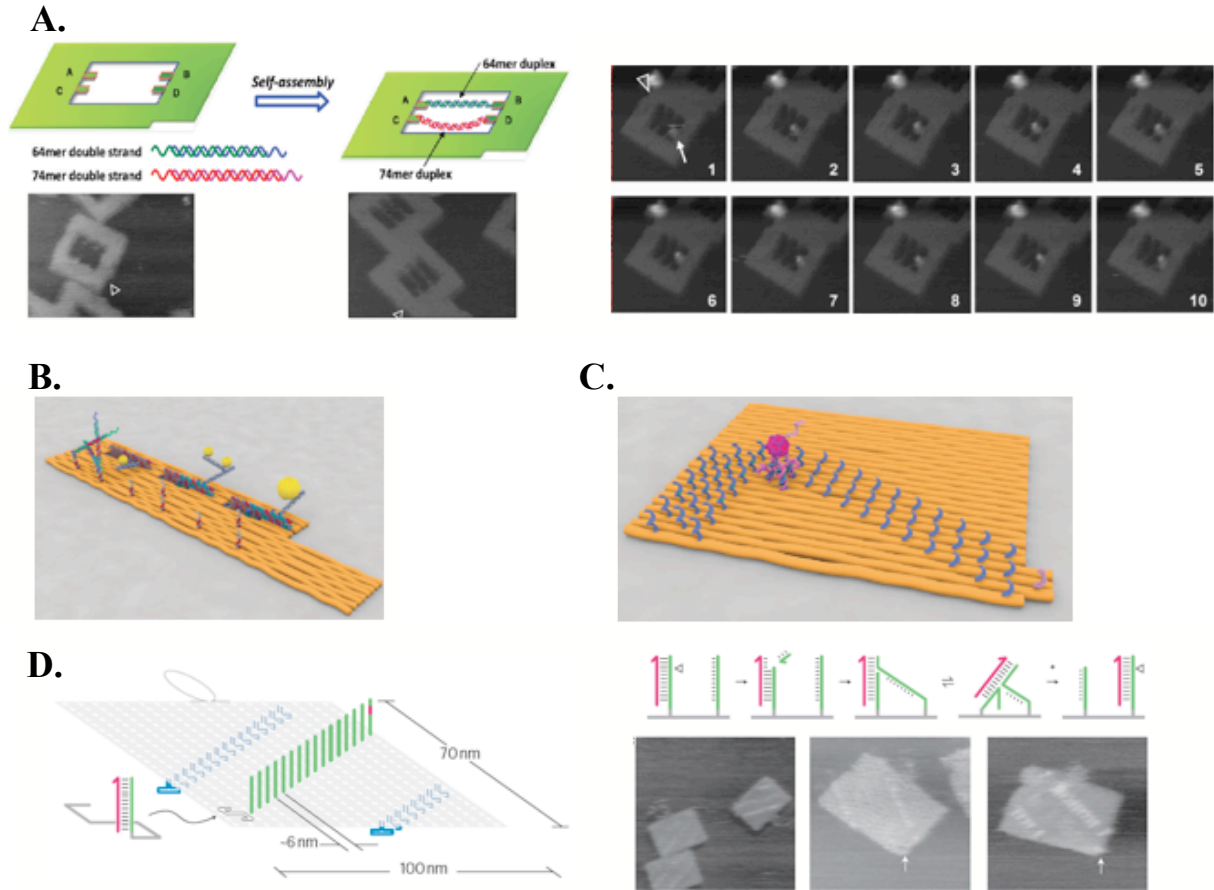


Figure 28. Étude et construction des systèmes dynamiques sur une plate-forme d'ADN origami
 (A) Observation en temps réel d'une réaction enzymatique d'une méthyltransférase (M.EcoRI) sur une surface d'ADN origami [191] : à gauche: dessin schématique du système utilisé et les images d'AFM correspondantes; à droite: visualisation de la réaction toutes les secondes par "fast-scanning AFM". (B) Représentation schématique du déplacement d'ADN moteur triangulaire sur une construction d'ADN origami comportant des stations de stockage de nanoparticules d'or [192] (image tirée de revue [193]). (C) Dessin schématique de la route de l'"araignée moléculaire" [194] (image tirée de revue [193]). (D) Mouvement d'un ADN moteur sur l'ADN origami rectangulaire [195] : à gauche: schéma de la construction d'ADN origami utilisé; à droite: schéma explicatif de la réaction enzymatique permettant le déplacement de l'ADN moteur et image d'AFM de son déplacement en temps réel.

c) Système dynamique

Au-delà des structures statiques, l'ADN origami présente une plate-forme pour construire et étudier des systèmes dynamiques. Cela permet des progrès considérables dans les domaines des robots moléculaires et des études de processus enzymatiques. Le développement concomitant de "fast-scanning AFM" a contribué au progrès de ces travaux.

En 2010, l'équipe de Sugiyama a étudié les réactions enzymatiques d'une méthyltransférase (M.EcoRI) [191] et de deux enzymes de réparation de l'ADN (hOgg1 et PDG) [196] en utilisant une surface d'ADN origami appelée "*DNA frame*". La structure de l'ADN origami contient une cavité interne et quatre sites de connexions permettant la fixation de deux doubles hélices de longueurs différentes, et donc de tensions différentes (Figure 28A). Le "fast-scanning AFM" a révélé les mouvements dynamiques des ADN double-brins et a permis de suivre la réaction enzymatique sur les deux duplexes. Cette étude a clairement montré que la flexibilité des doubles hélices est favorable à la réaction catalytique.

Plusieurs recherches passionnantes sur l'ADN robotique ont utilisé des plates-formes d'ADN origami. Seeman et son équipe ont décrit un système d'ADN moteur triangulaire qui s'est déplacé sur une construction d'ADN origami comportant des stations de stockage de nanoparticules d'or [192]. L'ADN moteur se déplace de station en station pour récupérer des particules d'or et ces mouvements sont uniquement contrôlés par des interactions moléculaires (Figure 28B). Stojanovic, Yan, Walter, Winfree et leurs collègues ont mis en place une "araignée moléculaire" contenant un corps de streptavidine et trois pattes de déoxyribozymes [194]. La route de l'"araignée moléculaire" est assemblée sur une surface d'origami en introduisant dans les "agrafes" des séquences spécifiques pour la reconnaissance et le clivage des pattes de ribozymes (Figure 28C). Plus récemment, une collaboration entre les équipes de Sugiyama et de Turberfield a permis d'observer le mouvement d'un ADN moteur par l'AFM en temps réel [195]. Ce mécanisme est facilité par un enchaînement de restrictions enzymatiques (Figure 28D).

d) ADN origami et G-quadruplexes

La seule étude qui s'intéresse à la formation de G-quadruplexes sur une structure d'ADN origami a été réalisée par l'équipe de Sugiyama en 2010 [197]. Le groupe a suivi en temps réel la formation et la dissociation d'un G-quadruplexe intemoléculaire en ajoutant et en retirant le potassium du tampon. La structure d'ADN origami utilisée est la même que celle

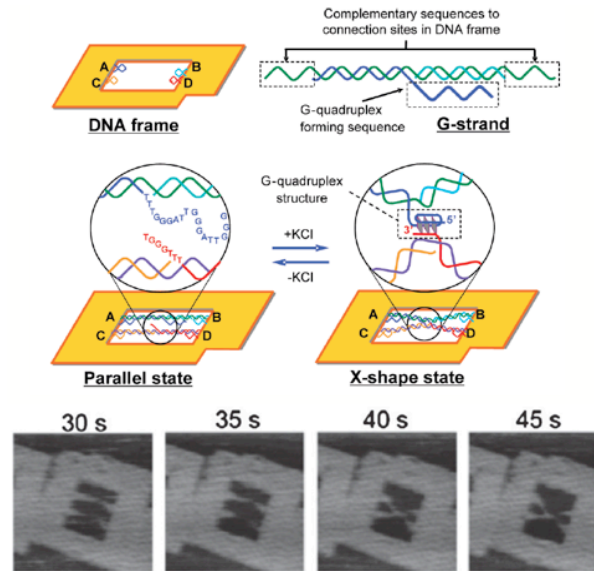


Figure 29. Observation de la formation de G4 en temps réel sur surface d'ADN origami [197]

En haut: schéma du système utilisé. La formation de G4 est facilitée par addition de K^+ et la réaction étudiée est réversible. *En bas*: images enregistrées toutes les 5 secondes par "fast-scanning AFM".

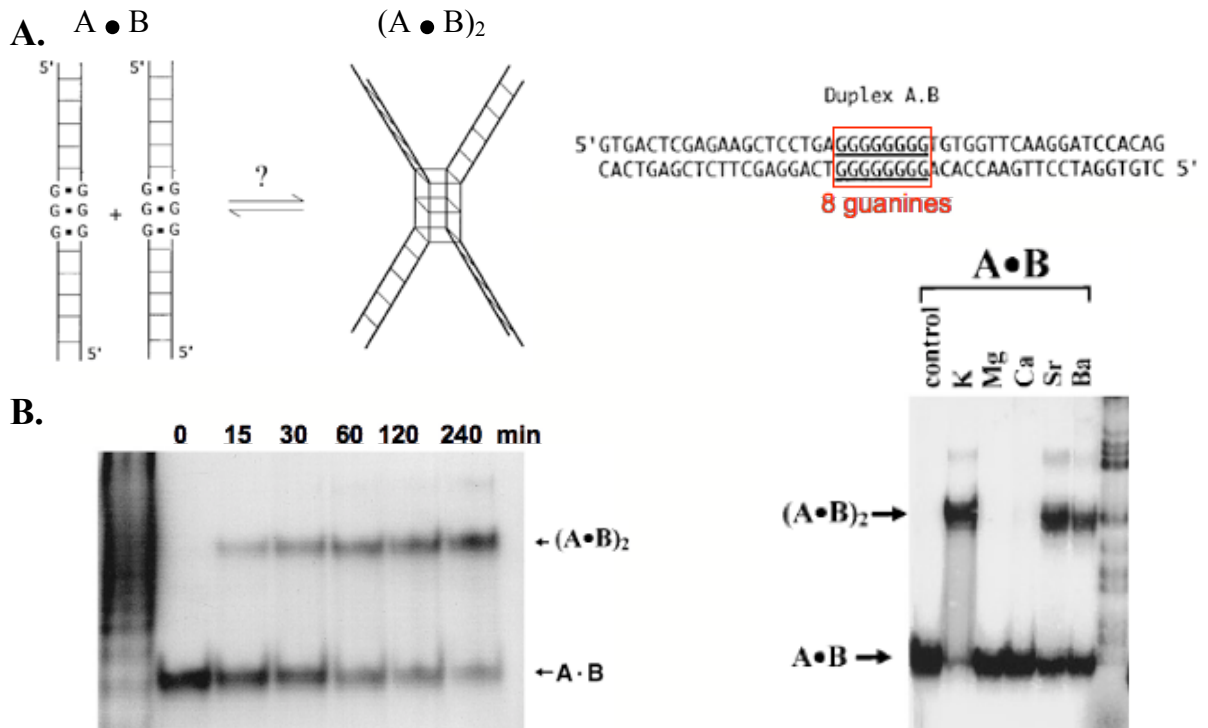


Figure 30. Propriétés de l'ADN synaptique [198, 199].

(A) *A gauche*: schéma de formation d'ADN "synaptique" par deux double-hélices contenant un défaut d'appariement au niveau des paires de bases de guanines consécutives; *à droite*: séquence du duplexe permettant la formation de l'ADN synaptique. (B) Formation lente des ADN synaptiques observée par l'électrophorèse en condition native. (C) Influence des cations monovalents (en particulier K^+) et divalents (notamment Mg^{2+}). Formation d'ADN synaptique en présence de K^+ mais pas de formation de la structure détectable en présence du Mg^{2+} .

décrit dans la figure 28A [191]. La cavité interne contenant des sites de connexions permet l'hybridation de deux double-hélices (Figure 29). Chaque double-hélice contient une cassure et une extension simple-brin riche en guanines. En absence de potassium, aucune interaction entre les deux double-hélices n'est observée, les deux duplex sont parallèles (= état parallèle). En présence de potassium, ils observent la formation de G-quadruplexe à partir des extensions simple-brins de deux duplexes formant ainsi un "X" (= état de forme "X"). L'étude a permis de déterminer les séquences idéales pour les interactions et de suivre la dynamique moléculaire des interactions en temps réel par "fast-scanning AFM".

Néanmoins, ce système est peu stable, seuls 44% de formes "X" sont observées dans les conditions optimales de formation et cela nécessite l'introduction des extensions simple-brins. Nous reportons, ci-dessous, un système plus simple de formation de G4 qui ne nécessite pas l'introduction des extensions simple-brins, il s'agit de la formation d'ADN "synaptique", en utilisant la même nanostructure. Les résultats préliminaires présentés ici ont été obtenus lors d'un stage de formation (5 semaines) à Kyoto (Japon) dans le laboratoire du professeur Sugiyama.

1.2 "ADN synaptique", G4 tétramoléculaire antiparallèle

L'ADN "synaptique" résulte de l'interaction de deux double-hélices contenant un défaut d'appariement dû à l'insertion des "paires de bases" de guanines (Figure 30A). Cette structure a été décrite par l'équipe de Dipankar Sen, en 1996 [198]. L'interaction des deux duplexes a lieu au niveau de 8 paires de bases de guanines via la formation de G-quartets. Le G-quadruplexe tétramoléculaire formé au niveau de la "synapse" possède une orientation des brins dite antiparallèle. Comme le G4 tétramoléculaire parallèle, cette structure a une formation lente, mais une fois formée, elle est très stable (Figure 30B). Leur formation dépend également de la nature et de la concentration ionique (cations monovalents ou divalents) (Figure 30C) [199]. La formation de l'ADN "synaptique" est possible entre deux duplexes identiques (homo dimère) ou entre deux différentes séquences de double-hélice (hétéro dimère). Les monomères de chaque duplexe peuvent également former des G-quadruplexes tétramoléculaires mais avec une orientation de brins parallèle.

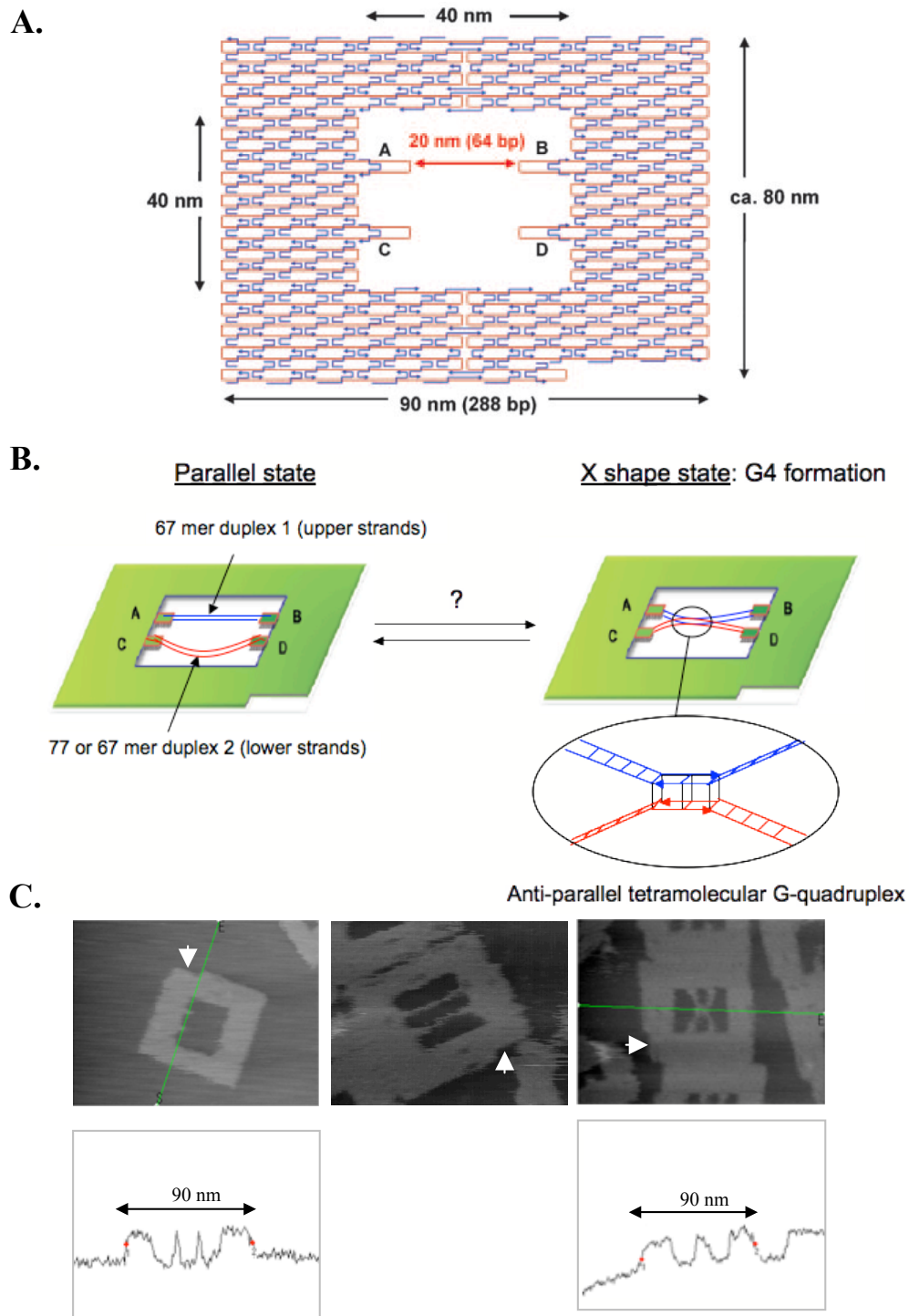


Figure 31. Système de visualisation d'ADN synaptique sur une nanostructure d'ADN

(A) Dessin schématique avec les dimensions de la nanostructure d'ADN rectangulaire utilisée [191] : en rouge: brin d'échafaudage; en bleu: les séquences d'"agrafes". (B) Schéma représentatif du système utilisé. (C) À gauche: images d'AFM de la structure d'ADN origami et la hauteur de la coupe transversale; au milieu: image d'AFM montrant l'insertion des deux duplexes sur la nanostructure; à droite: images d'AFM révélant la formation d'ADN synaptique et la hauteur de la coupe transversale.

2) Visualisation de l'ADN synaptique sur une nanostructure d'ADN

2.1 Système d'étude et préparation des échantillons

Comme mentionné dans l'introduction de ce chapitre (cf. : §I.1.1.3 d), la nanostructure d'ADN utilisée est rectangulaire et possède une cavité centrale avec quatre sites de liaisons à l'ADN (A, B, C et D) (Figure 31A). Nous fixons sur ces sites deux duplexes (duplexes 1 et 2) de mêmes (67 mer) ou de différentes (67 et 77 mer) longueurs (Figure 31B). Ce système nous permet d'observer la formation de l'ADN "synaptique" quand on passe d'un état parallèle à un état de forme "X". Les séquences étudiées disposent un nombre de guanines consécutives variables (3, 4, 6 ou 8 guanines) et sont présentées dans la table 6. Chacun des 2 duplexes possède un brin plus long comportant une extension simple brin de 16 nucléotides de chaque extrémité. Ces extensions permettent les interactions avec les sites de connexions.

Sequences (5' → 3')		
Upper strand (= duplex 1)	3Gs	67nts AGTGAGGAGCAACGCGCACCCACCGCCTTGAA GGG TTACGGCCATCAGGAACACTCTAGAGTCTCC
		99nts CTGTAGCTCATCATGT GGAGACTCTAGAGTGTTCCTGATGGCCGTGAA GGG TTCAAGGCGGTGGGTGCGCGTTGCT
	4Gs	67nts AGTGAGGAGCAACGCGCACCCACCGCCTTGAA GGG TTACGGCCATCAGGAACACTCTAGAGTCTCC
		99nts CTGTAGCTCATCATGT GGAGACTCTAGAGTGTTCCTGATGGCCGTGAA GGG TTCAAGGCGGTGGGTGCGCGTTGCT
	6Gs	67nts AGTGAGGAGCAACGCGCACCCACCGCCTTGAA GGGG TTACGGCCATCAGGAACACTCTAGAGTCTCC
		99nts CTGTAGCTCATCATGT GGAGACTCTAGAGTGTTCCTGATGGCCGTGAA GGGG TTCAAGGCGGTGGGTGCGCGTTGCT
	8Gs	67nts AGTGAGGAGCAACGCGCACCCACCGCCTTGAA GGGGGG TTACGGCCATCAGGAACACTCTAGAGTCTCC
		99nts CTGTAGCTCATCATGT GGAGACTCTAGAGTGTTCCTGATGGCCGTGAA GGGGGG TTCAAGGCGGTGGGTGCGCGTTGCT
	Short lower strand (= short duplex 2)	67nts GGAGACTCTAGAGTGTTCCTGATGGCCGTGAA GGG TTCAAGGCGGTGGGTGCGCGTTGCTCCTCACT
		99nts CGACAATAAAACAACAT AGTGAGGAGCAACGCGCACCCACCGCCTTGAA GGG TTACGGCCATCAGGAACACTCTA
		67nts GGAGACTCTAGAGTGTTCCTGATGGCCGTGAA GGG TTCAAGGCGGTGGGTGCGCGTTGCTCCTCACT
		99nts CGACAATAAAACAACAT AGTGAGGAGCAACGCGCACCCACCGCCTTGAA GGG TTACGGCCATCAGGAACACTCTA
		67nts GGAGACTCTAGAGTGTTCCTGATGGCCGTGAA GGGG TTAAGGCGGTGGGTGCGCGTTGCTCCTCACT
		99nts CGACAATAAAACAACAT AGTGAGGAGCAACGCGCACCCACCGCCTTA GGGG TTACGGCCATCAGGAACACTCTA
Long lower strand (= long duplex 2)	3Gs	77nts CGAGCGGAGACTCTAGAGTGTTCCTGATGGCCGTGAAAGGGTTCAAGGCGGT GGG TGCGCGTTGCTCCTCACTTCCTC
		109nts CGACAATAAAACAACAT GAGGAAGTGAGGAGCAACGCGCACCCACCGCCTTGAA GGG TTACGGCCATCAGGAACA
	4Gs	77nts CTCTAGAGTCTCCGCTCG AGCAAAACAAGAGAATC
		109nts CGACAATAAAACAACAT GAGGAAGTGAGGAGCAACGCGCACCCACCGCCTTGAA GGG TTACGGCCATCAGGAACA
	6Gs	77nts CTCTAGAGTCTCCGCTCG AGCAAAACAAGAGAATC
		109nts CGACAATAAAACAACAT GAGGAAGTGAGGAGCAACGCGCACCCACCGCCTTA GGGG TTACGGCCATCAGGAACA
	8Gs	77nts CTCTAGAGTCTCCGCTCG AGCAAAACAAGAGAATC
		109nts CGACAATAAAACAACAT GAGGAAGTGAGGAGCAACGCGCACCCACCGCCTT GGGGGG TTACGGCCATCAGGAACA
		77nts CGAGCGGAGACTCTAGAGTGTTCCTGATGGCCGTGAAAGGGTTCAAGGCGGT GGG TGCGCGTTGCTCCTCACTTCCTC
		109nts CGACAATAAAACAACAT GAGGAAGTGAGGAGCAACGCGCACCCACCGCCTT GGGGGG TTACGGCCATCAGGAACA
		77nts CTCTAGAGTCTCCGCTCG AGCAAAACAAGAGAATC
		109nts CGACAATAAAACAACAT GAGGAAGTGAGGAGCAACGCGCACCCACCGCCTT GGGGGG TTACGGCCATCAGGAACA

Table 6. Liste des séquences d'ADN testées

surligné en jaune: les nucléotides interagissant avec les sites de connexions de la nanostructure;
en rouge: les guanines impliquées dans la formation d'ADN synaptique.

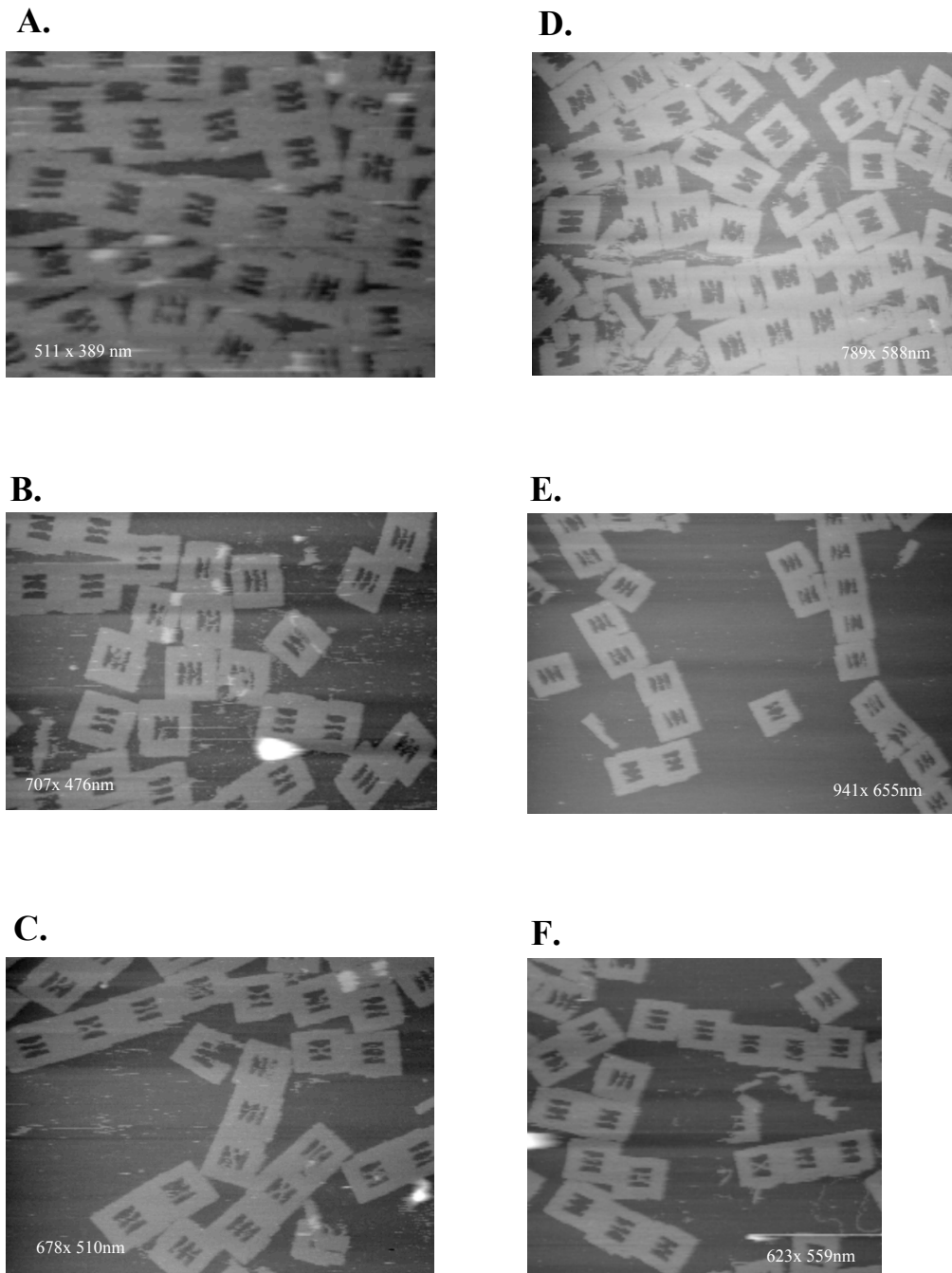


Figure 32. Images d'AFM des structures obtenues avec des duplexes courts

Images d'AFM obtenues avec les séquences à 3, 4 et 6 guanines en absence de K^+ (A, B et C respectivement) et en présence de 100 mM K^+ (D, E et F respectivement). Nous observons une majorité de formes "parallèles" (pas de formation d'ADN synaptique) dans les deux cas avec et sans K^+ .

La nanostructure d'ADN est préformée en utilisant la méthode classique d'ADN origami présentée précédemment (cf. : §II.4). En même temps, nous préformons séparément les deux duplexes dans le même tampon que celui utilisé pour l'ADN origami (20 mM Tris-HCl, 10 mM MgCl₂, 1mM EDTA à pH 7,6) en chauffant les échantillons à 85°C pendant 5 minutes et en les refroidissant lentement à 15°C avec un gradient de température de 1°C.min⁻¹. Les duplexes sont ensuite mélangés avec l'ADN origami préformé (2 équivalents de chaque duplexe pour 1 équivalent d'origami). Pour hybrider les duplexes sur la structure, nous chauffons le mélange à 50°C puis le refroidissons à 15°C avec un gradient de température de 0,5°C.min⁻¹. Pour les échantillons avec K⁺, nous ajustons simplement la concentration de K⁺ dans le mélange précédent. Les structures sont ensuite observées au microscope à force atomique (AFM) (Figure 31C).

2.2 Résultats et discussions

Le rendement de la formation de forme “X” est calculé pour les différentes séquences dans deux conditions avec ou sans K⁺. L'ensemble des résultats est récapitulé dans la table 7. Les images d'AFM des structures obtenues sont montrées dans la figure 32 et 33.

X shape yield (%)				
duplex 2	short		long	
	without K ⁺	+ 100 mM K ⁺	without K ⁺	+ 100 mM K ⁺
3 Gs	~ 0 %	0 %	4 %	2 %
4 Gs	0 %	0 %	17 %	3 %
6 Gs	15 %	5 %	25 %	11 %
8 Gs	n.d.	n.d.	86 %	40 %

Table 7. Tableau récapitulatif de l'ensemble des résultats obtenus
n.d.: non déterminé. Les ADN origamis ont été analysés dans chaque cas.

D'après les résultats obtenus, l'introduction d'un duplexe plus long permet d'augmenter le pourcentage de forme “X” formé, et donc la formation de l'ADN “synaptique” (Table 7). En effet, un duplexe plus long permet une plus grande flexibilité de la double-hélice et favorise donc des interactions inter-brins.

Nous observons également une augmentation du pourcentage de forme “X” quand nous augmentons le nombre de guanines consécutives. Ce résultat est attendu d'après nos connaissances sur les G4, l'augmentation du nombre de guanines consécutives favorise la formation et stabilise les G-quadruplexes.

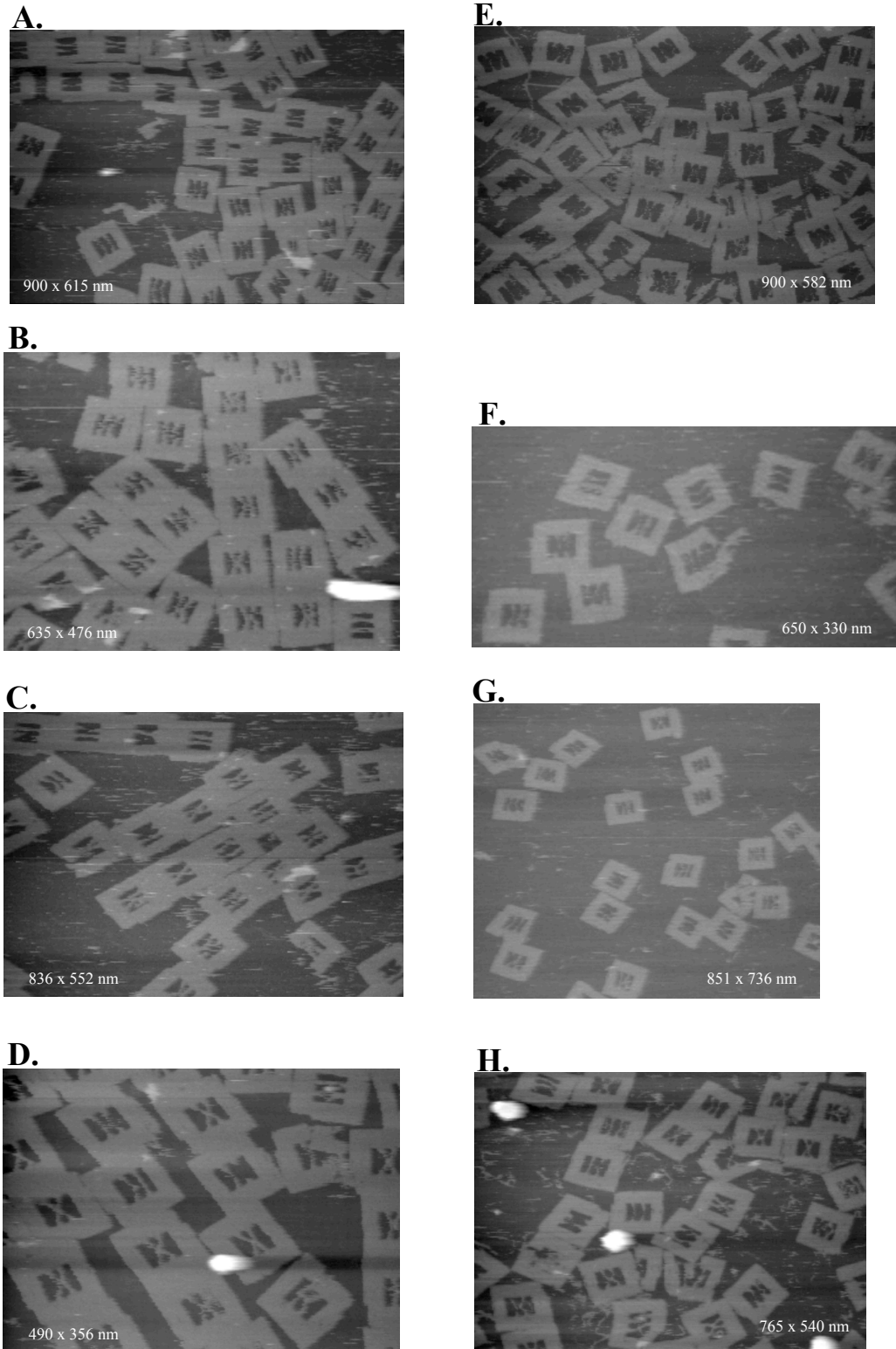


Figure 33. Images d'AFM des structures obtenues avec des duplexes longs

Images d'AFM obtenues avec les séquences à 3, 4, 6 et 8 guanines en absence de K^+ (A, B, C et D respectivement) et en présence de 100 mM K^+ (E, F, G et H respectivement). Nous observons une majorité de formes "parallèles" pour les séquences à 3, 4 et 6 guanines dans les deux cas avec et sans K^+ et une majorité de formes "X" (~86%) pour la séquence à 8 guanines mais seulement en l'absence de K^+ .

De manière surprenante, la formation de l'ADN "synaptique" ne nécessite pas la présence de K^+ (86% de forme "X" observé pour 8Gs). Au contraire, nous trouvons un rendement de forme "X" plus faible en présence qu'en absence de K^+ ! Pour toutes les séquences testées, nous observons une réduction au moins de moitié du pourcentage de forme "X" quand nous ajoutons 100 mM de K^+ . Une incubation plus longue (24 ou 48 heures dans K^+) ou une augmentation de la concentration de K^+ (200 mM) ne semble pas favoriser la formation de forme "X" (résultats non montrés). Nous ne pouvons pas non plus augmenter la concentration de K^+ au-delà de 200 mM ni trop diminuer la concentration du Mg^{2+} (pas de fixation à 1 mM Mg^{2+}) car cela pourrait compromettre la fixation de la structure sur la surface de mica. Comme mentionné dans l'introduction, le Mg^{2+} est nécessaire pour la fixation de l'ADN origami sur une surface de mica, et la charge positive du K^+ contrebalance cette adhésion [200].

La formation de G4 observée en absence de K^+ pourrait être stabilisée par le Mg^{2+} présent dans le tampon [14]. En ajoutant 100 mM de K^+ , nous devrions stabiliser la structure G4, cependant moins de formes "X" sont observées en AFM par comparaison aux images obtenues sans K^+ . Selon certains résultats obtenus dans le laboratoire du Pr. Sugiyama, ceci peut être dû à une différence d'adhésion entre la forme "X" stabilisée par K^+ et la forme "parallèle"; ou le G4 stabilisé par K^+ serait déstabilisé une fois fixé sur la surface de mica. Si c'est le cas, nous pourrions résoudre le problème en suivant la formation de G4 en temps réel, c'est-à-dire que nous fixerions d'abord la forme parallèle sur la surface de mica avant de suivre la formation de forme "X" en ajoutant K^+ . Cependant, ces résultats surprenants sont seulement préliminaires et devront être confirmés.

Avec 8 guanines consécutives, nous avons une formation de G4 majoritaire (86%) et très stable en absence de K^+ . Pour exploiter ce système simple et stable, nous pouvons étudier la dissociation du G4 tétramoléculaire antiparallèle en injectant un compétiteur comme $d(G_8)$ ou $d(G_8X_nG_8)$ (où $X = dT, dC$ ou dA) qui permettrait la formation de deux G4 tétramoléculaires séparés sur les deux duplexes. Nous pouvons également observer les interactions de certaines protéines (hélicases, Topoisomérase, NCp du VIH-1 [75] ou protéine Huntingtin [77] ou des ligands de G4 fluorescents avec ce système stable.

La nanostructure utilisée peut également servir pour observer d'autres structures G4, notamment un G4 tétramoléculaire parallèle (cf. : §VII.3.1) mise en place par Liliya Yatsunyk, Professeur au Swarthmore College, près de Philadelphie (Etats-Unis) qui a passé une année sabbatique dans notre équipe (résultats non montrés).

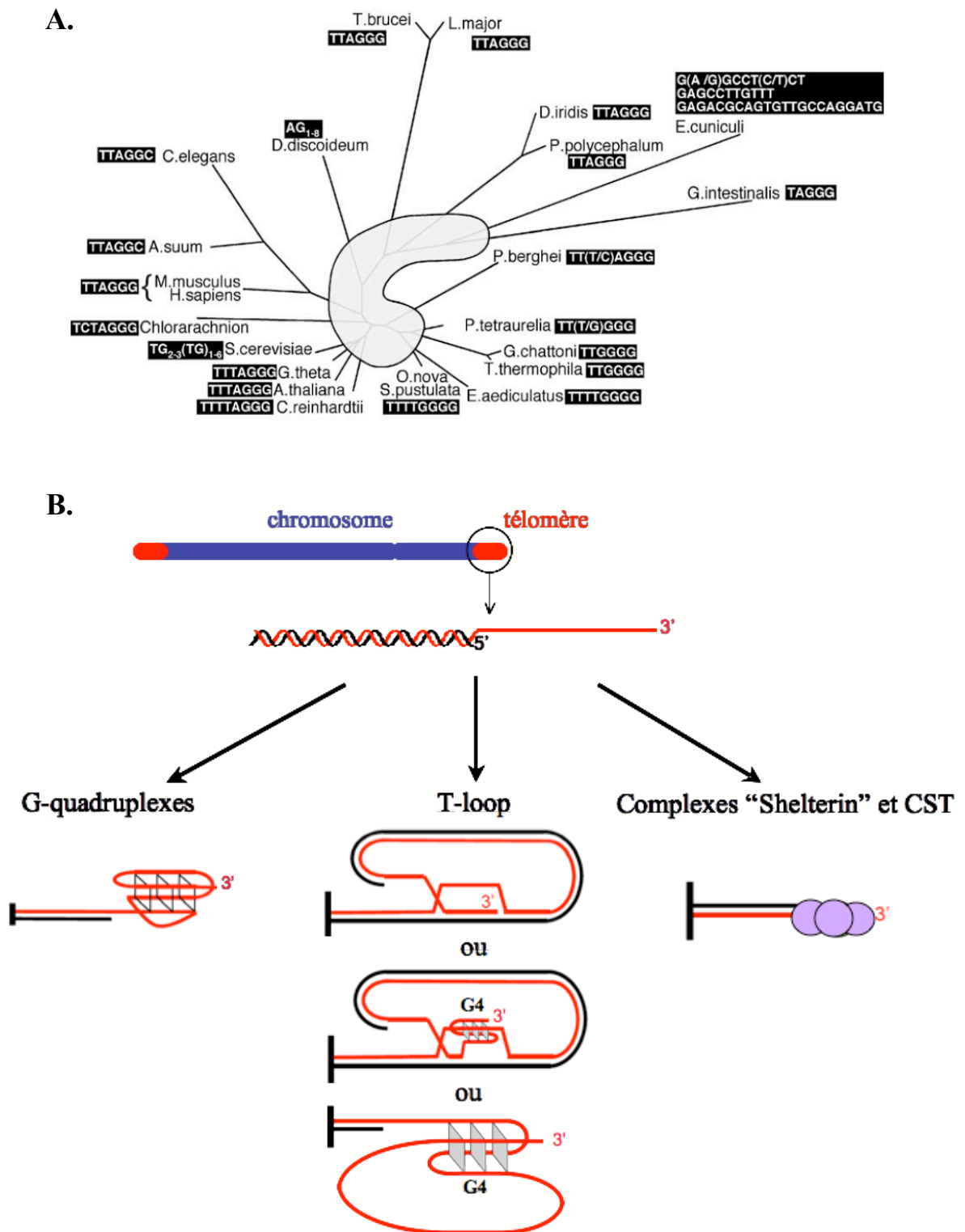


Figure 34. ADN et structures télomériques chez les eucaryotes

(A) Arbre phylogénique permettant de voir la conservation des motifs télomériques chez les eucaryotes et leur richesse en blocs de G [201]. (B) Structures adoptées par l'extrémité 3' télomérique permettant sa protection: G-quadruplexes, T-loop ou combinaison G4 et T-loop ou encore complexes "Shelterin", CST (pour Cdc13-Stn1-Ten1) pour les levures bourgeonnantes, TEBP ("Telomere End Binding Proteins") pour les ciliés.

Chapitre V. Formation de G4 par des séquences biologiques, en particulier les séquences télomériques

1) Introduction: Rôle, Structure et Maintenance des télomères

1.1 Structure du télomère et protection télomérique

Dès la fin des années 30, Hermann Muller [202] (Prix Nobel en 1946) et Barbara McClintock [203, 204] (Prix Nobel en 1983) ont observé que les extrémités des chromosomes (de drosophile et de maïs respectivement), devaient posséder des structures particulières qui les stabilisaient. Cette structure spéciale est nommée télomère (du grec “*telos*” pour “fin” et “*meros*” pour “partie”) par H. Muller. Plus tard, il a été démontré que ces télomères sont des structures nucléoprotéiques qui forme une structure “coiffante” (en anglais “*capping structure*”) stabilisant les extrémités chromosomiques chez les organismes eucaryotes et les empêchant d’être reconnues comme des cassures double-brins.

1.1.1 Séquences d’ADN télomériques

Le premier pas vers la compréhension de la nature des télomères a été accompli par E.H. Blackburn and J.G. Gall. Ces deux chercheurs ont montré que les extrémités des mini-chromosomes du cilié *Tetrahymena thermophila* étaient constituées d’une courte séquence répétée en tandem d(GGGGTT/AACCC) [205]. Comme mentionné dans l’introduction générale, les ciliés possèdent un nombre très important de mini-chromosomes dans leur macronoyau, ce qui rend aisé l’étude de leurs télomères.

Les études successives ont révélé que dans la majorité des eucaryotes, l’ADN télomérique est constitué de la répétition en tandem d’une séquence de 5 à 8 nucléotides, incluant 2, 3 ou 4 guanines consécutives [206]. Cette séquence répétée, dite motif télomérique, dépend des organismes (Figure 34A). La caractéristique de motifs télomériques chez la plupart des eucaryotes est la présence de 2-6 guanines consécutives (Table 8). Parmi les motifs télomériques, on observe une prévalence considérable pour le motif d(G₃T₂A), caractéristique de beaucoup d’organismes distants phylogénétiquement, dont les vertébrés et certains invertébrés aquatiques. Il y a, par ailleurs, différents motifs télomériques à 3 guanines consécutives (motifs en G₃): d(G₃TA) chez le parasite *Giardia intestinalis*; d(G₃T₃A) dans une variété de plantes comme *Arabidopsis thaliana* ; d(G₃T₄A) chez l’algue verte

	Organismes	Motif télomérique	Référence
Protozoaires	<i>Tetrahymena thermophila</i>	d(G ₄ T ₂)	[205]
	<i>Oxytricha nova</i>	}	[207]
	<i>Stylonicchia pustulata</i>		
	<i>Emplotes aediculatus</i>		
	<i>Paramecium</i>	d(G ₄ T ₂) et d(G ₃ T ₃)	[208, 209]
	<i>Plasmodium</i>	d(G ₃ T ₃ A) et d(G ₃ T ₂ CA)	[210, 211]
	<i>Giardia intestinalis</i>	d(G ₃ TA)	[212]
	<i>Trypanosoma brucei</i>	d(G ₃ T ₂ A)	[213, 214]
Levures	<i>Candida guilliermondi</i>	d(G ₂ TGTACT)	[215]
	<i>Schizosaccharomyces pombe</i>	d(G ₂₋₅ T ₂ AC) et d(G ₂₋₅ T ₂ ACA)	[216, 217]
	<i>Saccharomyces cerevisiae</i>	d(TG ₂₋₃ (TG) ₁₋₆)	[218, 219]
	<i>Candida glabrata</i>	d(CTGTG ₄ TCTG ₃ TG)	[215]
Plantes et algues	<i>Arabidopsis thaliana</i>	}	[220]
	<i>Silene latifolia</i>		
	<i>Chlamydomonas reinhardtii</i>	d(G ₃ T ₄ A)	[221]
	<i>Cyanidioschyzon merolae</i>	d(G ₆ A ₂ T)	[222]
	<i>Lycopersicon esculentum</i>	d(G ₃ T ₃ A) et d(G ₃ T ₂ A ₂)	[223]
Invertébrés	<i>Bombyx mori</i>	}	[224]
	et autres insectes		
	<i>Ascaris Lumbricoides</i>	}	[225-227]
	<i>Caenorhabditis elegans</i> et autres nématodes		
Vertébrés	<i>Homo sapiens</i> et autres vertébrés	d(G ₃ T ₂ A)	[228, 229]

Table 8. Motifs télomériques de différents organismes

unicellulaire *Chlamydomonas reinhardtii*. Il existe également des motifs à 2 guanines consécutives (motifs en G_2) : d(G_2T_2A) chez le ver à soie *Bombyx mori* et beaucoup d'autres insectes ; d(G_2CT_2A) chez *Ascaris lumbricoides*, *Caenorhabditis elegans* et d'autres nématodes. Les motifs à 4 guanines consécutives (motifs en G_4) se retrouvent chez les ciliés comme *Tetrahymena thermophila* (G_4T_2) et *Oxytricha nova* (G_4T_4). À la différence des autres eucaryotes, les levures ont des séquences télomériques très variées [215]: leurs motifs télomériques varient entre 8 et 26 nucléotides et leurs répétitions télomériques peuvent être régulières ou hétérogènes, selon les levures.

La longueur moyenne des télomères (région double-brin et extrémité 3' simple-brin) varie considérablement selon les organismes: de 50 à 150 kb chez la souris *Mus musculus*, de 5 à 15 kb chez l'homme, de 1 à 2 kb chez *Plasmodium falciparum*, et inférieure à 50 pb sur les minichromosomes de ciliés. Le nombre de répétitions des motifs télomériques varie également au sein d'une même espèce et même entre les différents chromosomes d'une même cellule. Chez les eucaryotes unicellulaires, les télomères sont maintenus à une longueur constante. Chez l'homme, les télomères sont plus longs dans les cellules germinales que dans les tissus somatiques, reflétant ainsi les différences dans leur processus de maintenance (cf. : §V.1.2).

Les télomères se terminent par une extrémité 3' simple-brin correspondant au brin riche en guanines (Figure 34B). L'extrémité 3' simple-brin sortante (en anglais “*overhang*”) a été observée chez plusieurs organismes présentés dans la table 8.

La présence de blocs de guanines consécutives rend l'ADN télomérique susceptible d'adopter des structures en G-quadruplexes, dans la région 3' simple-brin et également dans la région double-brin des télomères, impliquant, dans ce dernier cas, l'ouverture de la structure en double-hélice.

1.1.2 G-quadruplexes télomériques

En 1988, Sen et Gilbert ont montré la possibilité de formation de G-quadruplexes au sein des CSR (pour “*Class Switch Recombinations*”) des immunoglobulines, ils leur attribuent également un rôle potentiel au niveau des télomères, et notamment lors de la méiose [230]. En effet, une extrémité 3' simple-brin sortante et la présence de motifs répétés riches en blocs de guanines font, en théorie, des télomères des régions privilégiées pour la formation de G-quadruplexes.

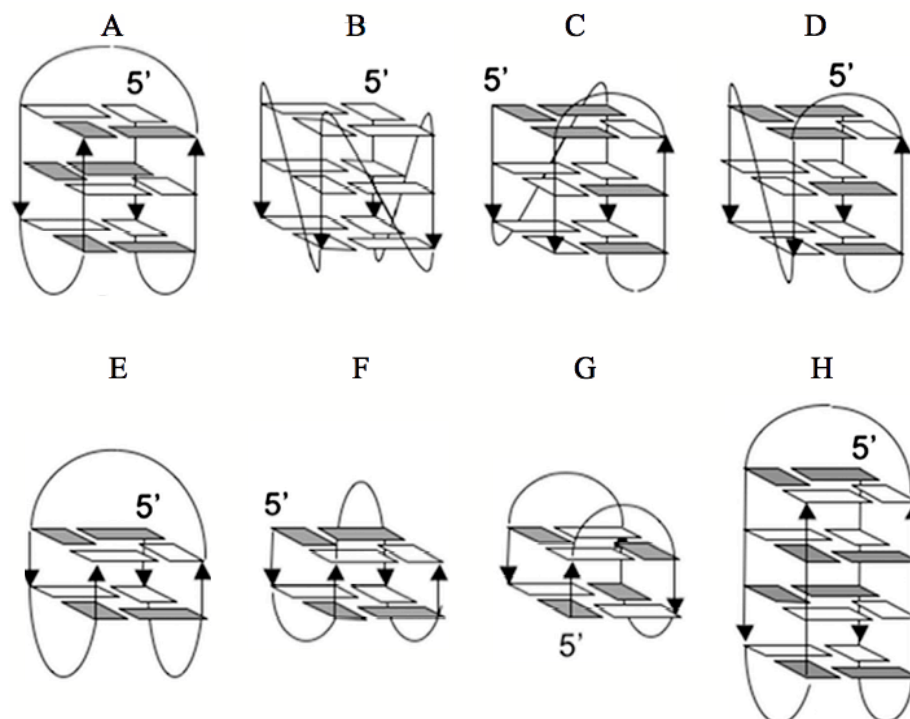


Figure 35. Structures G-quadruplexes intramoléculaires résolues à partir des motifs télomériques de quelques organismes

(A) Structure G4 formée par la séquence télomérique humaine dans Na^+ (résultat obtenu par RMN) [231]. (B-E) Structures G4 formées par la séquence télomérique humaine dans K^+ (B obtenu par cristallographie aux rayons X [232] et C-D obtenu par RMN [44, 233-238]). (B et F) Structures schématisées de G4 formées par les séquences télomériques de *Giardia intestinalis* [239] et (G) de *Bombyx mori* [240] dans K^+ . (D) G4 intramoléculaires formés par les séquences télomériques de *Tetrahymena thermophila* [43] et (H) de *Oxytricha nova* [241] dans Na^+ . Les guanines en conformation anti et syn sont représentées par des rectangles blancs et gris, respectivement.

Organisme[44]	Séquence télomérique	Condition de formation	Méthode	PDB	Référence
<i>Homo sapiens</i>	G4 “antiparallèle”	d[A(G ₃ T ₂ A) ₃ G ₃]	Na^+	RMN	143D [231]
	G4 “parallèle”	d[A(G ₃ T ₂ A) ₃ G ₃]	K^+	Rayon X	1KF1 [232]
	G4 “parallèle”	d[TA(G ₃ T ₂ A) ₃ G ₃]	K^+ + PEG ₂₀₀	RMN	2LD8 [242]
	G4 “hybride-1”	d[TA(G ₃ T ₂ A) ₃ G ₃]	K^+	RMN	2JSM [237]
	G4 “hybride-2”	d[TA(G ₃ T ₂ A) ₃ G ₃ T ₂]	K^+	RMN	2JSL [237]
	G4 “forme 3”	d[(G ₃ TTA) ₃ G ₃ T]	K^+	RMN	2KF8 [44]
<i>Giardia intestinalis</i>	d(TAG ₃) ₄ (2 quartets)	K^+	RMN	2KOW	[239]
<i>Tetrahymena thermophila</i>	d(T ₂ G ₄) ₄	Na^+	RMN	201D	[43]
<i>Oxytricha nova</i>	d[(G ₄ T ₄) ₃ G ₄]	Na^+	RMN	186D	[241]

Table 9. Etudes structurales menées sur les séquences télomériques de différents organismes

La formation de G4 bimoléculaires à partir des courtes extrémités 3' sortantes (16 nucléotides) des télomères de ciliés *O.nova* ou *S.lemnae* a été démontrée par cristallographie [243] et par RMN [244]. Leurs structures présentent des brins en orientation antiparallèle 2 à 2 avec des boucles diagonales (Figure 5B). L'existence de ces structures a d'ailleurs été mise en évidence *in vivo*, à l'aide d'anticorps dirigés contre ce type de G-quadruplexes (cf. : §I.2.4) [98, 245]. Il a été suggéré que ces G-quadruplexes intermoléculaires pourraient être impliqués dans l'appariement des chromosomes des macronoyaux (Figure 7A) et dans la protection ou la structure des télomères des ciliés.

Chez l'Homme, la présence d'extrémités comportant 100 à 200 nucléotides rend possible la formation de G4 intramoléculaires, qui nécessitent la présence d'au moins 4 blocs de guanines.

La première structure d'un G-quadruplexe intramoléculaire pour la séquence télomérique humaine (d[A(GGGTTA)₃GGG]) a été résolue, en 1993, par RMN par Wang et Patel en présence d'ions sodium (Figure 35A et table 9) [231]. Cette structure présente une orientation antiparallèle des brins, reliés par des boucles successivement latérale-diagonale-latérale.

La détermination de structure G-quadruplexe de la même séquence en potassium, plus pertinent du point de vue de la composition du milieu intracellulaire, a été plus complexe dû en raison de l'important polymorphisme des G4 formés par cette séquence en potassium. En 2002, Gary Parkinson *et al.* sont parvenus à cristalliser et à résoudre la structure en G4 formée par cette séquence en potassium. La structure obtenue présente une orientation parallèle des brins, reliés par des boucles en "forme de N". Cette structure s'avère cependant assez controversée, notamment à cause des conditions de cristallisation qui semblent influencer le repliement du G-quadruplexe (Figure 35B et table 9), notamment la présence de polyéthylène glycol et d'un ligand de G4, le BRACO19 [232]. Il est d'ailleurs à noter qu'une structure similaire à celle obtenue par Parkinson *et al.* a été récemment résolue par RMN en solution en présence de potassium et de polyéthylène glycol [242].

En modifiant légèrement les extrémités de la séquence télomérique humaine, des nouvelles structures G4 ont été obtenues par RMN. Les structures observées présentent 3 brins parallèles et 1 brin antiparallèle, appelées généralement des conformations "hybrides" (ou (3+1)). Il existe deux formes d'"hybrides": "hybride-1" est obtenu majoritairement à partir de la séquence d[TA(GGGTTA)₃GGG] dans K⁺ et présente successivement une boucle

en “forme de N” et 2 boucles latérales (Figure 35C et table 9) [233, 235-237], tandis que la séquence d[TA(GGGTTA)₃GGGTT] en potassium permet d’observer essentiellement la forme d’“hybride-2”, contenant successivement 2 boucles latérales et une boucle en “forme de N” (Figure 35D et table 9) [234, 237, 238]. En présence de K⁺, les G4 télomériques humains sont en équilibre dynamique entre 2 conformations, “hybride-1” et “hybride-2”. La modification des extrémités de la séquence télomérique suffit à faire basculer la balance en faveur d’une des 2 structures (pour revue voir [246]). En 2009, A.T. Phan a obtenu une nouvelle structure G4 intramoléculaire (“forme 3”) formé par la séquence télomérique humaine d[(GGGTTA)₃GGGT] dans K⁺ [44]. Cette structure antiparallèle en panier ne comporte que deux G-quartets (Figure 35E et table 9), mais apparaît plus stable que les deux formes “hybrides” présentées précédemment.

Les conformations “hybrides” et “parallèles” observées en présence d’ion K⁺ permettent notamment d’imaginer l’empilement de plusieurs G-quadruplexes tout au long de l’extrémité 3’ sortante (Figure 36A-C). Dès 2006, N. Sugimoto a caractérisé la structure et la stabilité des G4 multiples formés par des séquences télomériques longues de l’homme (d(TTAGGG)_n où n = 4 à 12) [247]. Et en 2009, M. Komiyama a clairement observé en AFM la formation *in vitro* de G4 consécutives avec la séquence télomérique longue (d(GGGTTA)₁₆) [248] (Figure 36D).

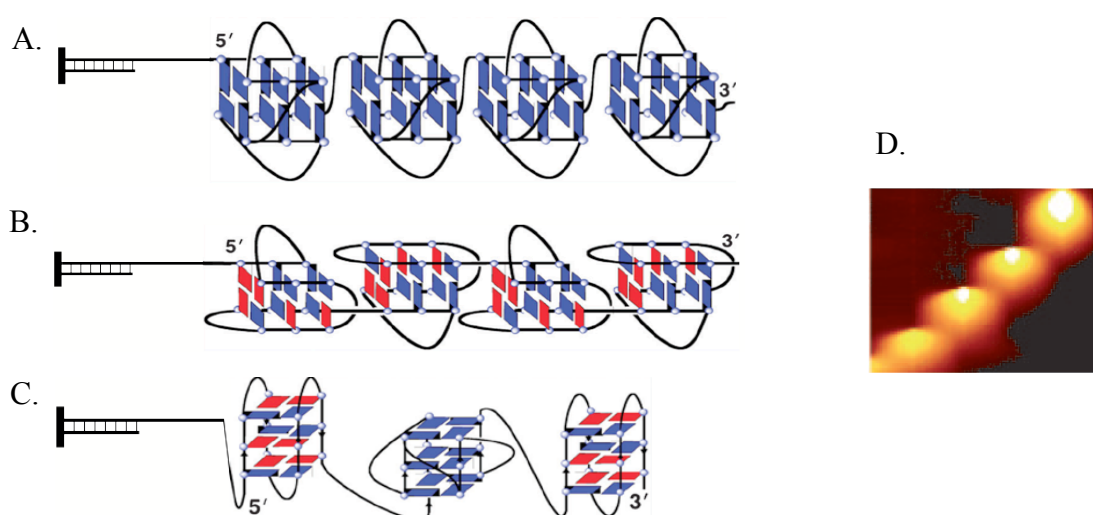


Figure 36. Enchaînement hypothétique de G-quadruplexes intramoléculaires à l’extrémité télomérique 3’ sortante humaine
Empilement de G4 “parallèles” (A), “hybrides” (B) ou “mixtes” (C) (revue [249]). (D) Formation de G4 consécutives par la séquence télomérique longue d(GGGTTA)₁₆ (observé en AFM) [248].

Les structures d'autres G4 télomériques intramoléculaires ont été résolues par RMN (Figure 35 et table 9) : la structure formée par le motif télomérique de nombreux insectes (GGTTA), celles formées par le motif de *Giardia intestinalis* (GGGTA) et celles formées par les motifs des ciliés *Tetrahymena thermophila* (GGGGTT) et *Oxytricha nova* (GGGGTTTT). Ces structures illustrent le polymorphisme de ces édifices nucléiques.

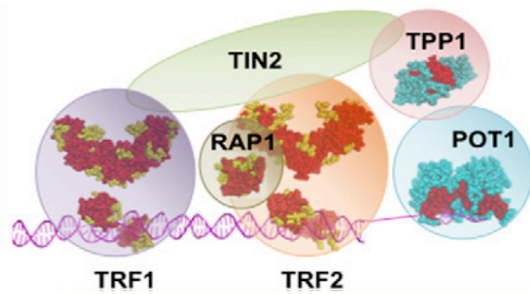
Nous avons vu précédemment que la conformation des G4 varie énormément avec la séquence primaire. L'ajout, l'enlèvement ou la substitution d'un nucléotide suffisent souvent pour changer complètement la structure du G4 formé. Certes, un certain nombre de structures de G4 télomérique ont été résolues, néanmoins, les données sur la structure et stabilité des séquences télomériques des différentes eucaryotes demeurent encore incomplètes et dispersées. De plus, la formation de G4 intramoléculaires par des séquences télomériques de levures n'est pas encore été systématiquement évaluée. Dans l'**article 3** présenté dans ce chapitre, nous allons présenter une étude comparative de la structure et de la stabilité des séquences télomériques de différents organismes, y compris des levures (cf.: §V.2 et §V.3).

Bien que la formation *in vitro* de G4 par des séquences télomériques est indiscutable, l'existence de ces structures au niveau des télomères reste encore à démontrer. Si l'existence de G4 est bien argumentée chez les ciliés [98], leur existence aux télomères des cellules humaines reste souvent contestée. Des résultats obtenus en présence de ligands de G-quadruplexes supportent, de façon indirecte, la possibilité de formation de G-quadruplexes aux télomères des cellules humaines [250]. De plus, des résultats obtenus sur des cellules humaines suggèreraient que la formation de telles structures peut perturber la réplication des télomères [251, 252]. En revanche, une étude conduite sur la levure *S. cerevisiae* indiquerait, de façon indirecte, que la formation de G4 protégerait les télomères dans certains mutants de cet organisme [253].

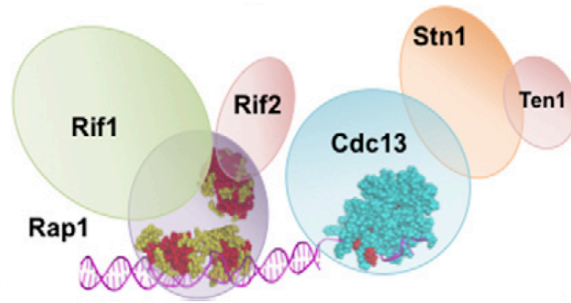
1.1.3 T-loop

La *T-loop*, pour boucle télomérique (en anglais "*telomeric-loop*") est observée pour la première fois en 1999 [254]. Cette structure a été mise en évidence par pontage au psoralène de cellules humaines (normales et tumorales) et de souris, et visualisé au microscope électronique. La *T-loop* est formée par le repliement du télomère sur lui-même et l'invasion de l'extrémité 3' simple-brin riche en guanine dans la région double-brin, formant ainsi une boucle terminale (Figure 34B). La formation de cette conformation a été montrée chez

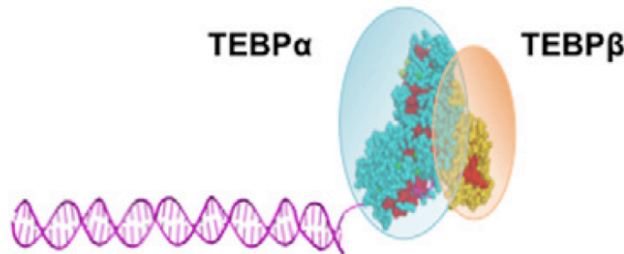
A. Complexe "Shelterin" (Vertébrés)



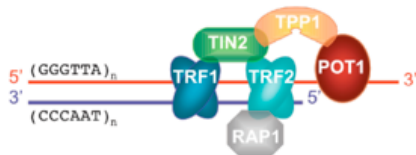
B. Complexe CST (*S. cerevisiae*)



C. Complexe télomérique (*O. nova*)



D. état "ouvert" ou "décapuchonné"



état "fermé" ou T-loop

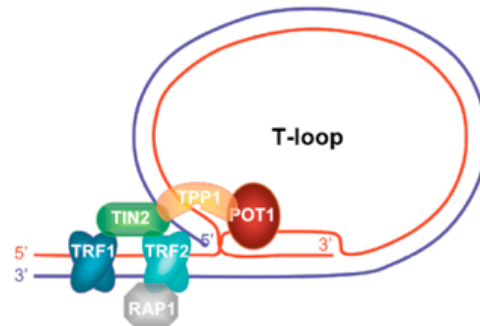


Figure 37. Complexes protéiques des télomères permettant leur protection

(A) Six protéines du complexe "Shelterin" trouvées chez les vertébrés ; (B) trois protéines du complexe CST (Cdc13-Stn1-Ten1) observées chez *S. cerevisiae* et (C) deux protéines télomériques des ciliés *O. nova* (Images extraites de [259]). (D) Modèles de régulation de la longueur et la protection du télomère en fonction de l'état du complexe "Shelterin": états "ouvert" et "fermé" (images de la thèse d'Anne De Cian). L'état "fermé" ou T-loop est un état dans lequel l'extrémité 3' sortante n'est pas accessible aux réactions enzymatiques de dégradation, de réparation ou d'élongation par la télomérase (cf. : §V.1.2.1 a). L'état "ouvert" est un état dans lequel les protéines du télosome sont présentes et protègent le télomère tout en laissant l'accès de la télomérase à l'extrémité 3' simple-brin. Cet état peut être présent durant les étapes d'élongation et de réplication du télomère (cf. : §V.1.2). L'état "décapuchonné" est un état où certaines protéines du complexe sont altérées, ce qui conduit à une dysfonction du télomère, entraînant une réponse aux dommages de l'ADN liée à la reconnaissance de cassures double-brins et la formation de TIFs ("dysfunctional Telomere-Induced Foci").

d'autres espèces telles que le trypanosome [255], le petit pois [256], les chromosomes polytènes de ciliés [257] et les cellules de poulet et de souris [258]. Ce modèle de la structure télomérique de certains organismes permettrait de masquer les extrémités simple-brins du télomère et contribuerait ainsi à sa protection.

1.1.4 G-quadruplexes et T-loop

La formation de G-quadruplexes et de *T-loop* a été proposée dans la même région, au niveau des télomères. Cela a conduit D. Patel [260, 261] et H. Sugiyama [262, 263] à s'intéresser à un possible modèle combinant les deux structures (Figure 34B). En 2005, D. Patel et son équipe ont suggéré que la formation de G-quadruplexe intermoléculaire en sodium pourrait participer à la formation de *T-loop*. En 2008, l'équipe de H. Sugiyama a mis en évidence par des méthodes biophysiques et par électrophorèse que la structure de G4 peut stabiliser une structure comme la *T-loop*.

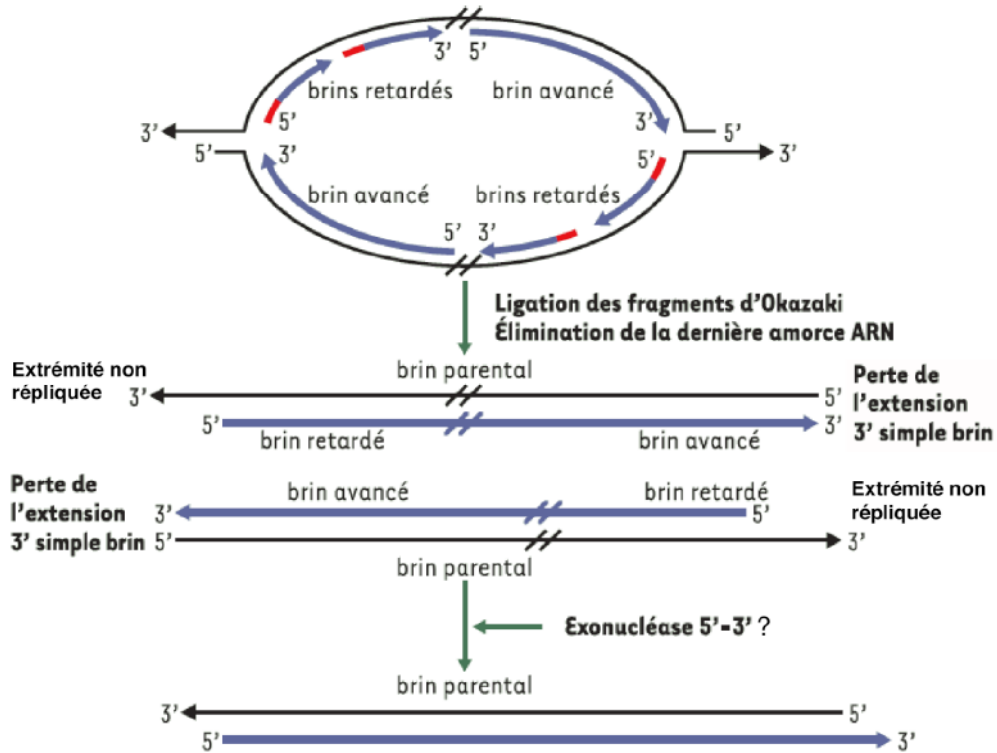
Comme mentionnée au début de ce chapitre, la structure "coiffante" des télomères est une structure nucléoprotéique, et donc formée d'ADN et de protéines. En effet, la formation de la *T-loop* décrite précédemment nécessite la participation de protéines télomériques, dites protéines du "télosome" [264] ou complexe "*Shelterin*" (pour protection) [265] (paragraphe suivant).

1.1.5 Complexe "*Shelterin*" ou protéines du "télosome"

Chez les mammifères, le complexe "*Shelterin*" est composé de six protéines, dont trois se lient directement à l'ADN télomérique (TRF1, TRF2 et POT1) (Figure 37A) et les trois autres (TIN2, TPP1 et RAP1) servent de connexion entre protéines.

TRF1 et TRF2 (TRF = "*Telomeric Repeat-Binding Factors*") se lient directement à la région double-brin de la séquence télomérique. TRF2 a été étudiée intensivement en particulier pour son rôle essentiel au sein du complexe "*Shelterin*". Il a été montré que la présence de TRF2 était suffisante pour la formation de *T-loop in vitro* [254, 266]. La protéine TRF2 pourrait également reconnaître *in vitro* les structures G4 et stimulerait un changement de conformation de ces structures [267]. Contrairement à TRF1 et TRF2, POT1 ("*Protection Of Telomeres 1*") se lie à l'extrémité 3' simple-brin du télomère et perturberait donc la formation de G4 dans cette région [91].

A.



B.

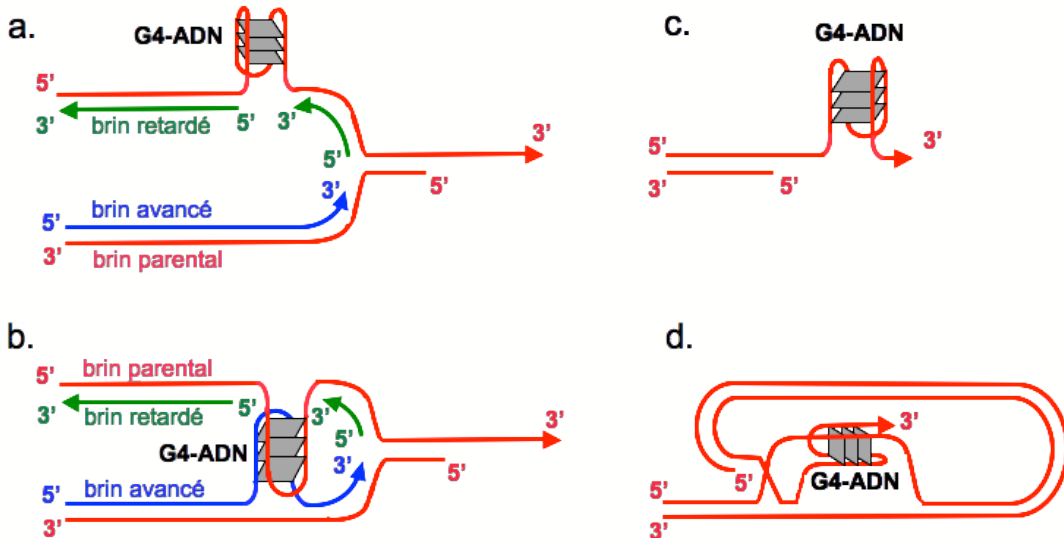


Figure 38. Problème de réplication télomérique

(A) Le problème de réplication terminale se situe au niveau du brin retardé ("lagging strand"), encore appelé brin discontinu. L'élimination de la dernière amorce ARN en 5' (qui n'est pas remplacée) génère une extrémité 3' sortante, mais ceci conduit également à un raccourcissement progressif des télomères au cours des divisions cellulaires, en absence d'une machinerie spécifique. Par ailleurs une activité exonucléase 5'-3' pourrait expliquer la génération d'extrémités 3' sortantes au niveau du brin continu, ou avancé ("leading strand").[269](B) Potentiel de formation de G4 au niveau des télomères. (a) G4 formé par l'ADN simple brin riche en G (entre les fragments d'Okazaki) pourrait causer un raccourcissement des répétitions télomériques; (b) G4 formé entre le brin continu et un brin parental pourrait bloquer la réplication; (c) G4 formé à l'extrémité 3' sortante riche en G pourrait inhiber l'extension du télomère par la télomérase; (d) G4 formé dans la T-loop pourrait empêcher son ouverture.

TIN2 (“*TRF2 and TRF1-Interacting Nuclear Protein 2*”) lie TRF1 et TRF2 au complexe TPP1-POT1, reliant ainsi la région double-brin à la partie simple-brin du télomère. En liant TRF1, TRF2 et TPP1, TIN2 occupe une position centrale du complexe “*Shelterin*”. Elle permet d’augmenter l’association de TRF1 à l’ADN télomérique [268].

RAP1 humain (“*Repressor-Activator Protein 1*”) ne se lie pas directement à l’ADN, elle est liée au télomère par l’intermédiaire de TRF2. Cependant, RAP1 de levure pourrait se lier à l’ADN télomérique afin de permettre la formation de G4 parallèle [270] (Figure 37B). Dans les cellules humaines, le complexe TRF2-RAP1 serait à l’origine de l’inhibition de la réparation par NHEJ (“*Non-Homologous End Joining*”) aux télomères [271], protégeant ainsi les télomères des mécanismes de réparation des cassures double-brins.

TPP1 (“*TIN2 and POT1 interacting Protein 1*”), comme son nom indique, interagit avec TIN2 et POT1. L’interaction entre TPP1 et POT1 est cruciale pour le recrutement de POT1 aux télomères [272, 273]. Des résultats publiés en 2007 ont également suggéré un rôle important de TPP1 dans le recrutement de la télomérase aux télomères [274].

Ce complexe “*Shelterin*”, composé seulement de 6 protéines, présente un rôle important dans la régulation et la protection télomérique. Il a été montré que l’interférence fonctionnelle de chacune de ces protéines (par ARNi ou expression d’un dominant négatif) affectait la taille ou la protection des télomères, démontrant ainsi leur rôle dans la maintenance télomérique [265]. Titia de Lange a proposé, en 2005, un modèle de régulation de la longueur et la protection du télomère en fonction de l’état du complexe “*Shelterin*” avec l’ADN télomérique (état “fermé”, “ouvert” ou “décapuchonné”) (Figure 37D) [265].

Les télomères des levure bourgeonnantes (“*budding yeast*”), tels que *S. cerevisiae*, sont protégés par un complexe trimérique nommé CST (formé par Cdc13, Stn1 and Ten1) qui s’associe à l’extrémité 3’ simple-brin [275, 276](Figure 37B). À la différence du *S. cerevisiae*, la levure fissipare (“*fission yeast*”) *S. pombe* présente un complexe de six protéines du type “*Shelterin*”, qui est composé, entre autre, d’orthologues de TRF1/2 et de POT1. Enfin, les télomères des ciliés ne contiennent apparemment que deux protéines dans leur complexe “*Shelterin*”, appelées TEBP (“*Telomere End Binding Proteins*”) sont protégés par un hétérodimère formé par deux protéines : TEBP α (homologue de POT1) et TEBP β (qui présente une certaine similitude de séquence et de structure avec TPP1) (Figure 37C) [277].

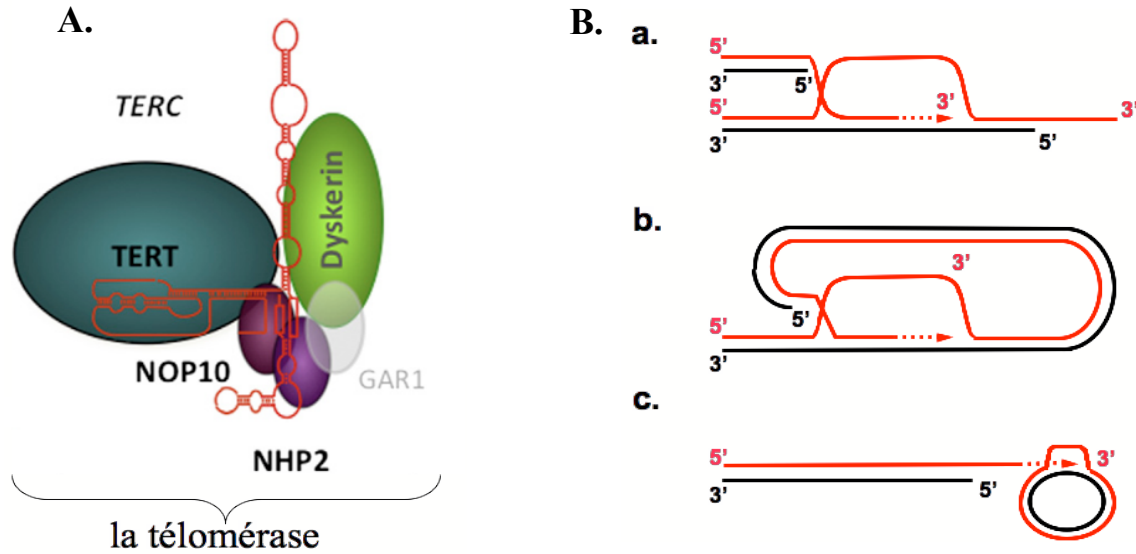


Figure 39. Mécanismes de maintenance télomérique

(A) Schéma représentatif de la télomérase, une ribonucléoprotéine [280]. (B) Modèles alternatifs d'allongement des télomères (ALT): (a) par recombinaison homologue inter-télomérique, (b) par synthèse sur la T-loop et (c) par synthèse sur les cercles télomériques (pour revue voir [281]).

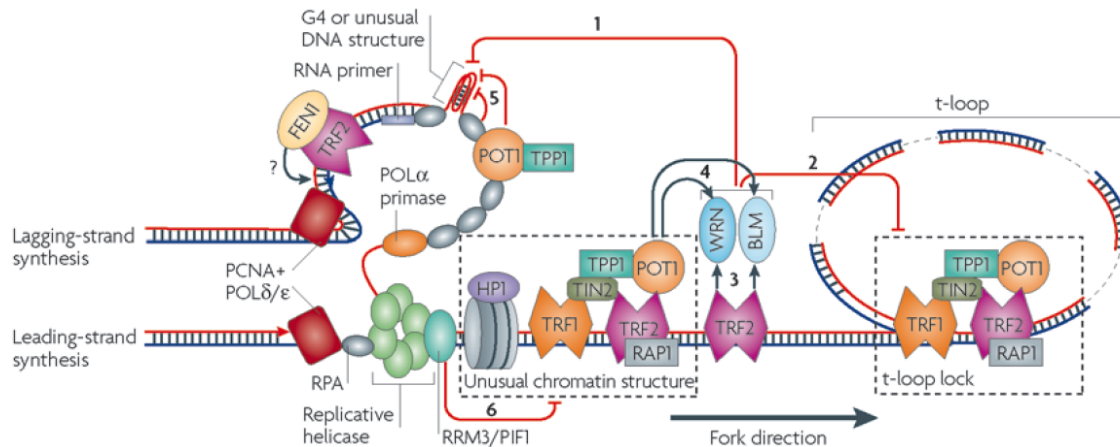


Figure 40. Modèle de progression de la fourche de réplication à l'extrémité des chromosomes de mammifères [282]

Ce modèle décrit le rôle des différentes activités permettant de se débarrasser des structures secondaires pouvant se former durant la réplication du télomère, particulièrement au niveau du brin discontinu riche en guanines. On y observe les rôles de WRN pour dissocier de telles structures (G-quadruplexe (flèche 1) et T-loop (flèche 2)), et leur stimulation par TRF2 et POT1 (flèches 3 et 4). POT1 et RPA (Replication Protein A) sont également susceptibles d'inhiber la formation de ces structures par leur liaison à l'ADN simple-brin (flèche 5). Enfin, d'après des données obtenues sur les levures, les hélicases RRM3 et PIF1 permettraient de faciliter la progression de la fourche de réplication au travers de complexes nucléoprotéiques inhabituels (flèche 6) [283]. Ces hélicases pourraient être associée à la fourche de réplication. HP1 : hétérochromatine protéine-1 ; PCNA : "Proliferating Cell Nuclear Antigen" ; POL δ/ϵ : ADN polymérases- δ/ϵ ; FEN1 : "5' flap endonuclease-1" (dégradation des fragments d'Okazaki).

1.2 Réplication télomérique et protéines impliquées

En 1961, Leonard Hayflick montrait que des cellules humaines issues de tissus embryonnaires avaient, en culture, un potentiel de réplication limité, d'environ 50 divisions (limite de Hayflick) [278]. Au début des années 1970, suite à la découverte du mécanisme de réplication de l'ADN et du "problème de la réplication télomérique" décrite par James Watson [279], A. Olovnikov proposa un mécanisme expliquant la réplication limitée des cellules par érosion des télomères à chaque division cellulaire (Figure 38A) [284].

En 1990, C. Harley montra que la taille moyenne des télomères diminue effectivement avec le nombre de divisions cellulaires dans les fibroblastes humains [285]. Il relia les télomères au vieillissement cellulaire et parla d'horloge mitotique.

Dans les cellules somatiques normales, les télomères raccourcissent à chaque division jusqu'à atteindre une taille critique qui entraîne un arrêt du cycle cellulaire et la sénescence ou la mort des cellules. Dans certaines cellules, notamment les cellules souches, germinales ou tumorales, des mécanismes de maintenance spécifiques sont mis en place afin d'allonger les télomères au cours du cycle cellulaire, et donc de prolonger la vie des cellules.

Par ailleurs, les télomères présentent d'autres obstacles potentiels à la réplication à l'ADN. Nous avons mentionné au début de ce chapitre les structures secondaires susceptibles de se former aux télomères, notamment les G-quadruplexes. La formation de G4 est également favorisée durant la réplication de l'ADN (Figure 38B). Une telle structure G4 pourrait bloquer la réplication, en particulier la réplication télomérique, et inhiber l'extension des télomères par les mécanismes de maintenance télomérique.

Nous allons vous décrire, ci-dessous, les mécanismes de maintenance permettant l'élongation télomérique et les protéines impliquées dans la réplication fidèle des télomères.

1.2.1. Extension télomérique par des mécanismes de maintenance spécifiques

a) Télomérase

Un des mécanismes d'élongation télomérique est réalisé grâce à une transcriptase inverse appelé télomérase. L'enzyme est identifiée pour la première fois en 1985 dans le cilié *Tetrahymena thermophila* par Elisabeth Blackburn et Carol Greider [286]. La télomérase compense le raccourcissement des télomères en ajoutant des répétitions télomériques à l'extrémité 3' simple-brin. L'ajout des répétitions est réalisé par l'intermédiaire d'un ARN matrice qui est rétro-transcrit. Depuis sa découverte, l'existence d'un tel mécanisme a été

mise en évidence dans un certain nombre d'organismes allant des ciliés aux humains en passant par les levures et les plantes (pour revue voir [287]).

Chez l'homme, la télomérase est un complexe enzymatique composé d'une sous-unité catalytique (hTERT, de 127 kDa), d'une matrice ARN (hTERC, de 451 nucléotides) et d'un complexe de 4 protéines associé à l'ARN (dyskérine, NHP2, NOP10 et GAR1) (Figure 39A) (pour revue voir [280]).

Dans les cellules présentant une activité télomérase importante, la taille des télomères est maintenue globalement constante. On parle d'homéostasie de la longueur des télomères. Nous remarquerons que pour les cellules cancéreuses, la taille moyenne maintenue est relativement courte, de 2-10 kb. Afin de maintenir cette longueur, la structure du télomère et l'accès de la télomérase à son substrat doivent être finement régulés. Le modèle du complexe "Shelterin" et de *T-loop* ouverte ou fermée décrite précédemment permet d'expliquer en partie cette régulation. Nous avons également mentionné que certaines protéines télomériques pourraient participer au recrutement de la télomérase, et donc à la régulation de ce mécanisme, notamment POT1 et TPP1. Cependant, ce mécanisme de maintenance est complexe et nécessite d'autres régulations qui restent encore assez mal connues.

La télomérase est active dans 85 à 90% des cellules tumorales. Il existe au moins un autre mécanisme de maintenance de la longueur des télomères, appelé ALT (pour "Alternative Lengthening of Telomeres").

b) Mécanisme ALT ("Alternative Lengthening of Telomeres")

Dans une minorité de tumeurs (10-15%), les cellules maintiennent leurs télomères par un mécanisme indépendant de la télomérase, appelé ALT ("Alternative Lengthening of Telomeres") [288-290].

Le mécanisme ALT est décrit pour la première fois chez les levures [291]. La première étude montrant l'existence d'un tel mécanisme dans des cellules humaines immortalisées est publiée en 1994 [292]. Les cellules utilisant la voie ALT sont caractérisées par des télomères très longs et hétérogènes (leur longueur varie entre 2 et 50 kb) [288, 293] (pour revue voir [294]).

Ce mécanisme de maintien de la longueur télomérique implique des événements de recombinaison inter- ou intra-télomérique. Plusieurs modèles de maintenance télomérique sont proposés. L'allongement des télomères dans ces cellules ALT pourrait se produire par

recombinaison homologue entre deux télomères ou au sein d'un même télomère en utilisant la T-loop ou des cercles télomériques (Figure 39B).

1.2.2. Protéines permettant une réplication fidèle des télomères

Comme mentionné précédemment, les télomères sont susceptibles de former des structures secondaires, notamment les G-quadruplexes ou la *T-loop*. Sans intervention des protéines, ces structures pourraient perturber la réplication télomérique.

La famille d'hélicases RecQ (notamment WRN et BLM) semble jouer un rôle clé dans la réplication des télomères. Des études biochimiques ont clairement mis en évidence la capacité des hélicases RecQ à faciliter la progression de la fourche de réplication. *In vitro*, il a été montré que TRF2 et POT1 pouvaient stimuler WRN et BLM afin de libérer la *T-loop* et de permettre ainsi l'avancée de la réplication jusqu'aux extrémités des chromosomes [295-298]. En plus de la *T-loop*, les hélicases RecQ peuvent dérouler les structures G4 [83, 85, 299, 300]. Il est donc possible d'imaginer que les protéines télomériques recrutent les hélicases RecQ au niveau des télomères afin de dérouler les structures secondaires (G-quadruplexes ou *T-loop*) et permettent ainsi la progression de la réplication. Par ailleurs, certaines protéines télomériques et POT1 peuvent aussi entraîner la déstabilisation des G4 en se fixant sur l'ADN simple-brin [91, 245]. Un modèle de progression de la fourche de réplication à l'extrémité des chromosomes de mammifères a été proposé par Eric Gilson et Vincent Geli en 2007 et présenté dans la figure 40.

Le recrutement de protéines de réponses aux dommages ("*DNA damage checkpoint*") et les effecteurs de la réparation de l'ADN au niveau des télomères sont également essentiels pour assurer une réplication complète et fidèle des télomères. Le recrutement de ces facteurs serait une première étape permettant la reprotection du télomère après l'ouverture de la *T-loop* et sa réplication [301, 302].

Enfin, nous pouvons citer un rôle potentiel de la protéine WRN dans la répression des recombinaisons inter-télomériques. En effet, la déficience en télomérase et WRN dans un modèle murin induit de nombreux échanges par recombinaisons entre chromatides soeurs, conduisant à l'immortalisation des cellules et l'engagement dans un mécanisme de maintenance ALT. Le rétablissement de WRN élimine ces recombinaisons, suggérant son rôle dans la répression des échanges télomériques entre chromatides sœurs [303]. De plus, BLM et WRN facilitent *in vitro* la séparation des jonctions Holliday formées lors de la recombinaison homologue [304-306].

1.3 Dysfonctionnement télomérique et maladies humaines associées

L'absence des mécanismes de protection et de maintenance de la longueur des télomères entraîne un dysfonctionnement télomérique et à un raccourcissement des télomères au fil des divisions cellulaires. Plusieurs maladies humaines sont associées à un raccourcissement télomérique telles que les maladies cardiaques, la recto-colite ulcéro-hémorragique, les cirrhoses hépatiques, l'athérosclérose et les syndromes de vieillissement précoce [307, 308]. La dyskératose congénitale est la première maladie identifiée. Les patients atteints de la dyskératose congénitale présentent des télomères raccourcis résultant de la mutation de constituants de la télomérase, en particulier hTERC [309, 310]. L'anémie aplasique a été reportée comme une maladie associée à des mutations des composants du complexe "*Shelterin*" [311, 312]. Récemment, un raccourcissement télomérique accéléré a été trouvé dans quelques cas de fibrose pulmonaire idiopathique [313]. De plus, il a été démontré que dans les deux syndromes du vieillissement précoce (Werner et Bloom), un déficit en hélicases WRN et BLM cause un raccourcissement accéléré des télomères et mène à l'apparition prématurée de cancer (voir revue [314]).

La protection et la maintenance des télomères jouent donc un rôle crucial dans la survie cellulaire. Une bonne compréhension des structures de l'ADN télomérique de différents modèles d'études (levures, ciliées, plantes, vertébrés, etc ...) sera précieuse pour le progrès de la recherche appliquée dans le domaine. Nous avons tenu à présenter dans l'**article 3** une étude comparative de la structure et de la stabilité des séquences télomériques de différents organismes allant des protozoaires aux vertébrés en passant par les levures, les plantes et les invertébrés.

2) Article 3: “Stability of telomeric G-quadruplexes”. *Nucleic Acids Research* - 2011; 39(8): 3282-94.

Tran, P.L.T., Mergny, J.L. & Alberti, P.

Résumé en français :

Chez la plupart des eucaryotes, l'ADN télomérique est constitué d'un court motif répété en tandem avec des guanines consécutives qui pourrait adopter une structure en G-quadruplexe. Les télomères des levures bourgeonnantes présentent de longs motifs répétés et montrent une variation du degré de répétition. Si la formation de G4 *in vitro* par des séquences télomériques de plusieurs organismes a déjà été montrée, il n'existe pas d'étude consacrée à la comparaison de différentes conformations et stabilité des G4 formés par les séquences télomériques connues. De plus, à notre connaissance, la formation de G4 intramoléculaires par des séquences télomériques de levures n'a pas encore été systématiquement évaluée. En utilisant des méthodes biophysiques et biochimiques, nous avons étudié des séquences synthétiques mimant les motifs télomériques de différents organismes. Ces séquences comportent 4 blocs de guanines consécutives permettant la formation des G4 intramoléculaires. Afin d'étudier et comparer le potentiel G4 des motifs télomériques connus à ce jour, nous avons mené une étude de la structure et stabilité de séquences télomériques de différents eucaryotes. Nous avons remarqué que la formation de G4 au niveau des télomères n'était pas une caractéristique conservée chez les levures. Cependant, toutes les séquences télomériques connues des autres organismes eucaryotes sont susceptibles de former des quadruplexes. Parmi ceux qui forment des G4, des motifs G₃T₁₋₄A (trouvés chez la plupart des organismes) et G₄T_{2,4} (trouvés chez les ciliées) forme des G4 stables, tandis que les G4 formés par les motifs G₂T₂A et G₂CT₂A (trouvés chez beaucoup d'insectes et chez les nématodes, respectivement) semblent être en équilibre avec d'autres structures non-G4 (comme des structures en épingle à cheveux).

Article 3

Stability of telomeric G-quadruplexes

Phong Lan Thao Tran^{1,2}, Jean-Louis Mergny^{1,2,*} and Patrizia Alberti^{1,*}

¹INSERM, U565, Acides nucléiques: dynamique, ciblage et fonctions biologiques, Muséum National d'Histoire Naturelle (MNHN), CNRS, UMR7196, Département de "Régulations, développement et diversité moléculaire", 43 rue Cuvier, CP26, Paris Cedex 5, F-75231 and ²INSERM, U869, ARNA Laboratory, European Institute of Chemistry and Biology, University of Bordeaux, 2 rue Robert Escarpit, Pessac, F-33607, France

Received July 13, 2010; Revised December 1, 2010; Accepted December 2, 2010

ABSTRACT

In most eukaryotes, telomeric DNA consists of repeats of a short motif that includes consecutive guanines and may hence fold into G-quadruplexes. Budding yeasts have telomeres composed of longer repeats and show variation in the degree of repeat homogeneity. Although telomeric sequences from several organisms have been shown to fold into G-quadruplexes *in vitro*, surprisingly, no study has been dedicated to the comparison of G-quadruplex folding and stability of known telomeric sequences. Furthermore, to our knowledge, folding of yeast telomeric sequences into intramolecular G-quadruplexes has never been investigated. Using biophysical and biochemical methods, we studied sequences mimicking about four repetitions of telomeric motifs from a variety of organisms, including yeasts, with the aim of comparing the G-quadruplex folding potential of telomeric sequences among eukaryotes. G-quadruplex folding did not appear to be a conserved feature among yeast telomeric sequences. By contrast, all known telomeric sequences from eukaryotes other than yeasts folded into G-quadruplexes. Nevertheless, while G₃T₁₋₄A repeats (found in a variety of organisms) and G₄T_{2,4} repeats (found in ciliates) folded into stable G-quadruplexes, G-quadruplexes formed by repetitions of G₂T₂A and G₂CT₂A motifs (found in many insects and in nematodes, respectively) appeared to be in equilibrium with non-G-quadruplex structures (likely hairpin-duplexes).

INTRODUCTION

Telomeres are specialized nucleoprotein complexes that cap and protect the extremities of linear eukaryotic

chromosomes. In most eukaryotes telomeric DNA consists of a tandem array of a short motif of 5–8 nt, that includes two, three or four consecutive guanines, with an impressive prevalence of the G₃T₂A motif (1).

The G₃T₂A motif is found in many phylogenetically distant organisms, including vertebrates (2,3), *Trypanosoma brucei* (4,5), several fungi (6), slime moulds (7,8) and several aquatic invertebrates (9). Variants of this motif are found in many other organisms: G₃TA in the parasite *Giardia intestinalis* (10); G₃T₃A in a variety of plants (11), such as *Arabidopsis thaliana* (12); G₃T₄A in the unicellular green alga *Chlamydomonas reinhardtii* (13); G₂T₂A in the silkworm *Bombyx mori* and in many other insects (14); G₂CT₂A in the nematodes *Ascaris lumbricoides* (15), *Caenorhabditis elegans* and in other nematodes (16,17). Telomeric motifs bearing four consecutive guanines are found in ciliates: G₄T₂ in holotrichous ciliates such as *Tetrahymena* (18) and G₄T₄ in hypotrichous ciliates such as *Oxytricha*, *Euplotes* and *Stylonychia* (19). Degenerated telomeric motifs have been reported for several organisms: GGGTTTA and GGGTTCA in *Plasmodium* species (20,21); GGGTTTA and GGGTTAA in *Lycopersicon esculentum* (tomato plant) (22); GGGTT and GGGTTT in *Paramecium* species (23,24), where a single telomerase RNA, encoding G₄T₂ repeats, has been identified (25).

Unlike in most eukaryotes, telomeres have rapidly evolved among budding yeasts; these organisms possess the most variable set of telomeric repeats (26,27). A number of budding yeast species have telomeres composed of long repeats (8–26 bp) and show variations in the degree of repeat homogeneity (28,29). The budding yeast *Saccharomyces cerevisiae* has short heterogeneous telomeric repeats of the general form TG₂₋₃(TG)₁₋₆ (30,31), encoded by a single telomerase RNA (32); the most abundant type of repeat is TG₃TG (29). G₃T₂A repeats are found in the immediate subtelomeric region of *S. cerevisiae*, suggesting that G₃T₂A-like telomeres were present prior to yeast divergence (26). Despite

*To whom correspondence should be addressed. Tel: +33 1 40 79 37 27; Fax: +33 1 40 79 37 05; Email: alberti@mnhn.fr
Correspondence may also be addressed to Jean-Louis Mergny. Tel: +33 5 4000 30 22; Fax: +33 5 4000 30 04; Email: jean-louis.mergny@inserm.fr

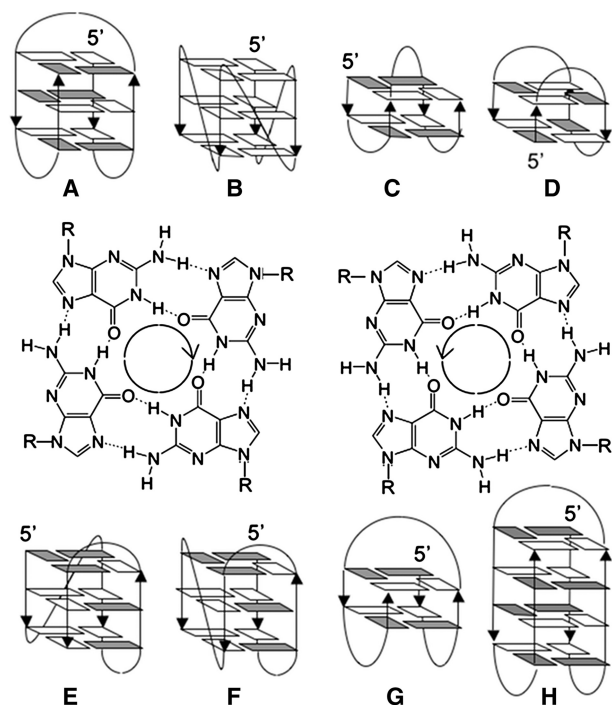


Figure 1. Schematic structures of intramolecular G-quadruplexes formed by about four repetitions of telomeric motifs from (A) vertebrates in Na^+ (51), (B, E–G) vertebrates in K^+ (52–57), (B and C) *Giardia* in K^+ (60), (D) *Bombyx* in K^+ (59), (F) *Tetrahymena* in Na^+ (50) and (H) *Oxytricha* in Na^+ (49). Anti and syn guanines are in white and grey, respectively. Structures of *Giardia* and *Bombyx* G-quadruplexes were obtained with modified sequences. The two G-quartets illustrate the possible donor-to-acceptor hydrogen-bond orientations (left: clockwise, right: anticlockwise).

length and sequence diversity, a TG-rich motif (a potential binding site for the repressor-activator protein 1 RAP1) appears to be a conserved feature of telomeric motifs among a variety of budding yeasts (28,29). Heterogeneous telomeric repeats, encoded by a single telomerase RNA (33,34), are also observed in the fission yeast *Schizosaccharomyces pombe*; $\text{G}_2\text{T}_2\text{AC}$ and $\text{G}_2\text{T}_2\text{ACA}$ are by far the most frequently found motifs (33–36).

Bearing runs of consecutive guanines, 5' to 3' telomeric strands from most eukaryotes may potentially fold into G-quadruplexes. G-quadruplexes are a family of nucleic acid secondary structures based on the formation and stacking of quartets of coplanar guanines (G-quartets) (Figure 1); they are stabilized by several cations, in particular potassium and sodium (37). G-quadruplexes can form either by the folding back of a single strand (intramolecular) or by the association of two (bimolecular) or four (tetramolecular) different strands (38–40). Ciliate telomeric repeats were the first to be studied for their G-quadruplex folding potential *in vitro* (38,39,41,42), and ciliates are the first organisms for which direct involvement of G-quadruplexes in telomere biology has been unambiguously demonstrated (43–45). Evidence for G-quadruplex formation by vertebrate telomeric repeats has been gathered *in vitro*, but evidence for G-quadruplex formation at telomeres *in vivo* is mostly indirect. The

G-quadruplex folding potential of telomeric motifs has been emphasized since early studies (46,47). Several G-quadruplex structures formed by telomeric repeats from different organisms have been resolved, notably the intramolecular G-quadruplexes formed by *Oxytricha* and *Tetrahymena* telomeric sequences in sodium (48–50), by human (vertebrate) telomeric sequences in sodium and in potassium (51–58) and by *Bombyx mori* and *Giardia* modified telomeric sequences in potassium (59,60) (Figure 1).

Although several telomeric sequences have been shown to fold into G-quadruplexes, data have been acquired with different techniques and under different conditions, hence do not allow a straightforward comparison of G-quadruplex folding and stability of telomeric sequences among eukaryotes. Furthermore, to our knowledge, folding of yeast telomeric sequences into G-quadruplexes has never been proved, with the exception of sequences mimicking two or three repetitions of *S. cerevisiae* heterogeneous telomeric motifs (46,61). Using biophysical and biochemical methods, we undertook an investigation of sequences consisting of about four repetitions of telomeric motifs from a variety of organisms (the minimal length required to allow formation of intramolecular G-quadruplexes); the aim was to (i) probe their propensity to fold into G-quadruplexes, (ii) provide some information about their folding and (iii) study/compare their stability. The studied sequences are listed in Table 1. Among yeasts, we chose to limit our study to *S. cerevisiae* and *S. pombe* (which are model organisms in the study of telomeres), to *Candida glabrata* and *Candida guilliermondii*. Among the diverse telomeric sequences of budding yeasts (other than species of the genus *Saccharomyces*), *C. glabrata* and *C. guilliermondii* telomeric sequences present the shortest loops (6 nt) connecting consecutive guanines (28,29); they should hence be the most prone to fold into G-quadruplexes. Actually, several studies on model sequences indicate that increasing loop length leads to a decrease in G-quadruplex stability (62–64). *S. cerevisiae* and *S. pombe* have heterogeneous telomeric sequences; we chose to study repetitions of the most frequently found motifs. In a previous study, we provided evidence of folding into stable G-quadruplexes by repetitions of *Plasmodium* degenerated telomeric motif (GGGT TYA, where Y is T or C) under the same conditions used here (65); melting temperatures and thermodynamic data from this work were included in Table 1.

MATERIAL AND METHODS

Oligonucleotides

Reverse-Phase Cartridge•GoldTM (RPC) purified oligonucleotides were purchased from Eurogentec (Seraing, Belgium), dissolved in bi-distilled water at a concentration of 200 μM and stored at -20°C . Concentrations were determined by ultraviolet (UV) absorption using the extinction coefficients provided by the manufacturer. For spectroscopic measurements, RPC-purified oligonucleotides were used without further purification; for non-denaturing PAGE experiments, oligonucleotides

Table 1. Sequences, melting temperatures and thermodynamic parameters for G-quadruplex folding of the studied oligonucleotides

Sequence name (<i>species</i>)	Sequence (from 5' to 3')	NaCl			KCl				
		T_m (°C)	ΔH° (kcal mol ⁻¹)	ΔS° (cal K ⁻¹ mol ⁻¹)	ΔG° (310 K) (kcal mol ⁻¹)	T_m (°C)	ΔH° (kcal mol ⁻¹)	ΔS° (cal K ⁻¹ mol ⁻¹)	ΔG° (310 K) (kcal mol ⁻¹)
G₂ motifs									
Bom17 (<i>Bombyx</i>)	(GGTTA) ₃ GG	a				a			
Asc20 (<i>Ascaris</i>)	(GGCTTA) ₃ GG	a				a			
Spom20 (<i>S. pombe</i>)	(GGTTAC) ₃ GG	no G4 ^b				nd ^c			
Spom23 (<i>S. pombe</i>)	(GGTTACA) ₃ GG	no G4 ^b				no G4 ^b			
Cgl26 (<i>C. guilliermondii</i>)	(GGTGTA CT) ₃ GG	no G4				no G4			
G₃ motifs									
Gial8 (<i>Giardia</i>)	(GGGTA) ₃ GGG	60	-53 ± 7	-157 ± 19	-3.7 ± 0.6	68	-63 ± 4 ^f	-185 ± 12 ^f	-5.7 ± 0.4 ^f
Hum21 (<i>H. sapiens</i>)	(GGGTTA) ₃ GGG	59	-55 ± 4	-167 ± 12	-3.7 ± 0.4	65	-60 ± 3	-177 ± 9	-5.0 ± 0.3
Par21 (<i>Paramecium</i> ^d)	(GGGTTT) ₃ GGG	50	-53 ± 8	-162 ± 23	-2.2 ± 0.5	64	-64 ± 10 ^f	-189 ± 29 ^f	-5.2 ± 0.9 ^f
Scer21 (<i>S. cerevisiae</i>)	(GGGTGT) ₃ GGG	50	-48 ± 5	-148 ± 15	-1.9 ± 0.2	68	-68 ± 8 ^f	-200 ± 25 ^f	-6.2 ± 0.9 ^f
Ara24 (<i>Arabidopsis</i>)	(GGGTTTA) ₃ GGG	57	-60 ± 5	-179 ± 14	-3.6 ± 0.4	64	-73 ± 10	-216 ± 28	-5.7 ± 0.9
<i>Plasmodium</i> ^e	(GGGTTYA) ₃ GGG, Y = T or C	55 ± 2	-53 ± 4	-158 ± 11	-2.8 ± 0.4	64 ± 1	-63 ± 5	-186 ± 14	-4.9 ± 0.6
Tom24 (<i>L. esculentum</i> ^d)	(GGGTTAA) ₃ GGG	51	-54 ± 4	-168 ± 13	-2.3 ± 0.2	64	-77 ± 5	-228 ± 13	-6.1 ± 0.4
Chla27 (<i>Chlamydomonas</i>)	(GGGTTTTA) ₃ GGG	46	-56 ± 4	-176 ± 13	-1.6 ± 0.3	57.5	-68 ± 8	-207 ± 24	-4.2 ± 0.6
G₄ motifs									
Tet22 (<i>Tetrahymena</i>)	(GGGGTT) ₃ GGGG	64	-54 ± 8	-161 ± 24	-4.4 ± 0.7	>80			
Oxy28 (<i>Oxytricha</i>)	(GGGGTTTT) ₃ GGGG	66	-82 ± 13	-242 ± 38	-6.9 ± 1.3	>80			
G₃/G₄ motif									
Gla26 (<i>C. glabrata</i>)	G ₄ T(CTG ₃ TGCTGTG ₄ T)CTG ₃	48	-53 ± 12	-164 ± 36	-1.8 ± 0.5	63	-70 ± 12	-206 ± 34	-5.4 ± 0.8

The number in the sequence name denotes the sequence length (in nucleotides); telomeric motifs are in italic font.

Melting temperatures (*T_m*) reported in this table depended neither on wavelength nor on oligonucleotide strand concentration (3, 10 and 30 μM).

Melting curves at 295 nm were analysed according a two-state equilibrium model and assuming linear low- and high-temperature absorbance baselines. Standard enthalpy and entropy changes (ΔH° and ΔS°) for folding were determined by linear fitting $\ln K$ versus $1/T$, where K is the equilibrium constant between the folded and the unfolded state; standard Gibbs free energy changes (ΔG°) were extrapolated at 310 K, from the relation $\Delta G^\circ(T) = \Delta S^\circ - T\Delta S^\circ$. ΔH° , ΔS° and ΔG° reported in this table are the mean values obtained from analysis of melting curves at 3 and 30 μM strand concentration upon varying linear low- and high-temperature absorbance baselines ± maximum deviation.

^aThe presence of a G-quadruplex and of a non-G-quadruplex competing structure made *T_m* determination not straightforward.

^bA very minor fraction of oligonucleotide may be folded into G-quadruplexes at low temperatures.

^cNot determined: *T_m* of Spom20 in KCl could not be determined accurately because of the incertitude in low-temperature absorbance baseline.

^d*Paramecium* and *L. esculentum* (tomato plant) have degenerated telomeric motifs: GGGKT in *Paramecium* (K = T or G) and GGGTTWA (W = A or T) in tomato plant; for W = T, the tomato telomeric motif comes back to the *Arabidopsis* one; for K = G, the *Paramecium* telomeric motif comes back to the *Tetrahymena* one.

^eData from a previous study (65) (mean values from analysis of the eight possible variant sequences); for Y = T, the *Plasmodium* telomeric sequence comes back to the *Arabidopsis* one.

^fFor Gial8, Par21 and Scer21 in KCl at 30 μM strand concentration a two-state model is not appropriate, as revealed by non-denaturing PAGE; nevertheless linear van't Hoff graphs were obtained.

were purified by PAGE according to standard protocols. For Asc20, Gia18 and Tet22, spectroscopic measurements were carried out on two different batches of RPC purified oligonucleotides; for Gia18, spectroscopic measurements were also carried out on a third batch, PAGE purified by the manufacturer: data reproducibility was excellent. The quality of Gia18 oligonucleotides (RPC and PAGE purified) was checked by denaturing PAGE.

Thermal difference spectra, UV-melting curves and circular dichroism

UV-absorbance measurements were acquired on an Uvikon 940 (Kontron) spectrophotometer or, for tomato plant and *Paramecium* variant sequences, on an Uvikon XL (Secomam). For melting experiments, samples were heated at 92°C for a few minutes, cooled from 92 to 2°C at a rate of 0.2°C min⁻¹, kept at 2°C for 30 min and heated from 2 to 92°C at a rate of 0.2°C min⁻¹. The temperature was varied with a circulating water bath; evaporation at high temperatures and condensation at low temperatures were prevented by a layer of mineral oil and by a dry air flow in the sample compartment, respectively. Temperature was measured with an inert glass sensor immersed into a water-filled quartz cell or, for tomato plant and *Paramecium* variant sequences, with a sensor embedded in a solid cuvet. The absorbance was monitored at 245, 260, 273, 295 and 405 nm; the absorbance at 405 nm was subtracted from each wavelength. For each melting curve, the melting temperature (T_m) was graphically determined as the intercept between the melting curve and the median line between low-temperature and high-temperature absorbance linear baselines (66). For each sample, the thermal difference spectrum (TDS) was obtained by subtracting the absorbance spectrum at 2°C from the one at 92°C (67). The spectrum at 92°C was recorded after heating the sample at 92°C for a few minutes, whereas the spectrum at 2°C was recorded after annealing from 92 to 2°C at a rate of 0.2°C min⁻¹. Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectropolarimeter or, for tomato plant and *Paramecium* variant sequences, on a Jasco J-815, at 4°C, after annealing from 92°C at a rate of 0.2°C min⁻¹. Each spectrum was obtained by averaging three scans at a speed of 500 nm min⁻¹ and by subtracting the contributions of the solvent and quartz cell. All experiments (melting, TDS and CD) were carried out at 3, 10 and 30 µM oligonucleotide strand concentration ($\pm 5\%$), in a cacodylic acid buffer (10 mM) at pH 7.2 (adjusted with LiOH), containing KCl (100 mM) or NaCl (100 mM), using 1 cm (for 3 and 10 µM strand concentration) or 0.2 cm (for 30 µM strand concentration) path-length quartz cells. Stock solutions (200 µM) were heated at 90°C for a few minutes immediately before sample preparation. TDS, UV-melting and CD experiments were carried out at least twice.

Non-denaturing PAGE

Stock solutions (200 µM) of PAGE-purified oligonucleotides were heated at 90°C for a few minutes immediately before sample preparation. Oligonucleotides (30 µM

strand concentration) were annealed from 92 to 2°C at a rate of 1°C min⁻¹ in the same solvents used for spectroscopic measurements (10 mM cacodylic acid buffer, pH 7.2 adjusted with LiOH, 100 mM NaCl or 100 mM KCl). Samples, supplemented with sucrose (11.5%), were then loaded (10 µl) into a 16.5 cm × 14.5 cm × 0.1 cm 12% polyacrylamide gel (acrylamide:bisacrylamide 19:1, prepared in a TBE buffer supplemented with 20 mM NaCl or KCl). Electrophoresis was run in a cold room at 4°C, for 2 h and 30 min, at 3 W per gel, in a TBE buffer supplemented with NaCl (20 mM) or KCl (20 mM); the temperature attained in the gels was about 15°C. DNA was detected by UV-shadow at 254 nm with a G:BOX (Syngene). An oligothymidylate marker (dT₂₁) and two double-stranded markers of 9 and 12 bp (dx9: 5'GCGTATCGG^{3'} + 5'CCGATACGC^{3'}; dx12: 5'GCGTGACTTCGG^{3'} + 5'CCGAAGTCACGC^{3'}) were loaded into the gels to provide internal migration markers. We point out that dT_n oligonucleotides are not appropriate markers for single-strand DNA migration in PAGE (68).

Fluorescence resonance energy transfer melting curves

The melting of the two double-dye labelled oligonucleotides FBom17T and FAsc20T was followed by fluorescence resonance energy transfer (FRET) on a FluoroMax-3 spectrofluorimeter (FBom17T: FAM-5'GGTTAGGTTAGGTTAGG^{3'}-TAMRA, FAsc20T: FAM-5'GGCTTAGGCTTAGGCTTAGG^{3'}-TAMRA, FAM: 6-carboxyfluorescein, TAMRA: carboxytetramethylrhodamine, purchased from Eurogentec, Seraing, Belgium). The temperature was raised from 5 to 90°C with a circulating water bath; FAM emission (excitation at 470 nm, emission at 520 nm) was recorded at intervals of 1°C, after an equilibration time of 30 s. FAM emission versus temperature plots were normalized between 0 and 1 and the temperature of half-dissociation ($T_{1/2}$), corresponding to an emission value of 0.5, was determined. FRET-melting experiments were carried out at 0.2 µM oligonucleotide strand concentration, in a cacodylic acid buffer (10 mM) at pH 7.2 (adjusted with LiOH), containing NaCl (100 mM) or KCl (100 mM).

RESULTS

Thermal difference spectra

A simple method to have an insight into the structure adopted by a nucleic acid sequence is to record its absorbance spectra at high temperature (unfolded form) and at low temperature (folded form). The difference between these spectra, called thermal difference spectrum (TDS), provides a spectroscopic signature for nucleic acid structures (67,69).

TDS spectra are shown in Figure 2. With the exception of the *C. guilliermondii* (Cgi26) and of the *S. pombe* (Spom20 and Spom23) sequences, all the studied telomeric sequences displayed TDS signatures characteristic of G-quadruplex structures, both in NaCl and in KCl: two major positive peaks around 240 and 270 nm and a negative peak around 295 nm.

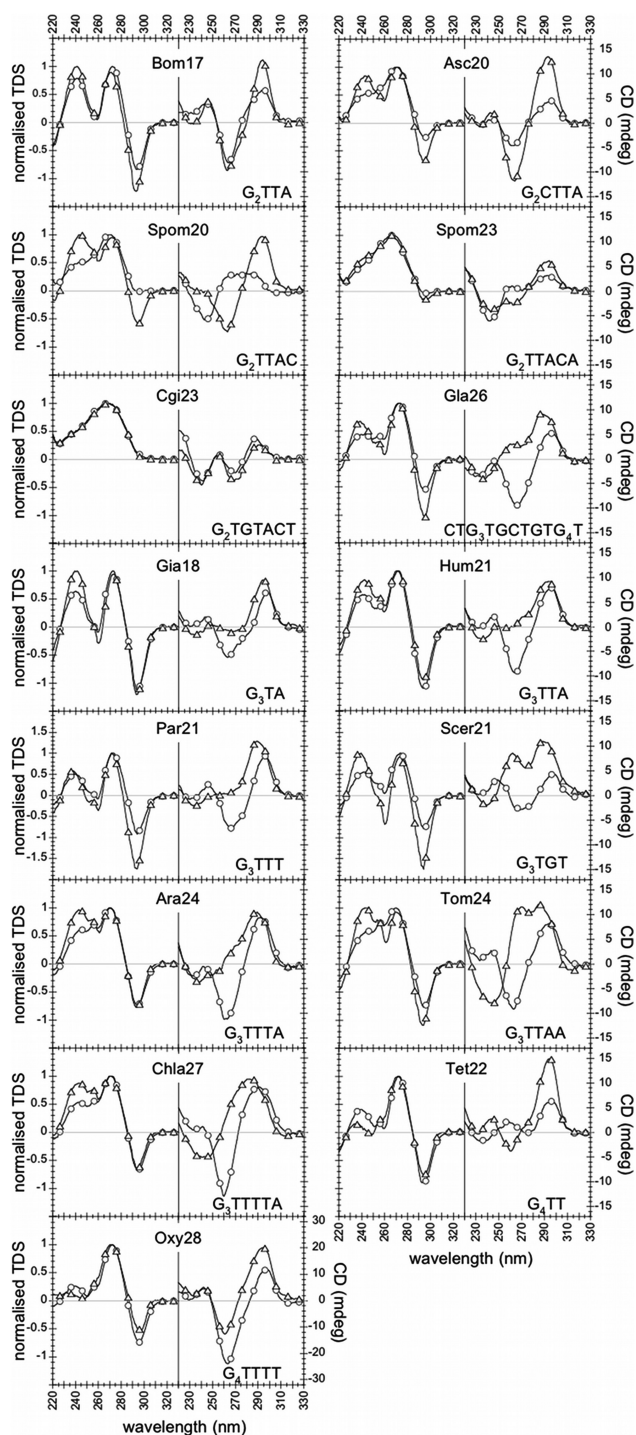


Figure 2. CD spectra at 4°C (right panels) and normalized TDS (left panels) of all the sequences listed in Table 1, in NaCl (circles) and KCl (triangles), at 3 μ M oligonucleotide strand concentration. The CD axis scale is the same for all sequences (–17/+17 mdeg), with the exception of Oxy28 (–32/+32 mdeg). TDS and CD spectra did not depend on strand concentration (3, 10 and 30 μ M), with the exception of Gia18, Tet22 and Scer21 CD spectra in KCl.

The *C. guillermontii* sequence (Cgi26) neither in NaCl nor in KCl displayed a G-quadruplex-type TDS signature. Concerning the *S. pombe* sequences (Spom20 and Spom23), an unambiguous TDS G-quadruplex signature

was obtained only for Spom20 in KCl. TDS of Spom20 in NaCl and Spom23 in NaCl and KCl exhibited a very small negative peak at 295 nm (autoscaled TDS are shown in Figure S1).

CD spectra

CD spectra were recorded at 4°C, a temperature at which the equilibrium was strongly shifted towards the folded forms, as shown by thermal melting profiles (Figures S2 and S3). CD spectra at 3 μ M oligonucleotide strand concentration are shown in Figure 2. The *Bombyx* (Bom17), *Ascaris* (Asc20) and *Oxytricha* (Oxy28) sequences exhibited a negative band around 260 nm and a positive band around 290 nm (a ‘Type II’ signature), both in NaCl and in KCl. All the other G-quadruplex folding sequences (i.e. sequences displaying G-quadruplex TDS) exhibited distinct CD signatures in NaCl and in KCl: a Type II signature in NaCl (with the exception of Tet22), and a positive peak around 290 nm and a positive peak or shoulder in the 250–270-nm region in KCl; this supports structural differences in NaCl and in KCl.

With the exception of the *Giardia* (Gia18), *Tetrahymena* (Tet22) and *S. cerevisiae* (Scer21) sequences in KCl, the CD spectra of all the other sequences were independent from oligonucleotide strand concentration (3, 10 or 30 μ M), both in NaCl and in KCl. Conversely, the CD spectra of *Giardia* (Gia18), *Tetrahymena* (Tet22) and *S. cerevisiae* (Scer21) sequences in KCl strongly depended on oligonucleotide strand concentration (Figure 3). For Gia18 and Tet22 increasing the oligonucleotide concentration from 3 to 30 μ M led to the appearance of a CD positive peak at about 260 nm (Figure 3A and B). For Scer21, increasing the concentration led to a change in the ratio r between the CD peak at about 260 nm and the one at about 290 nm: ($r < 1$ for 3 μ M, $r \approx 1$ for 10 μ M and $r > 1$ for 30 μ M) (Figure 3C). The dependence of Gia18, Tet22 and Scer21 CD spectra on strand concentration in KCl supports the presence of multi-stranded structures.

Concerning *C. guillermontii* (Cgi26) and *S. pombe* (Spom20 and Spom23) sequences, only the *S. pombe* sequence Spom20 in KCl displayed an unambiguous Type II G-quadruplex CD signature, consistently with the results obtained by TDS.

Hence, TDS provided evidence of G-quadruplex folding for most of the sequences listed in Table 1; CD allowed (i) highlighting structural differences in NaCl and in KCl for most of the sequences (at least for Gia18, Hum21, Par21, Scer21, Ara24, Tom24, Chla27, Tet22 and Gia26) and (ii) providing evidence of the presence of multi-stranded structures for a few sequences (Gia18, Tet22 and Scer21 in KCl). Individual CD spectra will not be discussed in detail for the following reason. Empirically, two types of CD spectra are considered indicative of DNA G-quadruplexes: spectra with a negative band around 240 nm and a dominant positive band around 260 nm (‘Type I’ spectra), and spectra with a negative band around 260 nm and a positive band around 290 nm (‘Type II’ spectra). Theoretical calculations give an insight into the origin of Type I and Type II signals:

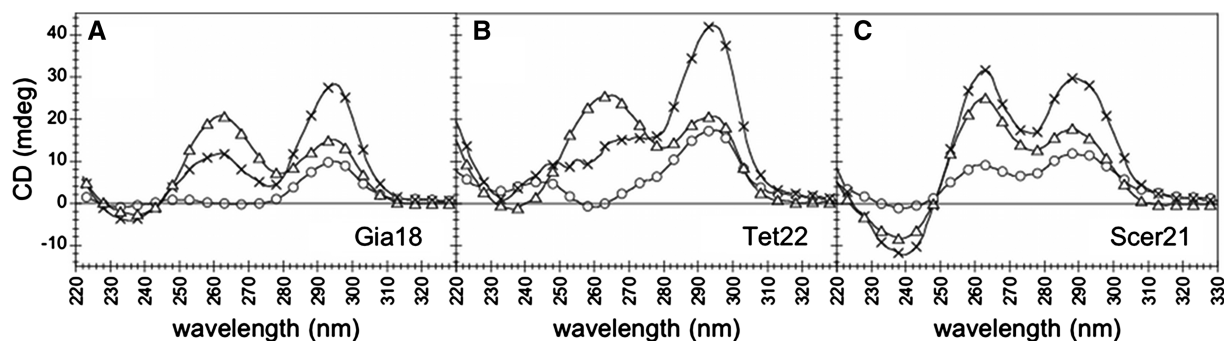


Figure 3. Circular dichroism spectra at 4°C of (A) Gia18, (B) Tet22 and (C) Scer21 in KCl at 3, 10 and 30 μ M oligonucleotide strand concentration (circles, crosses and triangles, respectively).

stacking of two G-quartets with the same hydrogen-bond orientation originates a Type I signal, stacking of two G-quartets with opposite hydrogen-bond orientation originates a Type II signal (70,71) (G-quartet hydrogen-bond orientation is illustrated in Figure 1). Nevertheless, G-quadruplexes are highly polymorphic and these empirical and theoretical evidences do not allow an unambiguous interpretation of CD spectra of potential G-quadruplex folding sequences.

UV melting experiments

TDS allowed verifying G-quadruplex folding at low temperature. To study stability, we recorded melting profiles at wavelengths of absorbance maximum variations upon folding, determined by TDS, i.e. 245, 273 and 295 nm.

All melting profiles were reversible (melting curves at 295 nm are reported in Supplementary Figures S2 and S3). Melting transitions of all G-quadruplex folding sequences presented a single transition (a single first derivative peak) and occurred (with the exception of Spom20 in KCl) in a relatively narrow range of temperature (about 30°C for Bom17 and Asc20 and about 20°C for the other sequences), thereby the graphically determined melting temperatures (T_m s) did not significantly depend on the choice of the high- and low-temperature absorbance linear baselines; T_m s could then be determined with a precision of $\pm 1^\circ\text{C}$ or better.

T_m s of sequences bearing G_3 and G_4 motifs (included Gla26) were higher in potassium than in sodium: T_m s ranged from 46°C (Chla27) to 64–66°C (Tet22 and Oxy28) in NaCl and from 57.5°C (Chla27) to >80°C (Tet22 and Oxy28) in KCl; they did not depend on wavelength (245, 273 and 295 nm) and are reported in Table 1. The flatness of the *S. pombe* Spom20 melting profile in KCl at 295 nm (Supplementary Figures S2 and S3) prevented us from determining an accurate T_m (incertitude in the determination of a low-temperature absorbance baseline); its T_m was roughly estimated to be around 27°C. *Bombyx* (Bom17) and *Ascaris* (Asc20) differed from all the other G-quadruplex folding sequences for several features. (i) Surprisingly, the graphically determined T_m s at 295 nm were higher in sodium than in potassium: 46°C in NaCl and 44°C in KCl for Bom17, 41°C in NaCl and 39°C in KCl for Asc20. The maxima of the first derivatives of their melting curves (T_{max}) showed the same trend than melting temperatures determined by graphical analysis:

$T_{max}(\text{NaCl}) > T_{max}(\text{KCl})$. (ii) Asc20 T_m s depended on wavelength (T_m s at 245 nm were around 45°C both in sodium and potassium); this feature suggests the presence of a competing structure. (iii) The amplitude of Asc20 melting transition at 295 nm in NaCl was about half of the amplitude in KCl (Figure 4A); Bom17 melting transitions at 295 nm in NaCl and in KCl had nearly equal amplitudes (Figure 4B). Conversely, for all G-quadruplexes with $T_m(\text{NaCl}) < T_m(\text{KCl})$, the amplitudes of melting transitions at 295 nm in NaCl were larger than in KCl (Figures S2 and S3) (this feature appears to be characteristic of many other G-quadruplexes).

Finally, for all the sequences, T_m did not depend on oligonucleotide strand concentration (3, 10 and 30 μ M), not even for those sequences displaying concentration-dependent CD spectra, i.e. Gia18 and Scer21 in KCl (Tet22 T_m in KCl was too high to be determined). In particular, we verified that T_m of Gia18 in KCl did not vary upon increasing oligonucleotide strand concentration up to 100 μ M (data not shown). The independence of T_m from oligonucleotide strand concentration suggests formation of intramolecular structures.

FRET melting experiments

The ‘non-classical’ behaviour of *Bombyx* (Bom17) and *Ascaris* (Asc20) prompted us to a deeper investigation. We studied the double-dye labelled sequences FAsc20T (FAM-5′(GGCTTA)₃GG^{3′}-TAMRA) and FBom17T (FAM-5′(GGTTA)₃GG^{3′}-TAMRA). FAsc20T neither in NaCl nor in KCl folded into G-quadruplexes, as revealed by CD spectroscopy (Figure 4C); it folded into a structure where the two dyes were close to each other, the stability of which ($T_{1/2} = 54^\circ\text{C}$) did not depend on the cation (very likely a hairpin-duplex), as revealed by thermal melting followed by FRET (Figure 4E). Conversely, FBom17T folded into G-quadruplexes, as revealed by CD spectroscopy (Figure 4D), more stable in potassium ($T_{1/2} = 55^\circ\text{C}$) than in sodium ($T_{1/2} = 45^\circ\text{C}$) (Figure 4F).

Gel electrophoresis

Non-denaturing PAGE was carried out at 30 μ M oligonucleotide strand concentration, i.e. the highest concentration used for spectroscopic measurements. DNA was detected by UV-shadow (Figure 5). The temperature of

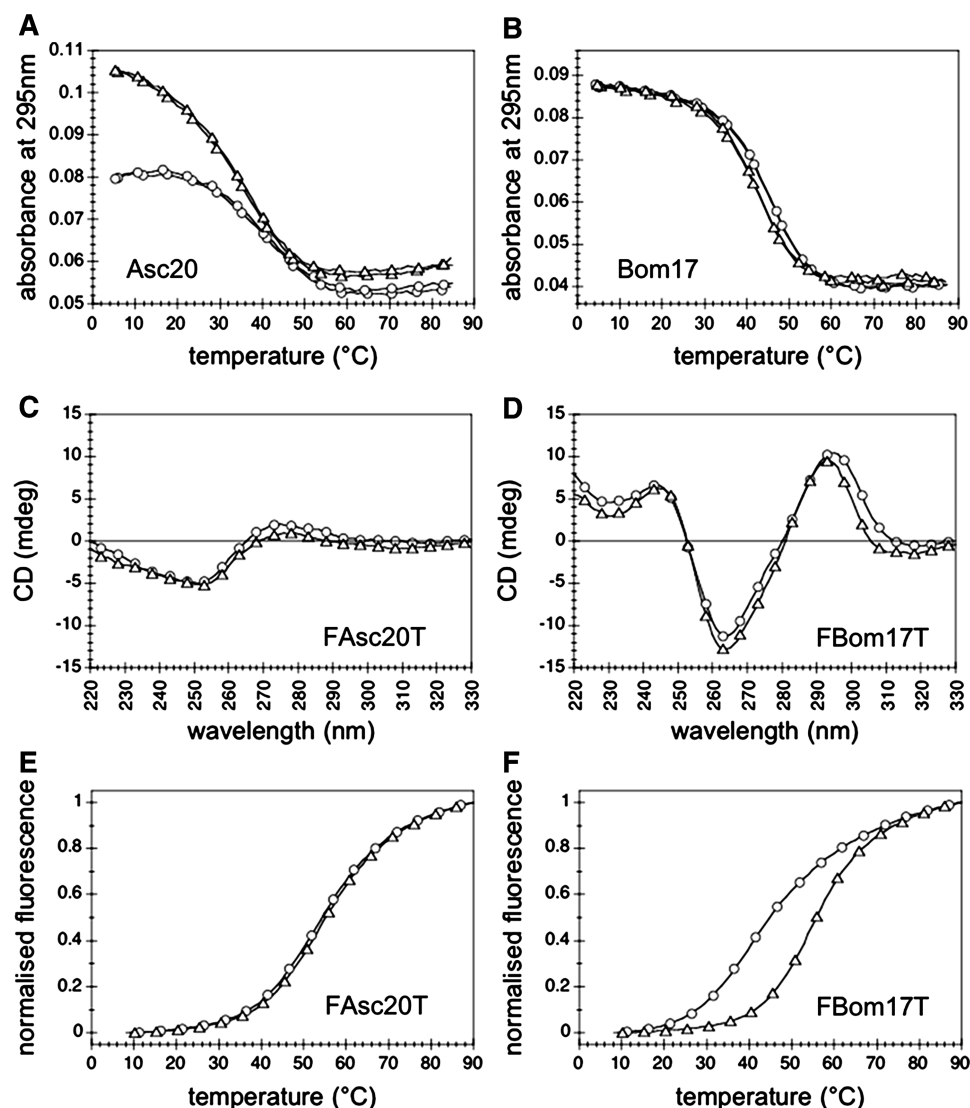


Figure 4. (A and B) Thermal melting (cooling and heating) followed by absorbance at 295 nm of (A) Asc20 and (B) Bom17 in NaCl (circles) and KCl (triangles), at 3 μ M oligonucleotide strand concentration. (C and D) CD spectra at 4°C of (C) FAsc20T and (D) FBom17T in NaCl (circles) and KCl (triangles), at 3 μ M oligonucleotide strand concentration. (E and F) Thermal melting (heating) followed by FRET (excitation at 470 nm, emission at 520 nm) of (E) FAsc20T and (F) FBom17T in NaCl (circles) and KCl (triangles), at 0.2 μ M oligonucleotide strand concentration.

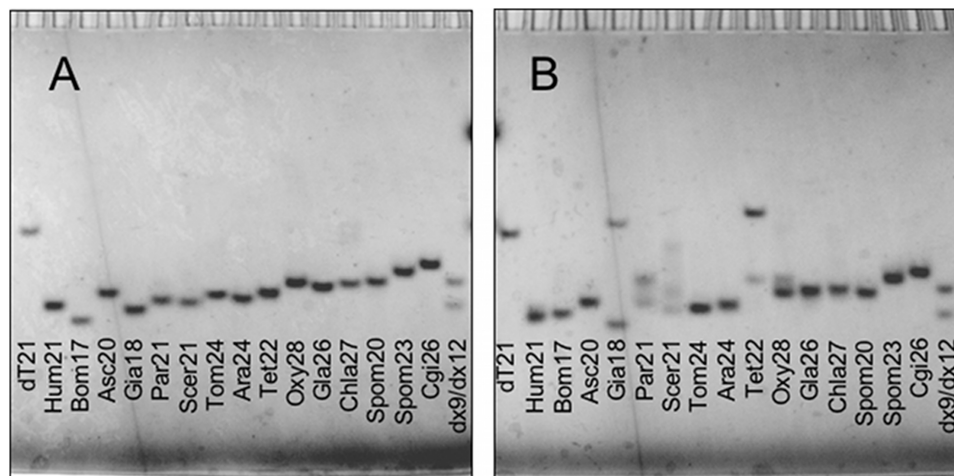


Figure 5. Non-denaturing PAGE in (A) NaCl and (B) KCl at 30 μ M oligonucleotide strand concentration. dT₂₁ is an oligothymidylate marker; dx9 and dx12 are two double-stranded markers of 9 and 12 bp, respectively. Oligonucleotides were detected by UV-shadow.

the gels was about 15°C; at this temperature, all the G-quadruplex folding sequences (i.e. all the sequences except Spom20, Spom23 and Cgi26 in NaCl and Spom23 and Cgi26 in KCl) were folded both in KCl and in NaCl, as could be inferred by the inspection of melting profiles (Supplementary Figures S2 and S3). We used an oligothymidylate marker (dT₂₁) and 9- and 12-bp duplexes (dx9 and dx12, respectively) as internal migration standards.

In NaCl, a single band was detected for all the telomeric sequences (Figure 5A). All the oligonucleotides displayed a higher mobility than dT₂₁; most of them had a mobility intermediate between the ones of dx9 and dx12. In NaCl, the mobility of the sequences displaying a TDS G-quadruplex signature (i.e. all the sequences except Spom20, Spom23 and Cgi26) slightly decreased with the length of the sequence, with the exception of Asc20 and Tet22. The yeast Spom20, Spom23 and Cgi26 sequences, which did not fold into G-quadruplexes in NaCl, as revealed by TDS, were the slowest migrating species.

In KCl, the migration patterns were more complex (Figure 5B). Most of the sequences (Hum21, Bom17, Asc20, Tom24, Ara24, Gia26, Chla27, Spom20, Spom23 and Cgi26) migrated as single bands and faster than dT₂₁. Among these single-band migrating sequences, Spom23 and Cgi26 had the lowest electrophoretic mobility; according to spectroscopic measurements in KCl, these two sequences are not folded into G-quadruplexes. The *Giardia* (Gia18) and *Tetrahymena* (Tet22) sequences displayed a fast- and a slow-migrating band, the latter migrating slower than the internal migration standard dT₂₁. For the *Paramecium* (Par21) and the *Oxytricha* (Oxy28) sequences, two bands were detected, both fast migrating as compared to dT₂₁. In particular, the two bands of Oxy28 (a principal and a secondary one) were very close to each other. Finally, three bands were detected for the *S. cerevisiae* sequence (Scer21): two fast migrating bands (between dx9 and dx12) and a slow migrating band (between dx12 and dT₂₁).

The migration pattern obtained in KCl for Gia18, Tet22, Scer21, Par21 and Oxy28 provides evidence of the presence of at least two (or three, for Scer21) different structures at the tested oligonucleotide strand concentration (30 µM). For Gia18, Tet22 and Scer21, the slow migrating band detected in KCl may be ascribed to the presence of multi-stranded species; this hypothesis is supported by changes in their CD spectra upon increasing oligonucleotide strand concentration from 3 to 30 µM (Figure 3).

DISCUSSION

In the present work we studied telomeric sequences from different organisms, with the aim of probing their propensity to fold into G-quadruplex structures, providing some information about their folding and studying their stability.

G-quadruplex folding

Thermal difference spectra provided evidence of the formation of G-quadruplex structures, both in sodium and in

potassium, for all the studied telomeric sequences, with the exception of the *C. guillermoidii* (Cgi26) and *S. pombe* (Spom20 and Spom23) sequences. By contrast, TDS and CD in sodium and in potassium did not support formation of G-quadruplexes by Cgi26. Spom20 in potassium displayed unambiguous TDS and CD G-quadruplex signatures. Finally, the very small negative peak at 295 nm in TDS of Spom20 in sodium and Spom23 in both sodium and potassium suggests that a very minor fraction of these sequences folds into G-quadruplexes; this fraction may contribute to the observed CD spectra.

Molecularity

For the telomeric sequences displaying a TDS G-quadruplex signature, three lines of evidence supported folding into intramolecular structures in sodium: the independence of (i) melting temperatures and (ii) CD spectra from oligonucleotide strand concentration; (iii) the presence of a single and fast migrating band (as compared to dT₂₁) in a non-denaturing PAGE experiment. The same conclusion could be drawn for most of the sequences in potassium (Bom17, Asc20, Hum21, Ara24, Tom24, Gia26, Chla27 and Spom20).

For the *Paramecium* (Par21) and *Oxytricha* (Oxy28) sequences in potassium, two bands were detected by PAGE; the mobility of these bands (as compared to the internal markers) and the independence of Par21 *T_m* and Par21–Oxy28 CD spectra from oligonucleotide strand concentration suggest folding into intramolecular structures of different electrophoretic mobilities (Oxy28 *T_m* in potassium were too high to be determined).

The PAGE migration pattern of *Giardia* (Gia18), *Tetrahymena* (Tet22) and *S. cerevisiae* (Scer21) sequences in potassium provided evidence of the presence of dramatically different structures (at least two for Gia18 and Tet22 and at least three for Scer21). For these three sequences, the slow migrating band may be ascribed to multi-stranded structures. This hypothesis is supported by the change in Gia18, Tet22 and Scer21 CD spectra upon increasing oligonucleotide strand concentration. Furthermore, the PAGE migration pattern of Gia18 in KCl was shown to be concentration-dependent: the slow migrating band was not detected at 100-fold lower oligonucleotide strand concentration (300 nM) (gel detected by Syber Gold staining, data not shown). This demonstrates that the slow migrating band of Gia18 corresponds to multi-stranded species.

On the other hand, the independence of Gia18 and Scer21 *T_m*s from oligonucleotide strand concentration (up to 100 µM for Gia18) does not support folding into intermolecular structures, such as bimolecular or tetramolecular G-quadruplexes (Tet22 *T_m* in potassium were too high to be determined). This puzzling behaviour might be explained by folding into intramolecular G-quadruplexes (monomers) and formation of stacked monomers upon increasing oligonucleotide strand concentration from 3 to 30 µM. The appearance of a positive CD band around 260 nm might arise from stacking of terminal G-quartets with the same hydrogen-bond orientation (70,71). The hypothesis of stacked intramolecular

G-quadruplexes is not appropriate (or is only partially appropriate) for Scer21, which present a more complex non-denaturing PAGE migration pattern.

Structural differences in sodium and in potassium

Besides providing evidence of a concentration-dependent behaviour of Gia18, Tet22 and Scer21 in potassium, CD allowed highlighting structural differences in sodium and potassium.

The CD spectra of most of the telomeric sequences (Gia18, Hum21, Ara24, Chla27, Par21, Scer21, Tom24 and Tet22) strongly depended on the nature of the cation present in solution (sodium or potassium), suggesting different G-quadruplex conformations or a different equilibrium among multiple possible G-quadruplex conformations according to the nature of the cation. Only the *Bombyx* (Bom17), *Ascaris* (Asc20) and *Oxytricha* (Oxy28) sequences displayed a similar (although not strictly identical) CD signature in sodium and in potassium.

It would be unwarranted to speculate about CD signatures and G-quadruplex folding topologies, for several reasons. (i) The simple empirical rule often used to interpret CD spectra (Type I spectrum = parallel folding, Type II spectrum = antiparallel folding) is based on a few resolved structures and is not of general validity (71). (ii) Interpretation of CD spectra in terms of relative strand orientation is not formally correct (70,71). (iii) Interpretation of CD spectra in terms of folding topologies relies on the implicit assumption that a sequence folds into a single G-quadruplex structure. A single sequence may fold into multiple G-quadruplex structures in the same solution; this is the case, for example, of human and *Giardia* telomeric sequences in potassium (Figure 1). Structure determination requires high-resolution structural studies, such as nuclear magnetic resonance (NMR) or X-ray crystallography, and is beyond the scope of this work.

Stability of G₃ and G₄ motifs

All sequences bearing three or more consecutive guanines folded into G-quadruplexes more stable in potassium than in sodium (higher T_m s), as expected for G-quadruplexes. Sequences bearing three consecutive guanines (including Gia26, for a total of eight motifs) had T_m s ranging from 46 to 60°C in sodium and from 57.5 to 68°C in potassium. Notably, for the G₃T_nA motif sequences, T_m slightly decreased with n for $n = 1, 2$ and 3 (Gia18, Hum21 and Ara24, respectively); increasing n to 4 (Chla27) led to a dramatic decrease in T_m (about 10°C). The *Tetrahymena* (Tet22) and *Oxytricha* (Oxy28) sequences, bearing four consecutive guanines, were the most stable; they had roughly the same T_m s in sodium (64 and 66°C, respectively) and T_m s higher than 80°C in potassium.

We carried out a thermodynamic analysis of melting transitions (Table 1). A two-state equilibrium model, assuming linear low- and high-temperature absorbance baselines, resulted in linear van't Hoff graphs. Non-denaturing PAGE experiments revealed that such a model is not appropriate for Gia18, Par21 and Scer21 in KCl, nevertheless linear van't Hoff graphs could

be obtained even for these sequences in KCl. Thermodynamic values did not depend on oligonucleotide strand concentration (3 and 30 μ M). For G₃ and G₃/G₄ motifs, ΔH° of folding varied between -60 and -48 kcal mol⁻¹ in NaCl and between -77 and -60 kcal mol⁻¹ in KCl; ΔG° of folding, extrapolated at 37°C, varied between -3.7 and -1.6 kcal mol⁻¹ in NaCl and between -6.2 and -4.2 kcal mol⁻¹ in KCl. The analysis of G₄ motifs was possible only in NaCl: standard enthalpy change of Oxy28 folding ($\Delta H^\circ = -82 \pm 13$ kcal mol⁻¹) was more negative than standard enthalpy changes of Tet22 folding ($\Delta H^\circ = -54 \pm 8$ kcal mol⁻¹) and of G₃ and G₃/G₄ motifs. For Tet22, G₃ and G₃/G₄ motifs in NaCl, T_m and ΔG° were linearly correlated (linear regression coefficient $R = 0.9877$). In KCl, T_m and ΔG° correlation was poorer; this might be ascribed to a greater error in ΔG° determination in KCl than in NaCl (due to a greater uncertainty in high-temperature baseline determination and to extrapolation of ΔG° at a temperature far below the temperature ranges where melting transitions occurred).

The influence of G-tract length, loop length and loop sequence on G-quadruplex structure and stability has been explored on model sequences (62–64,72–77); nevertheless, no general rules has emerged (78). In particular, the number of consecutive guanines does not necessarily correspond to the number of G-quartets (50,57,60). NMR studies of the *Oxytricha* 5'-(G₄T₄)₃G₄3' and of the *Tetrahymena* 5'-(T₂G₄)₄3' telomeric sequences in sodium revealed four stacked G-quartets in the *Oxytricha* G-quadruplex (48,49) (Figure 1H) and only three stacked G-quartets in the *Tetrahymena* G-quadruplex, connected by a GTTG lateral loop, a central TTG lateral loop and a TT double-chain-reversal loop (50) (Figure 1F). Our data do not allow drawing conclusions about the number of stacked G-quartets in the *Tetrahymena* telomeric sequence Tet22, which differed from the one used in the NMR study by two nucleotides (TT). Nevertheless, in sodium, the amplitude in absorbance of the 295 nm melting transition of Tet22 (0.047 variation in absorbance recorded at 3 μ M oligonucleotide strand concentration and 1 cm optical path length) was very similar to the amplitudes of the sequences bearing three consecutive guanines (0.049 ± 0.008 on average) and significantly smaller than the amplitude of Oxy28 (0.072) (Figure S2). This feature, along with ΔH° values discussed above, might argue stacking of three G-quartets in Tet22 and of four G-quartets in Oxy28.

For a comprehensive analysis of tomato plant and *Paramecium* degenerated telomeric sequences, we studied the sixteen following variant sequences: 5'-(GGG TTWA)₃GGG3' (where W is A or T) (tomato) and 5'-(GGGKTT)₃GGG3' (where K is T or G) (*Paramecium*) (Table 2). Not surprisingly, all these sequences folded into stable G-quadruplexes (TDS are shown in Supplementary Figures S4 and S5). T_m did not strongly vary among tomato variant sequences: it ranged from 55 to 58°C in NaCl and from 64 to 69°C in KCl (Table 2). T_m s of *Paramecium* variant sequences spanned a broader range: from 52 to 60°C in NaCl and from 68 to 76°C in KCl (Table 2). Two correlations between sequence and stability

Table 2. Tomato plant and *Paramecium* variant telomeric sequences and T_m s determined by analysis of melting curves recorded at 295 nm, at 3 μ M oligonucleotide strand concentration

Sequence name	Sequence (from 5' to 3')	T_m ($^{\circ}$ C) in NaCl	T_m ($^{\circ}$ C) in KCl
Tomato plant			
Tom24	G ₃ TTAAG ₃ TTAAG ₃ TTAAG ₃	55	67
TAA	G ₃ TTTAG ₃ TTAAG ₃ TTAAG ₃	58	65
ATA	G ₃ TTAAG ₃ TTTAG ₃ TTAAG ₃	57	69
AAT	G ₃ TTAAG ₃ TTAAG ₃ TTTAG ₃	58	67
TTA	G ₃ TTTAG ₃ TTTAG ₃ TTAAG ₃	58	66
TAT	G ₃ TTTAG ₃ TTAAG ₃ TTTAG ₃	57	64
ATT	G ₃ TTAAG ₃ TTTAG ₃ TTTAG ₃	56	67
Ara24	G ₃ TTTAG ₃ TTTAG ₃ TTTAG ₃	58	66
Paramecium			
Par21	G ₃ TTTG ₃ TTTG ₃ TTTG ₃	52	68
GTT	G ₃ TTTG ₃ TTTG ₃ TTTG ₃	54	72
TGT	G ₃ TTTG ₃ TTTG ₃ TTTG ₃	52	68
TTG	G ₃ TTTG ₃ TTTG ₃ TTTG ₃	59	72
GGT	G ₃ TTTG ₃ TTTG ₃ TTTG ₃	54	71
GTG	G ₃ TTTG ₃ TTTG ₃ TTTG ₃	58	74
TGG	G ₃ TTTG ₃ TTTG ₃ TTTG ₃	58	73
GGG	G ₃ TTTG ₃ TTTG ₃ TTTG ₃	60	76

T_m s reported in this table were obtained on a different spectrophotometer from those reported in Table 1.

Temperature was measured with a sensor in a sealed cuvet provided by the manufacturer; this may explain the slightly higher T_m s reported in this table as compared to T_m s reported in Table 1 for Tom24, Ara24 and Par21.

TDS and CD spectra are reported in Supplementary Figures S4 and S5.

and between sequence and CD signature emerged. (i) Among *Paramecium* variant sequences, stability in NaCl correlated with the sequence of the third repeated unit (more precisely of nucleotides 16–18): G-quadruplexes formed by $5'(\text{G}_3\text{KTT})_2\text{G}_3\text{TTTG}_3$ sequences were less stable (T_m s of 52–54 $^{\circ}$ C) than the ones formed by $5'(\text{G}_3\text{KTT})_2\text{G}_3\text{GTTG}_3$ sequences (T_m s of 58–60 $^{\circ}$ C). (ii) Among tomato plant variant sequences, the sequence of the first loop connecting runs of guanines (nucleotides 4–7) correlated with CD spectra in KCl: $5'\text{G}_3\text{TTAA}(\text{G}_3\text{TTWA})_2\text{G}_3$ sequences had Tom24-like CD spectra, while $5'\text{G}_3\text{TTTA}(\text{G}_3\text{TTWA})_2\text{G}_3$ sequences had Ara24-like CD spectra (Figure S4).

Stability of G₂ motifs

Among the telomeric sequences bearing two consecutive guanines, only the *Bombyx* (Bom17) and the *Ascaris* (Asc20) sequences folded into G-quadruplexes both in sodium and in potassium. Permuting a cytosine from the first position (Asc20) to the last position (Spom20) of the G-tract connecting sequences abolished G-quadruplex folding in sodium. The analysis of Bom17 and Asc20 melting transitions at 295 nm revealed higher T_m s in sodium than in potassium (about 2 $^{\circ}$ C). This behaviour is unique to Bom17 and Asc20 sequences, since the model sequences $5'(\text{G}_2\text{T}_n)_3\text{G}_2$ with $n = 3$ and $n = 4$ folded into G-quadruplexes with higher T_m s in potassium than in sodium (data not shown). An apparent higher T_m at 295 nm in sodium than in potassium can result from the equilibrium between a G-quadruplex and a

non-G-quadruplex structure. Asc20 and Bom17 can in principle fold into (hairpin-) duplexes with, respectively, 2 out of 3 and 1 out of 3 nucleotides engaged in Watson–Crick base pairs. The presence of a competing structure is suggested by results obtained in LiCl and results obtained with fluorescent conjugates. Folding of Asc20 and Bom17 into non-G-quadruplex structures, melting at about 38 and 25 $^{\circ}$ C, respectively, was indeed observed in LiCl (Figure S6); folding into these structures did not result into absorbance variation at 295 nm. Results obtained with the double-dye-labelled sequences suggest that both Asc20 and Bom17 fold into two structures (a G-quadruplex and a non-G-quadruplex structure) and that the presence of the dyes completely shift the equilibrium of the *Ascaris* sequence toward the non-G-quadruplex structure and the equilibrium of the *Bombyx* sequence toward the G-quadruplex (a ‘classical’ G-quadruplex, i.e. more stable in potassium than in sodium). If a competing structure is present, then the T_m at 295 nm is not a ‘true’ G-quadruplex T_m , even if the competing structure does not contribute to melting profiles at 295 nm. More precisely, if f is the G-quadruplex folded fraction at low temperatures, then the graphically determined T_m at 295 nm is the temperature at which a fraction $f/2$ is folded into a G-quadruplex; this temperature, which depend on f (i.e. on the equilibrium between the G-quadruplex and the competing structure) cannot be used as a parameter for G-quadruplex stability. The presence of a non-G-quadruplex competing structure may also explain the small amplitude of Asc20 melting transition at 295 nm in sodium compared to the one in potassium (Figure 4A). This feature may result from a smaller G-quadruplex folded fraction in sodium than in potassium at low temperatures. This would mean that G-quadruplexes formed by Asc20 are less stable in sodium than in potassium (under the reasonable hypothesis that the stability of the competitive structure does not depend on the nature of the cation). This hypothesis is not in contradiction with higher T_m s at 295 nm in sodium than in potassium, as discussed above. Finally, the independence of Asc20 and Bom17 melting temperatures from oligonucleotide strand concentration (3, 10 and 30 μ M) supports the intramolecular nature of both the G-quadruplex and the non-G-quadruplex structures. The PAGE migration patterns of Asc20 and Bom17 are not inconsistent with the presence of G-quadruplexes and hairpin-duplexes, when compared to the internal migration standard dx9 and dx12.

The inability of the Cgi26 and Spom23 sequences to fold into G-quadruplexes or stable G-quadruplexes does not result just from the length of the loops separating the runs of guanines (6 and 5 nt, respectively). Actually the model sequence $5'(\text{G}_2\text{T}_5)_3\text{G}_2$ folded into G-quadruplexes both in sodium and in potassium (with a T_m of 16 and 25 $^{\circ}$ C, respectively), whereas the model sequence $5'(\text{G}_2\text{T}_6)_3\text{G}_2$ folded into stable G-quadruplexes in potassium only (with a T_m of 18 $^{\circ}$ C) (data not shown). Hence, the behaviour of the telomeric sequences bearing just two consecutive guanines (from Bom17 to Cgi26) arises from their peculiar loop sequences.

In summary, with regard to the ability of folding into intramolecular G-quadruplexes, our study highlights at least two different sets of telomeric motifs: (i) yeast telomeric motifs and (ii) telomeric motifs from other eukaryotes. On the one hand, the ability to fold into G-quadruplex structures does not appear to be a conserved feature among the exceptionally diverse yeast telomeric sequences: unlike *C. glabrata*, *C. guilliermondii* telomeric sequence did not fold into G-quadruplexes; unlike repetitions of the TG₃TG motif (frequently occurring in *S. cerevisiae* telomeres), repetitions of G₂TTAC(A) motifs (frequently occurring in *S. pombe* telomeres) were not prone to fold into stable G-quadruplexes. On the other hand, evolution of telomeric sequences in eukaryotes other than yeasts (repetitions of G₃T₁₋₄A in a variety of organisms, G₄T_{2,4} in ciliates, G₂T₂A in many insects and G₂CT₂A in nematodes) appears to have preserved the G-quadruplex folding potential, even if these structures pose a potential problem to telomeric G-strand replication (79,80). Notably, telomeric sequences with three or more consecutive guanines folded into G-quadruplexes with melting temperatures above temperatures where organisms known to have these telomeric sequences thrive. Curiously, no telomeric motif with an adenine following the run of guanines has been found, with the exception of the putative G₆AAT telomeric motif of the hot-spring unicellular red alga *Cyanidioschyzon merolae* (81). Intriguingly, an adenine as the first base in the loops connecting runs of two or three consecutive guanines is detrimental to G-quadruplex folding (47,77). *Cyanidioschyzon merolae* is an extremophilic organism, thriving in acid hot springs (pH 1.5, 45°C). Its telomeric sequence 5'(G₆AAT)₃G₆3' folded into very stable G-quadruplexes, melting at 70°C in NaCl and stable at temperatures as high as 90°C in KCl (Supplementary Figure S7). Unlikely, telomeric motifs of other thermophilic eukaryotes (such as fungi able to grow at 55–60°C, hot-spring algae other than *C. merolae*, the worms *Paralvinella sulfincola* and *Alvinella pompejana*, tardigrades) are not known, to our knowledge.

In conclusion, the present study provides an insight into intramolecular G-quadruplex folding potential of telomeric sequences from a variety of organisms. (i) All the telomeric sequences bearing three or more consecutive guanines (in particular G₃T₁₋₄A motifs and G₄T_{2,4} ciliate motifs) folded into stable G-quadruplexes both in potassium and in sodium. (ii) The G₂TTA telomeric motif sequence (found in many insects, such as *Bombyx mori*) and the G₂CTTA telomeric motif sequence (found in the nematodes *Ascaris lumbricoides* and *Caenorhabditis elegans*) folded into G-quadruplexes both in potassium and in sodium. Our results strongly suggest that, for these two sequences, the G-quadruplex structures are in equilibrium with non-G-quadruplex structures (likely hairpin-duplexes). (iii) G-quadruplex folding does not appear to be a characteristic feature of budding yeast telomeres, despite a widespread conserved TG-rich motif of about six base pairs (28). Actually, unlike repetitions of the *S. cerevisiae* frequent occurring motif TG₃TG and of the *Candida glabrata* CTGTG₄TCTG₃TG motif, the *Candida guilliermondii* telomeric sequence (TACTG₂TG

motif) was unable to fold into G-quadruplexes. (iv) The most frequent telomeric motifs in the fission yeast *Schizosaccharomyces pombe* are G₂TTAC and G₂TTAC A. Only repetitions of the GGTTAC motif folded into G-quadruplexes and just in potassium. In order to verify if the *S. pombe* expanded consensus telomeric motif ggggggGGTTACAc (where lower case letters denote rare insertions) (34,35) is prone to fold into G-quadruplex structures a deeper investigation of combinations of possible *S. pombe* telomeric units is required. (v) For most of the telomeric sequences folding into G-quadruplexes, structural differences between the sodium and the potassium forms were revealed by CD. (vi) Finally, circular dichroism and PAGE supported folding of the *Giardia intestinalis* GGGTA, of the *Tetrahymena* GGGGTT and of the *S. cerevisiae* TGGGT G motif sequences into intramolecular and multi-stranded structures in potassium; the independence of their *T_m*s from oligonucleotide strand concentration suggests that these multi-stranded structures may be stacked intramolecular G-quadruplexes.

Besides allowing comparison of G-quadruplex folding and stability of eukaryotic telomeric sequences, the present study enriches the dataset necessary to build and refine algorithms predicting G-quadruplex stability (82).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank A. Guédin (Pessac) for her help in a few experiments. We thank N. Smith (Pessac) for her careful reading of the manuscript. We thank L. Lacroix (Paris), and A. Bourdoncle (Pessac) for helpful discussions.

FUNDING

The Institut National de la Santé et de la Recherche Médicale (INSERM), Centre National de la Recherche Scientifique (CNRS), Muséum National d'Histoire Naturelle (MNHN); *Fondation pour la Recherche Médicale* (FRM); University of Bordeaux 2; the *Agence Nationale de la Recherche* (G4-Toolbox ANR-09-BLAN-0355); Région Aquitaine grants. Funding for open access charge: Institut National de la Santé et de la Recherche Médicale (INSERM).

Conflict of interest statement. None declared.

REFERENCES

- Wellinger, R.J. and Sen, D. (1997) The DNA structures at the ends of eukaryotic chromosomes. *Eur. J. Cancer.*, **33**, 735–749.
- 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. *Proc. Natl Acad. Sci. USA*, **85**, 6622–6626.

3. Meyne, J., Ratliff, R.L. and Moyzis, R.K. (1989) Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl Acad. Sci. USA*, **86**, 7049–7053.
4. Blackburn, E.H. and Challoner, P.B. (1984) Identification of a telomeric DNA sequence in *Trypanosoma brucei*. *Cell*, **36**, 447–457.
5. Van der Ploeg, L.H., Liu, A.Y. and Borst, P. (1984) Structure of the growing telomeres of Trypanosomes. *Cell*, **36**, 459–468.
6. Schechtman, M.G. (1990) Characterization of telomere DNA from *Neurospora crassa*. *Gene*, **88**, 159–165.
7. Emery, H.S. and Weiner, A.M. (1981) An irregular satellite sequence is found at the termini of the linear extrachromosomal rDNA in *Dictyostelium discoideum*. *Cell*, **26**, 411–419.
8. Forney, J., Henderson, E.R. and Blackburn, E.H. (1987) Identification of the telomeric sequence of the acellular slime molds *Didymium iridis* and *Physarum polycephalum*. *Nucleic Acids Res.*, **15**, 9143–9152.
9. Sinclair, C.S., Richmond, R.H. and Ostrander, G.K. (2007) Characterization of the telomere regions of scleractinian coral, *Acropora surculosa*. *Genetica*, **129**, 227–233.
10. Le Blancq, S.M., Kase, R.S. and Van der Ploeg, L.H. (1991) Analysis of a *Giardia lamblia* rRNA encoding telomere with [TTAGGG]_n as the telomere repeat. *Nucleic Acids Res.*, **19**, 5790.
11. Fuchs, J., Brandes, A. and Schubert, I. (1995) Telomere sequence localisation and karyotype evolution in higher plants. *Plant Syst. Evol.*, **196**, 227–241.
12. Richards, E.J. and Ausubel, F.M. (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell*, **53**, 127–136.
13. Petracek, M.E., Lefebvre, P.A., Silflow, C.D. and Berman, J. (1990) *Chlamydomonas* telomere sequences are A+T-rich but contain three consecutive G-C base pairs. *Proc. Natl Acad. Sci. USA*, **87**, 8222–8226.
14. Okazaki, S., Tsuchida, K., Maekawa, H., Ishikawa, H. and Fujiwara, H. (1993) Identification of a pentanucleotide telomeric sequence, (TTAGG)_n, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.*, **13**, 1424–1432.
15. Muller, F., Wicky, C., Spicher, A. and Tobler, H. (1991) New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in *Ascaris lumbricoides*. *Cell*, **67**, 815–822.
16. Cangiano, G. and La Volpe, A. (1993) Repetitive DNA sequences located in the terminal portion of the *Caenorhabditis elegans* chromosomes. *Nucleic Acids Res.*, **21**, 1133–1139.
17. Wicky, C., Villeneuve, A.M., Lauper, N., Codourey, L., Tobler, H. and Muller, F. (1996) Telomeric repeats (TTAGGC)_n are sufficient for chromosome capping function in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA*, **93**, 8983–8988.
18. Blackburn, E.H. and Gall, J.G. (1978) A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol.*, **120**, 33–53.
19. 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. *Proc. Natl Acad. Sci. USA*, **78**, 3015–3019.
20. Dore, E., Pace, T., Ponzi, M., Scotti, R. and Frontali, C. (1986) Homologous telomeric sequences are present in different species of the genus *Plasmodium*. *Mol. Biochem. Parasitol.*, **21**, 121–127.
21. Bottius, E., Bakhsis, N. and Scherf, A. (1998) *Plasmodium falciparum* telomerase: *de novo* telomere addition to telomeric and nontelomeric sequences and role in chromosome healing. *Mol. Cell. Biol.*, **18**, 919–925.
22. Ganai, M.W., Lapitan, N.L. and Tanksley, S.D. (1991) Macrostructure of the tomato telomeres. *Plant Cell*, **3**, 87–94.
23. Baroin, A., Prat, A. and Caron, F. (1987) Telomeric site position heterogeneity in macronuclear DNA of *Paramecium primaurelia*. *Nucleic Acids Res.*, **15**, 1717–1728.
24. Forney, J.D. and Blackburn, E.H. (1988) Developmentally controlled telomere addition in wild-type and mutant paramecia. *Mol. Cell. Biol.*, **8**, 251–258.
25. McCormick-Graham, M. and Romero, D.P. (1996) A single telomerase RNA is sufficient for the synthesis of variable telomeric DNA repeats in ciliates of the genus *Paramecium*. *Mol. Cell. Biol.*, **16**, 1871–1879.
26. Teixeira, M.T. and Gilson, E. (2005) Telomere maintenance, function and evolution: the yeast paradigm. *Chromosome Res.*, **13**, 535–548.
27. Lue, N.F. Plasticity of telomere maintenance mechanisms in yeast. *Trends Biochem. Sci.*, **35**, 8–17.
28. McEachern, M.J. and Blackburn, E.H. (1994) A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. *Proc. Natl Acad. Sci. USA*, **91**, 3453–3457.
29. Cohn, M., McEachern, M.J. and Blackburn, E.H. (1998) Telomeric sequence diversity within the genus *Saccharomyces*. *Curr. Genet.*, **33**, 83–91.
30. Shampay, J., Szostak, J.W. and Blackburn, E.H. (1984) DNA sequences of telomeres maintained in yeast. *Nature*, **310**, 154–157.
31. Wang, S.S. and Zakian, V.A. (1990) Sequencing of *Saccharomyces* telomeres cloned using T4 DNA polymerase reveals two domains. *Mol. Cell. Biol.*, **10**, 4415–4419.
32. Singer, M.S. and Gottschling, D.E. (1994) TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase. *Science*, **266**, 404–409.
33. Leonardi, J., Box, J.A., Bunch, J.T. and Baumann, P. (2008) TER1, the RNA subunit of fission yeast telomerase. *Nat. Struct. Mol. Biol.*, **15**, 26–33.
34. Webb, C.J. and Zakian, V.A. (2008) Identification and characterization of the *Schizosaccharomyces pombe* TER1 telomerase RNA. *Nat. Struct. Mol. Biol.*, **15**, 34–42.
35. Hiraoka, Y., Henderson, E. and Blackburn, E.H. (1998) Not so peculiar: fission yeast telomere repeats. *Trends Biochem. Sci.*, **23**, 126.
36. Trujillo, K.M., Bunch, J.T. and Baumann, P. (2005) Extended DNA binding site in Pot1 broadens sequence specificity to allow recognition of heterogeneous fission yeast telomeres. *J. Biol. Chem.*, **280**, 9119–9128.
37. Neidle, S. and Balasubramanian, S. (2006) *Quadruplex Nucleic Acids*. RSC Biomolecular Sciences, Cambridge.
38. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871–880.
39. Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
40. Sen, D. and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*, **344**, 410–414.
41. Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I. Jr and Blackburn, E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **51**, 899–908.
42. Oka, Y. and Thomas, C.A. Jr (1987) The cohering telomeres of *Oxytricha*. *Nucleic Acids Res.*, **15**, 8877–8898.
43. 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. *Proc. Natl Acad. Sci. USA*, **98**, 8572–8577.
44. Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D. and Lipps, H.J. (2008) Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat. Struct. Mol. Biol.*, **15**, 598–604.
45. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. and Lipps, H. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nat. Struct. Mol. Biol.*, **12**, 847–854.
46. Petracek, M.E. and Berman, J. (1992) *Chlamydomonas reinhardtii* telomere repeats form unstable structures involving guanine-guanine base pairs. *Nucleic Acids Res.*, **20**, 89–95.
47. Murchie, A.I. and Lilley, D.M. (1994) Tetraplex folding of telomere sequences and the inclusion of adenine bases. *EMBO J.*, **13**, 993–1001.
48. Smith, F.W. and Feigon, J. (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature*, **356**, 164–168.
49. Wang, Y. and Patel, D.J. (1995) Solution structure of the *Oxytricha* telomeric repeat d[G4(T4G4)3] G-tetraplex. *J. Mol. Biol.*, **251**, 76–94.

50. Wang, Y. and Patel, D.J. (1994) Solution structure of the Tetrahymena telomeric repeat d(T2G4)4 G-tetraplex. *Structure*, **2**, 1141–1156.
51. Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263–282.
52. Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
53. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
54. 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.
55. 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.
56. Dai, J., Carver, M., Punchihewa, C., Jones, R.A. and Yang, D. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4940.
57. 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.
58. Phan, A.T., Kuryavyi, V., Luu, K.N. and Patel, D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution. *Nucleic Acids Res.*, **35**, 6517–6525.
59. Amrane, S., Ang, R.W., Tan, Z.M., Li, C., Lim, J.K., Lim, J.M., Lim, K.W. and Phan, A.T. (2009) A novel chair-type G-quadruplex formed by a Bombyx mori telomeric sequence. *Nucleic Acids Res.*, **37**, 931–938.
60. Hu, L., Lim, K.W., Bouaziz, S. and Phan, A.T. (2009) Giardia telomeric sequence d(TAGGG)4 forms two intramolecular G-quadruplexes in K⁺ solution: effect of loop length and sequence on the folding topology. *J. Am. Chem. Soc.*, **131**, 16824–16831.
61. Venczel, E.A. and Sen, D. (1993) Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry*, **32**, 6220–6228.
62. Hazel, P., Huppert, J., Balasubramanian, S. and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
63. Guédin, A., De Cian, A., Gros, J., Lacroix, L. and Mergny, J.L. (2008) Sequence effects in single-base loops for quadruplexes. *Biochimie*, **90**, 686–696.
64. Guédin, A., Gros, J., Alberti, P. and Mergny, J.L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.*, **38**, 7858–7868.
65. De Cian, A., Grellier, P., Mouray, E., Depoix, D., Bertrand, H., Monchaud, D., Teulade-Fichou, M.P., Mergny, J.L. and Alberti, P. (2008) Plasmodium telomeric sequences: structure, stability and quadruplex targeting by small compounds. *ChemBiochem*, **9**, 2730–2739.
66. Mergny, J.L. and Lacroix, L. (2009) UV melting of G-quadruplexes. *Curr. Protoc. Nucleic Acids Chem.*, 17.1.1–17.1.15.
67. 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.
68. Kejnovska, I., Kypr, J. and Vorlickova, M. (2007) Oligo(dT) is not a correct native PAGE marker for single-stranded DNA. *Biochem. Biophys. Res. Commun.*, **353**, 776–779.
69. Mergny, J.L., Phan, A.T. and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
70. Gray, D.M., Wen, J.D., Gray, C.W., Repges, R., Repges, C., Raabe, G. and Fleischhauer, J. (2008) Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality*, **20**, 431–440.
71. 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.
72. Risitano, A. and Fox, K.R. (2003) Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry*, **42**, 6507–6513.
73. Rachwal, P.A., Findlow, I.S., Werner, J.M., Brown, T. and Fox, K.R. (2007) Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res.*, **35**, 4214–4222.
74. Rachwal, P.A., Brown, T. and Fox, K.R. (2007) Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry*, **46**, 3036–3044.
75. Bugaut, A. and Balasubramanian, S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry*, **47**, 689–697.
76. Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E.S., Bowers, M.T., De Pauw, E. and Gabelica, V. (2008) G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.*, **130**, 10208–10216.
77. Guédin, A., Alberti, P. and Mergny, J.L. (2009) Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic Acids Res.*, **37**, 5559–5567.
78. Kumar, N. and Maiti, S. (2008) A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **36**, 5610–5622.
79. Crabbe, L., Verdun, R.E., Haggblom, C.I. and Karlseder, J. (2004) Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science*, **306**, 1951–1953.
80. 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.
81. Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T. et al. (2007) A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. *BMC Biol.*, **5**, 28.
82. Stegle, O., Payet, L., Mergny, J.L., MacKay, D.J. and Huppert, J.L. (2009) Predicting and understanding the stability of G-quadruplexes. *Bioinformatics*, **25**, i374–i382.

Stability of telomeric G-quadruplexes

Phong Lan Thao Tran, Jean-Louis Mergny & Patrizia Alberti

Supplementary Data

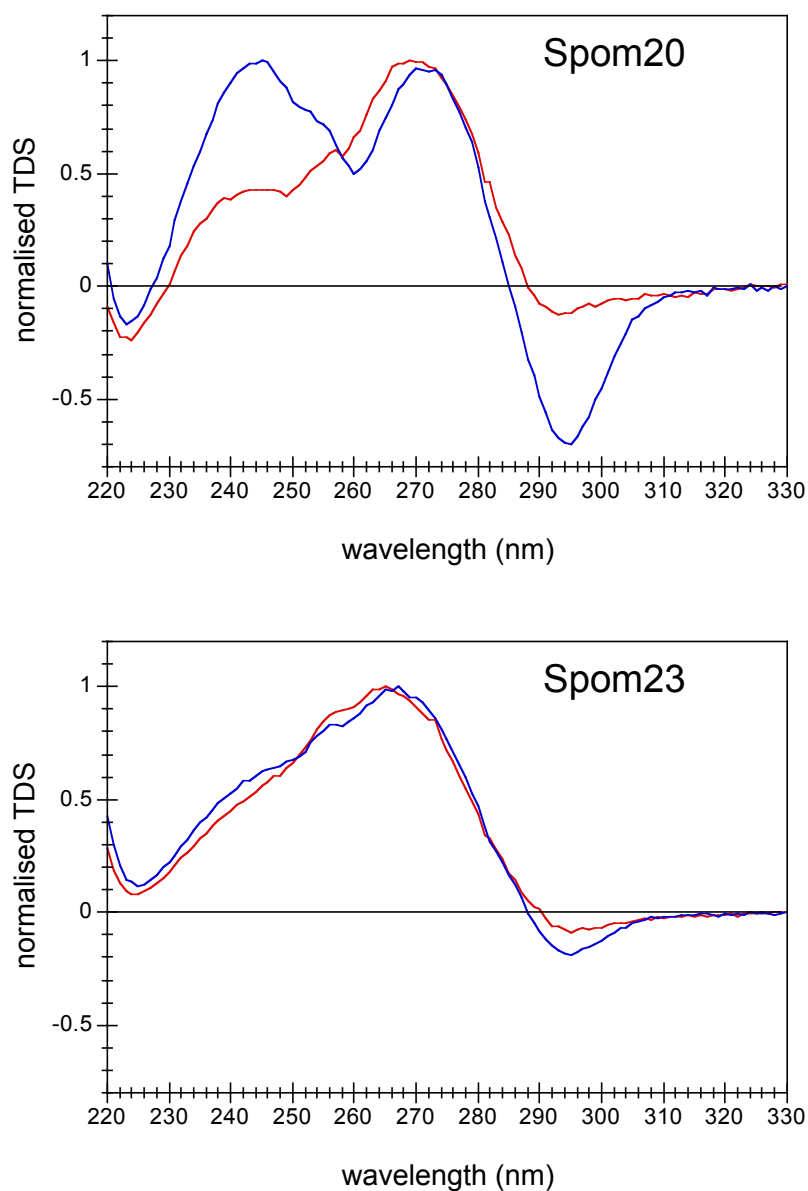


Figure S1

Normalised TDS of Spom20 and Spom23 in NaCl (red) and KCl (blue) (100 mM)
at 3 μ M oligonucleotide strand concentration.

For Spom20 in NaCl and Spom23 in NaCl and KCl, the small negative peak at 295 nm
(small compared to TDS of Spom20 in KCl and to other TDS of G-quadruplexes)
suggests that a very minor fraction of oligonucleotide folds into G-quadruplexes.

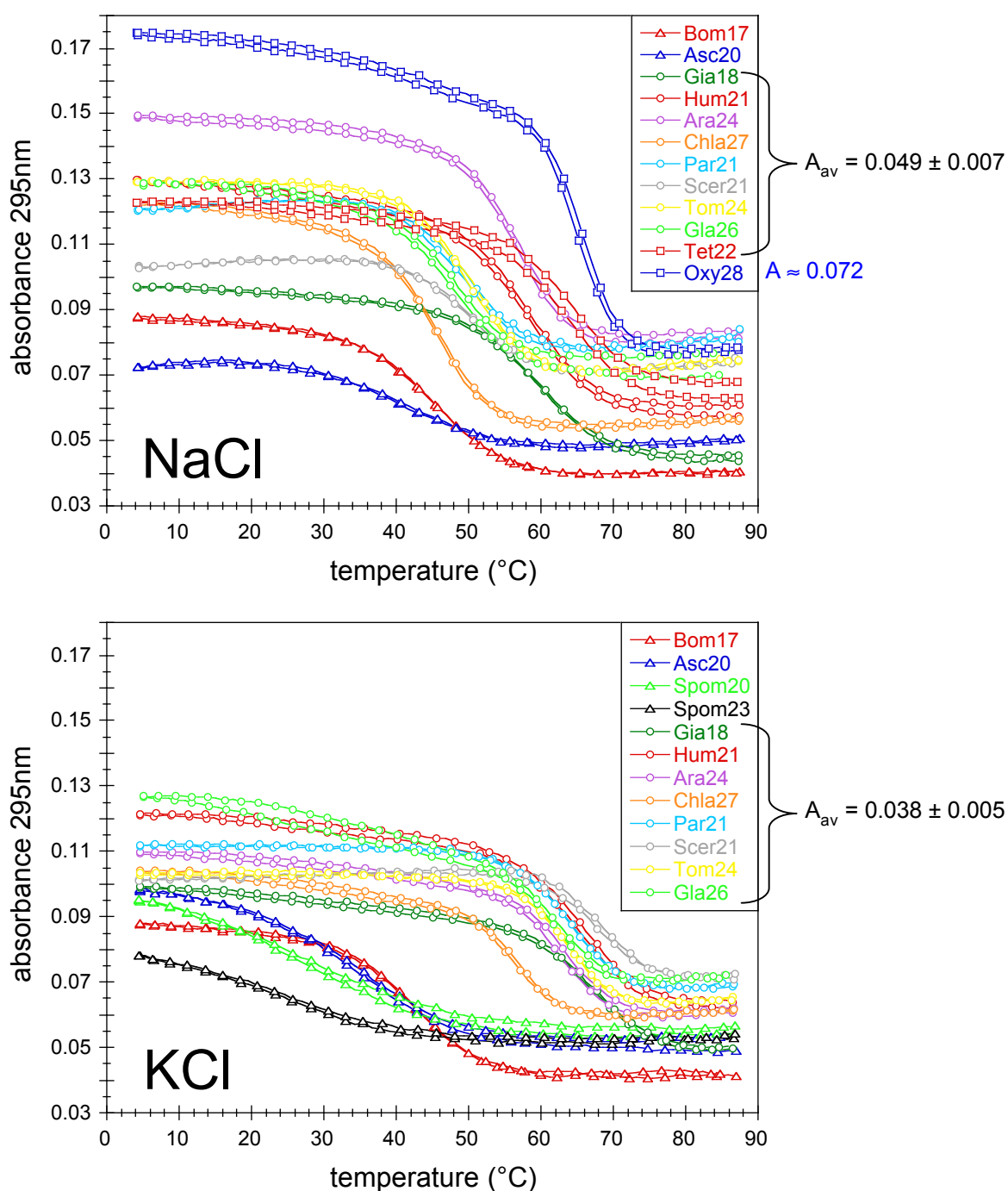


Figure S2

Melting curves (cooling and heating) at 295 nm in NaCl and KCl (100 mM),
at 3 μ M oligonucleotide strand concentration (1 cm optical path length).

The amplitudes (A) or average amplitude (A_{av}) of the transitions (in absorbance) are reported next to the graph legend.
 A were calculated at T_m , after low- and high-temperature absorbance linear baseline determination.

Spom20 and Spom23 did not fold into G-quadruplexes in NaCl.

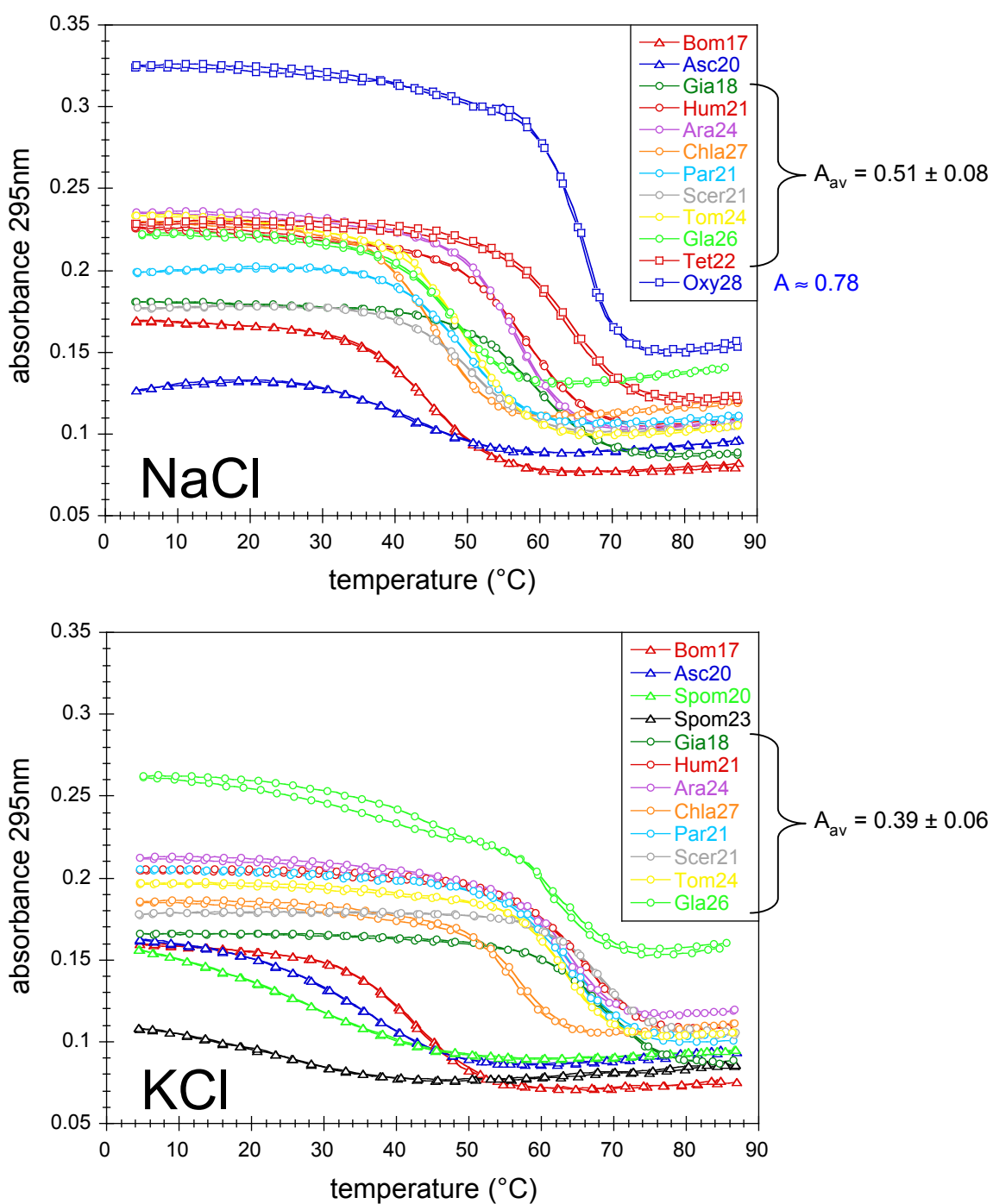
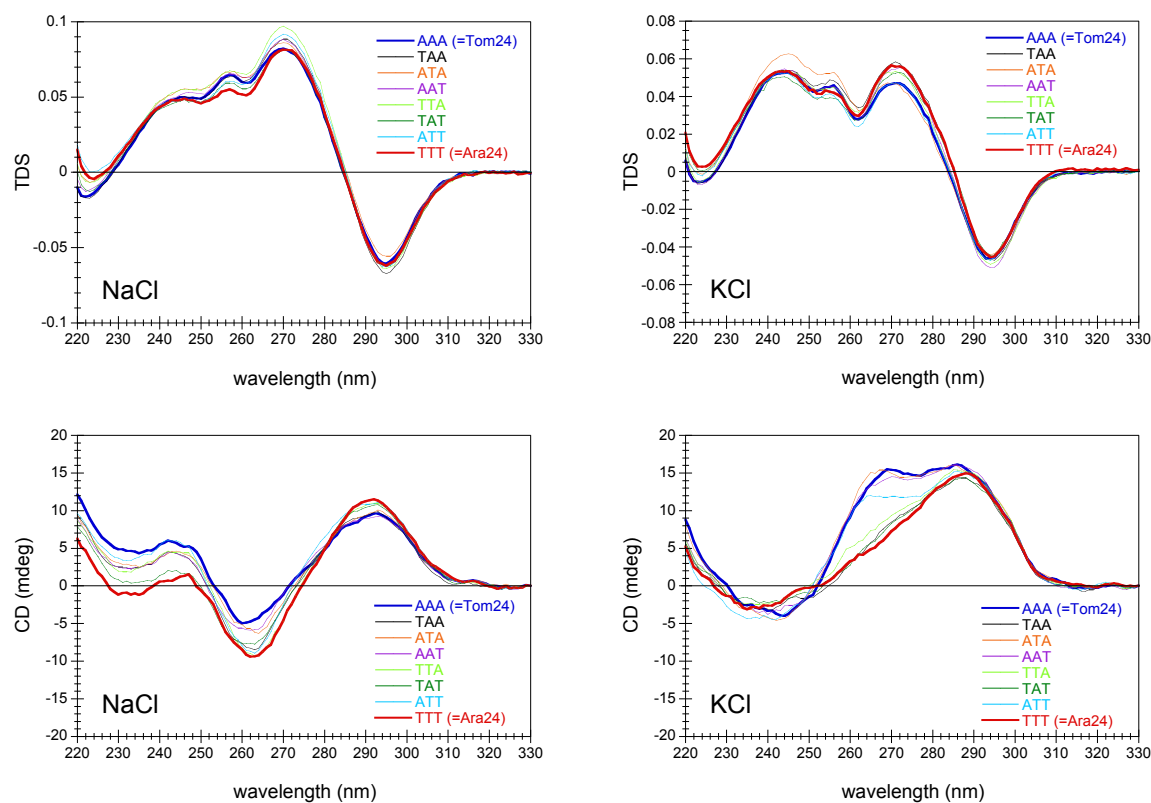


Figure S3

Melting curves (cooling and heating) at 295 nm in NaCl and KCl (100 mM), at 30 μ M oligonucleotide strand concentration (0.2 cm optical path length).

The amplitudes (A) or average amplitude (A_{av}) of the transitions (in absorbance) are reported next to the graph legend. A were calculated at T_m , after low- and high-temperature absorbance linear baseline determination. The amplitudes A and A_{av} were multiplied by 1/0.2 to take into account the optical path length and allow a straight comparison with the amplitudes at 3 μ M oligonucleotide strand concentration.

Spom20 and Spom23 did not fold into G-quadruplexes in NaCl.



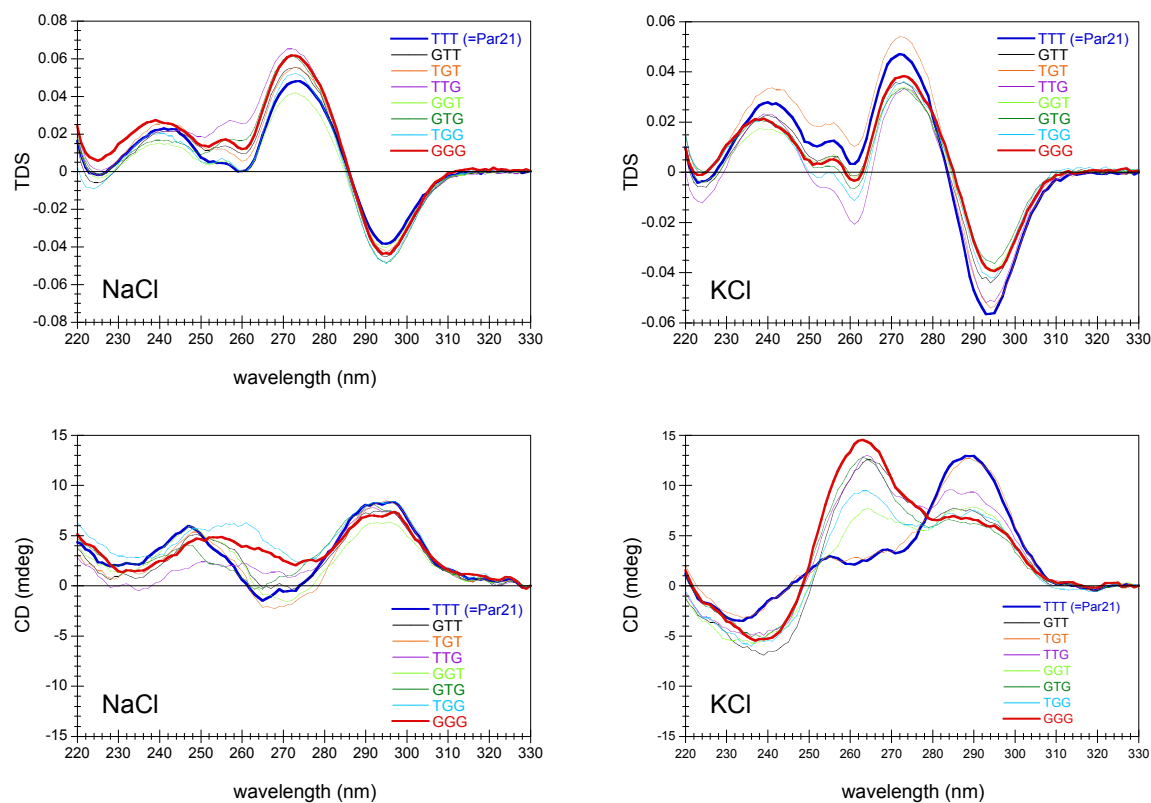
Name sequence	Sequence (from 5' to 3')
Tom24	GGGTTAAGGGTTAAGGGTTAAGGG
TAA	GGGTTTAGGGTTAAGGGTTAAGGG
ATA	GGGTTAAGGGTTTAGGGTTAAGGG
AAT	GGGTTAAGGGTTAAGGGTTTAGGG
TTA	GGGTTTAGGGTTTAGGGTTAAGGG
TAT	GGGTTTAGGGTTAAGGGTTTAGGG
ATT	GGGTTAAGGGTTTAGGGTTTAGGG
Ara24	GGGTTTAGGGTTTAGGGTTTAGGG

Figure S4

TDS and CD spectra of the possible combinations of 3.5 repetitions of the tomato plant degenerate telomeric motif in NaCl and in KCl (100mM) at 3 μ M oligonucleotide strand concentration.

CD spectra were acquired at 20°C on a Jasco-815 after annealing from 96°C at a rate of 1.3°C min⁻¹

For these measurements, new batches of Tom24 and Ara24 were used.
Melting temperatures are reported in Main Text.



Name sequence	Sequence (from 5' to 3')
Par21	GGGTTTGGGTTTGGGTTTGGG
GTT	GGGTTTGGGTTTGGGTTTGGG
TGT	GGGTTTGGGTTTGGGTTTGGG
TTG	GGGTTTGGGTTTGGGTTTGGG
GGT	GGGTTTGGGTTTGGGTTTGGG
GTG	GGGTTTGGGTTTGGGTTTGGG
TGG	GGGTTTGGGTTTGGGTTTGGG
GGG	GGGTTTGGGTTTGGGTTTGGG

Figure S5

TDS, CD spectra of the possible combinations of 3.5 repetitions of *Paramecium* degenerate telomeric motif in NaCl and in KCl (100 mM) at 3 μ M oligonucleotide strand concentration.

CD spectra were acquired at 20°C on a Jasco-815 after annealing from 96°C at a rate of 1.3°C min⁻¹.

For these measurements, a new batch of Par21 was used.
Melting temperatures are reported in Main Text.

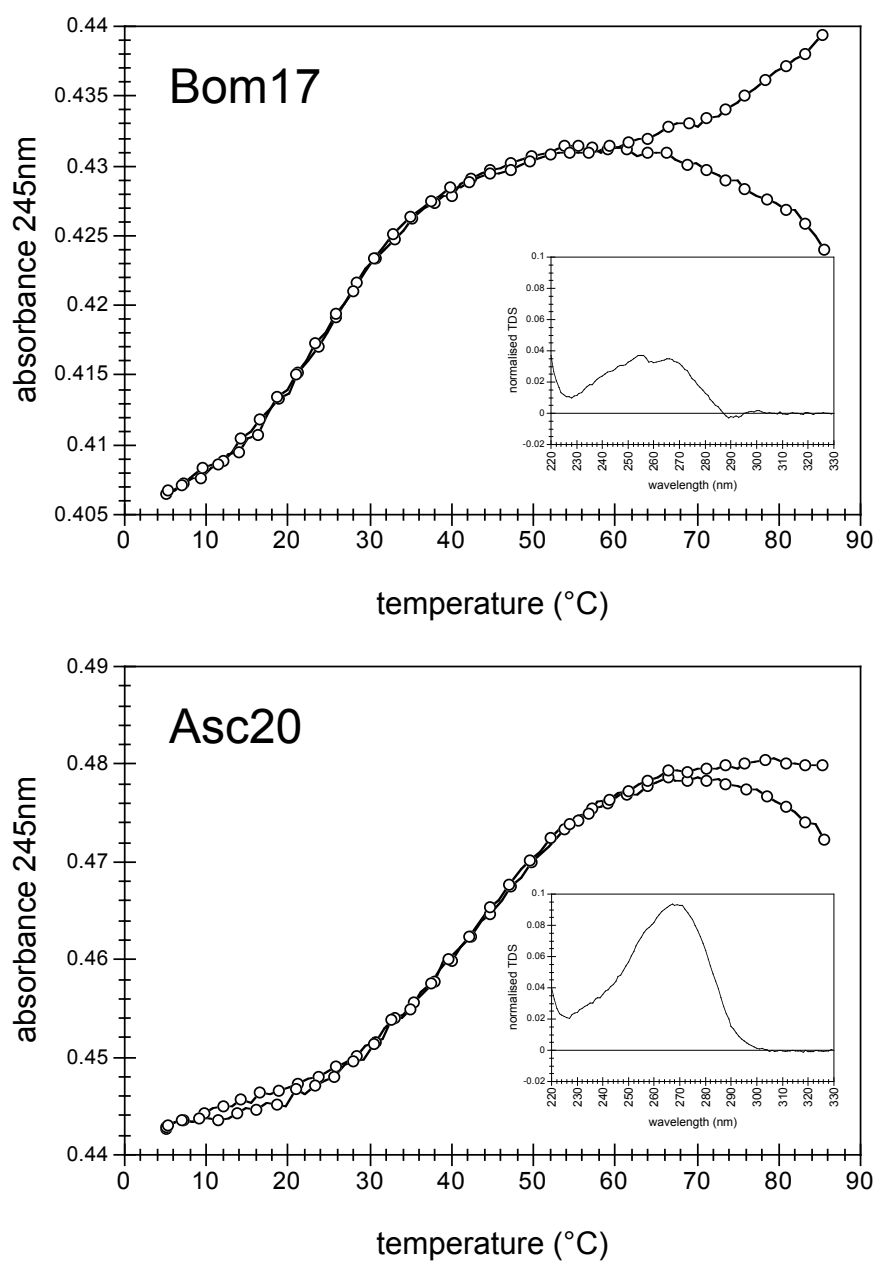


Figure S6

TDS and melting temperature at 245 nm of Bom17 and Asc20 in LiCl
(acid cacodylic 10 mM, pH 7.2 adjusted with LiOH, LiCl 100 mM)
at 3 μ M oligonucleotide strand concentration.

TDS indicated that Bom17 and Asc20 did not fold into G-quadruplexes in LiCl;
melting experiments revealed a transition around 38°C for Asc20 and around 25°C for Bom17,
suggesting folding into non-G-quadruplex structures.

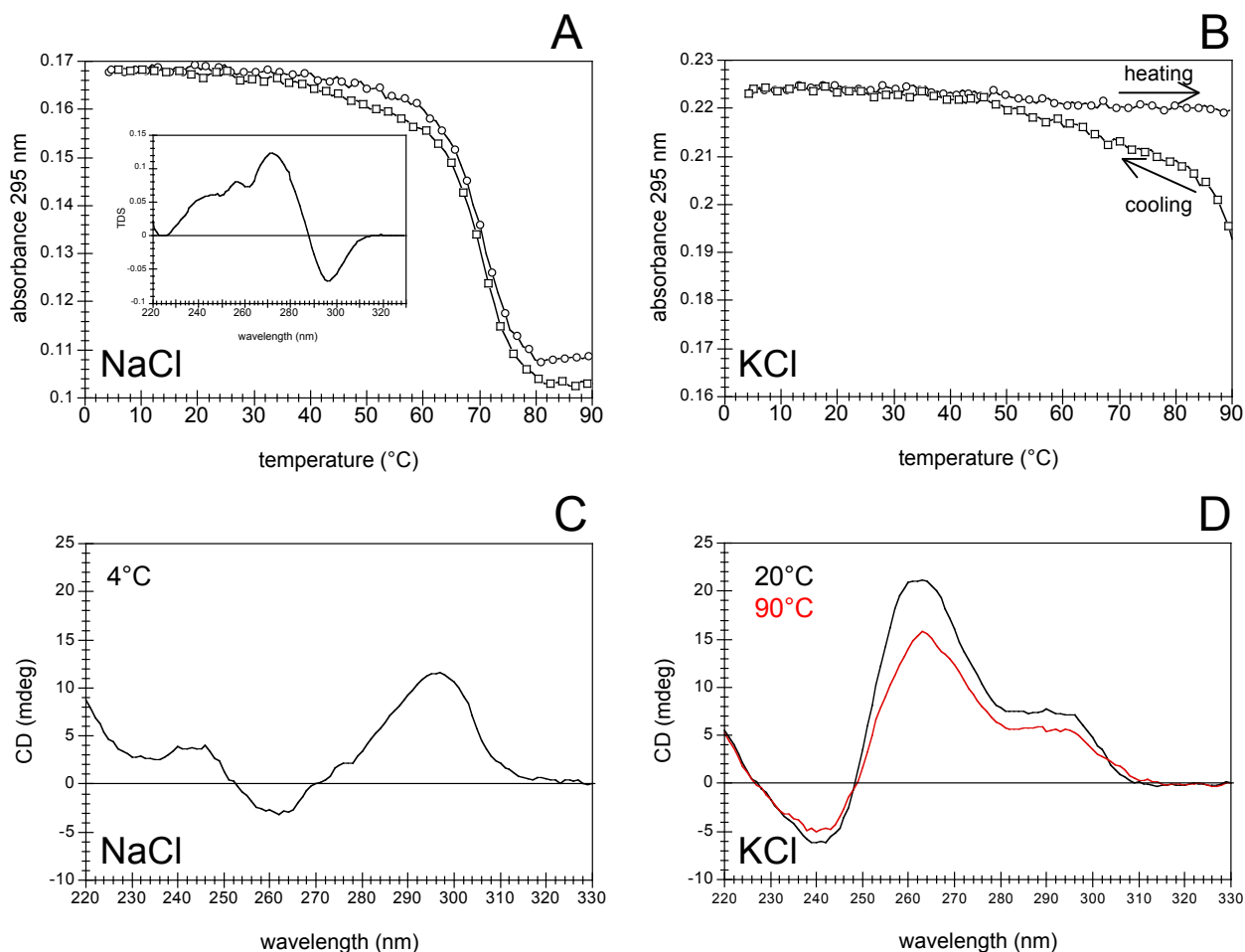


Figure S7

TDS, melting curves at 295 nm and CD spectra of the *C. merolae* telomeric sequence $5'(\text{G}_6\text{A}_2\text{T})_3\text{G}_6 3'$ at 3 μM strand concentration, in NaCl and KCl (100mM).

CD spectra at 4°C or 20°C were recorded after slowly cooling (in a metal block) at room temperature from 96°C.

In NaCl, the sequence folded into G-quadruplex(es), as revealed by TDS (Figure A, inset), with a T_m of 70°C (Figure A).

In KCl, no melting transition was observed either at 295 nm (Figure B) or 240 nm.

In KCl, the presence of structure(s) (very likely G-quadruplexes) was revealed by CD spectroscopy (Figure D).

The CD spectrum did not change upon heating the sample from 20°C to 90°C, indicating that the structure(s) did not melt.

3) *Résultats et Discussion : Chapitre V*

Cette étude sur les séquences télomériques de différents organismes nous a fourni des informations sur leur capacité à former des G-quadruplexes et sur la stabilité des structures formées.

Cette étude a révélé deux séries de motifs télomériques: les motifs des levures et ceux des autres eucaryotes. La capacité à former des G4 au niveau des télomères ne semble pas être une caractéristique conservée chez les levures. Contrairement aux motifs de *Saccharomyces cerevisiae* (TG₃TG) et de *Candida glabrata* (CTGTG₄TCTG₃TG), la séquence de *Candida guilliermondii* (TACTG₂TG) est incapable de former des G4. Parmi les deux motifs trouvés chez la levure fissipare *Schizosaccharomyces pombe* (G₂TTAC et G₂TTACA), seul le motif G₂TTAC forme des G4 mais seulement en présence de potassium.

Les autres séquences télomériques portant au moins trois guanines consécutives (en particulier les motifs G₃T₁₋₄A et G₄T_{2,4}) forment des G-quadruplexes stables aussi bien en présence de Na⁺ que de K⁺. La stabilité thermique des G4 formés par ces séquences télomériques augmente avec le nombre de guanines consécutives et dépend fortement de la longueur et de la séquence des boucles. A part l'algue rouge unicellulaire *Cyanidioschyzon merolae* (G₆AAT), aucun motif télomérique trouvé ne possède une adénine à la première position dans les boucles. En effet, il a été montré que l'adénine à cette position était défavorable à la formation de G4 [39, 315]. L'algue *Cyanidioschyzon merolae* est un organisme extrêmophile vivant dans les sources chaudes sulfureuses (pH 1,5 et 45°C) [222]. Son motif télomérique putatif est constitué de 6 guanines consécutives (G₆AAT) et forme des G4 extrêmement stables (T_m = 70°C dans Na⁺ et 90°C dans K⁺).

Les séquences télomériques contenant 2 guanines consécutives (G₂TTA trouvées chez certains insectes comme *Bombyx mori* et G₂CTTA trouvées chez les nématodes comme *Ascaris lumbricoides* et *Caenorhabditis elegans*) forment bien des G4 dans Na⁺ et dans K⁺. Cependant les résultats obtenus suggèrent que les G4 formés par ces deux séquences sont en équilibre avec d'autres structures non-G4 (comme des structures en épingle à cheveux).

Les spectres de CD de la plupart des séquences télomériques dépendent de la nature du cation utilisé (Na⁺ ou K⁺). Cela suggère que la nature du cation (Na⁺ ou K⁺) peut influencer la conformation de G4 ou l'équilibre des différentes conformations formées dans la solution. Seules les séquences de *Bombyx mori* (G₂TTA), d'*Ascaris lumbricoides* (G₂CTTA) et d'*Oxytricha nova* (G₄TTTT) présentent des signatures de CD similaires dans Na⁺ et dans K⁺. Cependant, il est difficile de déduire la topologie des G4 formés à partir des spectres de CD

pour des raisons cités dans l'**article 3** et dans le chapitre VII. Une étude structurale par RMN ou cristallographie au rayon X est nécessaire pour compléter ce travail.

Enfin, les profils CD à différentes concentrations en oligonucléotides et les résultats de PAGE soutiennent fortement la formation de G4 intramoléculaire et multi-brins par les séquences télomériques de *Giardia intestinalis* (G₃TA), de *Tetrahymena thermophila* (G₄TT) et de *Saccharomyces cerevisiae* (TG₃TG) dans K⁺. L'indépendance des T_m en fonction de la concentration en oligonucléotide laisse penser que ces structures multi-brins pouvaient correspondre à un empilement de G4 intramoléculaires.

Par ailleurs, cette étude comparative des G4 formés par des séquences télomériques des eucaryotes et de leur stabilité a enrichi les données nécessaires au développement d'un algorithme prédisant la stabilité des G4 [40] (et pour revue récente [316]).

Nous avons mentionné au début de ce chapitre (cf. : §V.1.1.2) que la taille de certaines extrémités 3' télomériques est assez longue et laisse imaginer un empilement de G-quadruplexes tout au long de la séquences. Il serait donc intéressant d'étudier la structure et la stabilité des G4 formés par des séquences télomériques longues, outre celles des vertébrés.

Chapitre VI. Développement de G4-FID sur plaque pour l'identification de ligands spécifiques d'une topologie particulière de G-quadruplexes.

1) Introduction : Études des ligands de G-quadruplexes

Nous avons mentionné dans l'introduction générale de ce manuscrit que les G-quadruplexes étaient susceptibles de se former dans de nombreuses régions du génome (télomères, promoteurs, ADN ribosomiaux, ARN messagers,...). Même si l'existence de G-quadruplexes *in vivo* n'a pas été formellement démontrée dans tous les cas, il est difficile d'ignorer les effets cellulaires (notamment au niveau de la régulation de l'expression de gènes, de la réplication, de la recombinaison et de la division des cellules) de certains ligands sélectifs de ces structures.

Le développement de ces ligands stabilisant ou induisant la formation de G-quadruplexes pourrait également être très prometteur pour les thérapies anticancéreuses. L'effet de certains ligands au niveau des télomères et de certains promoteurs a été étudié en particulier. Nous avons évoqué dans le chapitre précédent que la télomérase et la tumorigénèse étaient intimement liées, cette enzyme étant active dans 85 à 90% des cellules cancéreuses. En stabilisant la formation de G-quadruplexe au niveau des télomères, ces molécules entraînent un dysfonctionnement du télomère en empêchant la fixation des protéines télomériques et inhibent indirectement la télomérase. Cela induit, à plus long terme, un arrêt de prolifération accompagné de sénescence et un raccourcissement progressif des télomères dans diverses cellules cancéreuses traitées [317-321]. De plus, la stabilisation de G-quadruplexes dans le pré-ARNm de hTERT par des ligands entraîne une modification de l'épissage de l'ARNm et donc la production de sous-unité hTERT inactive [72]. Il a été également montré que la prolifération de cellules cancéreuses était liée à la surexpression de *c-myc*. La stabilisation de G-quadruplexes au niveau du promoteur de *c-myc* par **TMPyP4** peut réprimer l'expression de celui-ci [61](pour revue récente voir [322]).

Le **CX-3543** (aussi connu sous le nom de **quarfloxin**) est le premier ligand de G4 qui est entré en essais cliniques [323]. Bien qu'il y ait des problèmes de biodisponibilité, le profil de toxicité du ligand est très encourageant et il est bien toléré par les patients. Ces résultats encouragent le développement d'autres molécules ciblant spécifiquement les structures de G-quadruplexes.

Dans les paragraphes suivants, nous présenterons certaines familles de molécules de G-quadruplexes dont celles qui sont utilisées dans l'**article 4**. Nous discuterons ensuite des principales techniques disponibles pour l'étude de ces ligands. Et enfin, nous allons citer les avantages et les inconvénients de ces méthodes qui nous ont conduit à développer le test de criblage G4-FID sur plaque.

1.1 Classement des ligands de G-quadruplexes

La géométrie particulière de G-quadruplexes permet la reconnaissance spécifique de petits ligands. Il existe différents modes d'interactions entre les ligands et les G-quadruplexes (Figure 41) : interactions π sur les quartets terminaux du G-quadruplexe ("empilement terminal"), intercalation éventuelle des ligands entre les quartets, interactions avec les boucles, interactions avec les sillons du G-quadruplexe ou une combinaison de différents modes. Parmi les différents types d'interactions, l'intercalation des ligands entre les quartets semble peu probable, car elle demanderait un coût énergétique important pour "désempiler" deux quartets et chasser un cation coordonné entre deux quartets.

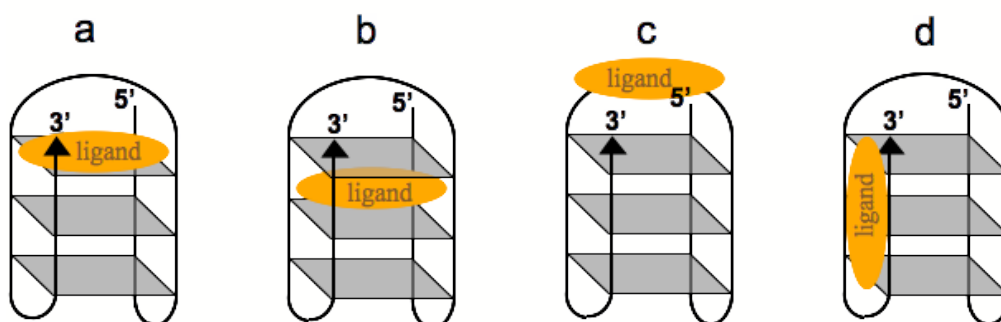


Figure 41. Schéma représentatif des différents sites d'interactions entre ligands et G-quadruplexe
(a) "Empilement terminal" ; (b) intercalation entre les quartets, jugée improbable ; (c) interactions avec les boucles ; (d) Interactions avec les sillons du G-quadruplexe.

1.1.1 Ligands basés sur des interactions d'empilement $\pi - \pi$

La surface plane et hydrophobe d'un quartet de guanines présente un site favorable à la fixation de molécules présentant une large surface aromatique (par empilement π - π). Cette dernière doit être plus étendue que celle des ligands de duplexes afin d'augmenter la sélectivité pour les G-quadruplexes. Cependant, il a fallu concilier cette aromaticité importante avec bonne solubilité du ligand. Ceci est souvent permis par l'ajout de charges à la molécule. De nombreux ligands de G4 ont été développés en ce sens. La stabilisation de quadruplexes peut être non seulement basée sur un empilement π - π mais aussi sur des interactions électrostatiques entre les charges positives de ligands et l'ADN G-quadruplexe

chargé négativement. La protonation des ligands peut être réalisée de différentes façons : (i) par ajout de chaînes latérales qui sont protonées *in situ*, (ii) par N-méthylation, (iii) par chélation de cation métallique (nickel, manganèse ou cuivre). Les ligands peuvent donc être classés en quatre catégories en se basant sur leur nature cationique [324].

a) Ligands à chaînes latérales protonées *in situ*

Nous avons mentionné précédemment que le développement des ligands devait concilier l'importante aromaticité qui permet des interactions π et la solubilité. Un moyen pour arriver à cela est d'introduire des chaînes latérales protonées autour d'un centre aromatique hydrophobe. Parmi les ligands utilisés dans l'article 4, **BRACO19**, **PIPER** et **TrisK3-NH** appartiennent à cette catégorie (Figure 42).

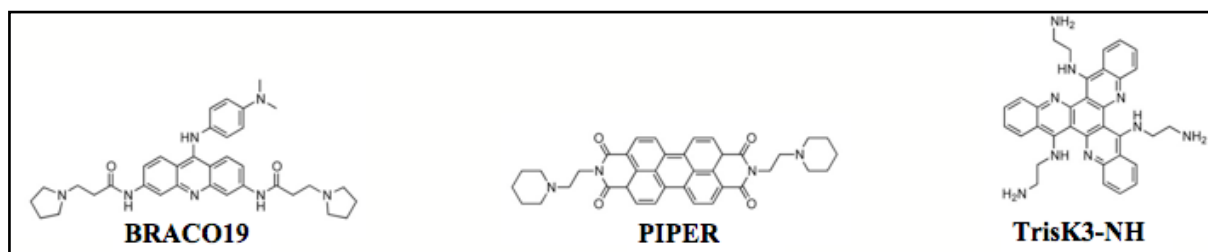


Figure 42. Exemples de ligands de G-quadruplexes à chaînes latérales protonables *in situ*

b) Ligands N-méthylés

Un autre moyen de protonation est d'utiliser des ligands N-méthylés. Ceci permet une bonne solubilité dans l'eau et augmente la capacité d'empilement π du ligand en réduisant la densité électronique de la partie aromatique. Les ligands utilisés pour le test FID (pour "Fluorescence intercalator displacement") sont souvent N-méthylés, notamment **BSU1051**, **12459**, **360A** (et ses dérivés : **307A**, **832A**), **Phen-DC3/6** et **Bipy-DC3/6** (Figure 43).

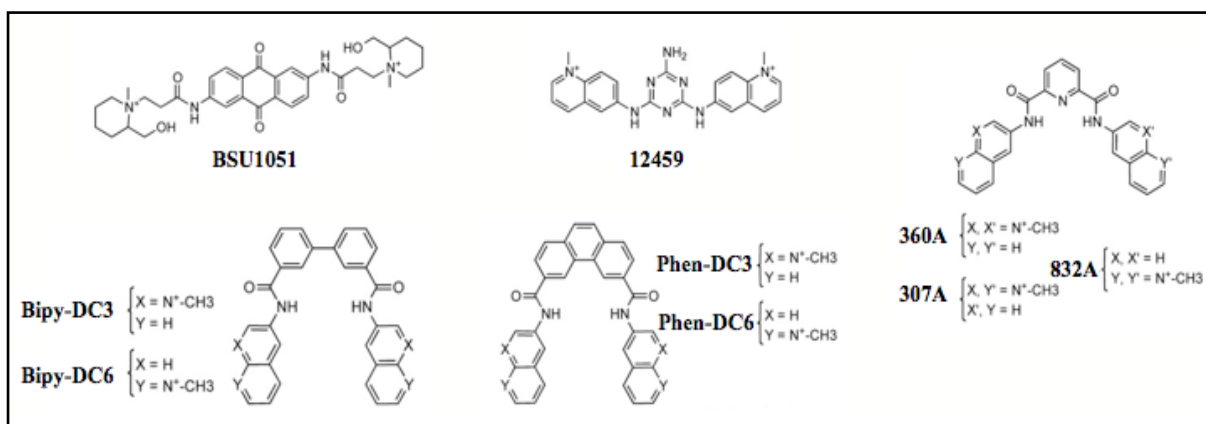


Figure 43. Exemples de ligands de G-quadruplexes protonés par N-méthylation

c) Ligands métallo-organiques

L'utilisation de molécules métallo-organiques est également envisagée pour le développement de ligands. Cette approche est basée sur l'insertion d'un métal central qui permettrait de le positionner sur le canal cationique du quadruplexe optimisant les interactions d'empilement $\pi - \pi$ [325]. La nature cationique de ces molécules permet d'améliorer l'association du ligand avec l'ADN G-quadruplexe chargé négativement. **Cu-ttpty** utilisé dans le test G4-FID représente bien cette catégorie (Figure 44).

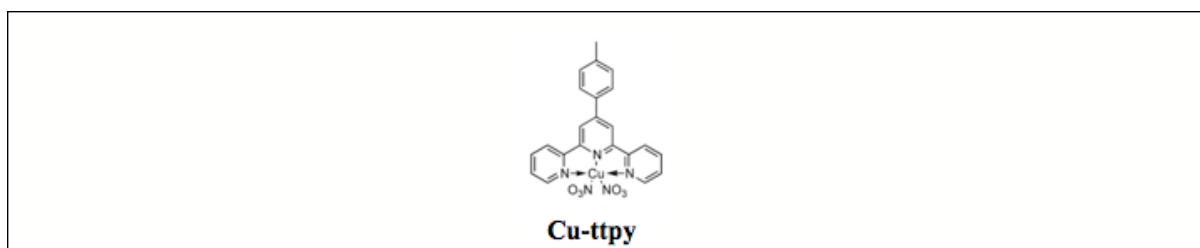


Figure 44. Exemple de ligands de G-quadruplexes métallo-organiques

d) Ligands neutres ou négativement chargés

Étonnamment, certains des ligands identifiés possèdent une bonne affinité pour les G-quadruplexes sans être chargés positivement. Ce type d'interaction serait principalement guidé par recouvrement du G-quartet par le ligand et des interactions de type π . Cette catégorie n'est pas très large. Les exemples les plus connus appartenant à cette catégorie sont la porphyrine **NMM**, la **télomestatine**, une molécule naturelle isolée de la bactérie *Streptomyces annulatus* en 2001 [326] et son dérivé heptacyclique, le **TOxaPy** qui sera discuté dans le chapitre VII (Discussion et Perspectives) (Figure 45).

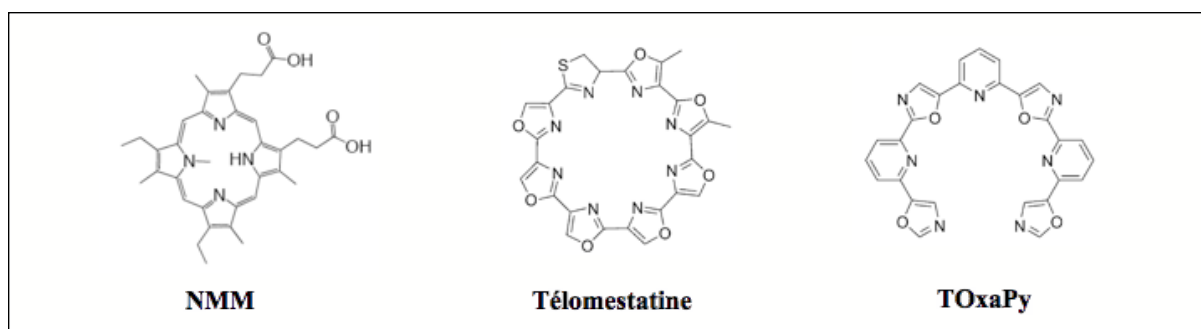


Figure 45. Exemples de ligands de G-quadruplexes neutres ou négativement chargés

1.1.2 Ligands des boucles et des sillons de G-quadruplexe

Les quatre sillons et les boucles de G-quadruplexes constituent également des sites potentiels pour le développement de ligands. Leur géométrie est non seulement différente de celles des duplexes d'ADN, mais elle varie également en fonction de la topologie du quadruplexe. On distingue trois types de sillons pour les G-quadruplexes: étroit, moyen ou large selon la géométrie *syn* / *anti* des guanines sur deux brins contigus. Ce type d'interaction semble être prometteur pour cibler spécifiquement les G-quadruplexes par rapport aux duplexes, mais également pour discriminer une topologie de G-quadruplexe en particulier, en fonction des géométries de ses sillons et de ses boucles. Ceci présente un intérêt particulier puisqu'il est éventuellement possible de discriminer certains types de G-quadruplexes, ce qui paraît peu envisageable avec des composés interagissant uniquement par empilement sur les quartets. Les exemples de ligands présentant ce mode d'interaction seraient les carbocyanines **DODC** et **DTDC** (Figure 46).

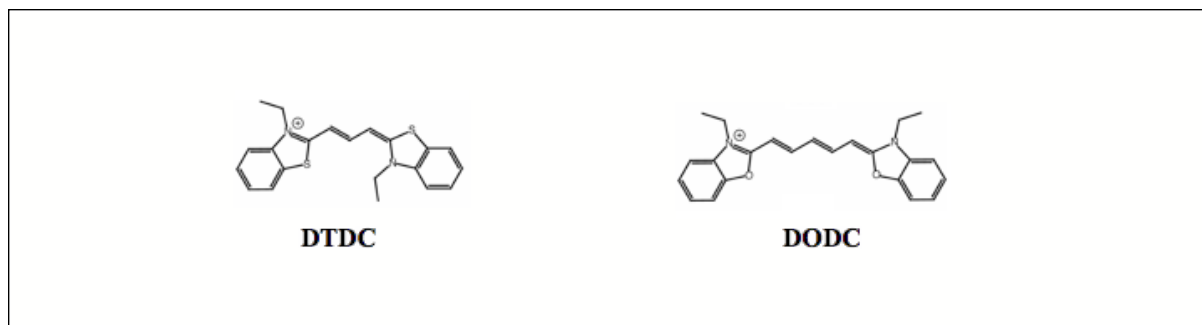


Figure 46. Exemples de ligands de G-quadruplexes interagissant avec les boucles et les sillons des G-quadruplexes

1.2 Techniques existantes pour l'étude des ligands

Afin d'étudier les interactions de ligands avec les structures en G-quadruplexe, de nombreuses techniques, plus ou moins directes, ont été développées. Certaines de ces méthodes seront discutées ci-dessous.

1.2.1 Spectroscopie en UV-Visible

Une méthode conventionnelle utilisée pour identifier les interactions ligand-G4 est la spectroscopie d'absorption UV-visible. Elle est liée à un effet bathochromique (augmentation de la longueur d'onde maximale, λ_{max}) et à un effet hypochromique (diminution de l'absorbance, A) lorsqu'un G-quadruplexe se lie à un chromophore, comme **TMPyP4**. M.W. Freyer et ses collègues ont étudié l'interaction de **TMPyP4** avec *c-myc* par cette approche

[327]. En présence du ligand, ils ont noté une importante déplacement de λ_{max} (de 422 à 440 nm) et une diminution remarquable de l'absorbance (~35%) (Figure 47). Ces observations pourraient indiquer une intercalation de **TMPyP4** dans la structure G-quadruplexe, ce que d'autres chercheurs contestent vigoureusement. Les données cristallographiques sur ce dérivé ne montrent pas d'intercalation dans un quadruplexe [328].

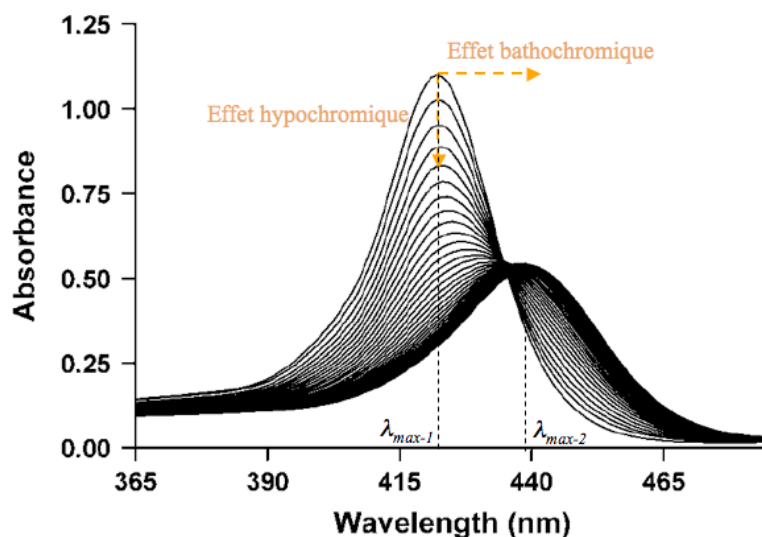


Figure 47. Interactions ligand-G4 observées en spectroscopie UV-visible [327].

L'ajout progressif de **TMPyP4** dans une solution contenant le quadruplexe *c-myc* entraîne une augmentation de λ_{max} de 422 à 440 nm (effet bathochromique) et une diminution d'environ 35% de l'absorbance (effet hypochromique).

1.2.2 Dichroïsme circulaire

Comme nous l'avons mentionné dans le chapitre II (Matériel et Méthodes), le dichroïsme circulaire fournit des informations sur la conformation de G-quadruplexes et donc également sur la structure du complexe G4-ligand. En général, les molécules sont non chirales et n'ont donc aucun signal CD. Une fois liée aux G-quadruplexes, la molécule non chirale induit un signal CD caractéristique [329, 330]. Ce signal appelé "ICD" (pour "Induced CD") correspond à l'absorption du ligand non chirale qui constitue une preuve de son interaction avec l'ADN. Un signal CD induit important serait plutôt en accord avec une fixation dans les sillons qu'un "empilement terminal".

1.2.3 Spectroscopie RMN et cristallographie aux rayons X

De nombreuses structures de G4 ADN ou ARN ont été obtenues par spectroscopie RMN (Résonance Magnétique Nucléaire) et par cristallographie aux rayons X. De la même

façon, ces techniques permettent l'étude structurale du complexe G4-ligand(s). Notamment en 2005, A.T. Phan a publié la structure RMN du complexe **TMPyP4** avec *c-myc* (d[$\text{TGAG}_3\text{TG}_2\text{TGAG}_3\text{TG}_4\text{A}_2\text{G}_2\text{A}$]) (PDB : 2A5R) [331] et en 2008, N.H. Campbell et G.N. Parkinson ont montré, par cristallographie aux rayons X, l'interaction de **BRACO19** avec la structure de G-quadruplexes parallèle formé par la séquence télomérique humaine (d[$\text{TAG}_3\text{T}_2\text{AG}_3$]₂) (PDB : 3CE5) [332] (Figure 48).

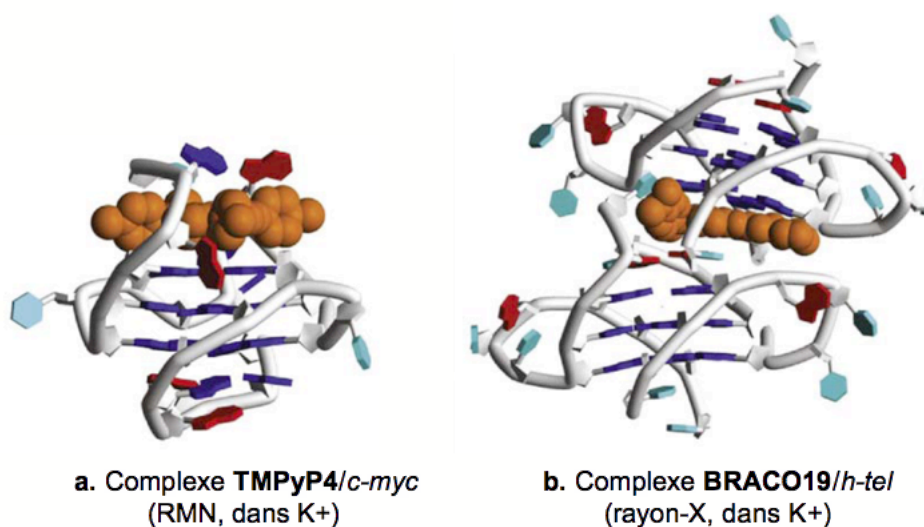


Figure 48. Dessins représentatifs des structures de complexes ligand-G4 [327]

a) Structure RMN du complexe **TMPyP4** avec *c-myc* (d[$\text{TGAG}_3\text{TG}_2\text{TGAG}_3\text{TG}_4\text{A}_2\text{G}_2\text{A}$]) (PDB : 2A5R) ; (b) Interaction de **BRACO19** avec la séquence télomérique humaine (d[$\text{TAG}_3\text{T}_2\text{AG}_3$]₂) observée par cristallographie aux rayons X (PDB : 3CE5).

1.2.4 Calorimétrie isothermale (ITC) et calorimétrie à balayage différentiel (DSC)

Des techniques calorimétriques telles que la calorimétrie isothermale (ITC pour “*Isothermal Titration Calorimetry*”) et la calorimétrie à balayage différentiel (DSC pour “*Differential Scanning Calorimetry*”) fournissent des informations quantitatives sans avoir recours à des modifications chimiques.

La calorimétrie isothermale (ITC) permet de mesurer la chaleur libérée lors d'une réaction. Dans le cas d'une interaction entre des molécules, cette technique permet d'évaluer la constante d'association, l'entropie et l'enthalpie spécifiques d'une réaction de liaison [333].

Le calorimétrie à balayage différentiel (DSC) permettent de mesurer la capacité thermique d'un échantillon en fonction de la température [24, 334, 335] et d'estimer l'enthalpie associée à une transition de phase ou un changement de conformation comme le dépliement ou le repliement coopératif d'une structure en présence d'un ligand.

1.2.5 Spectrométrie de masse à ionisation électrospray (ESI-MS)

Le couplage électrospray-spectrométrie de masse (ESI-MS) est une des techniques souvent sélectionnées pour l'étude des interactions non-covalentes entre une structure d'ADN et son ligand [336-340]. Elle permet d'obtenir directement la stœchiométrie du complexe formé, même pour les espèces mineures. La quantification des acides nucléiques liés et libres permet de déterminer la quantité de ligands liés, donc de déduire la constante d'association à l'équilibre, moyennant quelques hypothèses. La méthode permet également de mesurer la sélectivité en comparant avec d'autres cibles (duplexes d'ADN ou autres conformations de G-quadruplexes par exemple). Cette méthode ne nécessite pas d'oligonucléotide modifié, cependant il est nécessaire de remplacer les sels de NaCl ou KCl par de l'acétate ammonium "évaporable" permettant la formation de microgouttelettes sans concentration de sel. Ce changement d'ion peut entraîner, dans certains cas, un changement de conformation de G4.

1.2.6 Résonance des plasmons de surface (SPR)

La résonance des plasmons de surface (SPR pour "*Surface Plasmon Resonance*") permet l'observation des interactions intermoléculaires en temps réel. L'un des partenaires (le ligand ou l'oligonucléotide) est immobilisé sur une surface fonctionnalisée, la biopuce (sensorchip), puis l'autre partenaire, appelé l'analyte, est injecté en flux continu sur cette surface (Figure 49). Une lumière polarisée est dirigée sur la biopuce dans des conditions de réflexion interne totale. Les photons qui entrent en résonance avec les électrons libres d'une couche d'or, disposée sur la biopuce, entraînent une diminution d'intensité lumineuse du faisceau réfléchi analysé par un détecteur. La formation d'un complexe ligand-ADN entraîne une variation de l'angle auquel se produit cette résonance.

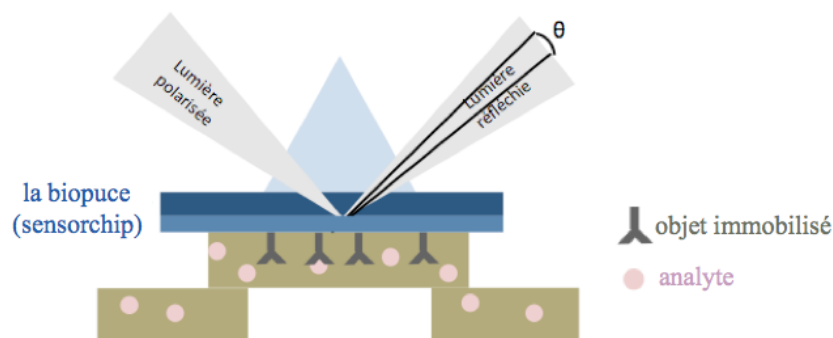


Figure 49. Schéma expliquant le principe de la résonance des plasmons de surface (SPR)

Les analytes sont injectés sur la biopuce sur laquelle le ligand ou l'oligonucléotide a été immobilisé au préalable. Le signal analysé correspond à une mesure de la variation d'angle de résonance en fonction du temps. Cet angle varie avec le changement de masse résultant de l'interaction à l'interface solide/liquide.

Cette technique permet de déterminer la constante d'association à l'équilibre, les constantes cinétiques d'association (k_{on}) et de dissociation (k_{off}) ainsi que les paramètres thermodynamiques (l'entropie et l'enthalpie d'une association biomoléculaire) [341-343]. L'automatisation de cette technique permet de tester les molécules à un débit relativement élevé. Elle ne nécessite que des quantités minimales de matériel, un atout pour étudier des molécules coûteuses et difficiles à produire. Cependant, l'analyse des données doit se faire avec beaucoup de précautions.

1.2.7 Retard sur gel (EMSA)

Le retard sur gel (EMSA pour “*Electrophoresis Mobility Shift Assay*”) est également utilisé pour étudier l'interaction de ligand avec le G-quadruplexe [344]. L'oligonucléotide en question est incubé avec le ligand testé. S'il y a interaction entre le ligand et l'ADN, nous devrions observer, sur gel en condition native, une bande retardée présentant le complexe G4-ligand qui devrait migrer moins vite que la structure sans ligand. Cependant, certains articles ont au contraire décrit une accélération de la migration d'un quadruplexe intramoléculaire en présence du ligand.

1.2.8 Dialyse à l'équilibre

La dialyse à l'équilibre est une technique permettant l'étude de l'interaction d'une molécule vis-à-vis de différentes conformations d'acides nucléiques [345-347]. Elle permet d'analyser la fixation d'un ligand sur un grand nombre d'acides nucléiques de structure distincte (simple-brins, duplexes, triplexes ou différentes conformations de quadruplexes).

Les solutions d'acides nucléiques sont placées dans des chambres de dialyse munies d'une membrane semi-perméable permettant le passage de petites molécules (ligands) et bloquant les macromolécules (oligo- et polynucléotides). À l'équilibre, le ligand se sera accumulé davantage dans la chambre de dialyse qui contient l'acide nucléique pour lequel il a le plus d'affinité. Cette méthode donne donc l'accès à des paramètres quantitatifs de l'association. Bien que cette technique soit relativement fiable et puisse être adaptée à haut débit, la préparation des stocks d'oligonucléotides reste assez longue.

1.2.9 Expérience de dénaturation suivi par Transfert d'Energie d'Excitation par Résonance (FRET)

Le FRET (pour "*Fluorescence Resonance Energy Transfer*") résulte du transfert d'énergie d'excitation d'un donneur de fluorescence vers un accepteur. Pour cela, il est nécessaire que les spectres d'émission du donneur et d'absorption de l'accepteur se recouvrent. Cette technique nécessite un couplage de deux fluorophores aux extrémités d'un oligonucléotide pouvant former un G-quadruplexe intramoléculaire. La formation de ce G-quadruplexe est accompagnée d'une diminution de la distance entre les deux fluorophores et d'un transfert d'énergie du donneur vers l'accepteur, ce qui se traduit par une extinction de l'émission du donneur. Au contraire, la dénaturation de l'oligonucléotide à haute température provoque une augmentation de l'émission du donneur qui est due à l'éloignement des deux fluorophores empêchant ainsi un transfert efficace (Figure 50). Cette approche est utile pour étudier l'effet des ligands sur la stabilité thermique des structures de G-quadruplexe. En général, la liaison des ligands de G-quadruplexe permet une stabilisation de la structure qui est traduite par une augmentation de $T_{1/2}$ [348]. Ces expériences peuvent être réalisées sur plaque utilisant un appareil de PCR en temps réel et permettent un criblage rapide d'un nombre important de ligands. Cependant, cette méthode présente aussi des inconvénients : elle nécessite des oligonucléotides modifiés coûteux et certains ligands pourraient interagir avec les fluorophores ou interférer avec le signal de fluorescence s'ils sont eux-même fluorescents.

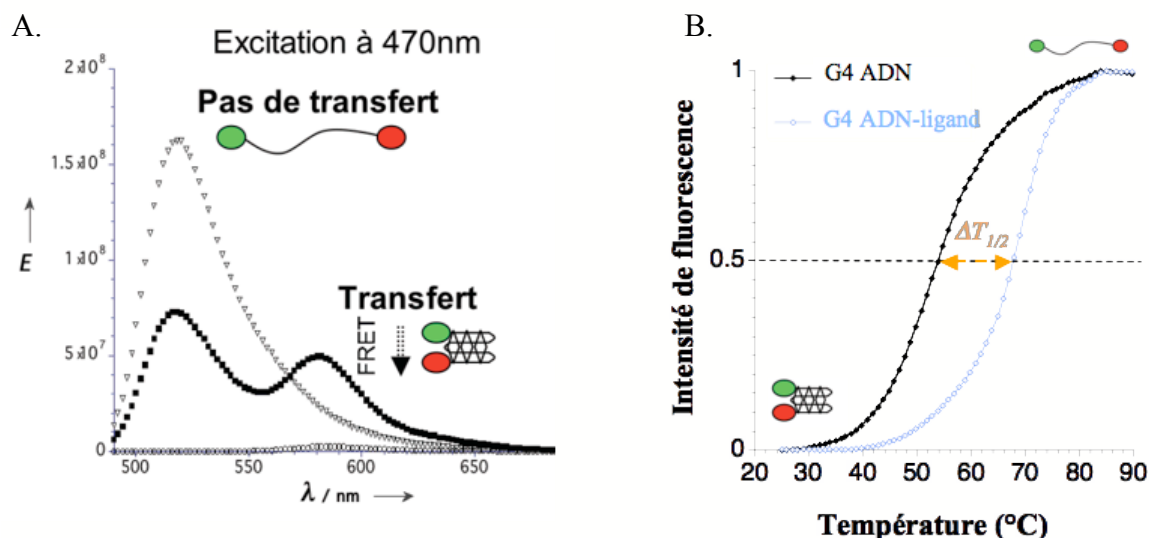


Figure 50. Schéma de la dénaturation de G-quadruplexes suivie par FRET

(A) Spectres de fluorescence de l'oligonucléotide F21T ($\lambda_{ex} = 470$ nm) sous sa forme déstructurée à haute température (pas de transfert d'énergie du donneur vers l'accepteur) et sous sa forme G-quadruplexe à basse température (transfert d'énergie). (B) Dénaturation thermique du G-quadruplexe formé par l'oligonucléotide F21T (5'-FAM-GGG(TTAGGG)₃-TAMRA-3' où FAM = 6-carboxyfluorescéine et TAMRA = tétraméthylrhodamine) suivie en fluorescence en absence (courbe noire) et en présence d'un ligand de G-quadruplexe (courbe bleue). $\Delta T_{1/2}$ est la différence entre $T_{1/2}$ obtenu en présence et celui obtenu en absence de ligand.

1.2.10 Test de criblage fluorescent basant sur la compétition duplexe-quadruplexe

Récemment, un test de criblage fluorescent basé sur la compétition duplexe-quadruplexe a été mis en place [344]. Ce test nécessite deux oligonucléotides marqués : une séquence riche en guanines portant un quencher et une séquence complémentaire avec un fluorophore. La formation de duplexe est associée à un signal de fluorescence faible (extinction de l'émission du fluorophore par le quencher), tandis que la stabilisation de G-quadruplexe par les ligands entraîne une augmentation du signal qui est due à l'éloignement du quencher. La sélectivité du ligand est obtenue en ajoutant des compétiteurs non-fluorescents tels que l'ADN simple-brin, l'ADN double-brin ou les autres conformations de G-quadruplexe (Figure 51).

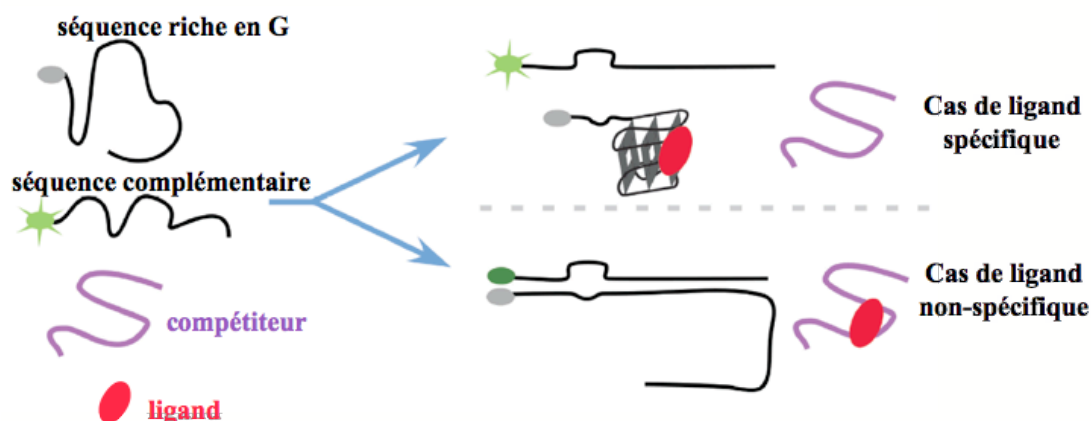


Figure 51. Schéma expliquant le principe du criblage fluorescent basé sur la compétition duplexe-quadruplexe [344].

Dans le cas d'un ligand spécifique de G-quadruplexes, on aura une stabilisation de la structure par le ligand et cela entraîne une augmentation du signal fluorescent. Dans le cas d'un ligand non spécifique reconnaissant le brin violet, ce dernier agissant comme compétiteur piègera le ligand, autorisant la formation de duplexe qui diminuera le signal de fluorescence.

1.2.11 Test de criblage combinant plusieurs molécules fluorescences

S. Paramasivan et P.H. Bolton ont mis en place, en 2008, un criblage utilisant différentes molécules fluorescentes (appelées molécules "reporteurs" : **NMM**, **DODC** et **DTDC**) qui se lient à différents sites de l'ADN G-quadruplexe (quartets terminaux, boucles et sillons) (cf. : §VI.1.1) [349]. En se liant à l'ADN, ces molécules "reporteurs" émettent une fluorescence. La fluorescence du complexe ADN - molécule "reporteur" est mesurée en absence et en présence de ligands testés. L'ajout d'un ligand peut entraîner une diminution ou une augmentation de la fluorescence. La diminution de fluorescence peut être associée à une compétition entre le ligand testé et la molécule "reporteur", montrant ainsi que le ligand et la molécule "reporteur" présentent le même mode d'interaction avec la structure de G-quadruplexes étudiée. Par contre, de nombreux mécanismes pourraient expliquer l'augmentation de la fluorescence en présence de ligands testés (par exemple : des interactions entre le ligand et la molécule "reporteur" ou des effets indirects tels que la neutralisation partielle des charges par l'ajout de ligand). Ce test, en combinant plusieurs molécules "reporteurs", donne des informations sur le mode d'interaction de ligands avec différentes conformations de G-quadruplexes.

1.3 Les raisons de cette étude

Les méthodes citées ci-dessus présentent des avantages et une fiabilité indiscutables. Cependant elles ont également des inconvénients. La plupart des méthodes citées nécessitent des équipements spécifiques, des oligonucléotides modifiés ou immobilisés, ou encore des conditions non isothermes. Elles sont souvent relativement coûteuses, lentes ou compliquées à réaliser. Cela nous a conduit à développer, dans l'**article 4**, un test de criblage de ligands en plaque 96 puits à partir du test G4-FID mis en place par l'équipe de M.P. Teulade-Fichou en 2006 [350]. Ce test est non seulement facile et rapide à réaliser, il ne nécessite pas d'oligonucléotides modifiés ni d'équipement spécifique ce qui réduit considérablement le coût des expériences. De plus, le test est réalisé en conditions isothermes.

Afin de développer des ligands efficaces pour l'étude thérapeutique, il faut non seulement des ligands sélectifs de G-quadruplexes par rapport aux duplexes, mais également des ligands qui sont capables de discriminer différentes conformations de G-quadruplexes. En effet, nous avons vu, dans le chapitre I (Introduction générale), que les G-quadruplexes sont très polymorphiques. Ces structures sont susceptibles de se former dans différentes régions du génome humain qui constituent des cibles potentielles pour le développement des médicaments. Néanmoins, ceci pose un problème de taille pour la sélectivité de la cible. Le grand défi d'aujourd'hui est de trouver des molécules qui peuvent cibler une région spécifique du génome. Il est nécessaire, pour cela, de développer de nouvelles méthodes permettant un criblage rapide de ligands avec différentes conformations de G-quadruplexes. Le test sur plaque présenté dans l'**article 4** est le premier criblage qui permet d'analyser la fixation d'un ligand sur 22 acides nucléiques de structure distincte. L'ensemble des structures étudiées ici comprend deux duplexes, un trinuécléotide et différentes conformations de G-quadruplexes (19 séquences qui présentent des intérêts biologiques).

Le principe de la technique G4-FID a été mentionné dans le chapitre II (Matériels et Méthodes). Les résultats obtenus et les limites de ce test seront discutés ci-dessous.

2) Article 4: “Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures”. *Biochimie* - 2011; 93(8): 1288-96.

Tran, P.L.T., Largy E., Hamon F., Teulade-Fichou M.P., Mergny J.L.

Résumé en français :

La formation de G-quadruplexes dans de nombreuses régions du génome pourrait constituer une cible attirante pour le développement des médicaments. Un grand nombre de petites molécules synthétisées ces dernières années sont capables de cibler préférentiellement les G-quadruplexes et les stabiliser. Pour cribler les ligands de G-quadruplexes, nous avons modifié un test G4-FID (“*G-quadruplex Fluorescent Intercalator Displacement*”). Ce test est basé sur le déplacement par des ligands d'une sonde fluorescente, le Thiazole Orange (TO), sur des oligonucléotides structurés en quadruplexes ou duplexes. Des mesures de sélectivité peuvent être obtenues facilement en comparant la capacité du ligand à déplacer le TO des structures quadruplexes et duplexes. Le G4-FID ne nécessite pas d'oligonucléotides modifiés, ni d'équipement spécifique et il est réalisé en conditions isothermes. Nous avons développé, dans cet article, une version "haut-débit" de ce test. Le test est effectué sur une plaque de 96 puits permettant la comparaison de plus de vingt structures différentes. Quinze ligands de G-quadruplexes connus appartenant à différentes familles sont testés. La plupart des molécules testées montre une bonne sélectivité pour les G-quadruplexes par rapport aux duplexes, mais aucun ne semble présenter de spécificité pour une séquence particulière de quadruplexe. La recherche de ligand de G4 “parfait” n'est donc pas terminée.

Article 4



Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures

Phong Lan Thao Tran^{a,b,c}, Eric Largy^d, Florian Hamon^d, Marie-Paule Teulade-Fichou^d, Jean-Louis Mergny^{a,b,*}^a University of Bordeaux, ARNA laboratory, F-33000 Bordeaux, France^b INSERM, U869, Laboratoire ARNA, IECB, F-33600 Pessac, France^c INSERM, U565, Acides nucléiques: dynamique, ciblage et fonctions biologiques, Muséum National d'Histoire Naturelle (MNHN) USM503, CNRS, UMR7196, F-75231 Paris, France^d Institut Curie, Section Recherche, CNRS UMR176, Centre Universitaire Paris XI, Bât. 110, F-91405 Orsay, France

ARTICLE INFO

Article history:

Received 5 April 2011

Accepted 15 May 2011

Available online xxx

Keywords:

Quadruplex

DNA–DNA ligand interactions

Fluorescence

Screening assay

ABSTRACT

The potential formation of G-quadruplexes in many regions of the genome makes them an attractive target for drug design. A large number of small molecules synthesized in recent years display an ability to selectively target and stabilize G-quadruplexes. To screen for G4 ligands, we modified a G4-FID (G-quadruplex Fluorescent Intercalator Displacement) assay. This test is based on the displacement of an “on/off” fluorescence probe, Thiazole Orange (TO), from quadruplex or duplex DNA matrices by increasing amounts of a putative ligand. Selectivity measurements can easily be achieved by comparing the ability of the ligand to displace TO from various quadruplex and duplex structures. G4-FID requires neither modified oligonucleotides nor specific equipment and is an isothermal experiment. This test was adapted for high throughput screening onto 96-well plates allowing the comparison of more than twenty different structures. Fifteen different known G4 ligands belonging to different families were tested. Most compounds showed a good G4 vs duplex selectivity but exhibited little, if any, specificity for one quadruplex sequence over the others. The quest for the “perfect” specific G4 ligand is not over yet!

© 2011 Elsevier Masson SAS. All rights reserved.

1. Introduction

G-quadruplexes are a polymorphic class of secondary DNA structures in which the structural unit is formed by a planar arrangement of four guanines, known as G-quartets or G-tetrads [1,2]. A vertical stacking of several G-quartets are stabilized by the presence of monovalent cations. These DNA structures can be formed by one, two or four G-rich strands. Tetramolecular quadruplexes generally adopt a well-defined structure, in which all guanines are in the *anti*-glycosidic conformation and all strands are parallel, and might be useful for biotechnology applications. Intramolecular G-quadruplexes formed by single DNA strand have attracted much interest because of their putative biological regulatory function at telomeres, oncogene promoters and other relevant regions of the genome [3–5]. In contrast to tetramolecular quadruplexes, intramolecular structures form faster and are more complex, showing great conformational diversity [3]. Different

sequences can adopt distinct topologies, but a given sequence can also fold into various different conformations, which may coexist. One of the best illustrations of this complexity are human telomere sequences, in particular in potassium conditions [6].

The potential formation of G-quadruplexes in many regions of the genome makes them an attractive target for drug design ever since quadruplex ligands were found to inhibit telomerase [7], to induce rapid apoptosis owing to the displacement of telomere-binding proteins and to regulate transcription or translation of genes, especially oncogenes [8–12]. A large number of small molecules synthesized in recent years display an ability to target and stabilize selectively G-quadruplexes [13,14]. Nowadays, the challenge is to develop small molecules capable of selective binding to one quadruplex type over another. To be effective, screening for selectivity has to be performed at the early stage of development and ideally this should be rapid and easy to implement. Several methods are currently used, such as the FRET-melting assays [15], SPR techniques [16], ITC experiments [17], ESI-MS analyses [18], multifluorescent probes assays [19] or equilibrium microdialysis [20]. Although these methods are reliable and powerful, they require specific equipment or buffer conditions, modified or immobilized oligonucleotides, and/or specific ligand properties (e.g. fluorescence).

* Corresponding author. INSERM, U869, Laboratoire ARNA, IECB, F-33600 Pessac, France. Tel.: +33 0 540 003 022; fax: +33 0 557 571 015.

E-mail address: jean-louis.mergny@inserm.fr (J.-L. Mergny).

In this context, we decided to use a simple alternative assay developed by Teulade-Fichou's team since 2006, G4-FID (G-quadruplex Fluorescent Intercalator Displacement) [21–24]. The principle of G4-FID assay is schematically presented in Fig. 1. This assay is based on the displacement of an “on/of” fluorescence probe, Thiazole Orange (TO), from quadruplex or duplex DNA matrices by increasing amounts of a putative ligand. TO being virtually non-fluorescent when free in solution but strongly fluorescent when bound to DNA, the ligand-induced displacement leads to a decrease of the fluorescence which is monitored as function of the ligand concentration. Therefore, the quadruplex-affinity of a candidate compound can be evaluated through its ability to displace TO from quadruplex DNA. Moreover, selectivity measurements can easily be achieved by comparing the ability of the ligand to displace TO from various quadruplex and duplex structures. Moreover, G4-FID requires neither modified oligonucleotides nor specific equipment and is an isothermal experiment. Initially, this assay was performed in a high total volume (3 mL), using quartz cells (referred hereafter as *standard G4-FID assay*). In this work, we developed a low-cost G4-FID assay for high throughput screening onto 96-well plates (which may be easily transposed into 384-well plates) where the total volume does not exceed 25 μ L. Notably, we decided to use a qPCR apparatus for all fluorescent measurements. Although it displays a relatively moderate fluorescence sensibility compared to dedicated fluorimeters or microplate readers, it is a widely spread device among research laboratories, already used for FRET-melting measurements.

In order to get more insight into the structure selectivity of the studied molecules, we have used a large number of oligonucleotides displaying various structures (quadruplexes, duplexes and single-strands). Thus, we have exhaustively characterized the structure and stability of 30 DNA and RNA sequences, by absorbance thermal difference spectra (TDS), circular dichroism (CD), UV melting and PAGE. We have determined which sequences could be reliably used in function of the fluorescence enhancement of bound TO. Finally, we have assayed 15 compounds, displaying different binding properties, using our screening method. All these ligands have already been tested by FRET-melting assay and/or standard G4-FID assay enabling a comparison of the results.

2. Experimental conditions

2.1. Relevant oligonucleotides and ligands

2.1.1. Oligonucleotides

Oligonucleotides were purchased from Eurogentec (Seraing, Belgium), dissolved in bi-distilled water and stored at -20°C . Concentrations of all oligonucleotides were determined by UV-absorption using the extinction coefficients provided by the manufacturer. Sequences are given in the 5' to 3' direction (Table S1). For spectroscopic measurements, Reverse-Phase Cartridge•Gold™ (RPC) purified oligonucleotides were used without further purification.

2.1.2. Ligands

TrisK3-NH, TrisQ [25], 360A, 832A, 307A, iPDC [26], Phen-DC3, Phen-DC6, Bipy-DC3, Bipy-DC6 [27], and the metallo-organic complex Cu-ttpy [28] have been synthesized and supplied by Teulade-Fichou's team (Fig. 2). BRACO-19 [29] and 12459 [30] were a kind gift from Dr. P. Mailliet (Sanofi). Piper [31] and BSU1051 [7] are commercially available. Stock solutions of these ligands (1–10 mM in DMSO) were used for G4-FID assay and are stored at -20°C . TO was purchased from Sigma–Aldrich and used without further purification.

2.2. Preparation of oligonucleotides

Oligonucleotides (intramolecular quadruplexes, duplexes and single-strands) were prepared by heating the corresponding oligonucleotides at 90°C for 5 min in a 10 mM lithium cacodylate buffer pH 7.2, 100 mM KCl then slowly cooling to room temperature for 2 h. Oligonucleotide structures were formed at 50 μ M strand concentration. Tetramolecular quadruplexes from TG₄T, TG₅T and UG₅U were pre-folded at high strand concentration (200 μ M), at 4°C for 48 h, in the same buffer.

2.3. Fluorescence enhancement measurements

Experiments were performed with 96-well microplates from Stratagene. Each condition was tested at least in duplicate, in a volume of 25 μ L for each sample. Measurements were performed at 25°C , all G-quadruplex sequences being folded in KCl at this temperature. Two equivalents of TO (2 μ M) were added in a solution of 1 μ M oligonucleotide. Samples were incubated in 10 mM lithium cacodylate buffer pH 7.2, 100 mM KCl. Fluorescence emission was collected at 516 nm with 8-fold gain after excitation at 492 nm in a real-time quantitative PCR (Stratagene Mx3005P instrument).

2.4. G4-FID protocol: isothermal assay

All experiments were performed with 96-well microplates. Each condition was tested in duplicate and at least 3 times, in a volume of 25 μ L for each sample. Samples were incubated in 10 mM lithium cacodylate buffer pH 7.2, 100 mM KCl. Pre-folded DNA target (0.5 μ M strand concentration for intramolecular G4, 2 μ M strand concentration for tetramolecular G4 and 1 μ M strand concentration for duplexes) was mixed with 1 μ M of Thiazole Orange. Five equivalents of each ligand (2.5 μ M) were added for each structure. The fluorescence of samples was measured at 25°C in a qPCR (Stratagene Mx3005P) instrument. The temperature was kept constant with a thermostat cell holder (Peltier). The Thiazole Orange was excited at 492 nm (± 5 nm) and the emission was collected at 516 nm (± 5 nm) with 8-fold gain.

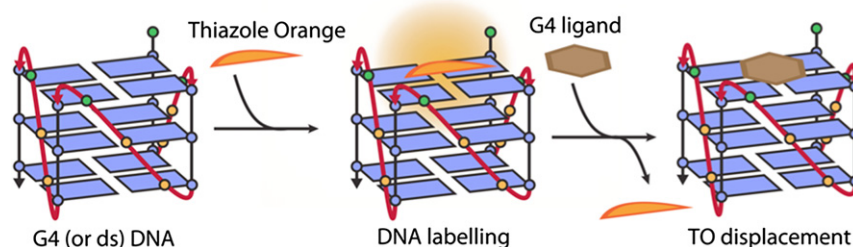


Fig. 1. Schematic representation of the two main steps of the G4-FID assay, i.e., (a) labelling of the DNA matrices (quadruplex or duplex DNA) by Thiazole Orange (TO) and (b) displacing the fluorescent probe from the DNA matrices by a small molecule candidate.

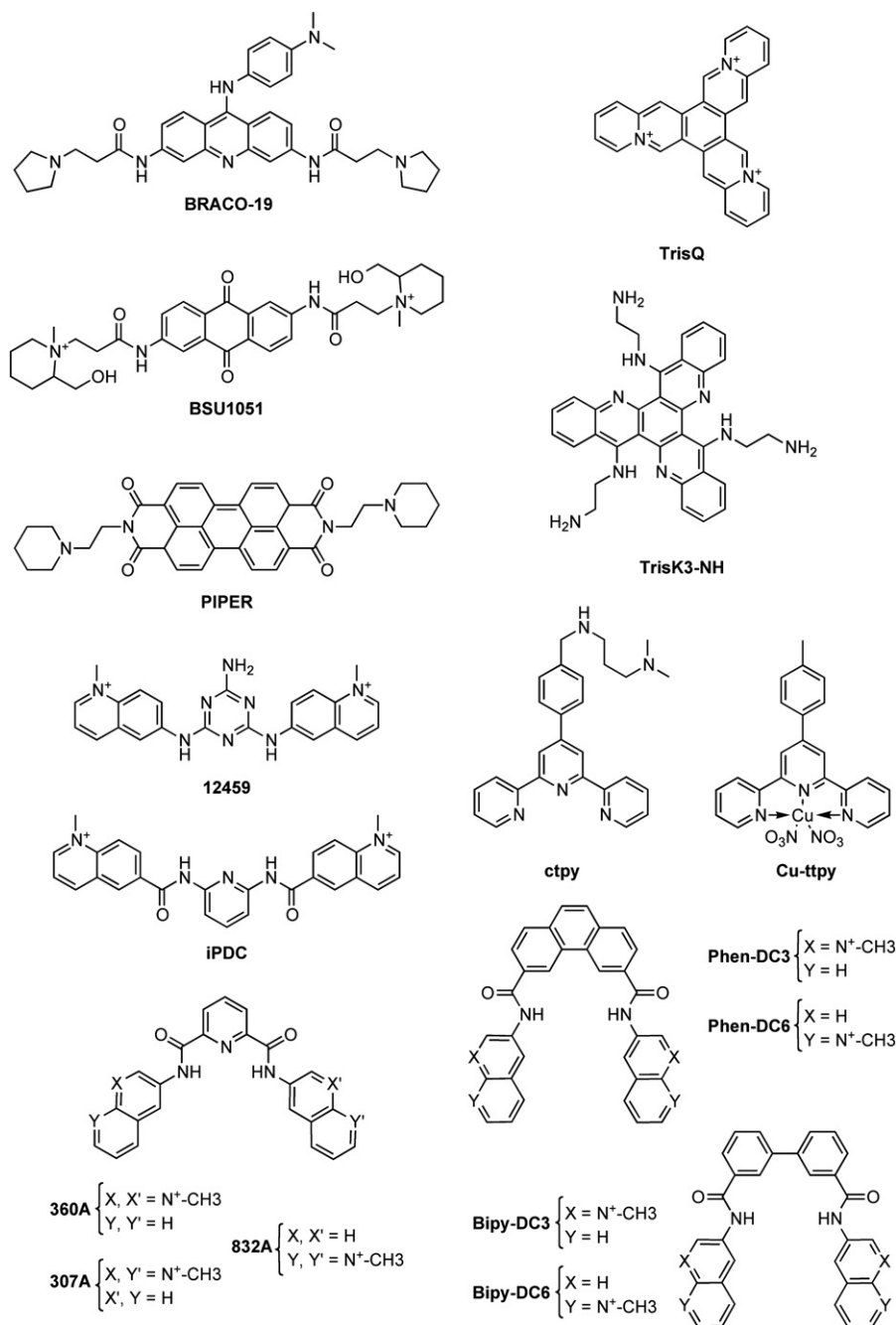


Fig. 2. Selected G-quadruplex ligands used in the G4-FID assay.

3. Analysis and interpretation of FID results

3.1. Choice of the oligonucleotides for FID test

3.1.1. Absorbance, circular dichroism measurement and PAGE of G-rich sequences

Initial experiments (UV–visible absorption, circular dichroism measurement and gel electrophoresis) were performed to confirm that all selected G-rich sequences could adopt a quadruplex structure. Except the trinucleotide CGG12, all the studied G-rich sequences displayed TDS (Thermal Difference Spectra) signatures characteristic of G-quadruplex structure in KCl (Fig. 3A) [32]. The trinucleotide CGG12 changed the conformation in presence of

ligand. In particular, the G4 formation of sequence CGG12 induced by 360A (Fig. S2).

Circular dichroism spectra recorded at 4 °C confirmed TDS results (Fig. 3B). Most of them exhibited a negative peak around 240 nm and a positive peak around 260 nm (a “Type I” signature) in 10 mM lithium cacodylate and 100 mM KCl (Fig. 3B, left). However, the thrombin binding aptamer (TBA), *Bombyx* (Bom17), *Ascaris* (Asc20) and mutant of human telomeric motifs (21CTA) [33] showed two maxima at 245 and 290 nm (a “Type II” signature) (Fig. 3B, right). The short (22Ag) and long (45Ag) human telomeric motif, *Plasmodium* (PlasC24), *Oxytricha* (Oxy 3.5) and promoter AKT1 displayed a positive peak around 290 nm and a positive peak or shoulder in the 260–270 nm region (Fig. 3B, right) in KCl

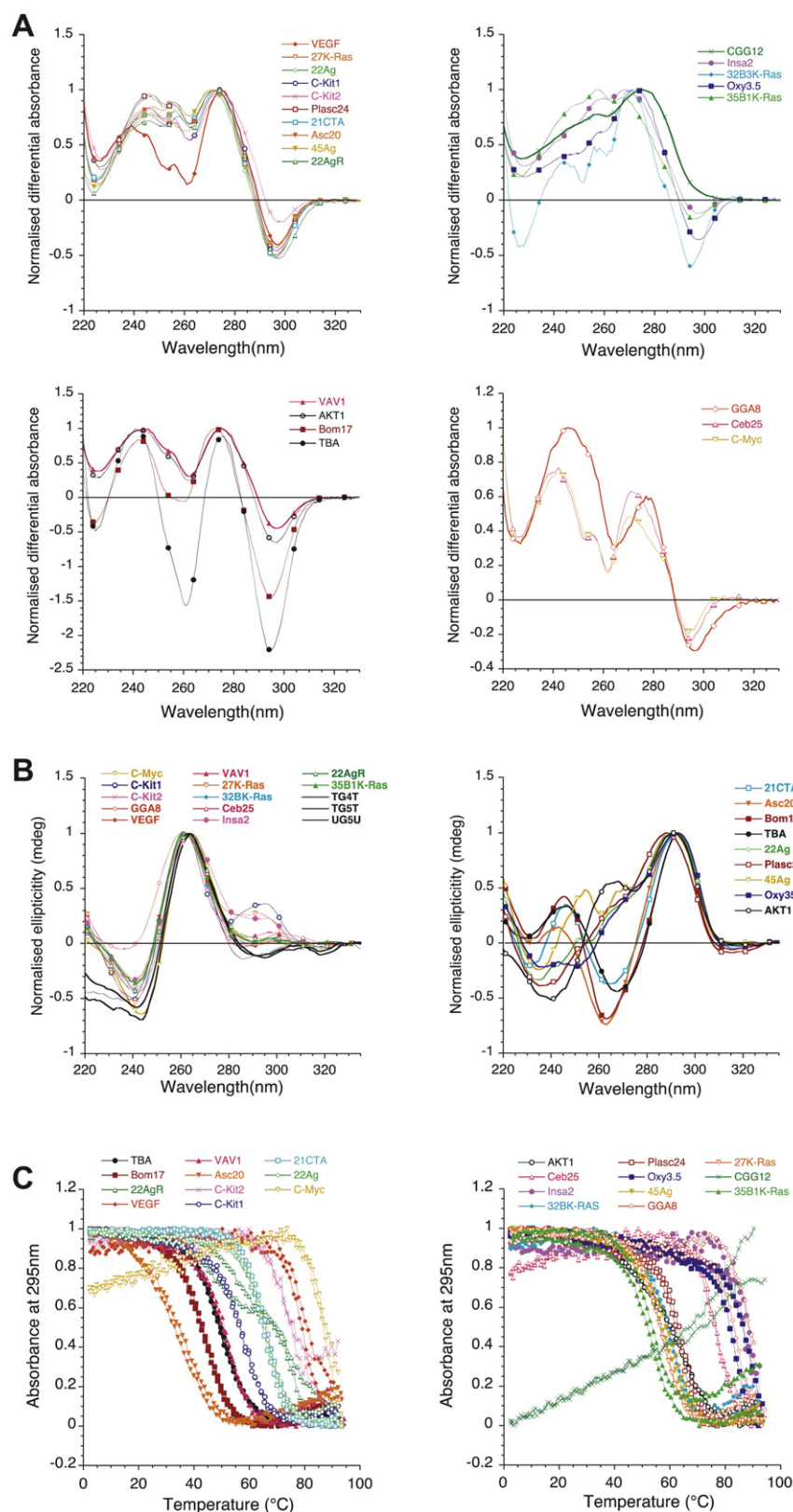


Fig. 3. UV melting curves, thermal difference spectra (TDS) and circular dichroism (CD). (A) Thermal difference spectra resulting from the difference between the absorbance recorded at high and low temperature in 100 mM KCl containing 10 mM lithium cacodylate at pH 7.2. (B) CD spectra recorded at 4 °C (in the same buffer). Data were normalized ($TDS_{norm} = TDS/\max(TDS)$ or $CD_{norm} = CD/\max(CD)$) over the 220–335 nm wavelength range. (C) Normalized UV melting curves (cooling and heating) at 295 nm in 100 mM KCl containing 10 mM lithium cacodylate (1 cm optical path length). Oligonucleotides were prepared at 5 μ M strand concentration.

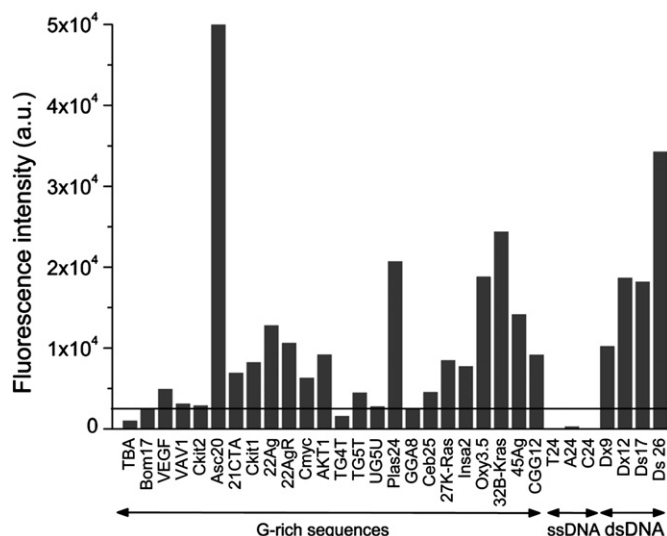


Fig. 4. Bar graph of fluorescence enhancement with 1 μ M of oligonucleotides and 2 μ M of TO. ssDNA: single-stranded DNA (C₂₄, A₂₄ or T₂₄). dsDNA: double-stranded DNA (9–26 bp-long).

condition. The CD spectra showed the diversity of G-quadruplex structures formed by selected sequences.

The stability of the quadruplexes was measured by UV melting experiments, recording absorbance at 295 nm [34]. Melting temperature of all G-quadruplex folding sequences were determined graphically [35]: T_m ranged from 37 °C to > 70 °C in KCl (Fig. 3C & Table S1). The formation of tetramolecular quadruplexes (TG₄T, TG₅T, UG₅U) was not studied by UV melting because of their high stability (>90 °C) and their slow association kinetics [36]. Consequently, in order to demonstrate the formation of tetramolecular quadruplex structures, we used non-denaturing polyacrylamide gel (PAGE) (Fig. S1). Non-denaturing gel electrophoresis allows separation of single-stranded oligonucleotides (indicated by “–”) from tetramolecular G-quadruplex structures (indicated by “+”).

Table 1
Selected oligonucleotides used in the G4-FID assay.

Name		Sequence (from 5' to 3')	T_m^a	Maxima CD	TO FI(2:1) ^b
VEGF	Promoter	G ₃ AG ₃ TTG ₄ TC ₃	>70	(+) 260 nm	5×10^3
VAV1	Promoter	G ₃ CAG ₃ AG ₃ AACTG ₃	51	(+) 260 nm	3×10^3
C-Kit2	Promoter	G ₃ CG ₃ CGCGAG ₃ AGG ₃	>70	(+) 260 nm	3×10^3
Asc20	Ascaris telomere	G ₂ CTTAG ₂ CTTAG ₂ CTTAG ₂	37	(+) 290 nm	50×10^3
21CTA	Telomere	G ₃ CTAG ₃ CTAG ₃ CTAG ₃	67	(+) 245 nm (+) 290 nm	7×10^3
C-Kit1	Promoter	G ₃ AG ₃ CGCTG ₃ AGGAG ₃	57.5	(+) 260 nm	8×10^3
22Ag	Human DNA-telomere	AG ₃ TTAG ₃ TTAG ₃ TTAG ₃	66	(+) 290 nm	13×10^3
22AgR	Human RNA-telomere	AG ₃ UUAG ₃ UUAG ₃ UUAG ₃	>70	(+) 260 nm	10×10^3
C-Myc	Promoter	TGAG ₃ TG ₃ TAG ₃ TG ₃ TAA	>70	(+) 260 nm	6×10^3
AKT1	Promoter	G ₃ CG ₃ CGGCTCCG ₃ CGCG ₃	62	(+) 290 nm	9×10^3
TG5T	DNA-G4 tetramolecular	TG ₅ T	>90	(+) 260 nm	4.5×10^3
PlasC24	Plasmodium telomere	G ₃ TTACG ₃ TTACG ₃ TTACG ₃	64	(+) 290 nm	20×10^3
Ceb25	Minisatellites	AG ₃ TG ₃ TGTAAGTGTG ₃ TG ₃ T	>70	(+) 260 nm	4.5×10^3
27KRas	Promoter	G ₃ CGGTGTG ₃ AAGA G ₃ AAGAGG ₃	52	(+) 260 nm	8.5×10^3
Insuline a2	Promoter	ACAG ₄ TGTG ₄ ACAG ₄ TGTG ₄	>70	(+) 260 nm	8×10^3
Oxy3.5	Oxytricha telomere	G ₄ TTTTG ₄ TTTTG ₄ TTTTG ₄	>70	(+) 290 nm	19×10^3
32B3 (K-Ras)	Promoter	AG ₃ CGGTGTG ₃ AAGAG ₃ AAGAG ₃ AGG	60	(+) 260 nm	25×10^3
35B1 (K-Ras)	Promoter	AG ₃ CGGTGTG ₃ AAGAG ₃ AAGAG ₃ AGGCAG	54	(+) 260 nm	19×10^3
45Ag	Human DNA-telomere	G ₃ TTAG ₃ TTAG ₃ TTAG ₃ TTAG ₃ TTAG ₃ TTAG ₃	59	(+) 290 nm	14×10^3
CGG12	Trinucleotide	(CGG) ₁₂	no G4	no G4	9×10^3
dx12	Duplex	(a)GCGGTGAGTTCGG(b)CCGAACCTCACGC	ND ^c	ND ^c	18×10^3
ds26	Duplex	CAATCGGATCGAATTCGATCCGATTG	ND ^c	ND ^c	34×10^3

^a T_m (in °C) determined by UV melting experiments using absorbance at 295 nm, in 100 mM KCl with 10 mM lithium cacodylate.

^b Fluorescence of TO with 1 oligo for 2 TO.

^c T_m no determined.

UV melting, TDS, CD spectra and PAGE confirmed that all selected G-rich sequences form stable quadruplexes. All of them melt at temperatures higher than 35 °C thereby enabling us to fix the temperature of FID experiments at 25 °C.

3.1.2. Fluorescence enhancement

Fluorescence enhancement of TO in presence of all the oligonucleotides was monitored in order to determine which sequences could be used in the G4-FID assay with a qPCR device. It is worth noting that the choice of wavelengths available with this experimental setting does not correspond to the maximum excitation and emission wavelengths of TO (501/534 nm), but to commonly available fluorescein filters (492/516). However, this is not detrimental to the measurements given that the bandwidths (± 5 nm) allow a good excitation and recovery of the emission.

We have measured fluorescence enhancement using 2 equivalents of TO vs DNA as it is the preferential stoichiometry for the G4-FID assay [23]. Results are gathered in Fig. 4. These measurements led us to eliminate oligonucleotides for which fluorescence intensity was not sufficient (<2500) to be reliably detected by a qPCR device, making the observation of TO displacement more difficult. Then from all previous experiments, the thrombin binding aptamer (TBA), *Bombyx* (Bom17), trinucleotide (GGA)₈, two tetramolecular quadruplexes (TG₄T and UG₅U), two double-strands (dx9 and ds17) and all single-strands (dT₂₄, dA₂₄ and dC₂₄) were eliminated for further studies. The sequences of the twenty remaining G4 oligonucleotides and of the two control duplex oligonucleotides (dx12, ds26) are shown in Table 1.

3.2. Analysis of FID results exemplified with 360A, a quadruplex ligand used as a benchmark

The G4-FID assay involves the following elements (Fig. 5A):

- A number of wells are dedicated to controls and calibration, such as:
 - buffer only (background level = F_b ; tested in 10 mM lithium cacodylate + 100 mM KCl),

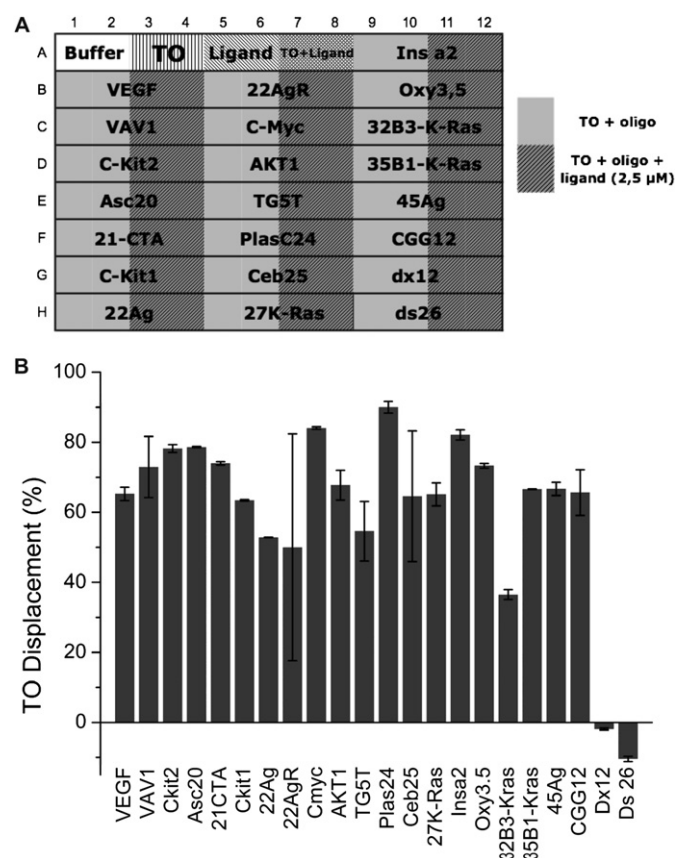


Fig. 5. G4-FID assay and the percentage displacement of TO. (A) Representation of a 96-well plate and disposition of wells. (B) Percentage displacement of TO for 360A.

2. *buffer* + *TO* should exhibit very low fluorescence, close to F_b , as no DNA is present and emission of free TO is very weak.
 3. *buffer* + *ligand* should also exhibit very low fluorescence, close to F_b , except if the ligand is naturally fluorescent.
 4. *buffer* + *TO* + *ligand* should also be close to F_b , except if the ligand interacts with TO.
- ii. The truly informative wells correspond to samples containing nucleic acids. Pre-folded DNA target (0.5 μ M of strand concentration for intramolecular G4, 2 μ M of strand concentration for tetramolecular G4 and 1 μ M of strand concentration for duplexes) is mixed with 1 μ M of Thiazole Orange. The fluorescence of TO bound to DNA is called F_{I0} .
 - iii. Five equivalents of each ligand (2.5 μ M) were added for each oligonucleotide. The fluorescence of TO after adding ligand is called F_I .

The initial fluorescence (F_I and F_{I0}) is subtracted by the background fluorescence ($F_A = F_I - F_b$ or $F_{A0} = F_{I0} - F_b$). We can then calculate the percentage of TO displacement by the formula as previously described:

$$\text{TO Displacement (\%)} = 100 - ((F_A/F_{A0}) \times 100)$$

We can see that 360A strongly binds to G-quadruplex structures (Fig. 5B), all the TO displacement percentages for G4 structures being superior or equal to 40. However, it does not bind to duplex-forming sequences as the TO displacement in those cases is close to 0. Consequently, the compound 360A can be considered as a potent and selective G-quadruplex binder, which is in full agreement with both FRET-melting and standard G4-FID results [23].

The use of 5 equivalents of ligands allows to screen the displacement of most compounds, including weakly binding compounds but,

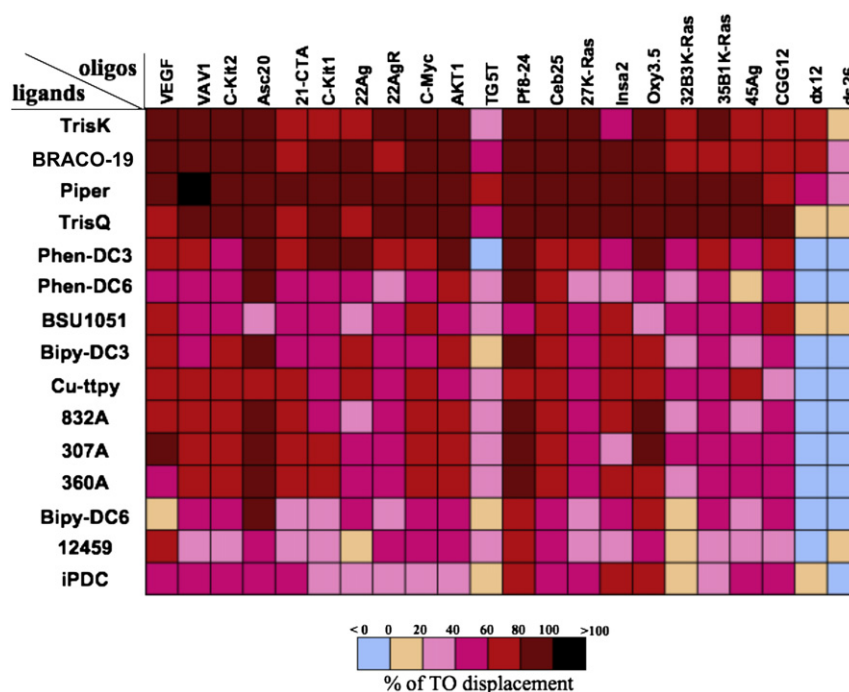


Fig. 6. Screening results are shown for the set of 22 oligonucleotides (20 quadruplexes + two duplexes on the rightmost columns) and all selected ligands. The colour chart representation is used to allow for visual comparison of the percentage of TO displacement obtained with different test molecules (colour code indicated below). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

as a drawback, does not allow to sort finely potent binders. Thus, the best compounds should be re-assayed at a lower concentration. Also, the use of a higher TO/oligonucleotide ratio as compared to the standard G4-FID assay (0.5 vs 0.25 μM in DNA) enables to use less sensitive fluorescence reader. Nevertheless, care should be taken not to use too high concentrations of TO since it could result in non-specific binding of the probe (by electrostatic interactions notably, unpublished results).

3.3. G4-FID screening results

The results are presented in Fig. 6 with blue indicating a “negative percentage” of TO displacement and hot colours, from light to dark, standing for 0–100% of TO displacement. The colour chart mode of presentation allows a direct and easy visual comparison of the results.

First of all, we verified that results obtained herein are in accordance with previously published results obtained by standard G4-FID assay and FRET-melting experiments. Well-known potent G4-DNA binders (Phen-DC compounds, BRACO-19, PIPER, etc.) were found to be very effective on most G4-DNA structures. Conversely, iPDC, which features an inverted connectivity of the two amide bondings as compared to 360A, exhibited the weakest TO displacement in comparison with most of the other tested compounds and remarkably as compared with 360A. This is fully consistent with results obtained both with the standard G4-FID assay (unpublished data) and by FRET-melting experiments [37].

Selective and non-selective compounds can also be easily distinguished. The four compounds TrisK, BRACO-19, Piper and TrisQ tend to bind quite well to duplex structures (although one should note that the fluorescence of BRACO-19 may interfere with the assay). Interestingly, we found that TrisQ is slightly more selective than TrisK3-NH as recently published [25]. As previously outlined, the use of 5 equivalents of ligands vs DNA may be detrimental for sorting highly potent binders and the use of a lower amount of compound may allow to better measure selectivity in this specific case. In that situation, it is thus recommended to re-assay the best compounds at a lower concentration, or to perform standard G4-FID assay to generate a displacement curve over a range of ligand concentration. The other compounds exhibit good to very good selectivity for G-quadruplex structures in comparison with the duplex structure.

Overall, we observed a weak affinity of compounds for TG₅T, which is not surprising since tetramolecular quadruplexes can accommodate more TO molecules than intramolecular structures [23]. Conversely, high TO-displacements were found for *Ascaris* 20, *Plasmodium* 24 and *Oxytricha* 3.5 which may be explained either by a weaker affinity of TO for those DNA matrices (and thus an easier displacement) or a higher affinity of the compounds for these structures.

Disappointingly, we were unable to identify any compound capable of discriminating one conformation of G-quadruplex in particular. However, these results were expected since, to the best of our knowledge, no compound able to selectively target one particular quadruplex structure has ever been reported.

4. Discussion and conclusion

The method presented here allows the screening of an important number of unmodified nucleic acids structures with widespread equipment i.e., a qPCR device. Nevertheless, the use of a moderately sensitive fluorescence reader does not allow to use all the 32 DNA structures of the initial panel since some of them lead to a weak increase in TO emission, which prevents an accurate determination of the displacement efficiency. The lower fluorescent enhancement

of the probe observed in certain cases may be the result of a non-planar conformation of TO upon interaction (especially with single-strands) and/or of a non sufficient isolation from the aqueous environment. Thus, out of the 30 sequences tested, we had to discard 8 motifs – including all single-strands – and we finally kept 22 sequences (20 quadruplexes and 2 duplexes). Also, the working concentration ratio in TO/oligonucleotide should be higher (>2) than in the standard G4-FID assay to circumvent the detection limits of the apparatus.

Additionally, we found that the screening results presented in Fig. 6 illustrate a general property of reported G4 binders: most of them do discriminate between duplexes and quadruplexes (the two rightmost columns exhibit a low competition level); on the other hand, none of the tested ligands have a clear selectivity for one quadruplex over the others as they efficiently displace TO from most of the G4 structures. A relative exception may be found for TG₅T: TO is notably harder to displace from this structure than from other quadruplexes. The absence of loops and therefore of potential additional hydrogen bonds or van der Waals interactions may explain these results. Also, flanking thymine nucleotides could render the external G-quartets accessibility more difficult than expected. In the same line, the remarkably low fluorescence of TO bound to this tetramolecular matrix might indicate a binding mode different from the classical stacking on G-quartets. For instance groove binding of the probe can be hypothesized thereby resulting in indirect competition with the ligands.

The total number of structures tested for each ligand, including oncogenes sequences, dimers and RNA, is within the same order as competition dialysis [38] and the duplex–quadruplex test we recently developed [39], and superior to what is commonly performed with FRET-melting [15] or surface plasmon resonance [40,41]. Moreover, virtually any sequence can be used provided that TO fluorescence enhancement is sufficient. Furthermore, this test is simple to implement, as it involves unmodified DNA/RNA sequences and an inexpensive dye, Thiazole Orange. Finally, widespread qPCR apparatus may be used to perform these measurements.

This test nevertheless has some limitations: i) compounds that are naturally fluorescent with excitation and emission maxima overlapping TO may interfere with the measurements. This is why it is necessary to evaluate prior to screening TO and compounds to be tested in absence of DNA, and compound + DNA with TO. Alternative fluorescent dyes are currently developed to circumvent this problem [43] ii) TO does not equally bind to all structures (different K_d) and has different quantum yields when bound (see above) iii) One cannot formally exclude that simultaneous binding of a compound and of TO is not formally exclusive: if a ternary complex is formed, then the test will underestimate the affinity of the dye to the structure as no decrease in TO fluorescence will be obtained despite binding. However, from the results shown in Fig. 6 and our unpublished results, this seems to be a relatively rare case. Even if TO and the tested compound have different binding modes, the quadruplex structure is a relatively small object, and simultaneous binding appears disfavoured iv) the “<0%” or “>100%” displacement values found for certain targets cannot easily be explained. One may propose a direct interaction between the dye and TO leading to total quenching (for >100%) or that the rigidification/conformational change of a nucleic acid structure induced by a ligand may alter TO properties.

To obtain an independent validation of this method, we performed binding studies by spectroscopic titrations with one of the compounds tested here. Such experiments are indeed possible for ligands that undergo a significant change of absorbance or fluorescence upon binding to nucleic acids. For sensitivity purposes, we performed such titrations using the fluorescence signal of a pyridine dicarboxamide derivative, 360A. Its weak emission is almost

completely quenched upon binding to DNA. This allowed us to compare fluorescence intensity as a function of DNA concentration. This titration is presented as [Supplementary Fig. S3](#). These experiments qualitatively confirm the results obtained by FID: such G4 ligand has a much weaker affinity for duplexes than for quadruplexes. Unpublished surface plasmon resonance data on other G4 ligands also confirmed that ligands that easily displace TO from quadruplex structures exhibit a high affinity (Kd in the nM range) for this conformation (data not shown).

To conclude, we have determined user-friendly experimental conditions in which a large number of DNA/RNA sequences, displaying various well-characterized structures (intra/tetramolecular quadruplexes, dimeric quadruplexes and duplexes), can be used in G4-FID screenings, employing widespread materials. The set of structures presented here is obviously biased towards G4 structures, but can be easily modified to accommodate other structures such triplexes, Z-DNA, i-motifs, trinucleotide repeats...provided that TO be sufficiently fluorescent when bound to these motifs [42]. Additional work enabling to perform full TO displacement titration curves in microplates in the standard assay conditions [43], as well as an example of screening of a chemical library will be presented elsewhere.

Acknowledgements

We thank members of the Mergny and Teulade-Fichou laboratories for helpful discussions. This work was supported by an ANR grant (G4-TOOLBOX, ANR-Blan-09-355) to J.L.M., F.H. and M.P.T.F. P.L.T.T. is the recipient of an Université Bordeaux Segalen – MESR PhD studentship. E.L. was supported by a BDI graduate student fellowship from the Institut Curie and the Centre National de la Recherche Scientifique. J.L.M. acknowledges support from the Région Aquitaine, the Fondation pour la Recherche Médicale (FRM), Association Française de l'Ataxie de Friedreich (AFAF) and INCA.

Appendix. Supplementary material

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.biochi.2011.05.011](https://doi.org/10.1016/j.biochi.2011.05.011).

References

- [1] S. Neidle, S. Balasubramanian, *Quadruplex Nucleic Acids*, RSC Biomolecular Sciences, Cambridge, 2006, pp. 301.
- [2] D. Sen, W. Gilbert, Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its applications for meiosis, *Nature* 334 (1988) 364–366.
- [3] D.J. Patel, A.T. Phan, V. Kuryavyi, Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics, *Nucleic Acids Res.* 35 (2007) 7429–7455.
- [4] Y. Qin, L.H. Hurley, Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions, *Biochimie* 90 (2008) 1149–1171.
- [5] B.S. Herbert, J.L. Huppert, F.B. Johnson, A.N. Lane, A.T. Phan, Meeting report: second international meeting on quadruplex DNA, *Biochimie* 91 (2009) 1059–1065.
- [6] A.T. Phan, Human telomeric G-quadruplex: structures of DNA and RNA sequences, *FEBS J.* 277 (2010) 1107–1117.
- [7] D. Sun, B. Thompson, B.E. Cathers, M. Salazar, S.M. Kerwin, J.O. Trent, T.C. Jenkins, S. Neidle, L.H. Hurley, Inhibition of human telomerase by a G-quadruplex-interactive compound, *J. Med. Chem.* 40 (1997) 2113–2116.
- [8] T.A. Brooks, S. Kendrick, L. Hurley, Making sense of G-quadruplex and i-motif functions in oncogene promoters, *FEBS J.* 277 (2010) 3459–3469.
- [9] J.L. Huppert, Structure, location and interactions of G-quadruplexes, *FEBS J.* 277 (2010) 3452–3458.
- [10] H.J. Lipps, D. Rhodes, G-quadruplex structures: in vivo evidence and function, *Trends Cell Biol.* 19 (2009) 414–422.
- [11] P. Wang, C.-H. Leung, D.-L. Ma, S.-C. Yan, C.-M. Che, Structure-based design of platinum(II) complexes as c-myc oncogene down-regulators and luminescent probes for G-quadruplex DNA, *Chem. Eur. J.* 16 (2010) 6900–6911.
- [12] Z.A.E. Waller, S.A. Sewitz, S.-T.D. Hsu, S. Balasubramanian, A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression, *J. Am. Chem. Soc.* (2009).
- [13] D. Monchaud, M.P. Teulade-Fichou, A hitchhiker's guide to G-quadruplex ligands, *Org. Biomol. Chem.* 6 (2008) 627–636.
- [14] A. De Cian, J.L. Mergny, Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation, *Nucleic Acids Res.* 35 (2007) 2483–2493.
- [15] A. De Cian, L. Guittat, M. Kaiser, B. Saccà, S. Amrane, A. Bourdoncle, P. Alberti, M.P. Teulade-Fichou, L. Lacroix, J.L. Mergny, Fluorescence-based melting assays for studying quadruplex ligands, *Methods* 42 (2007) 183–195.
- [16] E.W. White, F. Tanious, M.A. Ismail, A.P. Reszka, S. Neidle, D.W. Boykin, W.D. Wilson, Structure-specific recognition of quadruplex DNA by organic cations: influence of shape, substituents and charge, *Biophys. Chem.* 126 (2007) 140–153.
- [17] B. Pagano, C.A. Mattia, C. Giancola, Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes, *Int. J. Mol. Sci.* 10 (2009) 2935–2957.
- [18] F. Rosu, E. De Pauw, V. Gabelica, Electrospray mass spectrometry to study drug–nucleic acids interactions, *Biochimie* 90 (2008) 1074–1087.
- [19] S. Paramasivan, P.H. Bolton, Mix and measure fluorescence screening for selective quadruplex binders, *Nucleic Acids Res.* 36 (2008) e106.
- [20] P. Ragazzon, J.B. Chaires, Use of competition dialysis in the discovery of G-quadruplex selective ligands, *Methods* 43 (2007) 313–323.
- [21] D. Monchaud, C. Allain, M.-P. Teulade-Fichou, Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands, *Bioorg. Med. Chem. Lett.* 16 (2006) 4842–4845.
- [22] D. Monchaud, C. Allain, M.P. Teulade-Fichou, Thiazole orange: a useful probe for fluorescence sensing of G-quadruplex–ligand interactions, *Nucleosides, Nucleotides Nucleic Acids* 26 (2007) 1585–1588.
- [23] D. Monchaud, C. Allain, H. Bertrand, N. Smargiasso, F. Rosu, V. Gabelica, A. De Cian, J.L. Mergny, M.P. Teulade-Fichou, Ligands playing musical chairs with G-quadruplex DNA: a rapid and simple displacement assay for identifying selective G-quadruplex binders, *Biochimie* 90 (2008) 1207–1223.
- [24] D. Monchaud, M.-P. Teulade-Fichou, G4-FID: a fluorescent DNA probe displacement assay for rapid evaluation of quadruplex ligands, in: P. Baumann (Ed.), *G-quadruplex DNA*. Springer, Kansas City, 2010, pp. 257–271.
- [25] H. Bertrand, A. Granzhan, D. Monchaud, N. Saettel, R. Guillot, S. Clifford, A. Guédin, J.-L. Mergny, M.-P. Teulade-Fichou, Recognition of G-quadruplex DNA by triangular star-shaped compounds: with or without side chains? *Chem. Eur. J.* 17 (2011) 4529–4539.
- [26] A. Hittinger, T. Caulfield, P. Mailliet, E. Mandine, J.-L. Mergny, L. Guittat, J.-F. Riou, D. Gomez, C. Belmokhtar, Chemical Derivatives Binding Very Specifically with G-quadruplex DNA Structures and Use Thereof as a Specific Anti-cancer Agent. (2004), Patent WO2004072027.
- [27] A. De Cian, E. DeLemos, J.-L. Mergny, M.-P. Teulade-Fichou, D. Monchaud, Highly efficient G-quadruplex recognition by bisquinolinium compounds, *J. Am. Chem. Soc.* 129 (2007) 1856–1857.
- [28] H. Bertrand, D. Monchaud, A.D. Cian, R. Guillot, J.-L. Mergny, M.-P. Teulade-Fichou, The importance of metal geometry in the recognition of G-quadruplex-DNA by metal-terpyridine complexes, *Org. Biomol. Chem.* 5 (2007) 2555–2559.
- [29] M. Read, R.J. Harrison, B. Romagnoli, F.A. Tanious, S.H. Gowan, A.P. Reszka, W.D. Wilson, L.R. Kelland, S. Neidle, Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 4844–4849.
- [30] J.F. Riou, L. Guittat, P. Mailliet, A. Laoui, E. Renou, O. Petitgenet, F. Mégnin-Chanet, C. Hélène, J.L. Mergny, Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 2672–2677.
- [31] O.Y. Fedoroff, M. Salazar, H. Han, V.V. Chmeris, S.M. Kerwin, L.H. Hurley, NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA, *Biochemistry* 37 (1998) 12367–12374.
- [32] J.L. Mergny, J. Li, L. Lacroix, S. Amrane, J.B. Chaires, Thermal difference spectra: a specific signature for nucleic acid structures, *Nucleic Acids Res.* 33 (2005) e138.
- [33] K.W. Lim, P. Alberti, A. Guédin, L. Lacroix, J.F. Riou, N. Royle, J.L. Mergny, A.T. Phan, Sequence variant (CTAGGG)_n in the human telomere favors a G-quadruplex structure containing a G•C•G•C tetrad, *Nucleic Acids Res.* 37 (2009) 6239–6248.
- [34] J.L. Mergny, A.T. Phan, L. Lacroix, Following G-quartet formation by UV-spectroscopy, *FEBS Lett.* 435 (1998) 74–78.
- [35] J.L. Mergny, L. Lacroix, UV melting of G-quadruplexes, *Curr. Protoc. Nucleic Acid Chem.* 17 (2009) 17.11.11–17.11.15.
- [36] J.L. Mergny, A. De Cian, A. Ghelab, B. Saccà, L. Lacroix, Kinetics of tetramolecular quadruplexes, *Nucleic Acids Res.* 33 (2005) 81–94.
- [37] 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, F.D. Boussin, Preferential binding of a G-quadruplex ligand to human chromosome ends, *Nucleic Acids Res.* 33 (2005) 4182–4190.
- [38] J.S. Ren, J.B. Chaires, Rapid screening of structurally selective ligand binding to nucleic acids, in: J.B. Chaires, M.J. Waring (Eds.), *Drug Nucleic Acid Interaction*.

- Academic Press Inc, 525 B Street, Suite 1900, San Diego, CA 92101-4495, USA, 2001, pp. 99–108.
- [39] L. Lacroix, A. Séosse, J.L. Mergny, Fluorescence-based duplex–quadruplex competition test to screen for telomerase RNA quadruplex ligands, *Nucleic Acids Res.* 39 (2011) e21.
- [40] C. Carrasco, F. Rosu, V. Gabelica, C. Houssier, E. De Pauw, C. Garbay-Jauregui, B. Roques, W.D. Wilson, J.B. Chaires, M.J. Waring, C. Bailly, Tight binding of the antitumor drug ditercalinium to quadruplex DNA, *Chem-biochem* 3 (2002) 1235–1241.
- [41] M.P. Teulade-Fichou, C. Carrasco, L. Guittat, C. Bailly, P. Alberti, J.L. Mergny, A. David, J.M. Lehn, W.D. Wilson, Selective recognition of G-quadruplex telomeric DNA by a bis(quinacridine) macrocycle, *J. Am. Chem. Soc.* 125 (2003) 4732–4740.
- [42] D.P. Arya, New approaches toward recognition of nucleic acid triple helices, *Account. Chem. Res.* 44 (2011) 134–146.
- [43] E. Largy, F. Hamon, M.-P. Teulade-Fichou, Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures, *Anal. Bioanal. Chem.* (2011), doi:[10.1007/s00216-00011-05018-z](https://doi.org/10.1007/s00216-00011-05018-z).

Fluorescence Intercalator Displacement assay for screening G4 ligands towards a variety of G-quadruplex structures

Phong Lan Thao Tran *et al.*

Supplementary information

Name		Sequence (from 5' to 3')	T _m ^a	Maxima CD	TO FI (2:1) ^b
TBA	Aptamer	G ₂ TTG ₂ TGTG ₂ TTG ₂	51	(+) 290 nm (+) 245 nm	1 x 10 ³
Bom17	Bombyx telomere	G ₂ TTAG ₂ TTAG ₂ TTAG ₂	45	(+) 290 nm (+) 245 nm	2.5 x 10 ³
VEGF	Promoter	G ₃ AG ₃ TTG ₄ TG ₃	> 70	(+) 260 nm	5 x 10 ³
VAV1	Promoter	G ₃ CAG ₃ AG ₃ AACTG ₃	51	(+) 260 nm	3 x 10 ³
C-Kit2	Promoter	G ₃ CG ₃ CGCGAG ₃ AGG ₃	> 70	(+) 260 nm	3 x 10 ³
Asc20	Ascaris telomere	G ₂ CTTAG ₂ CTTAG ₂ CTTAG ₂	37	(+) 290 nm (+) 245 nm	50 x 10 ³
21CTA	telomere	G ₃ CTAG ₃ CTAG ₃ CTAG ₃	67	(+) 290 nm (+) 245 nm	7 x 10 ³
C-Kit1	Promoter	G ₃ AG ₃ CGCTG ₃ AGGAG ₃	57.5	(+) 260 nm	8 x 10 ³
22Ag	Human DNA-telomere	AG ₃ TTAG ₃ TTAG ₃ TTAG ₃	66	(+) 290 nm	13 x 10 ³
22AgR	Human RNA-telomere	AG ₃ UUAG ₃ UUAG ₃ UUAG ₃	> 70	(+) 260 nm	10 x 10 ³
C-Myc	Promoter	TGAG ₃ TG ₃ TAG ₃ TG ₃ TAA	> 70	(+) 260 nm	6 x 10 ³
AKT1	Promoter	G ₃ CG ₃ CGGCTCCG ₃ CGCG ₃	62	(+) 290 nm	9 x 10 ³
PlasC24	Plasmodium telomere	G ₃ TTCAG ₃ TTCAG ₃ TTCAG ₃	64	(+) 290 nm	20 x 10 ³
GGA8	Trinucleotide	G ₂ AG ₂ AG ₂ AG ₂ AG ₂ AG ₂ AG ₂ A	> 70	(+) 260 nm	2.5 x 10 ³
Ceb25	minisatellites	AG ₃ TG ₃ TGTAAGTGTG ₃ TG ₃ T	> 70	(+) 260 nm	4.5 x 10 ³
27KRas	Promoter	G ₃ CGGTGTG ₃ AAGA G ₃ AAGAGG ₃	52	(+) 260 nm	8.5 x 10 ³
Insuline a2	Promoter	ACAG ₄ TGTG ₄ ACAG ₄ TGTG ₄	> 70	(+) 260 nm	8 x 10 ³
Oxy3.5	Oxytricha telomere	G ₄ TTTTG ₄ TTTTG ₄ TTTTG ₄	> 70	(+) 290 nm	19 x 10 ³
32B3 (K-Ras)	Promoter	AG ₃ CGGTGTG ₃ AAGAG ₃ AAGAG ₃ AGG	60	(+) 260 nm	25 x 10 ³
35B1 (K-Ras)	Promoter	AG ₃ CGGTGTG ₃ AAGAG ₃ AAGAG ₃ AGGCAG	54	(+) 260 nm	19 x 10 ³
45Ag	Human DNA-telomere	G ₃ TTAG ₃ TTAG ₃ TTAG ₃ TTAG ₃ TTAG ₃ TTAG ₃ TTAG ₃	59	(+) 290 nm	14 x 10 ³
TG4T	DNA-G4 tetramolecular	TG ₄ T	> 90	(+) 260 nm	1.6 x 10 ³
TG5T	DNA-G4 tetramolecular	TG ₅ T	> 90	(+) 260 nm	4.5 x 10 ³
UG5U	RNA-G4 tetramolecular	UG ₅ U	> 90	(+) 260 nm	2.5 x 10 ³
CGG12	Trinucleotide	(CGG) ₁₂	no G4	no G4	9 x 10 ³
dx9	duplex	(a)GCGATACGG (b)CCGTATCGC	ND ^c	ND ^c	11 x 10 ³
dx12	duplex	(a)GCGTGAGTTCGG (b)CCGAACACGC	ND ^c	ND ^c	18 x 10 ³
ds17	duplex	(a)CCAGTTCGTAGTAACCC (b)GGGTTACTACGAACCTGG	ND ^c	ND ^c	29 x 10 ³
ds26	duplex	CAATCGGATCGAATTCGATCCGATTG	ND ^c	ND ^c	34 x 10 ³
Poly dT	Single strand	TTTTTTTTTTTTTTTTTTTTTTTTTT	ND ^c	ND ^c	-144
Poly dA	Single strand	AAAAAAAAAAAAAAAAAAAAAAAAAAAA	ND ^c	ND ^c	308
Poly dC	Single strand	CCCCCCCCCCCCCCCCCCCCCCCCCCC	ND ^c	ND ^c	-350

Table S1. Sequences of oligonucleotides used in preliminary test/initial experiments.

^a: T_m (in °C) determined by UV melting experiments using absorbance at 295 nm, in 100 mM KCl with 10 mM lithium cacodylate. ^b: fluorescence of TO with 1 oligo for 2 TO. ^c: T_m not determined. The highlighted oligonucleotides were eliminated for the G4-FID test.

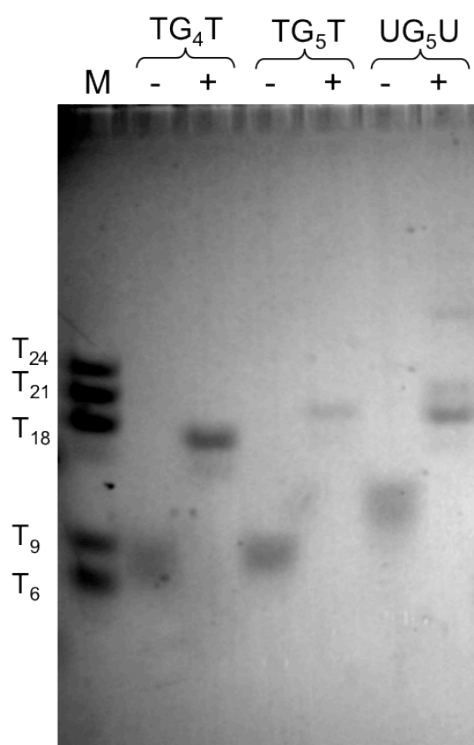


Figure S1. Behaviour of the tetramolecular G-quadruplex forming oligonucleotides on a non-denaturing gel. Samples were loaded on non-denaturing gels, and gels were run at 24°C. Oligonucleotides were revealed by UV shadowing. Each oligonucleotide was loaded at 60 μ M strand concentration. Lanes “-” correspond to the sequence pre-treated with LiOH (40 mM, 15min at 37°C) followed by neutralization by 40 mM HCl and immediate loading on a gel. Lanes “+” correspond to the sequence incubated in 10 mM lithium cacodylate 100 mM KCl for 48 h at 4°C. M corresponds to the migration markers (dT24, dT21, dT18, dT9 and dT6).

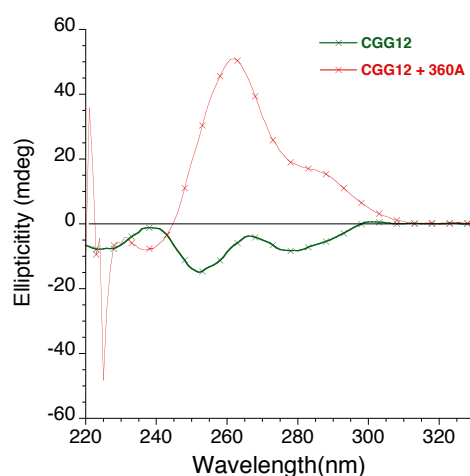


Figure S2. CD spectra of CGG12 (5 μ M) with and without ligand 360A (25 μ M). Spectra recorded at 4°C, in 10 mM lithium cacodylate (pH 7.2) supplemented with 100 mM KCl.

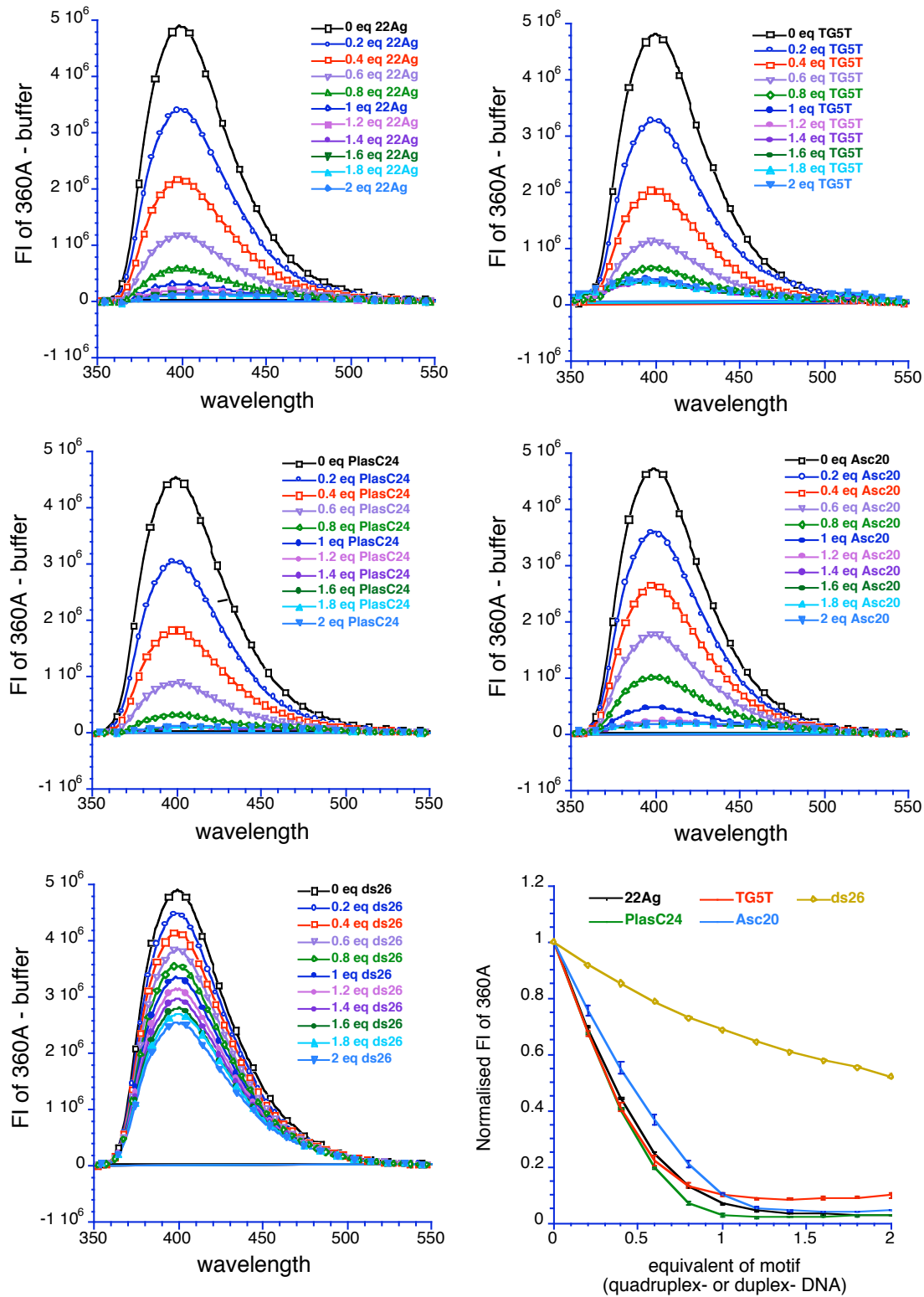


Figure S3. Fluorescence titrations of 360A (1 μ M) as a function of DNA concentration (excitation 325 nm, emission recorded on a Jobin-Yvon Fluoromax 3 spectrofluorimeter. Data quantified in the lower right panel – emission intensity at 400 nm plotted vs strand concentration with 5 nm bandwidth). Spectra recorded at 20°C, in 10 mM lithium cacodylate (pH 7.2) supplemented with 100 mM KCl.

3) Résultats et Discussion : Chapitre VI

Les résultats du criblage sont conformes aux résultats publiés obtenus par le test G4-FID classique ou par Tm-FRET. La plupart des molécules étudiées montre une discrimination entre les duplexes et les G-quadruplexes. Cependant aucun ligand testé ne semble montrer clairement une sélectivité pour une structure particulière. Les ligands testés déplacent le TO de la plupart des séquences de G-quadruplexes, sauf peut-être la séquence d(TG₅T). Nous notons un déplacement de TO plus faible sur le G-quadruplexe tétramoléculaire d(TG₅T). Ceci peut être dû à l'absence de boucle sur cette structure ou à la présence de thymine aux quartets terminaux qui pourrait rendre plus difficile l'accès à ces sites. Par ailleurs, en se liant aux G-quadruplexes tétramoléculaires, le TO présente une plus faible fluorescence par rapport aux autres structures. Cela pourrait indiquer un autre mode d'interaction de la sonde TO que l'empilement π sur les quartets, et rend plus délicate l'interprétation des résultats pour cette séquence.

Comme nous l'avons mentionné précédemment, la méthode décrite dans l'article permet un criblage d'un très grand nombre de structures d'acides nucléiques non-modifiés. Elle présente de nombreux avantages qui sont cités dans l'introduction de ce chapitre (cf. : §VI.1). Néanmoins, ce test a certaines limites et nécessite d'être amélioré.

(1) Les molécules fluorescences ayant des spectres d'émission et d'excitation qui se recouvrent avec ceux du TO ne peuvent pas être évalués par le test. Récemment, l'équipe de M.P. Teulade-Fichou a développé d'autres sondes fluorescentes (TO-PRO-3 et Hoechst 33258) [351] permettant de contourner ce problème.

(2) Le TO présente différentes constantes de dissociations (K_d) et différents rendements quantiques (= nombre de photons émis / nombre de photons absorbés) selon les structures liées. Ceci peut rendre difficile l'analyse des résultats et exclure l'évaluation de certaines séquences. Des structures exposant une augmentation d'émission de TO trop faible ne peuvent pas être utilisées pour le criblage puisque cela pourrait empêcher la détermination du déplacement de fluorescence. Pour ces structures qui présentent des constantes d'association faible pour le TO, l'augmentation du rapport de concentration TO/oligonucléotide (>2) pourrait régler le problème dans certains cas.

(3) Nous ne pouvons pas exclure une liaison simultanée du TO et du ligand sur la même structure. Cependant la formation d'un complexe ternaire semble peu probable vu la petite taille du quadruplexe.

(4) De plus, il est difficile d'expliquer les valeurs de déplacement "< 0%" et ">100%" trouvées pour certaines cibles.

(5) Enfin, ce test ne permet pas de déterminer facilement les paramètres quantitatifs ou cinétiques de l'association, il est donc nécessaire de compléter le test par d'autres techniques citées ci-dessus.

Malgré ces limites, la méthode est fiable, les résultats obtenus sont confirmés par des données issues d'autres techniques telles que le SPR et le FRET. Elle peut être très facilement utilisée pour déterminer la sélectivité du ligand pour les G-quadruplexes [352]. Le test peut également s'élargir à d'autres structures telles que des triplexes, l'ADN Z, des motifs i, des trinuécléotides, ... (sous réserve que ces complexes TO-structure permettent une fluorescence suffisante). Un criblage d'un grand nombre de ligands est également possible par cette technique [351].

Discussion et Perspectives

Chapitre VII. Discussion et perspectives

1) Effets des modifications sur la formation de G4 tétramoléculaires

Nous avons étudié, dans l'**article 1**, la formation de G4 tétramoléculaires formés par des séquences de type d(TG_nT) portant des guanines modifiées (8-méthyleguanine) à différentes positions. Notre étude a permis de montrer (i) que l'insertion d'une 8-méthyleguanine à l'extrémité (3' ou 5') favorise la formation d'un quartet tout-*syn* et (ii) que l'introduction de la guanine modifiée à l'extrémité 5' permet de stabiliser et d'augmenter la vitesse de formation des G4 tétramoléculaires.

La deuxième étude (**article 2**) a porté sur la formation de quadruplexes par des séquences d'ADN en série L (l'image de l'ADN canonique dans un miroir). Nous avons montré qu'il était possible de former des G4 tétramoléculaires composés uniquement de L-ADN avec une chiralité inversée par rapport aux G4 formés par les mêmes séquences en série D (D-ADN). Cependant, l'exclusion chirale empêche le mélange de deux énantiomères (D- et L-ADN) dans le même G4 tétramoléculaire.

Les résultats obtenus des études de G-quadruplexes modifiés (8-méthyleguanine et L-ADN) ont contribué à l'élucidation des mécanismes d'association de ces espèces et pourraient permettre de nombreuses applications en nanotechnologie et en biotechnologie. Ces aspects sont développés dans les paragraphes suivants.

1.1 Mécanismes réactionnels envisagés pour la formation de G4 tétramoléculaires

Les résultats de simulation de dynamique moléculaire publiés en 2003 par l'équipe de J. Sponer [353] suggèrent que l'association de G4 tétramoléculaires pourrait passer par plusieurs intermédiaires réactionnels stables. Ils décrivent un modèle "pas à pas" traduisant une association successive des brins (formation rapide de dimères et de trimères et enfin formation plus lente de tétramères) (Figure 52). Ils proposent également la formation d'espèces tétramoléculaires intermédiaires avec des appariements décalés (Figure 52). Ces conclusions sont confirmées par les résultats obtenues par l'équipe de J.L. Leroy en 2007 [354] et plus récemment par l'étude de spectrométrie de masse accomplie par F. Rosu et ses collègues [355]. Bardin et Leroy ont montré par RMN 1D que la formation de ces intermédiaires réactionnels mésappariés était rapide. La demi-vie de ces espèces étant assez longue, leur formation pourrait ralentir l'association du G4 tétramoléculaire "canonique".

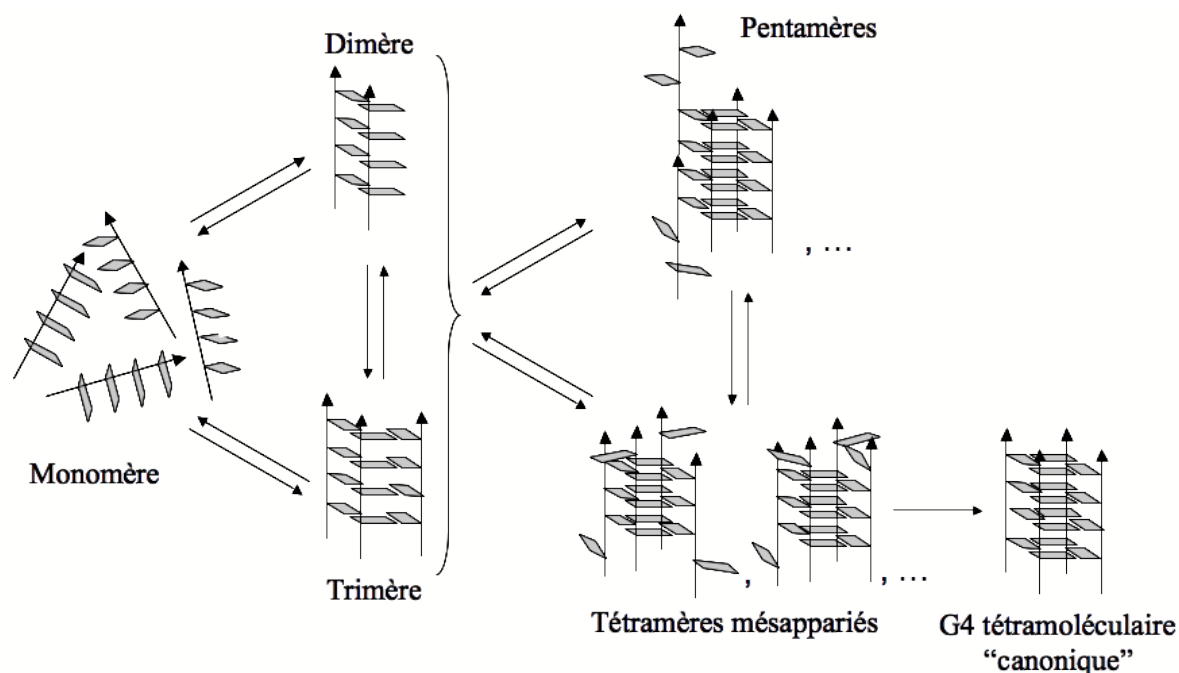


Figure 52. Schéma réactionnel proposé pour la formation de G-quadruplexe tétramoléculaire ($d(TG_nT)$) [355].

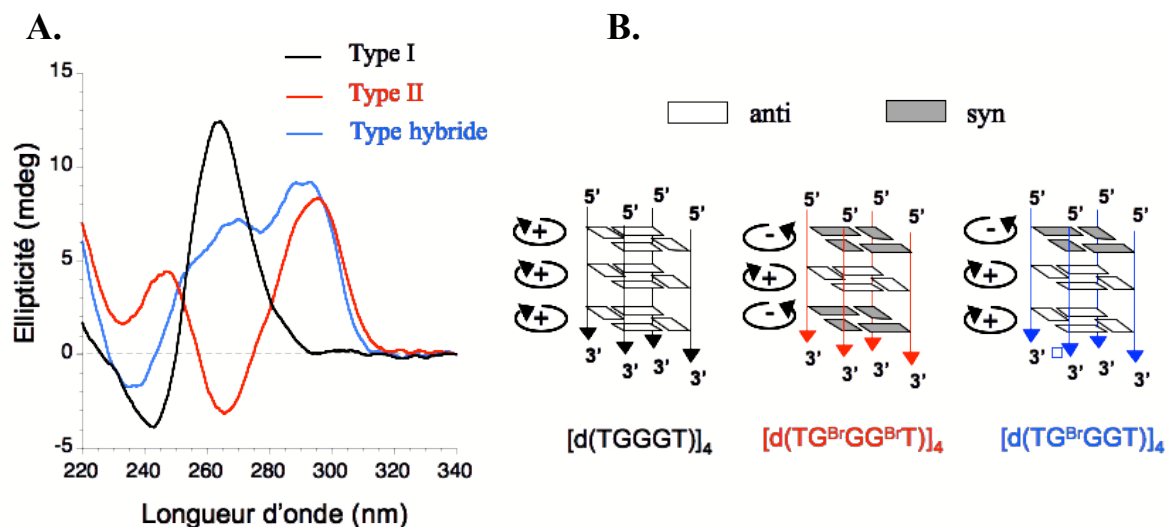


Figure 53. Signatures CD de G-quadruplexes

(A) Différents types de spectres CD: type I (en noir), type II (en rouge) et type hybride (en bleu). (B) Représentation schématique des enchaînements de polarités de G-quartets successifs qui pourraient permettre l'obtention de différents types de CD. La formation de G4 par les séquences $[d(TGGGT)]_4$ (en noir), $[d(TG^{Br}GG^{Br}T)]_4$ (en rouge), $[d(TG^{Br}GGT)]_4$ (en bleu) conduirait à des signaux CD de type I, de type II et de type hybride respectivement. G^{Br} correspond à la modification chimique 8-bromoguanine [127].

En effet, cela rend possible la détection de ces intermédiaires par les techniques utilisées. Rosu et ses collègues ont observé par spectrométrie de masse la présence de ces espèces intermédiaires, mais également d'une espèce formée de 5 brins (pentamère). Ils ont montré que le réarrangement des brins décalés et l'exclusion du cinquième brin étaient extrêmement lents, ce processus pouvant durer plus de 4 mois.

L'existence d'intermédiaires réactionnels ralentissant la formation de G4 tétramoléculaires "canoniques" est également observée dans notre étude de G4 avec des 8-méthyleguanines. En effet, pour certaines séquences étudiées (notamment, d(TGMGGT) et d(TGGGMT)), nous remarquons une différence entre le profil de CD enregistré à 48 heures et celui mesuré après un mois d'incubation (Figure S1 de l'**article 1**).

La position 8 d'une guanine pointe vers le sillon du quadruplexe. Le groupement méthyle porté par la 8-méthyleguanine est donc localisé dans le sillon du G4, tout comme la 8-bromoguanine et la 8-aminoguanine (Figure 21). Ces modifications sont en général relativement favorables pour la formation de G4, car elles ne perturbent pas les réseaux de liaisons H entre les guanines d'un quartet. Cependant, toutes les positions d'insertion de la modification sur la séquence d(TG₄T) ne sont pas favorables à l'association de la structure. La formation de G4 tétramoléculaires est accélérée de 15 fois quand la modification est introduite à l'extrémité 5' (d(TMGGT) et d(TMGGGT)). Ce résultat soutient l'hypothèse [141] que la guanine en *syn* à l'extrémité 5' participerait au processus de nucléation du complexe.

De plus, la formation d'un quartet tout-*syn* en 5' stabilise le complexe formé. En effet, des modèles moléculaires ont suggéré que l'empilement d'un quartet tout-*syn* sur un quartet tout-*anti* sous-jacent était plus favorable que l'empilement dans le sens inverse (quartet tout-*anti* sur quartet tout-*syn*).

1.2 Applications possibles de G4 modifiés, notamment 8-MeG et L-ADN

1.2.1 Étude de ligands spécifiques de G4

Nous pourrions étudier l'affinité de ligands de G4 avec les quadruplexes en série L et D. Certains ligands de G4 possédant un centre chiral, tels que la télomestatine, pourraient permettre une discrimination chirale [356]. Ce phénomène a été décrit auparavant pour les interactions avec les duplexes [357]. La comparaison de l'affinité pour les deux G-quadruplexes (L vs. D) pourrait nous renseigner sur le mode d'interaction du ligand. En effet, une interaction d'empilement terminal ne pourrait simplement expliquer une discrimination éventuelle entre les deux structures.

1.2.2 Étude de spectres de dichroïsme circulaire (CD) de G-quadruplexes

Précédemment, nous avons parlé de l'insertion d'une guanine modifiée en position 8 pour contrôler la formation de quartets tout-*syn* ou tout-*anti*. L'utilisation de ces analogues de G permet donc d'interpréter les variations de spectre de CD en fonction des changements induits dans l'enchaînement des G-quartets. Cela a été utilisé récemment par G.P. Spada *et al.* pour leur étude de variation de CD en fonction de l'empilement de quartets (*syn-anti*), indépendamment de l'orientation des brins du G4 [127].

En effet, dans des nombreux articles, les spectres CD de G4 sont interprétés en terme d'orientation de brins : un spectre avec un pic négatif autour de 240 nm et un pic positif autour de 260 nm est souvent interprété comme indicatif d'une structure G4 parallèle, tandis que un spectre avec un pic négatif autour de 260 nm et un pic positif autour de 295 nm est souvent interprété comme indicatif d'une structure G4 antiparallèle. Cette interprétation n'est pas appropriée.

En 2008, D. Gray *et al.* ont montré, par des calculs théoriques, l'origine de signaux de type I et de type II [126]. Ils ont montré que l'empilement de deux quartets avec la même polarité induit un signal de type I et, à l'inverse, l'empilement de deux quartets de polarités opposées donne un signal de type II.

Spada *et al.* ont étudié les spectres CD de G4 tétramoléculaires parallèles contenant des modifications chimiques (8-bromoguanine et 8-méthyleguanine) (cf. : §III.1.2.1 a), ainsi que les spectres théoriques résultants de l'empilement de quartets non reliés par un squelette phosphaté. Leur étude a confirmé les calculs de Gray *et al.* et a clairement montré que les signatures CD n'étaient pas liées à l'orientation des brins mais à la nature de l'empilement des quartets (cf. : §I.1.2.1 a). Ainsi, un G-quadruplexe possédant 3 quartets de polarités successives +/+ ou -/- (enchaînement *anti/anti/anti* ou *syn/syn/syn* sur un brin) conduirait à un signal CD de type I (Figure 53). Un signal de type II est obtenu avec un G4 présentant des quartets de polarités successives +/- ou -/+ (enchaînement *anti/syn/anti* ou *syn/anti/syn* sur un brin) (Figure 53). Enfin, un G4 avec des quartets de polarités successives -/+ ou +/+ (enchaînement *syn/anti/anti* sur un brin) induirait un signal CD de type hybride (2 pics positifs à 260 et 295 nm) (Figure 53).

1.2.3 Conception de sonde fluorescente interne

Les modifications de guanines en position 8 perturbant peu la formation de G-quadruplexes pourraient être utilisées pour construire des sondes fluorescentes internes. En effet, A. Dumas et N. Luedtke ont montré que l'incorporation de 8-(2-pyridyl)-guanine

(2PyG), de 8-(2-phenylethenyl)-guanine (StG) ou de 8-(2-[pyrid-4-yl]-ethenyl)-guanine (4PVG) dans une séquence télomérique humaine permettrait de détecter la formation de G4 par émission de fluorescence (Figure 54) [358]. Ces sondes fluorescentes internes montrent une grande sélectivité et sensibilité pour la détection de formation de G4, même à faible concentration en oligonucléotides (0,25 nM).

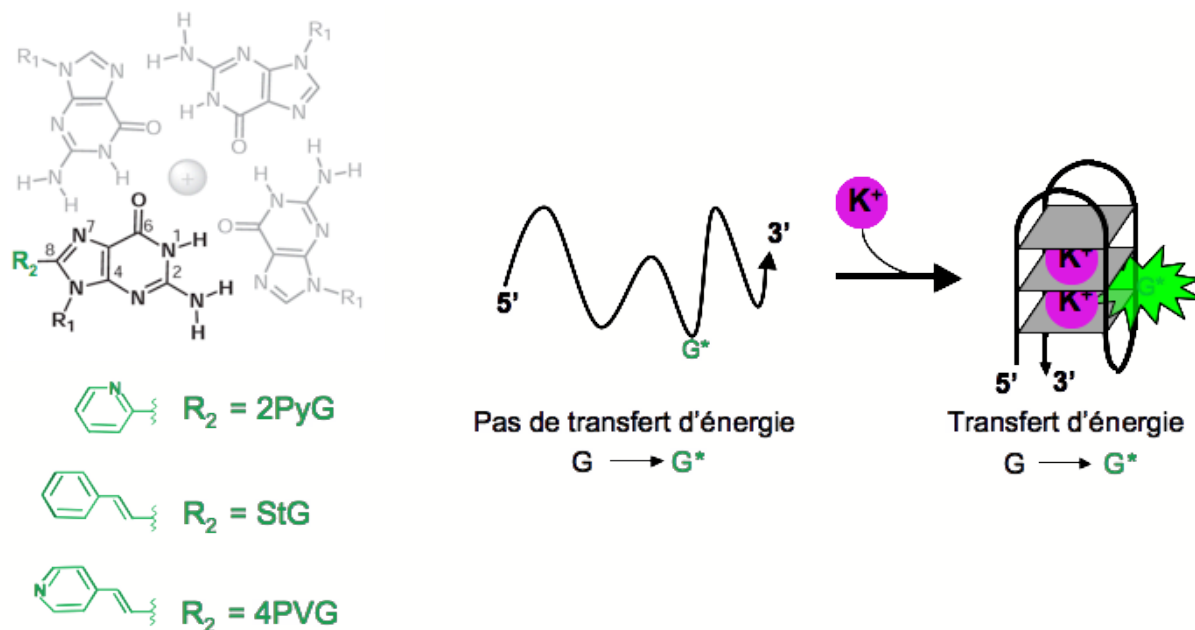


Figure 54. Schéma représentatif d'une sonde fluorescente interne [358]

(A) Introduction de différentes modifications (2PyG, StG ou 4PVG représenté en vert) à la position 8 du guanosine. (B) La formation de G4 permettrait un transfert d'énergie d'une G non-modifiée vers G modifiée (G^*), cela entraînerait une émission de fluorescence.

1.2.4 Application en nanotechnologie

Nous avons présenté, dans l'introduction générale, des applications possibles du G-quadruplexe, notamment les “G-wires” et les canaux ioniques à base de G4. Il est intéressant de noter qu'un transport efficace d'ions à travers ces canaux ioniques serait obtenu par rigidification des empilements de quartets en phase organique. La recherche de guanines modifiées (guanines lipophiles [116] et dérivés folates [117]) permettant la stabilisation de ces structures serait donc très utile pour cette application (cf. : §I.3.3).

Par ailleurs, les “G-wires” sont des structures dont la longueur, l'orientation et la molarité sont relativement difficiles à contrôler. Leur production et leur utilisation peuvent s'avérer complexes. Nous pouvons envisager l'insertion de modification permettant de mieux contrôler leur formation. Certaines études présentées dans l'introduction (cf. : §I.3.1) se sont intéressées à ce sujet [108, 114].

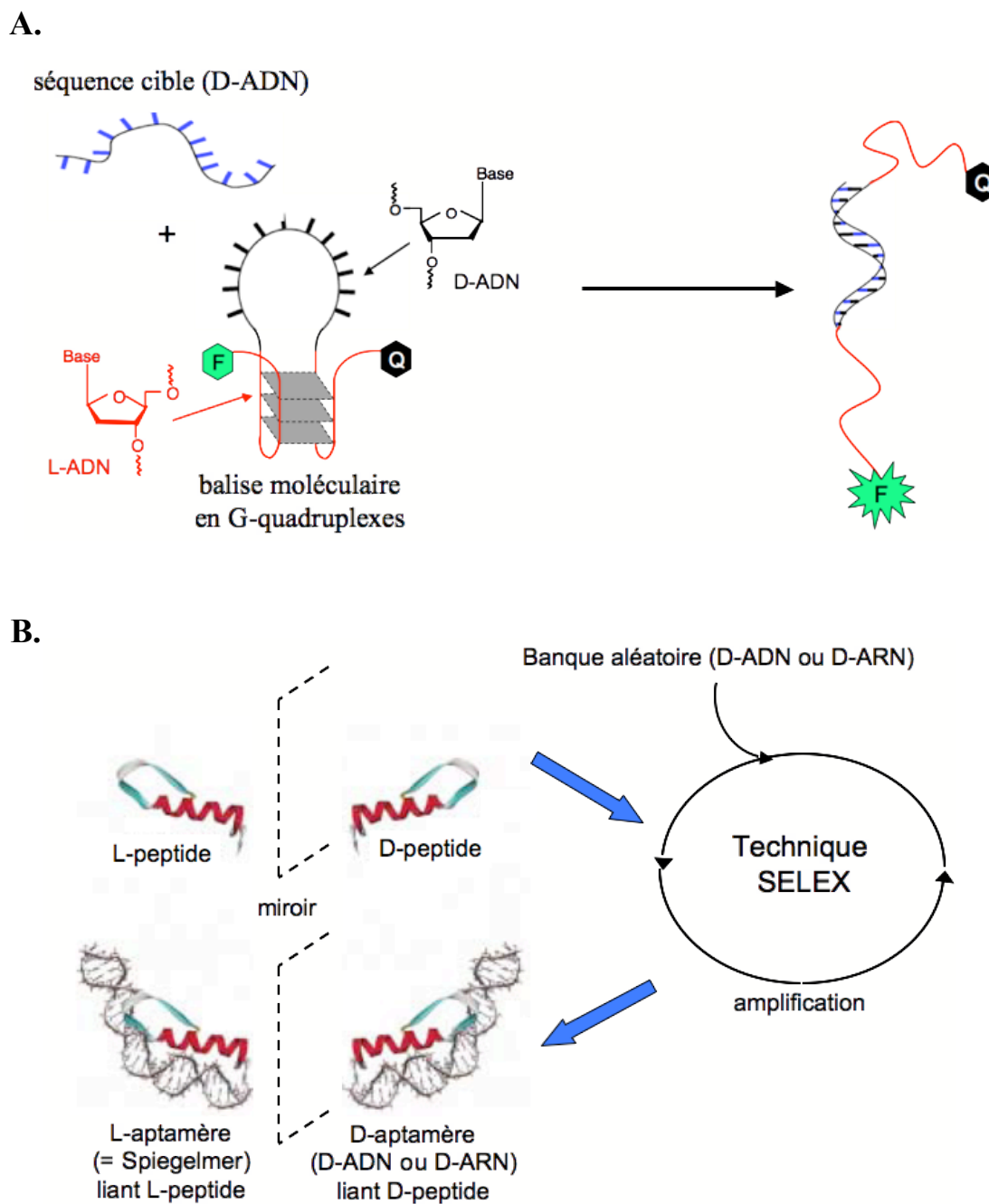


Figure 55. Applications potentielles de L-ADN en nanotechnologie et en biotechnologie
(A) Balise moléculaire (MB) fluorescente basée sur l'ouverture et la fermeture de G4. L'insertion de L-ADN dans la partie permettant la formation de G4 pourrait empêcher l'ouverture de la structure par des interactions non-spécifiques. **(B)** Principe de la technologie de "Spiegelmer" [359].

L'ADN en série L avec une chiralité inversée par rapport à son homologue en série D pourrait être utilisé pour la formation de structures supramoléculaires de G4. Lin et ses collègues ont réussi à contrôler l'hélicité d'un nanotube en introduisant des brins en L-ADN [360]. De plus, leur résistance aux nucléases permettrait à un cargo moléculaire à base de L-ADN transportant des molécules biologiquement actives de ne pas être dégradé dans un environnement cellulaire.

Nous pouvons également envisager l'utilisation de L-ADN pour construire une balise moléculaire à base de G4 (Figure 55A). L'équipe de W. Tan a inséré des L-acides nucléiques au niveau de la tige d'une balise moléculaire classique (en tige boucle) [361]. La séquence L-ADN ne forme pas de duplexe avec son homologue et empêche ainsi toute invasion de la tige par des interactions non-spécifiques. La boucle composant de D-ADN permet d'interagir avec la cible.

1.2.5 Application en biotechnologie

Nous avons mentionné précédemment que les L-acides nucléiques étaient résistants aux nucléases, cette caractéristique est très intéressante pour des applications en biotechnologie, notamment pour la recherche d'aptamères en série L, dit "*Spiegelmer*" ("*Spiegel*" signifie miroir en allemand). Le principe de la technologie "*Spiegelmer*" est simple, il s'agit de combiner la sélection *in vitro* d'aptamère par la méthode SELEX ("*Systematic Evolution of Ligands by EXponential enrichment*") et la reconnaissance chirale (Figure 55B). Il faut tout d'abord synthétiser l'image dans un miroir de la cible (protéines, ADN ou ARN), puis sélectionner *in vitro* un aptamère (oligonucléotide présentant une grande affinité avec la cible) contre cette cible synthétique. Cet aptamère en série D (ou D-aptamère) reconnaît spécifiquement la cible synthétique. Par symétrie, son image dans un miroir, un L-aptamère, interagirait efficacement avec la cible naturelle. Deux "*Spiegelmers*" sont entrés en essais cliniques [154, 362] (inhibiteurs de chimiokines MCP-1 et SDF-1) et ont montré un excellent profil de sécurité en phase I. Néanmoins cette technologie est utilisée, jusqu'à présent, contre des cibles peptidiques, nous pourrions donc envisager de sélectionner des "*Spiegelmers*" contre des cibles à bases d'acides nucléiques. Toutefois, cette approche est risquée pour des aptamères dirigés contre une cible ADN ou ARN non-structurée, car l'instabilité des duplexes L-D pourrait interdire la reconnaissance entre une cible (ADN ou ARN) en série L et une bibliothèque aléatoire d'oligonucléotides en série D. L'approche pourrait se révéler intéressante pour sélectionner des aptamères dirigés contre des structures G4.

2) Étude comparative de séquences télomériques et criblage de ligands sélectifs pour une conformation de G4 donnée

Dans l'**article 3**, nous avons présenté une étude comparative de la structure et de la stabilité des séquences télomériques de différents organismes. Cette étude a montré un potentiel de formation de G4 par toutes les séquences télomériques, à l'exception des séquences de levures. Ce travail a également révélé la formation possible d'un mélange de plusieurs structures en solution (G4 ou non-G4), ce point sera discuté dans le paragraphe ci-dessous.

2.1 Difficultés de l'étude de G4 intramoléculaires

Les techniques de haute résolution, comme la RMN ou la cristallographie, sont les seules qui permettent d'élucider des structures aussi variées que les G-quadruplexes. Mais, dans le cas des séquences formant plusieurs structures en solution, les techniques biophysiques et biochimiques utilisées dans l'**article 3** (spectroscopie d'absorption UV-visible, CD, gel d'électrophorèse en condition native) permettent d'obtenir des informations qui sont parfois perdues lors d'une étude de haute résolution. En effet, pour obtenir un spectre de RMN exploitable, il faut faire émerger une seule structure en solution, cela oblige, dans certains cas, à modifier la séquence pour favoriser une seule conformation.

Par exemple, la structure de la séquence télomérique de *Bombyx mori* (motif GGTTA) a été résolue par RMN[240] (cf. : §V.1.1.2). Cette structure a été obtenue en introduisant une modification dans la séquence étudiée. En revanche, l'analyse des courbes de dénaturation thermique (suivies par absorption et par FRET) et le dichroïsme circulaire ont permis, dans l'**article 3**, de mettre en évidence la présence d'au moins deux structures très différentes : une structure en G4 et une structure en épingle à cheveux. Cette information est importante puisqu'elle remet en question le potentiel de formation de G4 par cette séquence. Le même type de résultat a été obtenu avec le motif pour la séquence télomérique des nématodes (motif GGCTTA). Dans le cas de ces deux séquences (*Bombyx mori* et nématodes), nous observons un T_m à 295 nm plus important dans le sodium que dans le potassium. Ce résultat est inhabituel pour une séquence ne formant que des G4. La formation de deux structures (G4 et duplexe) par ces séquences explique ce comportement.

Nous observons également dans ce travail une variation des spectres CD avec la concentration en oligonucléotides pour certaines séquences (*Giardia intestinalis*, *Tetrahymena thermophila* et *Saccharomyces cerevisiae*) suggérant la formation de structures

intermoléculaires. Cependant les T_m de ces séquences ne varient pas avec la concentration en oligonucléotides : ce résultat laisse penser à une structuration de G4 intramoléculaires. L'hypothèse d'un empilement de G4 intramoléculaires permettant la formation de ces structures multi-brins pourrait expliquer ces résultats inusuels.

Le polymorphisme de structures formées par des séquences riches en guanines rend difficile la mise en place d'un algorithme prédisant la conformation de G4 à partir d'une séquence donnée.

Nous avons aussi remarqué que la vitesse de refroidissement peut influencer les conformations des G4 formées. Le spectre CD d'une solution issue d'un refroidissement rapide (chauffer à 90°C puis refroidir dans la glace) peut être différent de celui d'une solution issue d'un refroidissement plus lent (chauffé à 90°C puis refroidi avec un gradient de température de 1°C.min⁻¹ ou plus lent - jusqu'à 0,01°C.min⁻¹ pour R. Moriyama) (résultats non publiés). Un refroidissement rapide permettrait la formation de structures favorisées cinétiquement et un refroidissement lent favoriserait la formation de structures plus stables thermodynamiquement. Il est donc important d'indiquer la méthode de préparation de l'échantillon afin de garder les mêmes conformations de G4 structurées en solution. Cette étape est donc essentielle pour les oligonucléotides utilisés dans le test G4-FID sur plaque.

Par ailleurs, les séquences télomériques étudiées dans cette étude sont des séquences courtes (environ 4 répétitions de motif télomérique), ne pouvant former qu'un seul G-quadruplexe. Cependant, les séquences télomériques sont en général beaucoup plus longues et pourraient permettre la formation de G4 multiples. Suite à ces résultats, notre équipe a commencé à s'intéresser à la structure et à la stabilité des G4 multiples formés par des séquences télomériques longues, outre que celles des vertébrés.

2.2 Existe-il des ligands capables de discriminer différentes conformations de G4?

Compte tenu du grand nombre de PQS (*"Putative Quadruplex Sequences"*) dans différentes régions du génome, il n'est pas simple de montrer que l'effet biologique d'un ligand de G4 découle de son interaction avec une cible particulière.

Il existe, à présent, des ligands sélectifs pour les quadruplexes par rapport aux autres structures d'acides nucléiques, tels que les duplexes. Cependant, malgré de récents progrès dans ce domaine, nous n'avons, pour l'instant, qu'une seule molécule, le TOxapy (Figure 45), (Toxapy, patent EP 31 10 2008 /n° 08168104.1), capable de discriminer les différentes conformations de G4. Le TOxapy présente une grande affinité pour la séquence d(TG₅T) et le

G4 formé par la séquence télomérique humaine en Na^+ . À l'inverse, le ligand ne fixe pas le TBA (aptamère de la thrombine) et les G4 formés par la séquence télomérique humaine en K^+ . Les auteurs suggèrent une flexibilité du TOxapy qui permet au ligand de se fixer au niveau des sillons et de discriminer les différentes topologies de G4 [363]. L'étude d'interaction de ce dérivé heptacyclique neutre avec les G-quadruplexes serait très prometteuse pour le développement d'un ligand sélectif pour une topologie de G4 donnée. Ce ligand a été testé par criblage G4-FID. Il est intéressant de noter que le TOxapy n'est pas capable de déplacer totalement la fluorescence du TO de ces cibles (maximum de déplacement de TO = 60%). Ce déplacement partiel pourrait suggérer que le TOxapy et le TO n'aient pas le même site de fixation et qu'une liaison simultanée du TO et du ligand sur la même structure en G4 (complexe ternaire) pourrait se produire. Dans ce cas, le pourcentage de ligands liés à la structure serait donc sous-estimé. Ce résultat illustre bien une des limites de cette technique. Cela montre également que l'analyse des résultats de G4-FID doit se faire avec précaution.

Malgré des limites, le test G4-FID sur plaque développé dans l'**article 4** reste une méthode très rapide et facile à réaliser. Il pourrait être utilisé pour des tests préliminaires de ligands ou de protéines de G4, ou pour tester d'autres structures d'ADN (triplexes, ADN Z, motifs i ou trinuécléotides).

Par ailleurs, le développement d'un ligand "parfait" capable de cibler une région spécifique du génome s'avère difficile. Récemment, notre équipe a réalisé une SELEX contre le 360A, un ligand montrant une bonne sélectivité pour la structure en G4 par rapport au duplexe [364]. Le résultat est plutôt surprenant : seule une minorité des séquences d'aptamères sélectionnées (5 sur 80 séquences sélectionnées) pourrait former des G4. Cela suggère que le 360A, considéré comme un "bon" ligand de G4 n'est pas aussi sélectif qu'il y paraît, puisqu'il reconnaît non seulement les séquences de G4 mais également d'autres motifs d'acides nucléiques.

2.3 Structures G-quadruplexes d'ADN et d'ARN

Les télomères ont longtemps été considérés comme des régions non-transcrites de l'ADN. Cependant, des études ont récemment démontré que l'ADN télomérique des cellules de mammifères est transcrit en un ARN nommé TERRA (pour "Telomeric Repeat-Containing RNA") [365, 366]. Cet ARN est constitué d'un motif répété r(UUAGGG) de taille variable (entre 100 bases et 9kb). L'utilisation de la fluorescence de l'excimère du pyrène a permis de

mettre en évidence la formation *in vivo* de G-quadruplexes d'ARN par TERRA. Ces structures d'ARN sont localisées dans le noyau au niveau des télomères [367].

La découverte de TERRA et la présence de PQS dans des régions non-traduites d'ARNm (introns, 3' et 5'UTR) ont suscité l'intérêt pour l'étude structurale de séquences d'ARN riches en G. Les études de CD, de RMN, de spectrométrie de masse et de cristallographie aux rayons X ont montré que la séquence d'ARN télomérique humaine (TERRA) forme des structures de G4 parallèles très stable aussi bien en Na^+ qu'en K^+ (Figure 56A) [368-373]. Récemment, l'équipe de S. Balasubramanian a étudié l'effet des boucles sur la structure et stabilité des G4 intramoléculaires d'ARN [374]. Ils ont noté une diminution de la stabilité thermodynamique des G4 d'ARN lorsque la longueur des boucles est augmentée. Cependant la longueur des boucles semble ne pas ou peu affecter la structure du G4 d'après les spectres CD. Bien que les séquences d'ARN riche en G pourraient présenter un polymorphisme structural plus réduit, les études dans ce domaine et d'interaction de ligands de G4 avec ces structures méritent d'être approfondies.

Enfin, la localisation de TERRA au niveau des télomères soulève la question de savoir comment cet ARN s'associe à l'ADN télomérique. Pourrait-on supposer la formation d'un G4 hybride composé d'ADN et d'ARN (Figure 56A-B)? L'étude de Y. Xu *et al.*, en 2008, a montré que les séquences télomériques humaines d'ADN et d'ARN pouvaient s'hybrider pour former un G4 parallèle en présence de potassium [375].

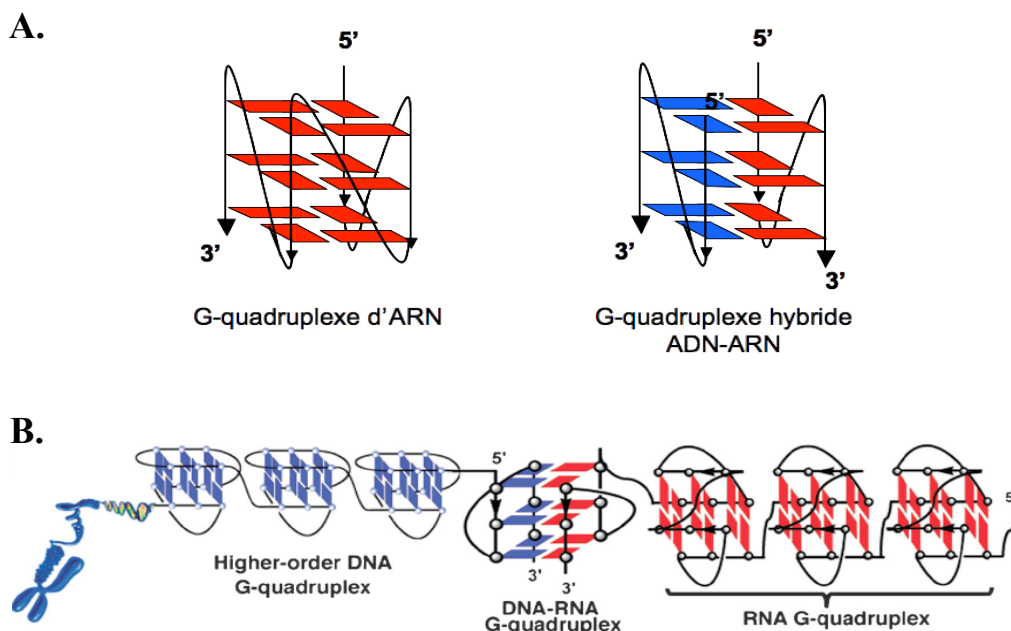
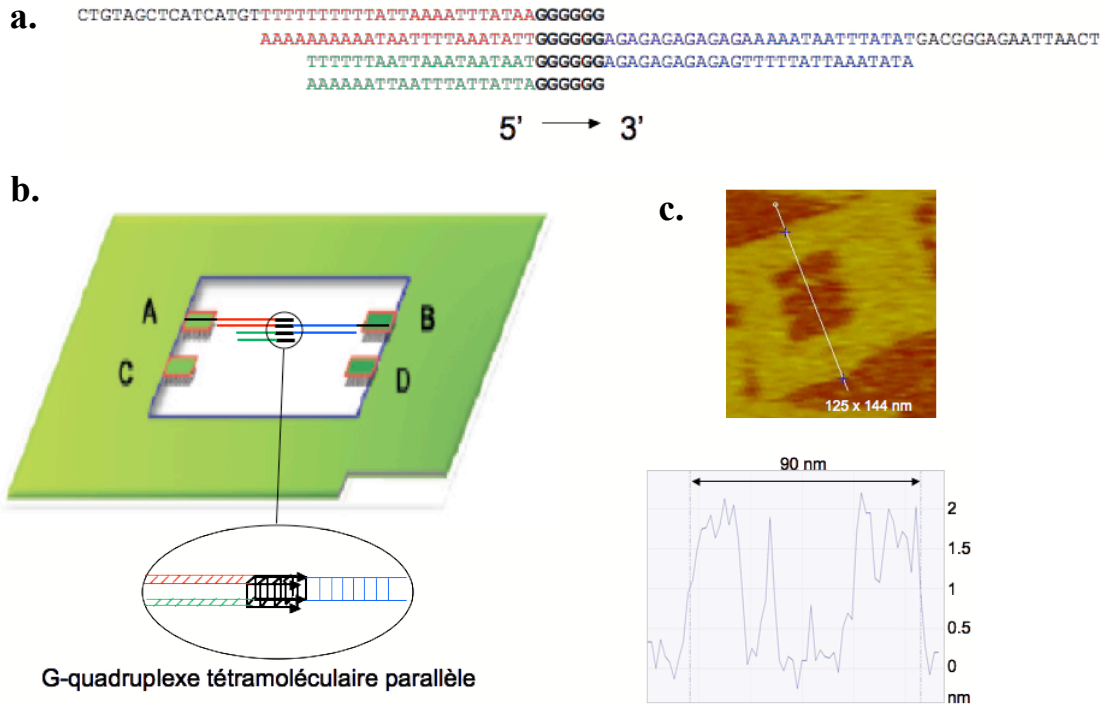


Figure 56. Schéma représentatif des structures de G4 d'ARN et de G4 hybride ARN-ADN
(A) Structure G-quadruplexe intramoléculaire parallèle de TERRA (à gauche) formé par la séquence $r[AGGG(UUAGGG)_3]$ [373] et structure putative de G-quadruplexe bimoléculaire parallèle composé d'ADN et d'ARN (à droite) [375]. (B) Représentation schématique de structure putative de G4 hybride ADN-ARN reliant l'ADN télomérique à TERRA [249].

A.



B.

Etat de forme "X": formation de G4

Etat parallèle

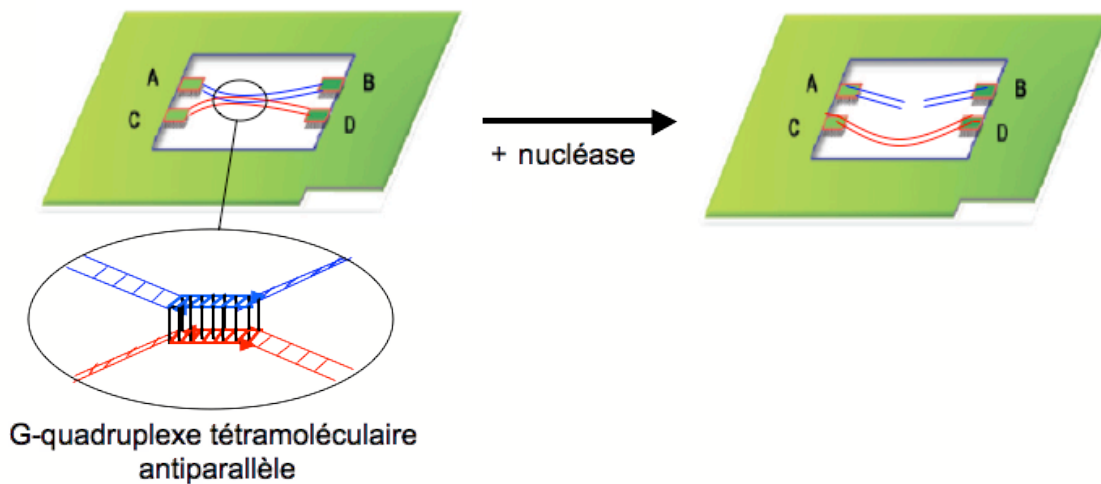


Figure 57. ADN origami et système d'observation de molécule unique

(A) Visualisation de G4 tétramoléculaire parallèle sur une nanostructure d'ADN: (a) Formation de G4 parallèle par association de duplexes parallèles entre les bases A-T; (b) Représentation schématique de l'insertion du G4 sur la nanostructure d'ADN origami; (c) Image d'AFM montrant l'insertion de la structure (haut) et mesure de hauteur de la coupe transversale (bas). (B) Schéma suggérant l'effet de nucléase sur le système de visualisation d'ADN "synaptique". En clivant le G4 tétramoléculaire, la nucléase permettrait le passage d'un état de forme "X" à un état parallèle.

3) L'ADN origami, un système d'observation de molécule unique

3.1 Système de visualisation de structures de G4

L'ADN origami présentant une surface plane permet d'attacher différentes molécules. Nous avons présenté dans le chapitre IV un système d'observation de molécules uniques à base d'ADN origami. Ce travail préliminaire, issu de 5 semaines de formation à Kyoto (Japon) dans le laboratoire du Pr. H. Sugiyama, a permis de visualiser la formation d'ADN "synaptique" (basé sur la formation d'un G4 tétramoléculaire antiparallèle) stable en absence de K^+ .

Ce système simple peut être utilisé pour la visualisation d'autres structures d'ADN, tel que le G-quadruplexe tétramoléculaire parallèle mise en place par Liliya Yatsunyk. La formation de ce G4 parallèle est contrôlée par la formation de duplexes parallèles composés de paires de bases A-T. En effet, il a été montré que les appariements dits "Watson-Crick inversé" entre les paires de bases A-T permettaient la formation de duplexes parallèles relativement stables en conditions physiologiques, en particulier en présence de Mg^{2+} [376, 377]. Le mélange de ces duplexes parallèles présentant deux blocs de G permet la formation de G4 tétramoléculaire parallèle (Figure 57A.a). Cette formation a été vérifiée par dichroïsme circulaire, par gel d'électrophorèse et par reconnaissance avec des ligands de G4 en spectroscopie UV-visible (cf. : §VI.1.2.1) (résultats non-communicés). La formation de cette structure a également été visualisée sur le système d'ADN origami présenté dans le chapitre IV (Figure 57A.b-c).

3.2 Système d'étude des interactions protéine / ligand – G4

Le système stable avec l'insertion de G4 tétramoléculaire antiparallèle (ADN "synaptique") ou parallèle (présenté dans le paragraphe précédent) permettrait d'étudier les interactions de certaines protéines (protéine Huntingtin, NCp du VIH-1 ou topoisomérase I humaine (cf. : §I.2.3)) ou de ligands fluorescents avec ces structures. Nous pourrions aussi envisager de suivre en temps réel la réaction enzymatique de certaines nucléases, comme GQN1 (cf. : §I.2.3), connues pour cliver le G4 tétramoléculaire (Figure 57B).

Le système avec la formation de G4 bimoléculaire par le motif télomérique humain [197] serait un système élégant pour l'étude des hélicases. Nous pourrions imaginer que l'interaction de l'hélicase avec la structure en G-quadruplexe (= état de forme "X") entraînerait la résolution de la structure (= état parallèle), nous aurions donc le passage d'un état de forme "X" à un état parallèle que l'on suivra par AFM.

Ce travail de thèse a permis d'aborder les différents aspects de l'étude de G-quadruplexes allant de la recherche fondamentale (la formation et la stabilité des G4 et l'interaction entre G4 et ligands) à l'application en nanotechnologie (système d'observation de molécule unique à base d'ADN origami). Ces études ont contribué : (i) à la description d'un mécanisme de formation de G4 tétramoléculaires avec l'étude de guanines modifiées, (ii) à l'enrichissement des données nécessaires au développement d'un algorithme prédisant la stabilité des G4 (cf. l'étude comparative des séquences télomériques), (iii) au développement d'une technique facile et rapide pour identifier des ligands spécifiques de G4 et (iv) à l'introduction d'un système simple et stable de visualisation de G4 unique sur une nanostructure d'ADN origami. Toutefois, de nombreux efforts sont encore à fournir pour anticiper et contrôler la formation des structures en G4, pour établir un algorithme permettant de prédire la conformation de G4 formé par une séquence donnée, pour identifier des ligands spécifiques d'une topologie de G4 et pour développer le système de visualisation de molécule unique à base d'ADN origami à d'autres applications.

Annexes

Annexe 1 : Séquence du phage M13

1-1000

AATGCTACTACTATTAGTAGAATTGATGCCACCTTTTCAGCTCGCGCCCCAAATGAAAATATAGCTAAACAGGTTATTGACCATTTCGCGAAATGTATCTA
ATGGTCAAACATAAATCTACTCGTTTCGAGCAATTTGGGAATCAACTGTTATATGGAATGAAACTTCCAGACACCGTACTTTAGTTGCATATTTAAACATGT
TGAGCTACAGCATTATATTACGAAATTAAGCTCTAAGCCATCCGCAAAATGACCTCTTATCAAAAGGAGCAATTAAGGTAAGTCTCTAATCCTGACCTG
TTGGAGTTTGGCTTCCGGTCTGGTTTCGCTTTGAAGCTCGAATTAAACCGCGATATTTGAAGTCTTTCGGGCTTCCCTTTAATCTTTTTGATGCAATCCCGT
TTGCTTCTGACTATATAAGTACAGGTAAGACCTGATTTTTGATTATAGGTGATTCGTTTTCTGAAGTCTTAAAGCAATTTGAGGGGATTTCAATGAA
TATTTATGACGATTCCGAGTATTGGACGCTATCCAGTCTAAACATTTTACTATTACCCCTCTGGCAAACTCTTTTGCAGAAAGCCTCTCGCTATTTT
GGTTTTTATCGTCGCTCGGTAACGAGGGTTATGATAGTGTGCTCTTACTATGCTCGTAATTCCTTTTGGCGTTATGATCTGCATTAGTTGAATGTG
GTATTCCTAAATCTCAACTGATGAATCTTTCTACCTGTAATAATGTTGTTCCGTTAGTTGCTTTTATTAACGTAGATTTTTCTTCCCAACGCTCCTGACTG
GTATAATGAGCCAGTTCTTAAATCGCATAGGTAATTCACAATGATTAAAGTTGAAATTAACCATCTCAAGCCCAATTTACTACTCTGTTCTGGTGT
CTCGTCAGGGCAAGCCTTATTCAGTGAATGAGCAGCTTTGTTACGTTGATTTGGGTAATGAATATCCGGTTCTTGTCAAGATTACTCTTGATGAAGGTCA

1001-2000

GCCAGCCTATGCGCTGGTCTGTACACCGTTTATCTGTCTCTTTCAAAGTTGGTCAGTTCGGTTCCCTTATGATTGACCGTCTGCGCTCGTTCCGGCT
AAGTAACATGGAGCAGGTCGCGGATTTTCGACACAATTTATCAGGCGATGATACAAATCTCCGTTGTACTTTGTTTCGCGCTTGGTATAATCGTGGGGT
CAAAGATGAGTGTGTTTAGTGATTTCTTTGGCTCTTTCTGTTTAGGTTGGTGCCTTCGTAGTGGCATTACGTTATTTACCCGTTTAAATGGAACTTCCCT
ATGAAAAGTCTTTAGTCCCTCAAAGCCTCTGTAGCCGTTGCTACCCCTCGTTCCGATGCTGTCTTTTCGCTGCTGAGGGTGACGATCCCGCAAAAGCGGCT
TTAATCCTCCTGCAAGCCTCAGCGACCGAATATATCGGTTATGCGTGGGCGATGGTTGTTGTGCTATTGTCGCGCACTATCGGTATCAAGCTGTTTAAAGAA
ATTCACCTCGAAAGCAAGCTGATAAACCGATACAATTAAGGCTCCTTTTGGAGCCTTTTTTTGGAGATTTTCAACGTGAAAAAATTTATTTTCGCAAT
TCCCTTAGTTTGTCTCTTCTATTTCTACTCCGCTGAAACTGTGAAAGTTGTTTAGCAAAATCCCATACAGAAAAATTTCTATTACGCTTTCGAAAGAC
GACAAAATTTAGATCGTTACGCTAATATGAGGCTGTCTGTGAATGCTACAGGCGTTGAGTTTGTACTGGTGACGAACTCAGTGTATACGGTACAT
GGGTTCTTATTGGGCTTGCTATCCCTGAAAAATGAGGTTGGTGGCTCTGAGGTTGGCGGTTCTGAGGTTGGCGGTTCTGAGGTTGGCGGTTACTAAACCTCC
TGAGTACGGTGAACACATTTCCGGGCTATCTTATATCAACCTCTCGACGGCACTTATCCGCTGGTACTGAGCAAAACCCCGCTAATCTAATCTCT

2001-3000

TCTCTTGAGGAGTCTCAGCCTCTTAATACTTTTACTGTTTTCAGAATAATAGGTTCCGAAATAGGCAGGGGGCATTAAGTGTTTATACGGGCACGTGTTACTC
AAGGCACTGACCCCGTTAAACTTATTACCAGTACACTCCTGTATCATCAAAGCCATGTATGACGCTTACTGGAACGGTAAATTCAGAGACTGCGCTTT
CCATTCCTGGCTTTAATGAGGATTTATTTGTTTGTGAATATCAAGGCCAATCGCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGT
GGTTCTGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAGGGAGGGCGGTTCCGGTGGTGGCTCTGGTTCCGGT
ATTTTGATTATGAAAGATGCGCAACGCTAATAAGGGGCTATGACCGAAATGCCGATGAAACGCGCTACAGTCTGACGCTAAAGGCAAACTTGATTC
TGTCGCTACTGATTACGGTGTGCTATCGATGGTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGTACTGGTGATTTTGTCTGGCTCTAAT
TCCCAAAATGGCTCAAGTCGGTGGGCTGATAATTCACCTTTAATGAATAATTTCCGTCATATTTTACCTTCCCTCCCTCAATGATGTCGCGCTT
TTGCTCTTGGCGCTGGTAAACCATATGAATTTTCTATTGATTGTGACAAAATAAATCTTATCCGTGGTGTCTTTGCGTTTCTTTTATATGTTGCCACCTT
TATGTATGATTTTCTACGTTTGCTAACATACTCGCTAATAAGGAGCTTAAATCATGCCAGTTCTTTTGGGTATTCCGTTATTATGGCTTTCTCCGTT
TCCTTCTGGTAATTTCTGCTCGCTATCTGCTTACTTTCTTAAAGAGTCTTCGTAAGATAGCTATTGCTATTTGTTCTTGTGCTTATTTATTTAGT

3001-4000

GCTTAACCTCAATTTCTGTGGGTATCTCTCTGATATTAGCGCTCAATTACCCTCTGACTTTGTTTCAGGGTGTTCAGTTAATTTCCCGCTCTAATGCGCTT
CCCTGTTTTTATGTTATTTCTCTCTGTAAAGGCTGCTATTTTCATTTTTTGACGTTAAACAAAAATCGTTTCTTATTGGATTGGGATAAAATATATGGCT
GTTTTATTTGTAACGGAATTAGGCTCTGGAAAGACGCTCGTTAGCCTTGGTAAGATTACAGGATAAAATGTAGCTGGTGCAAAATAGCAACTAATC
TTGATTTAAGGCTTCAAACCTCCCGCAAGTCGGGAGGTTTCGTAACAGCCTCCGCTTCTTAGAATACCGGATAAGCCTTCTATATCTGATTGCTTGC
TATTTGGGCGCGGTAATGATTCTACGATGAAAAATAAAGCGCTTGCTTGTCTCGATGAGTGGGTAAGTGGTTAATACCCGTTCTTGGAAATGATAAG
GAAAGACGCGGATTTATGATTGGTTTCTACATGCTCGTAAATTAGGATGGGATGATATTTTCTTGTTCAGGACTTATCTATTGTTGATAAACAGCGCG
GTTCTGCATTAGCTGAACATTTGTTTATTGTCGCTCGCTGAGCAGAAATTAATTTACCTTTTGTGCGGTACTTTATATCTTCTATTACTGGCTCGAAAA
GCCTCTGCTAAATACATGTTGGCGTTGTTAAATATGGCGATTCTCAATTAAGCCCTACTGTTGAGCGTTGGCTTTATCTGGTAAGAAATTTGTATAAC
GCATATGATACATAACAGGCTTTTCTAGTAATATGATTCGGGTGTTTATTTCTTATTTAAACGCTTATTTATCACAGCGTTCAGGATTTTCAACCACTAAC
ATTTAGGTCAGAAGATGAAATTAACATAAATATTTGAAAAAGTTTCTCGCTGCTTGTGCTTTCGCAATTGGATTGAGTTTACATATGATTAATGTTA

4001-5000

TATAACCCAACCTAAGCGGAGGTTAAAAAGGTAGTCTCTCAGACCTATGATTTTGATAAATCTACTATTGACTCTTCTCAGCGTCTTAATCTAAGCTAT
CGCTATGTTTTCAAGGATTTCAAGGAAAAATTAATTAATAGCGAGGATTACAGAGCAAGGTTATTCACACATATATTTGATTTATGACTGTTTCCAA
TTAAAAAGGTAATCAATGAAATTTGTTAAATGTAATTAATTTGTTTCTTGATGTTTGTTCATCATCTTCTTTTGGCTCAGGTAATGAAATGAATA
ATTCGCCCTCTCGCGGATTTGTTAATCTGGTATTCAAAGCAATCGGCAATTCGTTATTTGTTTCTCCCGATGTAAGAGGTAAGTCTTACTGATTTATTC
TGACGTTAAACCTGAAATCTACGCAATTTCTTTATTTCTGTTTTACGTGCAAAATAATTTTGATATGGTAGGTTCTAACCCCTTCCATTATTCAGAAGTAT
AATCCAAACATACAGGATTTATATTGATGAATGCGCATCATCTGATTAATCAGGAATATGATGATAATTCGCTCCTTCTGGTGGTTTCTTTGTTCCGCAAA
ATGATCAATTTACTCAAACTTTTAAATTAATAAGCTTCGGGCAAGGATTTAATACGAGTTGTCGAATTTGTTGTAAGCTCTAATACTCTCTAAATCTC
AAATGATTTATCTATTGACGGCTCTAATCTATTAGTTGTTAGTGCTCTTAAAGATATTTAGATAACCTTCTCAATTCCTTTCAACTGTTGATTGGCA
ACTGACAGATATTGATTGAGGGTTTGATATTTGAGGTTACAGACAGGTGATGCTTTAGATTTTCTATTGCTGCTGGCTCTCAGCGTGGCACTGTTGTCAG
CGGTTGTTAATACTGACCGCTCACCCTGTTTTATCTTCTGCTGTTGGTTGCTTCGTTATTTTAAAGGCGATGTTTAGGCGCTATCAGTTTCGCGCAAT

5001-6000

AAAGACTAATAGCCATTCAAAAATATTGTCGTGTCACGTAATTTCTACGTTTTCAGGTCAGAAGGGTTCTATCTCTGTTGGCCAGAAATGTCCCTTTTATT
ACTGGTCGTGTGACTGGTGAATTTGCGCAATGTAATAATCCATTTCAGACGATTGAGCGTCAAAATGTAGGTATTTCCATGAGCGTTTTCTGTTGCAAA
TGGCTGGCGGTAATATTGTTCTGGATATTACAGCAAGGCGGATAGTTTGTAGTTCTTCTACTCAGGCAAGTGATGTTATTAATCAAGAAAGTATTGC
TACAACGGTTAATTTGGCGTATGACAGACTCTTTTACTCGGTGGCCTCAGTATGATAAAAACTTCTCAGGATTTCTGGCGTACCGGTTCTGCTCTAA
ATCCCTTTAATCGCGCTCTCTGTTTAGCTCCCGCTCTGATTCTAACGAGGAAAGCAGTTATACGTGCTCGTCAAAAGCAACCATAGTACGCGCCCTTAGC
GGCGCATTAAGCGCGGCGGGTGTGGTGTGACGCGCAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCTCCCTCTCTT
TCGCGACGTTCCCGGCTTTTCCCGCTCAAGCTCTAAATCGGGGCTCCTCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAGAACTTGA
TTTGGGTGATGGTTACAGTAGTGGGCCATCGCCCTGATAGACGGTTTTTTCGCGCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA
ACTGGAACAACACTCAACCTTATCTCGGGCTATTCTTTGATTTAAGGGATTTTTCGCGATTTCGGAACCAACATCAACAGGATTTTTCGCTGCTGGG
GCAACACAGCTGGACCGCTTGCTGCAACTCTCTCAGGCGCAGGCGGTGAAGGCAATCAGTGTTCGCGCTCTACTGGTGAAAGAAAAACCAACCTG

6001-7000

GCGCCCAATACGCAAAACCGCTCTCCCCGCGGTTGGCGGATTCATTAATGACGCTGGCAGCAGGTTTCCGACTGGAAAGCGGGCAGTGAGCGCAAC
GCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTTACACTTTATGCTTCCGGCTCGTATGTTGTGGAATTTGTAGCGGATAACAATTTT
ACACAGGAAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGATCTCTAGAGTCGACCTGACGGCATGCAAGCTTTGGCACTGGCGCTCG
TTTTACAACGTCGTGACTGGGAAAACTTGGCGTTACCCAACTTAATCGCTTGCAGCAGATCCCCCTTTCGCGAGCTGGCGTAATAGCGAAGAGGCGCC
CACCGATCGCCCTTCCCAACGTTGCGCAGCTGAAATGGCGGCTTTGCTGGTTTCCGGCACCAGAGCGGTTCGCGAAGAGCTGGTGGAGTGC
GATCTTCTGAGGCGGATATGTCGTCGTTCCCTCAAACCTGGCAGATGACAGGTTACGATGCGGCCATCTACACCAACGTGACCTATCCCATTAACGGTCA
ATCCGCGTTGTTTCCACGAGAAATCCGACGGGTTGTACTCGCTCAGTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCAATTTATTT
TGATGGCGTTTCCATTTGGTTAAAAAATGAGCTGATTTAACAATAATTTAATGCGAATTTAACAATAATTAACGTTTACAAATTTAAATTTGCTTATA
CAATCTTCTGTTTGGGGCTTTTCTGATTATCAACCGGGTACATATGATTGACATGCTAGTTTACGATTACCGTTTATCGATTCTCTGTTGTTGCTC
CAGACTCTCAGGCAATGACCTGATAGCCTTTGTAGATCTCTCAAAAATAGCTACCCCTCTCCGGCATTAAATTTATCAGCTAGAACGGTTGAATATCATATT

7001-7429

GATGGTGATTGACTGTCTCCGGCTTTCTCACCCCTTTTGAATCTTTTACCTACACATTACTCAGGCATTGCATTTAAAAATATATGAGGGTTCTAAAAAT
TTTATCTTTCGCTTGAATAAAGGCTTCTCCGCAAAAGTATTACAGGCTCATATGTTTTTGGTACAACCGATTAGCTTTATGCTCTGAGGCTTTATTT
GCTTAATTTTGCTAATTTCTTGGCTTGCTGTATGATTATTTGGATGTT

Annexe 2 : Séquences d’“agraphes” utilisés

	Séquences	Longueur
1	TCGGTTTATCAGCTTGCTTTTCGAGTGCGCCGA	32
2	AAAAAAGGCTCCAAAAGGAGCCTTCATAACCG	32
3	AATAATAATTTTTTCACGTTGAAAAGGGAGTT	32
4	GAGAATAGAAAAGGAACAATAAGACCCTCAG	32
5	TGCTAAACAACCTTTCAACAGTTTCAGCGGAGT	32
6	ACGTTAGTAAATGAATTTTCTGTATACCGCCA	32
7	CGTAAACGATCTAAAGTTTGTGTCGTCGCCACC	32
8	TGTAGCATTCCACAGACAGCCCTCAGGGATAG	32
9	GTCACCAGTACAACTCGTAACAC	24
10	CAATGACAGAGGCAAAAAGAATACATACCAAGC	32
11	ATATATTCTGCCACTACGAAGGCATGTATCAT	32
12	AAAGGCCGGGAAGTTTCCATTAAACGCGACCT	32
13	CAGCGAAAAGAGGCTTTGAGGACTAGGCGCAG	32
14	TATCACCGTATAAGTATAGCCCGGGCCAGAAT	32
15	CCCTCAGAACCAGGCGGATAAGTGCGTTCCAG	32
16	CTCAGAGCGAAGGATTAGGATTAGGATACAGG	32
17	CAAGCCCATGAAAAGTATTAAGAGGGGGTCAG	32
18	TGAGTTTC TTT CTGCCTATTTGCGGAACAAACAGTT	36
19	CTCATCTTT TTT GCTTGATACCGATAGTGTGAATTTCTTAAACA	44
20	AAACGAAAACAACCATCGCCCCACGTAATTGTA	32
21	ATACGTAAGGTCGCTGAGGCTTGCATCTCCAA	32
22	TTTCATGACTTTTGCGGGATCGTCGAATTGCG	32
23	ACGGCTACGACAGCATCGGAACGAAATAGGTG	32
24	AGGGTTGATACTCAGGAGGTTTGTGTTGGGATT	32
25	TGCTCAGTACCGCCACCCTCAGAACTTTCCAG	32
26	CCTCAAGACACCACCCTCATTTTTCATAGTTAG	32
27	CTGAAACAATAGGAACCCATGTACACAACGCC	32
28	GCGAAACAACGTAACAAAGCTGTACACCAGA	32
29	CGCCTGATTGACAAGAACCGGATATGGTTTAA	32
30	GCTCCATGATAGGCTGGCTGACCTACCTTATG	32
31	ACGGTCAAGAGGACAGATGAACGGACCAGTCA	32
32	GGAAAGCGATAAATCCTCATTAATAATCAAGTT	32
33	TAAGCGTCGTCAGACGATTGGCCTCGTTTCA	32
34	AGTGTACTCCGCCGCCAGCATTGATTATTAGC	32
35	TGCCTTGACCTCAGAGCCGCCACCATCACCGG	32
36	AATGCCCT TTT ACCCTCAGAACCGCCACTCCCTCA	36
37	GAATAAGG TTT TGACCCCCAGCGATTACTAAAAACA	36
38	CCAAATCAAAGTACAACGGAGATTCCAACCTA	32
39	AGTAATCTAAATTGTGTCGAAATCCGGGTAAA	32
40	CCAGGCGCTTACTTAGCCGGAACGAAAGACTT	32
41	CTTTGAAATCATAAGGGAACCGAAGGGTAGCA	32
42	ACAAACAACAGTCTCTGAATTTACCGTCGAG	32
43	TGAGGCAGATACATGGCTTTTGATCGGGGTTT	32
44	CACCAGAGGGTAATAAGTTTAACTGAGACT	32
45	GCCACCACGTAACAGTGCCCGTATCTATTATT	32
46	ACGAGTAGGAGGCATAGTAAGAGCGATAAAAA	32
47	TTTCAACTTAATGCAGATACATAAAAAGAAGT	32
48	CGATTTTAATCAGTTGAGATTTAG	24
49	GGACGTTGACGGAACAACATTATTACAGGTAG	32
50	TGCCTTTAAATCAGTAGCGACAGATTAATAAAA	32
51	TCGGCATTTGTCACCAATGAAACCATCGATAGC	32
52	GTTTGCCAGTAGCACCATTACCATGGTTTACC	32
53	AACCAGAGCCATTTGGGAATTAGATTCAACCG	32
54	GAGCCGCC TTT TAAAGGTGAATTATCAGACGGAAA	36
55	CATAACCC TTT CTTGCCCTGACGAGAACATTAGT	36
56	GGAATTACTAAATTGGGCTTGAGATTTCATTAC	32
57	CATTCAACTAATCATTGTGAATTCATCAAG	32
58	AAAGATTCAGAACTGGCTCATTATTGTACAGA	32
59	ACGAACTAGGAAGAAAAATCTACGCTGACCAA	32
60	AGCACCGTGCGTCAGACTGTAGCGTGATATTC	32
61	TAGCAAGGCCGGAACCTTCGGTCATAGCCCCCAGGAGGT	40
62	AATCACCATCTTTTCATAATCAAAAAGAACCA	32
63	GACTTGAGCCACCACCGAACCGCCCCCTCAGA	32
64	CCAAAAATACATTGAATCCCCCTCACCATAAAT	32
65	TTTGCCAGGCGTCCAATACTGCGGTATTATAG	32
66	AGCGCCAAAATAGAAAATTCATATATAACGGA	32

67	ATTGAGGGACACCACGGAATAAGTAAGACTCC	32
68	TTATTCATTTTTAAAGGTGGCAACATAGTAGAAAA	36
69	AAACAGTTTTTCGTTTACCAGACGACAACACTAT	36
70	TAAATATTGCGAGAGGCTTTTGCACGCCAAAA	32
71	ACTGGATAAGGGGGTAATAGTAAAGAATACCA	32
72	TCACAATCAGACAAAAGGGCGACAGCCAGCAA	32
73	ACGCAAAAGAGGGAAGGTAAATATTCCGTCACC	32
74	CAAAAATCCGAGCTTCAAAGCGAATTAGAGAG	32
75	TCAGAAGCGCCCGAAAAGACTTCAAAGAGGTCA	32
76	ATACCCAACCGAGGAAACGCAATAGCGCTAAT	32
77	TTATTACGGCAGATAGCCGAACAATGAGTTAA	32
78	TACATACATTTTATCTTACCGAAGCCCATGAAATA	36
79	GAAGCAAAATTTTCAGAAAACGAGAATGAAATGCTTT	36
80	TTTAATTAGGTCTTTACCCTGACAATCGTCA	32
81	AAGAGGAAAAAGCGGATTGCATCAATGTTTAG	32
82	GAAGGAAAAAGAACTGGCATGATTTTATTTTG	32
83	AAAAGTAACAGTATGTTAGCAAACTAAAAGAA	32
84	TACCTTTATTTAGTTTGACCATTAAGCTATAT	32
85	TTTTTGCGACAGTTGATTCCCAATAGGTGGCA	32
86	ATTGCTGAATATAATGAAGTACGGTGTCTGGA	32
87	ATCAGAGAAGAGAATAACATAAAAACTCTGAA	32
88	GCCCAATAAACGTCAAAAATGAAATCCAGAG	32
89	GCAATAGCTTTTCCCAATCCAAATAAGACAGCCAT	36
90	TCGCAAAATTTTCTCCAACAGGTCAGGACCGACCG	36
91	CGAGTAGAATTGCTCCTTTTGATAATATCGCG	32
92	TCCATATAGATGGCTTAGAGCTTAAAAAGATT	32
93	ACAGGGAAGCGCATTAGTCAGAGGGTAATTGA	32
94	CTTTACAGGATAACCCACAAGAATAGTTACCA	32
95	TTTTGTTTATAAGAGCAAGAAACATTTTAAAG	32
96	TTTCATTTAGCAATAAAGCCTCAGTTATGACC	32
97	TCAATTCTCATACAGGCAAGGCAACTTTATTT	32
98	TCTTACCAACCCAGCTACAATTTTCGTAGGAA	32
99	CCTAATTTAAGCCTTAAATCAAGAAAAATCAGA	32
100	ATTATTTATTTTTAGCGAACCTCCCGAAAGAACGC	36
101	GCTAAATCTTTTGGTCAATAACCTGTTTGATACATT	36
102	CAAAATTAGGGGCGCGAGCTGAAATCTGCGAA	32
103	CAATAAAATACTAATAGTAGTAGCAAGTTTCAT	32
104	TATTTTGCACGCTAACGAGCGTCTATAGCAGC	32
105	AGGTTTGGCCAGTTACAAAATAAAAAACGATT	32
106	CTGTAATAATTCAAAAAGGGTGAGAGATATTCA	32
107	CAACGCAAAAATGCAATGCCTGAGCCGGAGAG	32
108	TCATTACCCCGTTTTTATTTTCATACAATAGA	32
109	TATAGAAGACCAAGTACCGCACTCATCCCATC	32
110	GAGGCGTTTTTCTTTCTTATCATTCATCAATA	36
111	AGACAGTCTTTTGGTTGTACCAAAAACAAGCATAAA	36
112	AGGTAAAGCTTTTGCGGGAGAAGCAGAATTAG	32
113	TATATTTGGATAAAAAATTTTAGTTAACATC	32
114	CAAGCAAGGCGCCCAATAGCAAGCTTAGTTGC	32
115	GGTATTAAGCTTATCCGGTATTCTCTTGCGGG	32
116	ACCGTTCTCCCAAAAACAGGAAGAAATATTTT	32
117	GGTAGCTATCATATGTACCCCGGTTTAAATCA	32
118	AAAGGCTATCAGGTCAGATGAACGGTAATCGT	32
119	TAAGTCCTCGACAAAAGGTAAAGTTATCATAT	32
120	CTAATTTATTTTCGAGCCAGTAATATAAAGCCA	32
121	ATCGGCTGTTTACAACGCCAACATGTAAGAATCGC	36
122	GCAAATATTTTAAATCACCATCAATATAAGGCCGG	36

Liste des publications scientifiques co-signées

1. **Tran, P.L.T.**, Mergny, J.L. & Alberti P. Stability of telomeric G-quadruplexes. **Nucleic Acids Research** - 2011; 39(8): 3282-94
2. **Tran, P.L.T.**, Virgilio, A. & Esposito, V., Citarella, G., Mergny, J.L. & Galeone, A. Effects of 8-methylguanine on structure, stability and kinetics of formation of tetramolecular quadruplexes. **Biochimie** - 2011; 93(3): 399-408
3. **Tran, P.L.T.**, Moriyama, R., Maruyama, A., Rayner, B. & Mergny, J.L. A mirror-image tetramolecular DNA quadruplex. **Chemical Communications** (Camb) - 2011; 47(19): 5437-9
4. **Tran, P.L.T.**, Largy E., Hamon F., Teulade-Fichou M.P., Mergny J.L. Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplexes. **Biochimie** –2011 Aug; 93(8): 1288-96.
5. Marcel, V., **Tran, P.L.T.**, Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.P., Hall, J. Mergny, J.L., Hainaut, P. & Van Dyck, E. G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. **Carcinogenesis** – 2011 Mar; 32(3): 271-8.
6. Smith, N.M., Labrunie, G., Corry, B., **Tran, P.L.T.**, Norret, M., Djavaheri-Mergny, M., Raston, C.L. & Mergny, J.L. Unraveling the relationship between structure and stabilization of triarylpyridines as G-quadruplex binding ligands. **Organic & Biomolecular Chemistry** – 2011 Sep; 9(17): 6154-62.
7. Mergny, J.L., Yatsunyk, L., Guédin, A., Gros, J., Renaud de la Faverie, A., Smith, N., Renciuk, D., **Tran, P.L.T.** & Bourdoncle, A., DNA quadruplexes for bio- and nano-technologies. *Chemistry of Nucleic Acid Components - Collection symposium series* – 2011; 12: 218-221.
8. Evrard, B., Merle, P., Petitjean, A., Lehn, J.M., Teulade-Fichou, M.P., Chautard, E., De Cian, A., Guittat, L., **Tran, P.L.T.**, Mergny, J.L., Verrelle, P. & Tchirkov, A. Telomere targeting with a new G4 ligand enhances radiation-induced killing of human glioblastoma cells. **Molecular Cancer Therapeutics** – 2011 Oct; 10(10): 1784-95.
9. Mandine, E., Mailliet, P., Walsh, J., Wolfrom, M., Hoffmann, D., Guédin, A., **Tran, P.L.T.** & Mergny, J.L. Targeting the KRAS promoter: LNAs vs quadruplex ligands (2011) – in preparation

Liste des communications

Présentation orale:

- Réunion de GDR-G4 – Paris (France) – 30 Mars 2011
- Journée Jeunes Chercheurs d'IFR 66 – Bordeaux (France) – 22 Juin 2010
- 3^{ème} Journées Jeunes Chercheurs à IECB – Bordeaux (France) – 27-28 Mai 2010
- 36^{ème} Forum des Jeunes Chercheurs – Nancy (France) – 25 - 27 Août 2009

Présentation de poster:

- 11^{ème} Congrès Général de la SFP (Société Française de Physique) – Bordeaux (France) – 4 - 8 juillet 2011
- 3^{ème} Congrès International sur “G-quadruplex and G-assembly” – Sorrente (Italie) – 28 juin - 1^{er} juillet 2011
- 18^{ème} Journée Jeunes Chercheurs de la SCT (Société de Chimie Thérapeutique) – Paris (France) – 4 février 2011
- Conférence organisé par EMBO sur les Télomères et la réponse aux dommages de l'ADN - Marseille (France) – 14 - 17 septembre 2010
- 2nd Summer School on Medecine “The whole process from ideas to drugs” – Toulouse (France) – 31 août - 10 septembre 2010
- 1st Training School – “Fundamentals of guanosine self-assembly and quadruplex formation” – Ischia (Italie) – 27 - 29 octobre 2009

Bibliographie

Bibliographie

1. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**(4356): p. 737-8.
2. Franklin, R.E. and R.G. Gosling, *Molecular configuration in sodium thymonucleate*. Nature, 1953. **171**(4356): p. 740-1.
3. Felsenfeld, G. and A. Rich, *Studies on the formation of two- and three-stranded polyribonucleotides*. Biochim Biophys Acta, 1957. **26**(3): p. 457-68.
4. Rhee, S., et al., *Structure of a triple helical DNA with a triplex-duplex junction*. Biochemistry, 1999. **38**(51): p. 16810-5.
5. Gellert, M., M.N. Lipsett, and D.R. Davies, *Helix formation by guanylic acid*. Proc Natl Acad Sci U S A, 1962. **48**: p. 2013-8.
6. Smith, F.W., F.W. Lau, and J. Feigon, *d(G3T4G3) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-syn-syn-anti and 5'-syn-anti-anti N-glycosidic conformations*. Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10546-50.
7. Smith, F.W. and J. Feigon, *Quadruplex structure of Oxytricha telomeric DNA oligonucleotides*. Nature, 1992. **356**(6365): p. 164-8.
8. Chantot, J. and W. Guschlbauer, *Physicochemical properties of nucleosides 3. Gel formation by 8-bromoguanosine*. FEBS Lett, 1969. **4**(3): p. 173-176.
9. Hud, N.V., et al., *The selectivity for K⁺ versus Na⁺ in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by 1H NMR*. Biochemistry, 1996. **35**(48): p. 15383-90.
10. Williamson, J.R., *G-quartet structures in telomeric DNA*. Annu Rev Biophys Biomol Struct, 1994. **23**: p. 703-30.
11. Hardin, C.C., et al., *Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG)*. Biochemistry, 1992. **31**(3): p. 833-41.
12. Wlodarczyk, A., et al., *Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies*. J Phys Chem B, 2005. **109**(8): p. 3594-605.
13. Wong, A. and G. Wu, *Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5'-monophosphate: a solid-state NMR study*. J Am Chem Soc, 2003. **125**(45): p. 13895-905.
14. Venczel, E.A. and D. Sen, *Parallel and antiparallel G-DNA structures from a complex telomeric sequence*. Biochemistry, 1993. **32**(24): p. 6220-8.
15. Wong, A., et al., *Disodium guanosine 5'-monophosphate self-associates into nanoscale cylinders at pH 8: a combined diffusion NMR spectroscopy and dynamic light scattering study*. J Am Chem Soc, 2005. **127**(19): p. 6990-8.
16. Patel, P.K., A.S. Koti, and R.V. Hosur, *NMR studies on truncated sequences of human telomeric DNA: observation of a novel A-tetrad*. Nucleic Acids Res, 1999. **27**(19): p. 3836-43.
17. Pan, B., et al., *Crystal structure of an RNA purine-rich tetraplex containing adenine tetrads: implications for specific binding in RNA tetraplexes*. Structure, 2003. **11**(7): p. 815-23.
18. Sket, P. and J. Plavec, *Tetramolecular DNA quadruplexes in solution: insights into structural diversity and cation movement*. J Am Chem Soc, 2010. **132**(36): p. 12724-32.
19. Patel, P.K. and R.V. Hosur, *NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by Saccharomyces cerevisiae telomere repeats*. Nucleic Acids Res, 1999. **27**(12): p. 2457-64.

20. Patel, P.K., N.S. Bhavesh, and R.V. Hosur, *NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG*. Biochem Biophys Res Commun, 2000. **270**(3): p. 967-71.
21. Deng, J., Y. Xiong, and M. Sundaralingam, *X-ray analysis of an RNA tetraplex (UGGGGU)(4) with divalent Sr(2+) ions at subatomic resolution (0.61 Å)*. Proc Natl Acad Sci U S A, 2001. **98**(24): p. 13665-70.
22. Cheong, C. and P.B. Moore, *Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures*. Biochemistry, 1992. **31**(36): p. 8406-14.
23. Pan, B., K. Shi, and M. Sundaralingam, *Base-tetrad swapping results in dimerization of RNA quadruplexes: implications for formation of the i-motif RNA octaplex*. Proc Natl Acad Sci U S A, 2006. **103**(9): p. 3130-4.
24. Mergny, J.L., et al., *Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes*. Nucleic Acids Res, 2006. **34**(8): p. 2386-97.
25. Kettani, A., R.A. Kumar, and D.J. Patel, *Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat*. J Mol Biol, 1995. **254**(4): p. 638-56.
26. Kettani, A., et al., *Solution structure of a Na cation stabilized DNA quadruplex containing G.G.G.G and G.C.G.C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA*. J Mol Biol, 1998. **282**(3): p. 619-36.
27. Bouaziz, S., A. Kettani, and D.J. Patel, *A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats*. J Mol Biol, 1998. **282**(3): p. 637-52.
28. Zavasnik, J., P. Podbevsek, and J. Plavec, *Observation of water molecules within the bimolecular d(GCTGC)G-quadruplex*. Biochemistry, 2011. **50**(19): p. 4155-61.
29. Lim, K.W., et al., *Sequence variant (CTAGGG)_n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad*. Nucleic Acids Res, 2009. **37**(18): p. 6239-48.
30. Webba da Silva, M., *Experimental demonstration of T:(G:G:G:G):T hexad and T:A:A:T tetrad alignments within a DNA quadruplex stem*. Biochemistry, 2005. **44**(10): p. 3754-64.
31. Zhang, N., et al., *Dimeric DNA quadruplex containing major groove-aligned A-T-A-T and G-C-G-C tetrads stabilized by inter-subunit Watson-Crick A-T and G-C pairs*. J Mol Biol, 2001. **312**(5): p. 1073-88.
32. Crnugelj, M., P. Sket, and J. Plavec, *Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations*. J Am Chem Soc, 2003. **125**(26): p. 7866-71.
33. Rachwal, P.A., et al., *Intramolecular DNA quadruplexes with different arrangements of short and long loops*. Nucleic Acids Res, 2007. **35**(12): p. 4214-22.
34. Guédin, A., et al., *How long is too long? Effects of loop size on G-quadruplex stability*. Nucleic Acids Res, 2010. **38**(21): p. 7858-68.
35. Bugaut, A. and S. Balasubramanian, *A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes*. Biochemistry, 2008. **47**(2): p. 689-97.
36. Rachwal, P.A., T. Brown, and K.R. Fox, *Sequence effects of single base loops in intramolecular quadruplex DNA*. FEBS Lett, 2007. **581**(8): p. 1657-60.
37. Vorlickova, M., et al., *Intramolecular and intermolecular guanine quadruplexes of DNA in aqueous salt and ethanol solutions*. Biopolymers, 2007. **86**(1): p. 1-10.
38. Guédin, A., et al., *Sequence effects in single-base loops for quadruplexes*. Biochimie, 2008. **90**(5): p. 686-96.
39. Guédin, A., P. Alberti, and J.L. Mergny, *Stability of intramolecular quadruplexes: sequence effects in the central loop*. Nucleic Acids Res, 2009. **37**(16): p. 5559-67.

40. Stegle, O., et al., *Predicting and understanding the stability of G-quadruplexes*. Bioinformatics, 2009. **25**(12): p. i374-82.
41. Okamoto, K., et al., *G-quadruplex structures of human telomere DNA examined by single molecule FRET and BrG-substitution*. Bioorg Med Chem, 2008. **16**(14): p. 6873-9.
42. Webba da Silva, M., et al., *Design of a G-quadruplex topology through glycosidic bond angles*. Angew Chem Int Ed Engl, 2009. **48**(48): p. 9167-70.
43. Wang, Y. and D.J. Patel, *Solution structure of the Tetrahymena telomeric repeat d(T2G4)4 G-tetraplex*. Structure, 1994. **2**(12): p. 1141-56.
44. Lim, K.W., et al., *Structure of the human telomere in K⁺ solution: a stable basket-type G-quadruplex with only two G-tetrad layers*. J Am Chem Soc, 2009. **131**(12): p. 4301-9.
45. Hud, N.V., et al., *Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex*. J Mol Biol, 1999. **285**(1): p. 233-43.
46. Sket, P., et al., *15NH₄⁺ ion movement inside d(G4T4G4)2 G-quadruplex is accelerated in the presence of smaller Na⁺ ions*. Org Biomol Chem, 2004. **2**(14): p. 1970-3.
47. Miyoshi, D., A. Nakao, and N. Sugimoto, *Molecular crowding regulates the structural switch of the DNA G-quadruplex*. Biochemistry, 2002. **41**(50): p. 15017-24.
48. Xue, Y., et al., *Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K⁺ solution under molecular crowding condition*. J Am Chem Soc, 2007. **129**(36): p. 11185-91.
49. Kumar, N., R. Basundra, and S. Maiti, *Elevated polyamines induce c-MYC overexpression by perturbing quadruplex-WC duplex equilibrium*. Nucleic Acids Res, 2009. **37**(10): p. 3321-31.
50. Yin, F., J. Liu, and X. Peng, *Triethylene tetraamine: a novel telomerase inhibitor*. Bioorg Med Chem Lett, 2003. **13**(22): p. 3923-6.
51. Campbell, N.H. and G.N. Parkinson, *Crystallographic studies of quadruplex nucleic acids*. Methods, 2007. **43**(4): p. 252-63.
52. Masse, J.E., et al., *Simple, efficient protocol for enzymatic synthesis of uniformly ¹³C, ¹⁵N-labeled DNA for heteronuclear NMR studies*. Nucleic Acids Res, 1998. **26**(11): p. 2618-24.
53. Phan, A.T. and D.J. Patel, *Differentiation between unlabeled and very-low-level fully ¹⁵N,¹³C-labeled nucleotides for resonance assignments in nucleic acids*. J Biomol NMR, 2002. **23**(4): p. 257-62.
54. Renciu, D., et al., *Arrangements of human telomere DNA quadruplex in physiologically relevant K⁺ solutions*. Nucleic Acids Res, 2009. **37**(19): p. 6625-34.
55. Yu, Z., et al., *ILPR G-quadruplexes formed in seconds demonstrate high mechanical stabilities*. J Am Chem Soc, 2009. **131**(5): p. 1876-82.
56. Huppert, J.L. and S. Balasubramanian, *Prevalence of quadruplexes in the human genome*. Nucleic Acids Res, 2005. **33**(9): p. 2908-16.
57. Huppert, J.L. and S. Balasubramanian, *G-quadruplexes in promoters throughout the human genome*. Nucleic Acids Res, 2007. **35**(2): p. 406-13.
58. Eddy, J. and N. Maizels, *Gene function correlates with potential for G4 DNA formation in the human genome*. Nucleic Acids Res, 2006. **34**(14): p. 3887-96.
59. Kumari, S., et al., *An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation*. Nat Chem Biol, 2007. **3**(4): p. 218-21.
60. Du, Z., Y. Zhao, and N. Li, *Genome-wide colonization of gene regulatory elements by G4 DNA motifs*. Nucleic Acids Res, 2009. **37**(20): p. 6784-98.

61. Siddiqui-Jain, A., et al., *Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription*. Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11593-8.
62. Grand, C.L., et al., *The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo*. Mol Cancer Ther, 2002. **1**(8): p. 565-73.
63. Bejugam, M., et al., *Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-kit oncogene expression*. J Am Chem Soc, 2007. **129**(43): p. 12926-7.
64. Palumbo, S.L., S.W. Ebbinghaus, and L.H. Hurley, *Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands*. J Am Chem Soc, 2009. **131**(31): p. 10878-91.
65. Qin, Y. and L.H. Hurley, *Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions*. Biochimie, 2008. **90**(8): p. 1149-71.
66. Kennedy, G.C., M.S. German, and W.J. Rutter, *The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription*. Nat Genet, 1995. **9**(3): p. 293-8.
67. Connor, A.C., et al., *Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide*. J Am Chem Soc, 2006. **128**(15): p. 4986-91.
68. Maleki, S., H. Cederberg, and U. Rannug, *The human minisatellites MS1, MS32, MS205 and CEB1 integrated into the yeast genome exhibit different degrees of mitotic instability but are all stabilised by RAD27*. Curr Genet, 2002. **41**(5): p. 333-41.
69. Ribeyre, C., et al., *The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo*. PLoS Genet, 2009. **5**(5): p. e1000475.
70. Piazza, A., et al., *Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae*. Nucleic Acids Res, 2010. **38**(13): p. 4337-48.
71. Marcel, V., et al., *G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms*. Carcinogenesis, 2011. **32**(3): p. 271-8.
72. Gomez, D., et al., *Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing*. Nucleic Acids Res, 2004. **32**(1): p. 371-9.
73. Gomez, D., et al., *A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells*. Nucleic Acids Res, 2010. **38**(20): p. 7187-98.
74. Huppert, J.L., et al., *G-quadruplexes: the beginning and end of UTRs*. Nucleic Acids Res, 2008. **36**(19): p. 6260-8.
75. Lyonnais, S., et al., *G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA*. Nucleic Acids Res, 2003. **31**(19): p. 5754-63.
76. Erlitzki, R. and M. Fry, *Sequence-specific binding protein of single-stranded and unimolecular quadruplex telomeric DNA from rat hepatocytes*. J Biol Chem, 1997. **272**(25): p. 15881-90.
77. Yerkes, S., J. Vesenka, and E.B. Kmiec, *A stable G-quartet binds to a huntingtin protein fragment containing expanded polyglutamine tracks*. J Neurosci Res, 2010. **88**(2): p. 335-45.
78. Giraldo, R. and D. Rhodes, *The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA*. EMBO J, 1994. **13**(10): p. 2411-20.

79. Muniyappa, K., S. Anuradha, and B. Byers, *Yeast meiosis-specific protein Hop1 binds to G4 DNA and promotes its formation*. Mol Cell Biol, 2000. **20**(4): p. 1361-9.
80. Arimondo, P.B., et al., *Interaction of human DNA topoisomerase I with G-quartet structures*. Nucleic Acids Res, 2000. **28**(24): p. 4832-8.
81. Bock, L.C., et al., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin*. Nature, 1992. **355**(6360): p. 564-6.
82. Li, W.X., et al., *A novel nucleotide-based thrombin inhibitor inhibits clot-bound thrombin and reduces arterial platelet thrombus formation*. Blood, 1994. **83**(3): p. 677-82.
83. Mohaghegh, P., et al., *The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases*. Nucleic Acids Res, 2001. **29**(13): p. 2843-9.
84. Kamath-Loeb, A.S., et al., *Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)_n trinucleotide repeat sequence*. J Biol Chem, 2001. **276**(19): p. 16439-46.
85. Fry, M. and L.A. Loeb, *Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)_n*. J Biol Chem, 1999. **274**(18): p. 12797-802.
86. Paeschke, K., J.A. Capra, and V.A. Zakian, *DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase*. Cell, 2011. **145**(5): p. 678-91.
87. Lopes, J., et al., *G-quadruplex-induced instability during leading-strand replication*. EMBO J, 2011. **30**(19): p. 4033-46.
88. Lipps, H.J. and D. Rhodes, *G-quadruplex structures: in vivo evidence and function*. Trends Cell Biol, 2009. **19**(8): p. 414-22.
89. Yang, Q., et al., *Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: II. The binding characterization with specific intramolecular G-quadruplex and the recognizing mechanism*. Nucleic Acids Res, 2010. **38**(3): p. 1022-33.
90. Fukuda, H., et al., *Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UPI*. Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12685-90.
91. Zaug, A.J., E.R. Podell, and T.R. Cech, *Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro*. Proc Natl Acad Sci U S A, 2005. **102**(31): p. 10864-9.
92. Salas, T.R., et al., *Human replication protein A unfolds telomeric G-quadruplexes*. Nucleic Acids Res, 2006. **34**(17): p. 4857-65.
93. Chung, I.K., et al., *Eukaryotic topoisomerase II cleavage of parallel stranded DNA tetraplexes*. Nucleic Acids Res, 1992. **20**(8): p. 1973-7.
94. Sun, H., A. Yabuki, and N. Maizels, *A human nuclease specific for G4 DNA*. Proc Natl Acad Sci U S A, 2001. **98**(22): p. 12444-9.
95. Liu, Z., et al., *Identification and characterization of a nuclease activity specific for G4 tetrastranded DNA*. Proc Natl Acad Sci U S A, 1993. **90**(8): p. 3157-61.
96. Liu, Z. and W. Gilbert, *The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of in vivo functions for this novel DNA structure*. Cell, 1994. **77**(7): p. 1083-92.
97. Liu, Z., A. Lee, and W. Gilbert, *Gene disruption of a G4-DNA-dependent nuclease in yeast leads to cellular senescence and telomere shortening*. Proc Natl Acad Sci U S A, 1995. **92**(13): p. 6002-6.

98. Schaffitzel, C., et al., *In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei*. Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8572-7.
99. Yang, Q., et al., *Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: I. recognizing mixed G-quadruplex in human telomeres*. Chem Commun (Camb), 2009(9): p. 1103-5.
100. Marsh, T.C. and E. Henderson, *G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures*. Biochemistry, 1994. **33**(35): p. 10718-24.
101. Lyonnais, S., et al., *Functionalization of DNA G-Wires for patterning and nanofabrication*. Nucleic Acids Symp Ser (Oxf), 2008(52): p. 689-90.
102. Protozanova, E. and R.B. Macgregor, Jr., *Frayed wires: a thermally stable form of DNA with two distinct structural domains*. Biochemistry, 1996. **35**(51): p. 16638-45.
103. Sen, D. and W. Gilbert, *Novel DNA superstructures formed by telomere-like oligomers*. Biochemistry, 1992. **31**(1): p. 65-70.
104. Marsh, T.C., J. Vesenska, and E. Henderson, *A new DNA nanostructure, the G-wire, imaged by scanning probe microscopy*. Nucleic Acids Res, 1995. **23**(4): p. 696-700.
105. Dai, T.Y., S.P. Marotta, and R.D. Sheardy, *Self-assembly of DNA oligomers into high molecular weight species*. Biochemistry, 1995. **34**(11): p. 3655-62.
106. Marotta, S.P., P.A. Tamburri, and R.D. Sheardy, *Sequence and environmental effects on the self-assembly of DNA oligomers possessing GxT2Gy segments*. Biochemistry, 1996. **35**(32): p. 10484-92.
107. Miyoshi, D., et al., *DNA nanowire sensitive to the surrounding condition*. Nucleic Acids Symp Ser (Oxf), 2005(49): p. 43-4.
108. Miyoshi, D., et al., *Artificial G-wire switch with 2,2'-bipyridine units responsive to divalent metal ions*. J Am Chem Soc, 2007. **129**(18): p. 5919-25.
109. Cohen, H., et al., *Polarizability of G4-DNA observed by electrostatic force microscopy measurements*. Nano Lett, 2007. **7**(4): p. 981-6.
110. Pavlov, V., B. Shlyahovsky, and I. Willner, *Fluorescence detection of DNA by the catalytic activation of an aptamer/thrombin complex*. J Am Chem Soc, 2005. **127**(18): p. 6522-3.
111. Bourdoncle, A., et al., *Quadruplex-based molecular beacons as tunable DNA probes*. J Am Chem Soc, 2006. **128**(34): p. 11094-105.
112. Zhu, J., et al., *G-quadruplex DNzyme based molecular catalytic beacon for label-free colorimetric logic gates*. Biomaterials, 2011. **32**(30): p. 7318-24.
113. Alberti, P. and J.L. Mergny, *DNA duplex-quadruplex exchange as the basis for a nanomolecular machine*. Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1569-73.
114. Borovok, N., et al., *Assembling of G-strands into novel tetra-molecular parallel G4-DNA nanostructures using avidin-biotin recognition*. Nucleic Acids Res, 2008. **36**(15): p. 5050-60.
115. Xiao, Y., et al., *Catalytic beacons for the detection of DNA and telomerase activity*. J Am Chem Soc, 2004. **126**(24): p. 7430-1.
116. Kaucher, M.S., W.A. Harrell, Jr., and J.T. Davis, *A unimolecular G-quadruplex that functions as a synthetic transmembrane Na⁺ transporter*. J Am Chem Soc, 2006. **128**(1): p. 38-9.
117. Sakai, N., et al., *Dendritic folate rosettes as ion channels in lipid bilayers*. J Am Chem Soc, 2006. **128**(7): p. 2218-9.
118. Hou, X., et al., *A biomimetic potassium responsive nanochannel: G-quadruplex DNA conformational switching in a synthetic nanopore*. J Am Chem Soc, 2009. **131**(22): p. 7800-5.

119. Alberti, P., et al., *DNA nanomachines and nanostructures involving quadruplexes*. Org Biomol Chem, 2006. **4**(18): p. 3383-91.
120. Ma, L., et al., *Large and stable transmembrane pores from guanosine-bile acid conjugates*. J Am Chem Soc, 2008. **130**(10): p. 2938-9.
121. Lubitz, I. and A. Kotlyar, *Self-assembled G4-DNA-silver nanoparticle structures*. Bioconjug Chem, 2011. **22**(3): p. 482-7.
122. Lubitz, I. and A. Kotlyar, *G4-DNA-Coated Gold Nanoparticles: Synthesis and Assembly*. Bioconjug Chem, 2011.
123. Mergny, J.L., A.T. Phan, and L. Lacroix, *Following G-quartet formation by UV-spectroscopy*. FEBS Lett, 1998. **435**(1): p. 74-8.
124. Mergny, J.L., et al., *Thermal difference spectra: a specific signature for nucleic acid structures*. Nucleic Acids Res, 2005. **33**(16): p. e138.
125. Mergny, J.L. and L. Lacroix, *Analysis of thermal melting curves*. Oligonucleotides, 2003. **13**(6): p. 515-37.
126. Gray, D.M., et al., *Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities*. Chirality, 2008. **20**(3-4): p. 431-40.
127. Masiero, S., et al., *A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures*. Org Biomol Chem, 2010. **8**(12): p. 2683-92.
128. Moriyama, R., et al., *DNA assembly and re-assembly activated by cationic comb-type copolymer*. Biomaterials, 2011. **32**(9): p. 2351-8.
129. Moriyama, R., et al., *The role of cationic comb-type copolymers in chaperoning DNA annealing*. Biomaterials, 2011. **32**(30): p. 7671-6.
130. Moriyama, R., et al., *Cationic comb-type copolymer as a nucleic acid chaperone for DNA quadruplex*. Nucleic Acids Symp Ser (Oxf), 2009(53): p. 61-2.
131. Moriyama, R., et al., *Poly (L-lysine)-graft-dextran acts as a nucleic acid chaperone for tetramolecular quadruplex formation*. Nucleic Acids Symp Ser (Oxf), 2008(52): p. 227-8.
132. Moriyama, R., et al., *The role of cationic comb-type copolymers in chaperoning DNA annealing*. Biomaterials, 2011.
133. Rothemund, P.W., *Folding DNA to create nanoscale shapes and patterns*. Nature, 2006. **440**(7082): p. 297-302.
134. Andersen, E.S., et al., *DNA origami design of dolphin-shaped structures with flexible tails*. ACS Nano, 2008. **2**(6): p. 1213-8.
135. Andersen, E.S., et al., *Self-assembly of a nanoscale DNA box with a controllable lid*. Nature, 2009. **459**(7243): p. 73-6.
136. Hu, D., et al., *A label-free, quadruplex-based functional molecular beacon (LFG4-MB) for fluorescence turn-on detection of DNA and nuclease*. Chemistry, 2011. **17**(5): p. 1635-41.
137. Prislán, I., et al., *Kinetically governed polymorphism of d(GTG) quadruplexes in K⁺ solutions*. Nucleic Acids Res, 2011. **39**(5): p. 1933-42.
138. Krishnan-Ghosh, Y., D. Liu, and S. Balasubramanian, *Formation of an interlocked quadruplex dimer by d(GGGT)*. J Am Chem Soc, 2004. **126**(35): p. 11009-16.
139. Phillips, K., et al., *The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution*. J Mol Biol, 1997. **273**(1): p. 171-82.
140. Mergny, J.L., et al., *Kinetics of tetramolecular quadruplexes*. Nucleic Acids Res, 2005. **33**(1): p. 81-94.
141. Gros, J., et al., *Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes*. Nucleic Acids Res, 2007. **35**(9): p. 3064-75.

142. De Cian, A. and J.L. Mergny, *Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation*. Nucleic Acids Res, 2007. **35**(8): p. 2483-93.
143. Esposito, V., et al., *Effects of the introduction of inversion of polarity sites in the quadruplex forming oligonucleotide TGGGT*. Bioorg Med Chem, 2009. **17**(5): p. 1997-2001.
144. Nielsen, J.T., K. Arar, and M. Petersen, *NMR solution structures of LNA (locked nucleic acid) modified quadruplexes*. Nucleic Acids Res, 2006. **34**(7): p. 2006-14.
145. Randazzo, A., et al., *NMR solution structure of a parallel LNA quadruplex*. Nucleic Acids Res, 2004. **32**(10): p. 3083-92.
146. Petraccone, L., et al., *Energetic aspects of locked nucleic acids quadruplex association and dissociation*. Biopolymers, 2006. **83**(6): p. 584-94.
147. Krishnan-Ghosh, Y., E. Stephens, and S. Balasubramanian, *A PNA4 quadruplex*. J Am Chem Soc, 2004. **126**(19): p. 5944-5.
148. Datta, B. and B.A. Armitage, *Hybridization of PNA to structured DNA targets: quadruplex invasion and the overhang effect*. J Am Chem Soc, 2001. **123**(39): p. 9612-9.
149. Datta, B., C. Schmitt, and B.A. Armitage, *Formation of a PNA2-DNA2 hybrid quadruplex*. J Am Chem Soc, 2003. **125**(14): p. 4111-8.
150. Marin, V.L. and B.A. Armitage, *RNA guanine quadruplex invasion by complementary and homologous PNA probes*. J Am Chem Soc, 2005. **127**(22): p. 8032-3.
151. Paul, A., et al., *Combining G-quadruplex targeting motifs on a single peptide nucleic acid scaffold: a hybrid (3+1) PNA-DNA bimolecular quadruplex*. Chemistry, 2008. **14**(28): p. 8682-9.
152. Roy, S., et al., *Kinetic discrimination in recognition of DNA quadruplex targets by guanine-rich heteroquadruplex-forming PNA probes*. Chem Commun (Camb), 2011. **47**(30): p. 8524-6.
153. Eulberg, D., et al., *The Aptamer Handbook: Functional Oligonucleotides and Their Applications*. 2006. **Chapter 18. Spiegelmers for Therapeutic Applications – Use of Chiral Principles in Evolutionary Selection Techniques**.
154. Sayyed, S.G., et al., *Podocytes produce homeostatic chemokine stromal cell-derived factor-1/CXCL12, which contributes to glomerulosclerosis, podocyte loss and albuminuria in a mouse model of type 2 diabetes*. Diabetologia, 2009. **52**(11): p. 2445-54.
155. Urata, H., et al., *Synthesis and properties of mirror-image DNA*. Nucleic Acids Res, 1992. **20**(13): p. 3325-32.
156. Garbesi, A., et al., *L-DNAs as potential antimessenger oligonucleotides: a reassessment*. Nucleic Acids Res, 1993. **21**(18): p. 4159-65.
157. Urata, H., et al., *Anti-HIV-1 activity of L-DNA quadruplex*. Nucleic Acids Res Suppl, 2002(2): p. 163-4.
158. Urata, H., et al., *Anti-HIV-1 activity and mode of action of mirror image oligodeoxynucleotide analogue of Zintevir*. Biochem Biophys Res Commun, 2004. **313**(1): p. 55-61.
159. Virgilio, A., et al., *Unprecedented right- and left-handed quadruplex structures formed by heterochiral oligodeoxyribonucleotides*. Biochimie, 2011. **93**(7): p. 1193-6.
160. Gros, J., et al., *8-Amino guanine accelerates tetramolecular G-quadruplex formation*. Chem Commun (Camb), 2008(25): p. 2926-8.
161. Virgilio, A., et al., *8-methyl-2'-deoxyguanosine incorporation into parallel DNA quadruplex structures*. Nucleic Acids Res, 2005. **33**(19): p. 6188-95.
162. Esposito, V., et al., *Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure [d(TGGGT)]4*. Org Biomol Chem, 2004. **2**(3): p. 313-8.

163. Shih, W.M., J.D. Quispe, and G.F. Joyce, *A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron*. Nature, 2004. **427**(6975): p. 618-21.
164. Seeman, N.C., *Nanomaterials based on DNA*. Annu Rev Biochem, 2010. **79**: p. 65-87.
165. Kuzuya, A. and M. Komiyama, *DNA origami: fold, stick, and beyond*. Nanoscale, 2010. **2**(3): p. 310-22.
166. Nangreave, J., et al., *DNA origami: a history and current perspective*. Curr Opin Chem Biol, 2010. **14**(5): p. 608-15.
167. Douglas, S.M., et al., *Rapid prototyping of 3D DNA-origami shapes with caDNAno*. Nucleic Acids Res, 2009. **37**(15): p. 5001-6.
168. Li, Z., et al., *Molecular behavior of DNA origami in higher-order self-assembly*. J Am Chem Soc, 2010. **132**(38): p. 13545-52.
169. Rajendran, A., et al., *Programmed two-dimensional self-assembly of multiple DNA origami jigsaw pieces*. ACS Nano, 2011. **5**(1): p. 665-71.
170. Liu, W., et al., *Crystalline two-dimensional DNA-origami arrays*. Angew Chem Int Ed Engl, 2011. **50**(1): p. 264-7.
171. Zhao, Z., Y. Liu, and H. Yan, *Organizing DNA origami tiles into larger structures using preformed scaffold frames*. Nano Lett, 2011. **11**(7): p. 2997-3002.
172. Douglas, S.M., et al., *Self-assembly of DNA into nanoscale three-dimensional shapes*. Nature, 2009. **459**(7245): p. 414-8.
173. Ke, Y., et al., *Scaffolded DNA origami of a DNA tetrahedron molecular container*. Nano Lett, 2009. **9**(6): p. 2445-7.
174. Han, D., et al., *DNA origami with complex curvatures in three-dimensional space*. Science, 2011. **332**(6027): p. 342-6.
175. Hogberg, B., T. Liedl, and W.M. Shih, *Folding DNA origami from a double-stranded source of scaffold*. J Am Chem Soc, 2009. **131**(26): p. 9154-5.
176. Jungmann, R., et al., *Isothermal assembly of DNA origami structures using denaturing agents*. J Am Chem Soc, 2008. **130**(31): p. 10062-3.
177. Pound, E., et al., *Polymerase chain reaction based scaffold preparation for the production of thin, branched DNA origami nanostructures of arbitrary sizes*. Nano Lett, 2009. **9**(12): p. 4302-5.
178. Zhao, Z., H. Yan, and Y. Liu, *A route to scale up DNA origami using DNA tiles as folding staples*. Angew Chem Int Ed Engl, 2010. **49**(8): p. 1414-7.
179. Endo, M., et al., *Inside Cover: Programmed-Assembly System Using DNA Jigsaw Pieces (Chem. Eur. J. 18/2010)*. Chemistry, 2010. **16**(18): p. 5228.
180. Endo, M., et al., *DNA prism structures constructed by folding of multiple rectangular arms*. J Am Chem Soc, 2009. **131**(43): p. 15570-1.
181. Ke, Y., et al., *Self-assembled water-soluble nucleic acid probe tiles for label-free RNA hybridization assays*. Science, 2008. **319**(5860): p. 180-3.
182. Voigt, N.V., et al., *Single-molecule chemical reactions on DNA origami*. Nat Nanotechnol, 2010. **5**(3): p. 200-3.
183. Rinker, S., et al., *Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding*. Nat Nanotechnol, 2008. **3**(7): p. 418-22.
184. Subramanian, H.K., et al., *The label-free unambiguous detection and symbolic display of single nucleotide polymorphisms on DNA origami*. Nano Lett, 2011. **11**(2): p. 910-3.
185. Kuzyk, A., K.T. Laitinen, and P. Torma, *DNA origami as a nanoscale template for protein assembly*. Nanotechnology, 2009. **20**(23): p. 235305.
186. Kuzuya, A., et al., *Precisely programmed and robust 2D streptavidin nanoarrays by using periodical nanometer-scale wells embedded in DNA origami assembly*. Chembiochem, 2009. **10**(11): p. 1811-5.

187. Maune, H.T., et al., *Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates*. Nat Nanotechnol, 2010. **5**(1): p. 61-6.
188. Ding, B., et al., *Gold nanoparticle self-similar chain structure organized by DNA origami*. J Am Chem Soc, 2010. **132**(10): p. 3248-9.
189. Pal, S., et al., *DNA-origami-directed self-assembly of discrete silver-nanoparticle architectures*. Angew Chem Int Ed Engl, 2010. **49**(15): p. 2700-4.
190. Numajiri, K., et al., *Stepwise and reversible nanopatterning of proteins on a DNA origami scaffold*. Chem Commun (Camb), 2010. **46**(28): p. 5127-9.
191. Endo, M., et al., *Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure*. J Am Chem Soc, 2010. **132**(5): p. 1592-7.
192. Gu, H., et al., *A proximity-based programmable DNA nanoscale assembly line*. Nature, 2010. **465**(7295): p. 202-5.
193. Torring, T., et al., *DNA origami: a quantum leap for self-assembly of complex structures*. Chem Soc Rev, 2011.
194. Lund, K., et al., *Molecular robots guided by prescriptive landscapes*. Nature, 2010. **465**(7295): p. 206-10.
195. Wickham, S.F., et al., *Direct observation of stepwise movement of a synthetic molecular transporter*. Nat Nanotechnol, 2011. **6**(3): p. 166-9.
196. Endo, M., et al., *A versatile DNA nanochip for direct analysis of DNA base-excision repair*. Angew Chem Int Ed Engl, 2010. **49**(49): p. 9412-6.
197. Sannohe, Y., et al., *Visualization of dynamic conformational switching of the G-quadruplex in a DNA nanostructure*. J Am Chem Soc, 2010. **132**(46): p. 16311-3.
198. Venczel, E.A. and D. Sen, *Synapsable DNA*. J Mol Biol, 1996. **257**(2): p. 219-24.
199. Fahlman, R.P. and D. Sen, *Cation-regulated self-association of "synapsable" DNA duplexes*. J Mol Biol, 1998. **280**(2): p. 237-44.
200. Pastre, D., et al., *Adsorption of DNA to mica mediated by divalent counterions: a theoretical and experimental study*. Biophys J, 2003. **85**(4): p. 2507-18.
201. Teixeira, M.T. and E. Gilson, *Telomere maintenance, function and evolution: the yeast paradigm*. Chromosome Res, 2005. **13**(5): p. 535-48.
202. Muller, H.J., *The remaking of chromosomes*. Collect. Net 13 p. 181-195.
203. McClintock, B., *The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis*. Proc Natl Acad Sci U S A, 1939. **25**(8): p. 405-16.
204. McClintock, B., *The Stability of Broken Ends of Chromosomes in Zea Mays*. Genetics, 1941. **26**(2): p. 234-82.
205. Blackburn, E.H. and J.G. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena*. J Mol Biol, 1978. **120**(1): p. 33-53.
206. Wellinger, R.J. and D. Sen, *The DNA structures at the ends of eukaryotic chromosomes*. Eur J Cancer, 1997. **33**(5): p. 735-49.
207. Klobutcher, L.A., et al., *All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus*. Proc Natl Acad Sci U S A, 1981. **78**(5): p. 3015-9.
208. Baroin, A., A. Prat, and F. Caron, *Telomeric site position heterogeneity in macronuclear DNA of Paramecium primaurelia*. Nucleic Acids Res, 1987. **15**(4): p. 1717-28.
209. Forney, J.D. and E.H. Blackburn, *Developmentally controlled telomere addition in wild-type and mutant paramecia*. Mol Cell Biol, 1988. **8**(1): p. 251-8.
210. Dore, E., et al., *Homologous telomeric sequences are present in different species of the genus Plasmodium*. Mol Biochem Parasitol, 1986. **21**(2): p. 121-7.

211. Bottius, E., N. Bakhsis, and A. Scherf, *Plasmodium falciparum telomerase: de novo telomere addition to telomeric and nontelomeric sequences and role in chromosome healing*. Mol Cell Biol, 1998. **18**(2): p. 919-25.
212. Le Blancq, S.M., R.S. Kase, and L.H. Van der Ploeg, *Analysis of a Giardia lamblia rRNA encoding telomere with [TAGGG]_n as the telomere repeat*. Nucleic Acids Res, 1991. **19**(20): p. 5790.
213. Van der Ploeg, L.H., A.Y. Liu, and P. Borst, *Structure of the growing telomeres of Trypanosomes*. Cell, 1984. **36**(2): p. 459-68.
214. Blackburn, E.H. and P.B. Challoner, *Identification of a telomeric DNA sequence in Trypanosoma brucei*. Cell, 1984. **36**(2): p. 447-57.
215. McEachern, M.J. and E.H. Blackburn, *A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts*. Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3453-7.
216. Leonardi, J., et al., *TER1, the RNA subunit of fission yeast telomerase*. Nat Struct Mol Biol, 2008. **15**(1): p. 26-33.
217. Trujillo, K.M., J.T. Bunch, and P. Baumann, *Extended DNA binding site in Pot1 broadens sequence specificity to allow recognition of heterogeneous fission yeast telomeres*. J Biol Chem, 2005. **280**(10): p. 9119-28.
218. Wang, S.S. and V.A. Zakian, *Sequencing of Saccharomyces telomeres cloned using T4 DNA polymerase reveals two domains*. Mol Cell Biol, 1990. **10**(8): p. 4415-9.
219. Shampay, J., J.W. Szostak, and E.H. Blackburn, *DNA sequences of telomeres maintained in yeast*. Nature, 1984. **310**(5973): p. 154-7.
220. Richards, E.J. and F.M. Ausubel, *Isolation of a higher eukaryotic telomere from Arabidopsis thaliana*. Cell, 1988. **53**(1): p. 127-36.
221. Petracek, M.E., et al., *Chlamydomonas telomere sequences are A+T-rich but contain three consecutive G-C base pairs*. Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8222-6.
222. Nozaki, H., et al., *A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae*. BMC Biol, 2007. **5**: p. 28.
223. Ganai, M.W., N.L. Lapitan, and S.D. Tanksley, *Macrostructure of the tomato telomeres*. Plant Cell, 1991. **3**(1): p. 87-94.
224. Okazaki, S., et al., *Identification of a pentanucleotide telomeric sequence, (TTAGG)_n, in the silkworm Bombyx mori and in other insects*. Mol Cell Biol, 1993. **13**(3): p. 1424-32.
225. Muller, F., et al., *New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in Ascaris lumbricoides*. Cell, 1991. **67**(4): p. 815-22.
226. Cangiano, G. and A. La Volpe, *Repetitive DNA sequences located in the terminal portion of the Caenorhabditis elegans chromosomes*. Nucleic Acids Res, 1993. **21**(5): p. 1133-9.
227. Wicky, C., et al., *Telomeric repeats (TTAGGC)_n are sufficient for chromosome capping function in Caenorhabditis elegans*. Proc Natl Acad Sci U S A, 1996. **93**(17): p. 8983-8.
228. Moyzis, R.K., et al., *A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes*. Proc Natl Acad Sci U S A, 1988. **85**(18): p. 6622-6.
229. Meyne, J., R.L. Ratliff, and R.K. Moyzis, *Conservation of the human telomere sequence (TTAGGG)_n among vertebrates*. Proc Natl Acad Sci U S A, 1989. **86**(18): p. 7049-53.

230. Sen, D. and W. Gilbert, *Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis*. Nature, 1988. **334**(6180): p. 364-6.
231. Wang, Y. and D.J. Patel, *Solution structure of a parallel-stranded G-quadruplex DNA*. J Mol Biol, 1993. **234**(4): p. 1171-83.
232. Parkinson, G.N., M.P. Lee, and S. Neidle, *Crystal structure of parallel quadruplexes from human telomeric DNA*. Nature, 2002. **417**(6891): p. 876-80.
233. Ambrus, A., et al., *Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution*. Nucleic Acids Res, 2006. **34**(9): p. 2723-35.
234. Dai, J., et al., *Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence*. Nucleic Acids Res, 2007. **35**(15): p. 4927-40.
235. Dai, J., et al., *Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation*. Nucleic Acids Res, 2007. **35**(7): p. 2440-50.
236. Luu, K.N., et al., *Structure of the human telomere in K⁺ solution: an intramolecular (3 + 1) G-quadruplex scaffold*. J Am Chem Soc, 2006. **128**(30): p. 9963-70.
237. Phan, A.T., et al., *Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution*. Nucleic Acids Res, 2007. **35**(19): p. 6517-25.
238. Phan, A.T., K.N. Luu, and D.J. Patel, *Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K⁺ solution*. Nucleic Acids Res, 2006. **34**(19): p. 5715-9.
239. Hu, L., et al., *Giardia telomeric sequence d(TAGGG)4 forms two intramolecular G-quadruplexes in K⁺ solution: effect of loop length and sequence on the folding topology*. J Am Chem Soc, 2009. **131**(46): p. 16824-31.
240. Amrane, S., et al., *A novel chair-type G-quadruplex formed by a Bombyx mori telomeric sequence*. Nucleic Acids Res, 2009. **37**(3): p. 931-8.
241. Wang, Y. and D.J. Patel, *Solution structure of the Oxytricha telomeric repeat d[G4(T4G4)3] G-tetraplex*. J Mol Biol, 1995. **251**(1): p. 76-94.
242. Heddi, B. and A.T. Phan, *Structure of human telomeric DNA in crowded solution*. J Am Chem Soc, 2011. **133**(25): p. 9824-33.
243. Haider, S., G.N. Parkinson, and S. Neidle, *Crystal structure of the potassium form of an Oxytricha nova G-quadruplex*. J Mol Biol, 2002. **320**(2): p. 189-200.
244. Schultze, P., F.W. Smith, and J. Feigon, *Refined solution structure of the dimeric quadruplex formed from the Oxytricha telomeric oligonucleotide d(GGGGTTTTGGGG)*. Structure, 1994. **2**(3): p. 221-33.
245. Paeschke, K., et al., *Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo*. Nat Struct Mol Biol, 2005. **12**(10): p. 847-54.
246. Yang, D. and K. Okamoto, *Structural insights into G-quadruplexes: towards new anticancer drugs*. Future Med Chem, 2010. **2**(4): p. 619-46.
247. Yu, H.Q., D. Miyoshi, and N. Sugimoto, *Characterization of structure and stability of long telomeric DNA G-quadruplexes*. J Am Chem Soc, 2006. **128**(48): p. 15461-8.
248. Xu, Y., et al., *Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends*. Angew Chem Int Ed Engl, 2009. **48**(42): p. 7833-6.
249. Xu, Y., *Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA*. Chem Soc Rev, 2011. **40**(5): p. 2719-40.

250. Granotier, C., et al., *Preferential binding of a G-quadruplex ligand to human chromosome ends*. Nucleic Acids Res, 2005. **33**(13): p. 4182-90.
251. Crabbe, L., et al., *Defective telomere lagging strand synthesis in cells lacking WRN helicase activity*. Science, 2004. **306**(5703): p. 1951-3.
252. Sfeir, A., et al., *Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication*. Cell, 2009. **138**(1): p. 90-103.
253. Smith, J.S., et al., *Rudimentary G-quadruplex-based telomere capping in Saccharomyces cerevisiae*. Nat Struct Mol Biol, 2011. **18**(4): p. 478-85.
254. Griffith, J.D., et al., *Mammalian telomeres end in a large duplex loop*. Cell, 1999. **97**(4): p. 503-14.
255. Munoz-Jordan, J.L., et al., *t-loops at trypanosome telomeres*. EMBO J, 2001. **20**(3): p. 579-88.
256. Cesare, A.J., et al., *Telomere looping in P. sativum (common garden pea)*. Plant J, 2003. **36**(2): p. 271-9.
257. Murti, K.G. and D.M. Prescott, *Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops*. Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14436-9.
258. Nikitina, T. and C.L. Woodcock, *Closed chromatin loops at the ends of chromosomes*. J Cell Biol, 2004. **166**(2): p. 161-5.
259. Giraud-Panis, M.J., et al., *Structural identity of telomeric complexes*. FEBS Lett, 2010. **584**(17): p. 3785-99.
260. Zhang, N., A.T. Phan, and D.J. Patel, *(3 + 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex*. J Am Chem Soc, 2005. **127**(49): p. 17277-85.
261. Patel, D.J., A.T. Phan, and V. Kuryavyi, *Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics*. Nucleic Acids Res, 2007. **35**(22): p. 7429-55.
262. Xu, Y., et al., *Stable lariat formation based on a G-quadruplex scaffold*. J Am Chem Soc, 2008. **130**(49): p. 16470-1.
263. Xu, Y., Y. Noguchi, and H. Sugiyama, *The new models of the human telomere d[AGGG(TTAGGG)3] in K⁺ solution*. Bioorg Med Chem, 2006. **14**(16): p. 5584-91.
264. Liu, D., et al., *Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins*. J Biol Chem, 2004. **279**(49): p. 51338-42.
265. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres*. Genes Dev, 2005. **19**(18): p. 2100-10.
266. Stansel, R.M., T. de Lange, and J.D. Griffith, *T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang*. EMBO J, 2001. **20**(19): p. 5532-40.
267. Pedroso, I.M., W. Hayward, and T.M. Fletcher, *The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures*. Nucleic Acids Res, 2009. **37**(5): p. 1541-54.
268. Kim, S.H., et al., *The human telomere-associated protein TIN2 stimulates interactions between telomeric DNA tracts in vitro*. EMBO Rep, 2003. **4**(7): p. 685-91.
269. LeBel, C. and R.J. Wellinger, *[The structures at the termini of chromosomes: what is there hidden under the cap ?]*. Med Sci (Paris). , 2004. **20**(2): p. 207-12.
270. Giraldo, R., et al., *Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study*. Proc Natl Acad Sci U S A, 1994. **91**(16): p. 7658-62.
271. Sarthy, J., et al., *Human RAP1 inhibits non-homologous end joining at telomeres*. EMBO J, 2009. **28**(21): p. 3390-9.

272. Hockemeyer, D., et al., *Telomere protection by mammalian Pot1 requires interaction with Tpp1*. Nat Struct Mol Biol, 2007. **14**(8): p. 754-61.
273. Loayza, D. and T. De Lange, *POT1 as a terminal transducer of TRF1 telomere length control*. Nature, 2003. **423**(6943): p. 1013-8.
274. Cristofari, G., K. Sikora, and J. Lingner, *Telomerase unplugged*. ACS Chem Biol, 2007. **2**(3): p. 155-8.
275. Grandin, N., C. Damon, and M. Charbonneau, *Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13*. EMBO J, 2001. **20**(5): p. 1173-83.
276. Grandin, N., S.I. Reed, and M. Charbonneau, *Stn1, a new Saccharomyces cerevisiae protein, is implicated in telomere size regulation in association with Cdc13*. Genes Dev, 1997. **11**(4): p. 512-27.
277. Xin, H., et al., *TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase*. Nature, 2007. **445**(7127): p. 559-62.
278. Hayflick, L. and P.S. Moorhead, *The serial cultivation of human diploid cell strains*. Exp Cell Res, 1961. **25**: p. 585-621.
279. Watson, J.D., *Origin of concatemeric T7 DNA*. Nat New Biol, 1972. **239**(94): p. 197-201.
280. Bessler, M., D.B. Wilson, and P.J. Mason, *Dyskeratosis congenita*. FEBS Lett, 2010. **584**(17): p. 3831-8.
281. de Lange, T., *T-loops and the origin of telomeres*. Nat Rev Mol Cell Biol, 2004. **5**(4): p. 323-9.
282. Gilson, E. and V. Geli, *How telomeres are replicated*. Nat Rev Mol Cell Biol, 2007. **8**(10): p. 825-38.
283. Ivesa, A.S., et al., *Saccharomyces Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA*. Genes Dev, 2002. **16**(11): p. 1383-96.
284. Olovnikov, A.M., *A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon*. J Theor Biol, 1973. **41**(1): p. 181-90.
285. Harley, C.B., A.B. Futcher, and C.W. Greider, *Telomeres shorten during ageing of human fibroblasts*. Nature, 1990. **345**(6274): p. 458-60.
286. Greider, C.W. and E.H. Blackburn, *Identification of a specific telomere terminal transferase activity in Tetrahymena extracts*. Cell, 1985. **43**(2 Pt 1): p. 405-13.
287. Autexier, C. and N.F. Lue, *The structure and function of telomerase reverse transcriptase*. Annu Rev Biochem, 2006. **75**: p. 493-517.
288. Bryan, T.M., et al., *Telomere elongation in immortal human cells without detectable telomerase activity*. EMBO J, 1995. **14**(17): p. 4240-8.
289. Bryan, T.M., et al., *The telomere lengthening mechanism in telomerase-negative immortal human cells does not involve the telomerase RNA subunit*. Hum Mol Genet, 1997. **6**(6): p. 921-6.
290. Bryan, T.M., et al., *Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines*. Nat Med, 1997. **3**(11): p. 1271-4.
291. Zakian, V.A., K. Runge, and S.S. Wang, *How does the end begin? Formation and maintenance of telomeres in ciliates and yeast*. Trends Genet, 1990. **6**(1): p. 12-6.
292. Murnane, J.P., et al., *Telomere dynamics in an immortal human cell line*. EMBO J, 1994. **13**(20): p. 4953-62.

293. Park, K.H., et al., *Telomerase activity and telomere lengths in various cell lines: changes of telomerase activity can be another method for chemosensitivity evaluation*. Int J Oncol, 1998. **13**(3): p. 489-95.
294. Bhattacharyya, S., A. Sandy, and J. Groden, *Unwinding protein complexes in ALternative telomere maintenance*. J Cell Biochem, 2010. **109**(1): p. 7-15.
295. Lillard-Wetherell, K., et al., *Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2*. Hum Mol Genet, 2004. **13**(17): p. 1919-32.
296. Opresko, P.L., et al., *Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases*. J Biol Chem, 2002. **277**(43): p. 41110-9.
297. Opresko, P.L., et al., *The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2*. Mol Cell, 2004. **14**(6): p. 763-74.
298. Opresko, P.L., et al., *POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates*. J Biol Chem, 2005. **280**(37): p. 32069-80.
299. Sun, H., R.J. Bennett, and N. Maizels, *The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs*. Nucleic Acids Res, 1999. **27**(9): p. 1978-84.
300. Sun, H., et al., *The Bloom's syndrome helicase unwinds G4 DNA*. J Biol Chem, 1998. **273**(42): p. 27587-92.
301. Verdun, R.E. and J. Karlseder, *The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres*. Cell, 2006. **127**(4): p. 709-20.
302. Verdun, R.E., et al., *Functional human telomeres are recognized as DNA damage in G2 of the cell cycle*. Mol Cell, 2005. **20**(4): p. 551-61.
303. Laud, P.R., et al., *Elevated telomere-telomere recombination in WRN-deficient, telomere dysfunctional cells promotes escape from senescence and engagement of the ALT pathway*. Genes Dev, 2005. **19**(21): p. 2560-70.
304. Wu, L. and I.D. Hickson, *The Bloom's syndrome helicase suppresses crossing over during homologous recombination*. Nature, 2003. **426**(6968): p. 870-4.
305. Karow, J.K., et al., *The Bloom's syndrome gene product promotes branch migration of holliday junctions*. Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6504-8.
306. Constantinou, A., et al., *Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest*. EMBO Rep, 2000. **1**(1): p. 80-4.
307. Oh, H., et al., *Telomere attrition and Chk2 activation in human heart failure*. Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5378-83.
308. Samani, N.J., et al., *Telomere shortening in atherosclerosis*. Lancet, 2001. **358**(9280): p. 472-3.
309. Mitchell, J.R., E. Wood, and K. Collins, *A telomerase component is defective in the human disease dyskeratosis congenita*. Nature, 1999. **402**(6761): p. 551-5.
310. Vulliamy, T., et al., *Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC*. Nat Genet, 2004. **36**(5): p. 447-9.
311. Yamaguchi, H., et al., *Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia*. N Engl J Med, 2005. **352**(14): p. 1413-24.
312. Marrone, A., et al., *Heterozygous telomerase RNA mutations found in dyskeratosis congenita and aplastic anemia reduce telomerase activity via haploinsufficiency*. Blood, 2004. **104**(13): p. 3936-42.
313. Alder, J.K., et al., *Short telomeres are a risk factor for idiopathic pulmonary fibrosis*. Proc Natl Acad Sci U S A, 2008. **105**(35): p. 13051-6.

314. Blasco, M.A., *Telomeres and human disease: ageing, cancer and beyond*. Nat Rev Genet, 2005. **6**(8): p. 611-22.
315. Murchie, A.I. and D.M. Lilley, *Tetraplex folding of telomere sequences and the inclusion of adenine bases*. EMBO J, 1994. **13**(4): p. 993-1001.
316. Huppert, J.L., *Structure, location and interactions of G-quadruplexes*. FEBS J, 2010. **277**(17): p. 3452-8.
317. Riou, J.F., et al., *Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands*. Proc Natl Acad Sci U S A, 2002. **99**(5): p. 2672-7.
318. Shammash, M.A., et al., *Telomerase inhibition and cell growth arrest by G-quadruplex interactive agent in multiple myeloma*. Mol Cancer Ther, 2003. **2**(9): p. 825-33.
319. Gomez, D., et al., *Resistance to senescence induction and telomere shortening by a G-quadruplex ligand inhibitor of telomerase*. Cancer Res, 2003. **63**(19): p. 6149-53.
320. Zhou, J.M., et al., *Senescence and telomere shortening induced by novel potent G-quadruplex interactive agents, quindoline derivatives, in human cancer cell lines*. Oncogene, 2006. **25**(4): p. 503-11.
321. Cookson, J.C., et al., *Pharmacodynamics of the G-quadruplex-stabilizing telomerase inhibitor 3,11-difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate (RHPS4) in vitro: activity in human tumor cells correlates with telomere length and can be enhanced, or antagonized, with cytotoxic agents*. Mol Pharmacol, 2005. **68**(6): p. 1551-8.
322. Gonzalez, V. and L.H. Hurley, *The c-MYC NHE III(1): function and regulation*. Annu Rev Pharmacol Toxicol, 2010. **50**: p. 111-29.
323. Drygin, D., et al., *Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis*. Cancer Res, 2009. **69**(19): p. 7653-61.
324. Monchaud, D. and M.P. Teulade-Fichou, *A hitchhiker's guide to G-quadruplex ligands*. Org Biomol Chem, 2008. **6**(4): p. 627-36.
325. Reed, J.E., et al., *Stabilization of G-quadruplex DNA and inhibition of telomerase activity by square-planar nickel(II) complexes*. J Am Chem Soc, 2006. **128**(18): p. 5992-3.
326. Shin-ya, K., et al., *Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus*. J Am Chem Soc, 2001. **123**(6): p. 1262-3.
327. Freyer, M.W., et al., *Biophysical studies of the c-MYC NHE IIII promoter: model quadruplex interactions with a cationic porphyrin*. Biophys J, 2007. **92**(6): p. 2007-15.
328. Parkinson, G.N., R. Ghosh, and S. Neidle, *Structural basis for binding of porphyrin to human telomeres*. Biochemistry, 2007. **46**(9): p. 2390-7.
329. Wu, J.Y., et al., *Structural isomers and binding sites of guanine-rich quadruplexes investigated by induced circular dichroism of thionin: loops and tails*. J Biomol Struct Dyn, 2003. **21**(1): p. 135-40.
330. Ishikawa, Y., et al., *Interaction of porphyrins bearing peripheral cationic heterocycles with G-quadruplex DNA*. Nucleic Acids Res Suppl, 2001(1): p. 107-8.
331. Phan, A.T., et al., *Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter*. Nat Chem Biol, 2005. **1**(3): p. 167-73.
332. Campbell, N.H., et al., *Structural basis of DNA quadruplex recognition by an acridine drug*. J Am Chem Soc, 2008. **130**(21): p. 6722-4.
333. Read, M.A., et al., *Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: structure-activity relationships*. J Med Chem, 1999. **42**(22): p. 4538-46.

334. Kankia, B.I. and L.A. Marky, *Folding of the thrombin aptamer into a G-quadruplex with Sr(2+): stability, heat, and hydration*. J Am Chem Soc, 2001. **123**(44): p. 10799-804.
335. Li, W., et al., *Structural competition involving G-quadruplex DNA and its complement*. Biochemistry, 2003. **42**(40): p. 11736-44.
336. Gabelica, V., et al., *Stabilization and structure of telomeric and c-myc region intramolecular G-quadruplexes: the role of central cations and small planar ligands*. J Am Chem Soc, 2007. **129**(4): p. 895-904.
337. Guittat, L., et al., *Ascididemin and meridine stabilise G-quadruplexes and inhibit telomerase in vitro*. Biochim Biophys Acta, 2005. **1724**(3): p. 375-84.
338. Rosu, F., et al., *Telomestatin-induced stabilization of the human telomeric DNA quadruplex monitored by electrospray mass spectrometry*. Chem Commun (Camb), 2003(21): p. 2702-3.
339. Carrasco, C., et al., *Tight binding of the antitumor drug ditercalinium to quadruplex DNA*. Chembiochem, 2002. **3**(12): p. 1235-41.
340. Rosu, F., et al., *Determination of affinity, stoichiometry and sequence selectivity of minor groove binder complexes with double-stranded oligodeoxynucleotides by electrospray ionization mass spectrometry*. Nucleic Acids Res, 2002. **30**(16): p. e82.
341. Liu, Y. and W.D. Wilson, *Quantitative analysis of small molecule-nucleic acid interactions with a biosensor surface and surface plasmon resonance detection*. Methods Mol Biol, 2010. **613**: p. 1-23.
342. Nguyen, B., F.A. Tanious, and W.D. Wilson, *Biosensor-surface plasmon resonance: quantitative analysis of small molecule-nucleic acid interactions*. Methods, 2007. **42**(2): p. 150-61.
343. Redman, J.E., *Surface plasmon resonance for probing quadruplex folding and interactions with proteins and small molecules*. Methods, 2007. **43**(4): p. 302-12.
344. Lacroix, L., A. Seosse, and J.L. Mergny, *Fluorescence-based duplex-quadruplex competition test to screen for telomerase RNA quadruplex ligands*. Nucleic Acids Res, 2011. **39**(4): p. e21.
345. Ren, J. and J.B. Chaires, *Sequence and structural selectivity of nucleic acid binding ligands*. Biochemistry, 1999. **38**(49): p. 16067-75.
346. Ragazzon, P.A., N.C. Garbett, and J.B. Chaires, *Competition dialysis: a method for the study of structural selective nucleic acid binding*. Methods, 2007. **42**(2): p. 173-82.
347. Ragazzon, P. and J.B. Chaires, *Use of competition dialysis in the discovery of G-quadruplex selective ligands*. Methods, 2007. **43**(4): p. 313-23.
348. De Cian, A., et al., *Fluorescence-based melting assays for studying quadruplex ligands*. Methods, 2007. **42**(2): p. 183-95.
349. Paramasivan, S. and P.H. Bolton, *Mix and measure fluorescence screening for selective quadruplex binders*. Nucleic Acids Res, 2008. **36**(17): p. e106.
350. Monchaud, D., C. Allain, and M.P. Teulade-Fichou, *Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands*. Bioorg Med Chem Lett, 2006. **16**(18): p. 4842-5.
351. Largy, E., F. Hamon, and M.P. Teulade-Fichou, *Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures*. Anal Bioanal Chem, 2011. **400**(10): p. 3419-27.
352. Smith, N.M., et al., *Unraveling the relationship between structure and stabilization of triarylpyridines as G-quadruplex binding ligands*. Org Biomol Chem, 2011. **9**(17): p. 6154-62.

353. Stefl, R., et al., *Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates*. Biophys J, 2003. **85**(3): p. 1787-804.
354. Bardin, C. and J.L. Leroy, *The formation pathway of tetramolecular G-quadruplexes*. Nucleic Acids Res, 2008. **36**(2): p. 477-88.
355. Rosu, F., et al., *Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry*. Nucleic Acids Res, 2010. **38**(15): p. 5217-25.
356. Doi, T., et al., *(S)-stereoisomer of telomestatin as a potent G-quadruplex binder and telomerase inhibitor*. Org Biomol Chem, 2011. **9**(2): p. 387-93.
357. Dose, C., et al., *Recognition of "mirror-image" DNA by small molecules*. Angew Chem Int Ed Engl, 2007. **46**(44): p. 8384-7.
358. Dumas, A. and N.W. Luedtke, *Highly fluorescent guanosine mimics for folding and energy transfer studies*. Nucleic Acids Res, 2011. **39**(15): p. 6825-34.
359. Eulberg, D. and S. Klussmann, *Spiegelmers: biostable aptamers*. Chembiochem, 2003. **4**(10): p. 979-83.
360. Lin, C., et al., *Mirror image DNA nanostructures for chiral supramolecular assemblies*. Nano Lett, 2009. **9**(1): p. 433-6.
361. Kim, Y., C.J. Yang, and W. Tan, *Superior structure stability and selectivity of hairpin nucleic acid probes with an L-DNA stem*. Nucleic Acids Res, 2007. **35**(21): p. 7279-87.
362. Darisipudi, M.N., et al., *Dual blockade of the homeostatic chemokine CXCL12 and the proinflammatory chemokine CCL2 has additive protective effects on diabetic kidney disease*. Am J Pathol, 2011. **179**(1): p. 116-24.
363. Hamon, F., et al., *An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-Quadruplex Topologies*. Angew Chem Int Ed Engl, 2011. **50**(37): p. 8745-9.
364. Renaud de la Faverie, A., et al., *Nucleic acids targeted to drugs: SELEX against a quadruplex ligand*. Biochimie, 2011. **93**(8): p. 1357-67.
365. Azzalin, C.M., et al., *Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends*. Science, 2007. **318**(5851): p. 798-801.
366. Schoeftner, S. and M.A. Blasco, *Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II*. Nat Cell Biol, 2008. **10**(2): p. 228-36.
367. Xu, Y., et al., *Telomeric repeat-containing RNA structure in living cells*. Proc Natl Acad Sci U S A, 2010. **107**(33): p. 14579-84.
368. Sacca, B., L. Lacroix, and J.L. Mergny, *The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides*. Nucleic Acids Res, 2005. **33**(4): p. 1182-92.
369. Xu, Y., K. Kaminaga, and M. Komiyama, *Human telomeric RNA in G-quadruplex structure*. Nucleic Acids Symp Ser (Oxf), 2008(52): p. 175-6.
370. Xu, Y., K. Kaminaga, and M. Komiyama, *G-quadruplex formation by human telomeric repeats-containing RNA in Na⁺ solution*. J Am Chem Soc, 2008. **130**(33): p. 11179-84.
371. Martadinata, H. and A.T. Phan, *Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K⁺ solution*. J Am Chem Soc, 2009. **131**(7): p. 2570-8.
372. Collie, G.W., et al., *A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex*. Nucleic Acids Res, 2010. **38**(16): p. 5569-80.

373. Collie, G.W., et al., *Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures*. J Am Chem Soc, 2010. **132**(27): p. 9328-34.
374. Zhang, A.Y., A. Bugaut, and S. Balasubramanian, *A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology*. Biochemistry, 2011. **50**(33): p. 7251-8.
375. Xu, Y., T. Kimura, and M. Komiyama, *Human telomere RNA and DNA form an intermolecular G-quadruplex*. Nucleic Acids Symp Ser (Oxf), 2008(52): p. 169-70.
376. Ramsing, N.B. and T.M. Jovin, *Parallel stranded duplex DNA*. Nucleic Acids Res, 1988. **16**(14A): p. 6659-76.
377. van de Sande, J.H., et al., *Parallel stranded DNA*. Science, 1988. **241**(4865): p. 551-7.
378. Granotier, C., et al., *Preferential binding of a G-quadruplex ligand to human chromosome ends*. Nucleic Acids Res, 2005. **33**(13): p. 4182-90.
379. Gros, J., et al., *Kinetics and thermodynamics of G-quadruplexes*. Nucleic Acids Symp Ser (Oxf), 2005(49): p. 61-2.
380. Kuzuya, A. and M. Komiyama, *Design and construction of a box-shaped 3D-DNA origami*. Chem Commun (Camb), 2009(28): p. 4182-4.

Quadruplexes de guanines : formation, stabilité et interaction

Les quadruplexes de guanines (G4) sont des structures non canonique d'acides nucléiques à quatre brins formées à partir de séquences ADN ou ARN riches en guanines. Ces structures reposant sur la formation et l'empilement de quartets de guanines sont très polymorphes, leur formation pourrait être envisagé dans de nombreux domaines d'application, aussi bien pour les biotechnologies que les nanotechnologies. L'étude de G4 tétramoléculaires modifiés présentée dans ce manuscrit a participé à la compréhension du mécanisme d'association de ces complexes. En particulier, nous avons montré que l'insertion de 8-méthyle-2'-déoxyguanosine à l'extrémité 5' de la séquence favorise l'association et la stabilité du G4. Par ailleurs, l'étude de l'ADN en série L (image de l'ADN naturel dans un miroir) a montré la formation d'un G4 tétramoléculaire avec les mêmes propriétés que son énantiomère, à l'exception de sa chiralité, qui est inversée. L'étude a révélé également une auto-exclusion de deux énantiomères (forme D et forme L) démontrant un assemblage contrôlé des brins parallèles. Ce travail de thèse a aussi permis d'introduire un système simple et stable de visualisation de G4 tétramoléculaire antiparallèle, appelé "ADN synaptique", sur une nanostructure d'ADN origami.

In vivo, ces structures pourraient être impliquées de façon transitoire dans de nombreux processus biologiques, en particulier au niveau des télomères. Nous avons réalisé, au cours de cette thèse, une étude comparative de la structure et de la stabilité des séquences télomériques connues de différents organismes. Cette étude a permis d'enrichir les données nécessaires au développement d'un algorithme prédisant la stabilité de G4. Enfin, nous avons développé une méthode facile et peu coûteuse de criblage (G4-FID) sur plaques 96 puits permettant d'identifier l'interaction de ligands avec différentes séquences biologiques pertinentes. La stabilisation du G4 dans certaines régions du génome via des ligands spécifiques pourrait limiter la prolifération de cellules tumorales et est donc intéressante pour les thérapies anticancéreuses.

Guanine quadruplexes : formation, stability and interaction

Guanine quadruplexes (G4) are non-canonical four-stranded nucleic acid structures formed by guanine-rich DNA and RNA sequences. These polymorphic structures are built from the stacking of several G-quartets and could be involved in many fields, in biotechnology as well as in nanotechnology. The study of modified tetramolecular G4 presented in this manuscript participated to the understanding of tetramolecular G4 formation. Especially, we showed that the insertion of 8-methyl-2'-deoxyguanosine at the 5'-end of the sequence accelerate G4 formation and increase its stability. Besides, we demonstrate here that short guanine rich L-DNA strands (mirror image of natural DNA) form a tetramolecular G4 with the same properties than their enantiomer, but with opposite chirality. The study revealed also self-exclusion between two enantiomers (D- and L- form), showing the controlled parallel self-assembly of different G-rich strands. This work introduced also a simple and stable system to observe tetramolecular antiparallel G4 formation, called "synaptic DNA", into a DNA origami nanostructure.

In vivo, such structures appear to be implicated in genome dynamics, and especially at telomeres. During this thesis, we dedicated a study to the comparison of G4 folding and stability of known telomeric sequences from different organisms. The present study allowed enriching the dataset necessary to build and refine algorithms predicting G4 stability. Last but not least, we developed a G4 ligand screening method onto 96-well plates allowing the comparison of different biological relevant sequences. The G4 stabilisation by specific ligands in some genome regions may prevent cancer cell proliferation, making it an attractive target for anticancer therapy.