

Mass Spectrometry Study of G-Quadruplex Nucleic Acids: folding Pathways and Ligand Binding Modes Adrien Marchand

▶ To cite this version:

Adrien Marchand. Mass Spectrometry Study of G-Quadruplex Nucleic Acids: folding Pathways and Ligand Binding Modes. Other [q-bio.OT]. Université de Bordeaux, 2016. English. NNT: 2016BORD0196. tel-01498459

HAL Id: tel-01498459 https://theses.hal.science/tel-01498459

Submitted on 30 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE PRÉSENTÉE POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DU VIVANT

SPÉCIALITÉ INTERFACE CHIMIE-BIOLOGIE

Par Adrien MARCHAND

Mass Spectrometry Study of G-Quadruplex Nucleic Acids: Folding Pathways and Ligand Binding Modes

Sous la direction de : Dr. Valérie GABELICA

Soutenue le 29 Novembre 2016

Membres du jury :

Pr. SCHMITTER Jean-Marie Pr. DEFRANCQ Eric Pr. AFONSO Carlos Pr. BIRKEDAL Victoria Pr. LIMBACH Patrick Dr. TEULADE-FICHOU Marie-Paule Université de Bordeaux Université Joseph Fourier Grenoble Université de Rouen Aarhus University, Denmark University of Cincinnati, USA Institut Curie, Orsay Président Rapporteur Rapporteur Examinateur Examinateur Invité

REMERCIEMENTS/ACKNOWLEDGEMENTS

Pour commencer, je tiens à remercier les Pr. Eric Defrancq et Pr. Carlos Afonso d'avoir accepté d'être rapporteurs de ce travail. I also wish to thank both Prof. Victoria Birkedal and Prof. Patrick Limbach for having accepted to examine this thesis. Je remercie également Marie-Paule Teulade-Fichou, non seulement pour faire partie de mon jury, mais aussi pour notre collaboration qui dure maintenant depuis plus de quatre ans. Merci pour le support, les discussions (et aussi les ligands !).

Cela semble évident mais je souhaite aussi remercier Valérie et Fred pour d'innombrables raisons. Tout d'abord, merci de m'avoir fait sortir de ma Belgique natale... Le belge est de nature casanière. Je dois l'avouer, le départ n'a pas été simple d'un point de vue personnel mais il m'a permis de découvrir de nouveaux horizons, de nouvelles perspectives, de découvrir de nouvelles choses. Mais cela n'est pas tout ! Je voudrais aussi vous remercier pour le coté scientifique et la formation que j'ai pu acquérir ici durant ma thèse. La biophysique, la chimie-physique, l'instrumentation, la modélisation moléculaire, la gestion d'étudiants, la création d'un laboratoire (eh oui nous n'étions que trois dans l'équipe au départ !)... Bref, je vais essayer d'utiliser à bien toutes ces compétences.

Merci aussi à Jean-Louis pour les discussions animées lors des "G4-meetings" qui m'ont permis d'aller plus loin dans mon raisonnement. Je pense que cela a contribué aux différents travaux présentés ciaprès.

Je souhaite remercier également les membres (présents et anciens) de notre équipe la « BioPhyMS Team ». Merci à Clémence et Stefano de prendre la relève ! Vous allez très bien vous en sortir j'en suis sûr ! Merci à Max pour ses discussions... intéressantes... Merci à Josephine, ma voisine de bureau pendant tout ce temps. Merci à Nina, nouvelle voisine de bureau. Courage pour le laser ! Merci à Sandrine pour les discussions autour du tuning de l'instrument. Merci à Solenne de nous faciliter la vie avec ses scripts ! Merci à Paul. Deux remerciments particuliers à Eric Largy et Valentina D'Atri, deux anciens. Je ne sais pas si c'est la nostalgie qui parle mais... Il était bien ce bon vieux temps où vous êtiez là ! J'espère qu'on va se recroiser vite.

Je tiens à remercier les étudiants (sous-fifres) qui ont réalisé des expériences liées aux projets auxquels j'ai participé. A very special thank to Mike, I really appreciated working with you. I am sure you will go far! Thank you again for the excellent weekend in CA. Thank you also Dominika, the hardest worker I have ever met. Great job with the copper complexes. Ce travail n'aurait pas pu se réaliser sans l'aide précieuse de plusieurs collègues de l'IECB, notamment de l'équipe Mergny. Merci à Nassima pour les discussions autour des ligands. Merci à Oscar, Samir, Gilmar, Anne, Carmelo.

Merci aussi aux amis. Merci Mathilde, William, Lionel et Aurore pour les bons moments passés ensemble, les cinés, les restos, les soirées...

Enfin, je souhaite remercier ma famille pour son support. Vous avez cru en moi, m'avez supporté et permis de finaliser ce travail. Merci à Emy d'avoir relu ce travail. Merci à Al, Paul, Clems.

Finalement, je ne pouvais pas finir sans remercier Amandine. Elle m'a permis de prendre du recul et de me changer les idées, de ne pas me noyer complètement sous la masse de travail. Parfois, il faut prendre du recul pour mieux avancer. Elle m'apporte un soutien au quotidien et je la remercie pour ça.

TABLE OF CONTENT

ABBREVIATIONS				
I. A	PERÇU EN FRANÇAIS	. 7		
I. S	UMMARY/OUTLINE/OBJECTIVES	13		
II. Ir	NTRODUCTION	17		
II.1.	NUCLEIC ACIDS	17		
II.1.	1 TRIDIMENSIONAL STRUCTURES	19		
II.2.	G-QUADRUPLEXES	21		
11.2.	1 WHAT IS A G-QUADRUPLEX?	21		
11.2.	2 WHERE ARE G-QUADRUPLEXES FOUND IN BIOLOGICAL SYSTEMS?	26		
11.2.	3 WHY IS IT IMPORTANT TO STUDY G-QUADRUPLEXES?	29		
II.2.	4 HOW TO STABILIZE G-QUADRUPLEXES?	32		
II.2.	5 WHICH IS THE TARGET?	35		
II.2.	6 HOW DO LIGANDS BIND TO G-QUADRUPLEXES?	39		
II.2. ⁻	7 HOW TO STUDY G-QUADRUPLEX FOLDING AND LIGAND BINDING MODES?	46		
II.3.	MASS SPECTROMETRY	53		
II.3.	1 NATIVE MASS SPECTROMETRY	53		
II.3.	2 ION MOBILITY SPECTROMETRY (IMS)	56		
II.3.	MASS SPECTROMETRY AS TOOL TO STUDY SOLUTION PHASE REACTIONS	58		
III. E	XPERIMENTAL APPROACHES	53		
III.1.	Mass spectrometry	53		
111.1	1 SAMPLE PREPARATION	63		
111.1	2 WHICH MASS SPECTROMETER SHOULD BE USED?	64		
111.1	3 MS INSTRUMENTAL SETTINGS	69		
111.1	4 ION MOBILITY SPECTROMETRY (IMS)	71		

111.1	.5 DATA TREATMENT	73
III.2.		
III.3.	THERMAL DENATURATION EXPERIMENTS	85
IV. N	ATIVE MS FROM POTASSIUM SOLUTIONS	
IV.1.		
IV.2.	Article 1	
IV.3.	ADDITIONAL WORK	
IV.3	.1 Additional structural and quantitative information from MS	
IV.3	.2 HIGHER SALT CONCENTRATIONS	
IV.3	.3 Other cations than K ⁺	
IV.4.	Conclusions	
V. P	OTASSIUM BINDING TO G -QUADRUPLEXES	109
V.1.		
V.2.	Article 2	111
V.3.	Additional work	127
V.3.	1. K ⁺ BINDING TO SYNTHETIC SEQUENCES	127
V.4.		129
V.5.	Perspectives	131
VI. G		
	-QUADRUPLEX LIGAND BINDING MODES	
VI.1.	-quadruplex ligand binding modes	133 133
VI.1. VI.2.	-QUADRUPLEX LIGAND BINDING MODES	133 133 135
VI.1. VI.2. VI.3.	-QUADRUPLEX LIGAND BINDING MODES INTRODUCTION ARTICLE 3 ARTICLE 4	133
VI.1. VI.2. VI.3. VI.4.	-QUADRUPLEX LIGAND BINDING MODES INTRODUCTION ARTICLE 3 ARTICLE 4 ADDITIONAL WORK	

Х.	ANNEXES	97
IX.	REFERENCES	71
VIII.	PH.D. THESIS PUBLICATIONS	69
VII.4	4. Outlook	67
VI	II.3.3 LIGAND BINDING MODES: STRATEGIES FOR FUTURE WORK	65
VI	II.3.2 Do ligands bind to G-quadruplexes by induced fit or conformational selection?.1	64
VI	II.3.1 The 1-K ⁺ antiparallel G-quadruplexes	63
VII.3	3. DISCUSSION ABOUT THE G-QUADRUPLEX RESULTS	63
VI	II.2.3 Sources of error in the treatment of the MS data1	61
VI	II.2.2 The importance of the buffer for MS of cation-binding biomolecules	60
VI	II.2.1 Mass spectrometry for biophysical studies1	59
VII.2	2. DISCUSSION ABOUT THE MS RESULTS1	59
VII.	1. SUMMARY OF THE RESULTS	57
VII.	GENERAL DISCUSSION AND CONCLUSION	57
VI.6	5. Perspectives	55
VI.5	5. CONCLUSIONS	53

ABBREVIATIONS

A	Adenine/Adenosine
С	Cytosine/Cytidine
CCS	Collision cross section
CD	Circular dichroism
CX-MS	Cross linking mass spectrometry
DNA	Deoxyribonucleic acid
(n)ESI	(nano) Electrospray ionization
G	Guanine/Guanosine
G4	G-quadruplex(es)
HDX	Hydrogen deuterium exchange
IMS	lon mobility spectrometry
ITC	Isothermal calorimetry
KCI	Potassium chloride
L	Ligand
m/z	Mass-to-charge ratio
MS	Mass spectrometry
NA	Nucleic acid(s)
NH₄OAc	Ammonium acetate
NMR	Nuclear magnetic resonance spectroscopy
PDB ID	Protein Data Bank identification number
Q	Quadrupole
RNA	Ribonucleic acid
SPR	Surface plasmon resonance
т	Thymine/Thymidine

T _m	Melting temperature
ТМА	Trimethylammonium
TMAA	Trimethylammonium acetate
TOF	Time of flight
ttpy	Tolylterpyridine
U	Uracil/Uridine
UV	Ultraviolet

I. APERÇU EN FRANÇAIS

Introduction

En plus de contenir l'information génétique, l'ADN et l'ARN peuvent être impliqués dans la régulation de processus biologiques. Une structure proposée comme régulateur est le G-quadruplex. Les G-quadruplexes sont des structures non-canoniques d'acides nucléiques formées par des séquences riches en guanines. Un G-quartet est formé par quatre guanines interagissant entre elles via des liaisons hydrogènes. L'empilement de deux (ou plus) de ces G-quartets forme le G-quadruplex, qui est stabilisé par ailleurs par des cations, comme le potassium, positionnés entre des G-quartets adjacents.

Les G-quadruplexes sont trouvés dans des régions clés du génome. En effet, des études bioinformatiques ont montré que des séquences riches en guanines se trouvent dans des promoteurs de gènes et aux télomères. Elles sont présentes dans le génome humain, chez les plantes, chez la souris, dans le génome de virus... Ces structures peuvent être formées durant la réplication de l'ADN ou hors du noyau, sous forme d'ARN. Certains aptamères (des oligonucléotides synthétiques capables de se lier avec haute affinité et sélectivité à leur cible) peuvent aussi former des G-quadruplexes.

L'une des stratégies proposées pour combattre le cancer consiste à stabiliser les Gquadruplexes formés dans les promoteurs de gènes et dans les régions télomériques. En effet, réguler l'expression d'oncogènes (gènes responsables du cancer) est une manière élégante de combattre le cancer car cela implique d'agir directement à la source en supprimant la production d'une protéine. De plus, empêcher la télomérase de se lier aux télomères est une stratégie envisagée pour réduire la prolifération des cellules cancéreuses.

Dans les deux cas (dans les promoteurs de gènes et aux télomères), le G-quadruplex agirait en tant que répresseur, c'est-à-dire que sa formation empêcherait soit la liaison de la machinerie de transcription soit la liaison de la télomérase. C'est pourquoi, il est souhaitable de vouloir stabiliser la forme G-quadruplexe par rapport à la forme duplexe. Pour ce faire, de nombreuses équipes synthétisent des ligands spécifiques et affins afin de reconnaitre et de

lier la structure G-quadruplexe. Ces ligands peuvent potentiellement être utilisés comme médicaments.

Objectifs

Le design de ligands de G-quadruplexes n'est cependant pas simple. En effet, il existe beaucoup de topologies différentes de G-quadruplexes dépendant de la molécularité du complexe, de l'orientation des brins dans le cœur du G-quadruplexe, de l'empilement des guanines, des types de boucles... Ce polymorphisme est illustré par la séquence télomérique humaine, pour laquelle au moins 10 structures différentes ont été obtenues expérimentalement.

Pour reconnaitre les G-quadruplexes, le design des ligands implique habituellement de larges plans aromatiques qui sont optimisés pour se lier par des interactions π - π sur les G-quartets extérieurs. Cependant, si ce type d'interaction était le seul mode de liaison, les ligands auraient des affinités similaires pour tous les G-quadruplexes. Or ce n'est pas le cas. D'autres modes de liaison sont en théorie possibles, comme l'intercalation entre des G-quartets, les liaisons avec les sillons ou les boucles. Cependant, aucune structure présentant uniquement un de ces modes de liaison n'a été rapportée.

Les objectifs de cette thèse consistent à déterminer quelles sont les complexes présents en solution, sans et avec ligands, et leurs quantités respectives.

Il existe beaucoup de techniques biophysiques permettant d'étudier les G-quadruplexes, leur repliement et la liaison de ligands. Cependant, ces techniques présentent un de ces deux désavantages : soit la technique ne permet pas d'obtenir d'information sur la structure du Gquadruplex, soit elle ne permet pas de quantifier chacune des structures présentes.

Afin de répondre à ces deux questions (structure et quantité), nous avons utilisé la spectrométrie de masse de type native. En effet, la spectrométrie de masse permet d'obtenir des informations sur la stœchiométrie de complexes non-covalents et de les quantifier. La spectrométrie de mobilité ionique couplée à la spectrométrie de masse apporte des informations sur la structure des complexes.

Résultats

Le noyau des cellules contient une haute concentration en ions potassium qui n'est pas directement compatible avec la spectrométrie de masse. Dans la littérature, des spectres de masse de G-quadruplexes sont typiquement obtenus à partir de solutions d'acétate d'ammonium. Cependant, il est important de replier les G-quadruplexes dans des solutions contenant un cation de nature biologiquement pertinente car ce cation va influencer la topologie de repliement des G-quadruplexes. Pour cette raison, nous avons développé une méthode de préparation d'échantillons nous permettant d'obtenir des spectres de masse à partir de solutions contenant du potassium. Dans notre méthode de préparation, la force ionique est fixée à l'aide d'acétate de triméthylammonium (TMAA). A cette solution est alors ajouté jusqu'à 1 mM de KCl, ce qui permet le repliement des G-quadruplexes. Tout en étant compatible avec la spectrométrie de masse, cette préparation permet aux G-quadruplexes d'être repliés selon des topologies similaires à celles des structures obtenues par résonance magnétique nucléaire (RMN), habituellement à 100 mM de KCl.

Grâce au développement d'un traitement mathématique des adduits non-spécifiques (adapté d'une méthode décrite par Klassen), nous avons pu déterminer les stœchiométries spécifiques en termes de nombre de cations liés, et les quantifier. Nous avons calculé des constantes d'équilibre de liaison du potassium aux G-quadruplexes télomériques humains et avons trouvé que des complexes ne contenant qu'un seul K⁺ au lieu de 2 (et donc ne possédant probablement que 2 G-quartets) étaient formés à basse concentration en KCl pour les séquences les plus courtes, ne contenant que le cœur d((GGGTTA)₃GGG). Plus des bases sont ajoutées aux extrémités de cette séquence, plus la liaison des deux ions potassium est coopérative et plus le complexe à 1-K⁺ devient minoritaire. En combinant les données de spectrométrie de masse et celles obtenues par dichroïsme circulaire (CD), nous avons obtenu les spectres CD déconvolués de chacune des espèces (à 0-, 1- et 2-K⁺). Les complexes à 0- et 2-K⁺ présentent respectivement des signatures CD typiques de structures dépliées et de Gquadruplexes hybrides. Le complexe à 1-K⁺ est quant à lui un G-quadruplexe antiparallèle. Ces résultats peuvent être étendus à d'autres séquences, comme nous l'avons démontré pour les séquences synthétiques 222 et 222T, d((GGGTT)₃GGG) et d(T(GGGTT)₃GGGT), respectivement.

Afin d'en étudier le mécanisme de repliement, la liaison des ions potassium aux Gquadruplexes a également été suivie au cours du temps. Nous avons trouvé que six ensembles cinétiques étaient nécessaires pour obtenir rendre compte des observations : un ensemble ne contenant pas de K⁺, deux ensembles en contenant un seul et trois ensembles contenant 2-K⁺. Le meilleur modèle cinétique implique la présence d'un chemin correspondant à des espèces « mal repliées », incluant notamment un G-quadruplex antiparallèle avec un seul K⁺ et donc deux G-quartets. La spectrométrie de mobilité ionique (IMS) couplée à la spectrométrie de masse a confirmé la présence d'au moins trois espèces à 2-K⁺. Ce mécanisme de repliement, obtenu pour les séquences télomériques humaines, est généralisable à d'autres systèmes. Nous avons en effet démontré qu'une séquence d'un promoteur (c-myc) se replie aussi selon le même mécanisme. Ces résultats suggèrent que le chemin de repliement impliquant un piège cinétique formé par des G-quadruplexes antiparallèles est un principe général du repliement des G-quadruplexes.

Finalement, nous avons étudié la liaison de ligands à des séquences formant des Gquadruplexes. Premièrement, à travers un criblage sur plus de 20 séquences d'ADN différentes, nous avons montré que deux molécules Cu*-ttpy* se lient avec une haute affinité et coopérativité aux G-quadruplexes formés par la séquence télomérique humaine : le complexe 2:1 (L:DNA) est majoritaire et le complexe 1:1 est quasiment inexistant. Cette coopérativité inégalée est attribuée à des changements de conformation de la séquence télomérique humaine. Grâce au dichroïsme circulaire (CD), nous avons montré que les structures des Gquadruplexes sont affectées et, avec la liaison des deux ligands Cu*-ttpy*, les G-quadruplexes deviennent de topologies antiparallèles. La stœchiométrie en termes de nombre de K⁺ est aussi affectée : les ligands se lient à des structures à 2-K⁺.

Deuxièmement, nous avons étudié la liaison de certains ligands parmi des plus affins et sélectifs pour les G-quadruplexes (Phen-DC3, 360A et PDS). Ces ligands se lient avec une haute affinité sur ces séquences, en éjectant un potassium, formant un complexe 1:1:1 (DNA:L:K⁺). Le complexe est antiparallèle et présente les mêmes caractéristiques que celui observé soit à l'équilibre à basse concentration en KCl, soit à temps court après l'addition de KCl. Ces découvertes peuvent avoir des implications dans le domaine du design de ligands. En effet, une potentielle manière d'obtenir des ligands spécifiques d'une séquence formant des G-quadruplexes pourrait être de cibler des intermédiaires de repliement.

En conclusion, nos résultats démontrent l'importance de caractériser finement la stœchiométrie des complexes par spectrométrie de masse afin de déterminer et d'interpréter correctement les constantes d'équilibre et de caractériser les structures. Dans plusieurs cas, nous avons observé que des changements structuraux étaient accompagnés d'un changement du nombre d'ions potassium liés. La spectrométrie de masse de type native est une méthode de choix pour caractériser ces complexes car elle permet de déterminer et quantifier différentes espèces en solution.

I. SUMMARY/OUTLINE/OBJECTIVES

Introduction

In addition to the storage of genetic information, nucleic acids are involved in the regulation of biological processes. G-quadruplexes, non-canonical structures formed by G-rich nucleic acids, are proposed as biological regulators. These structures are formed by G-quartets, macrocyclic planes formed by four guanines interacting *via* hydrogen bonds. Two or more G-quartets stacked on top of each other form the G-quadruplex, furthermore stabilized by cations, such as potassium, positioned in between adjacent G-quartets.

G-quadruplexes are found in key regions of the genome. Bioinformatics studies showed that G-rich sequences are found in gene promoters or at the telomeres. These sequences are present in the human genome but also in mice, in viruses... These structures can be formed during DNA replication or in RNA outside the nucleus. Some aptamers (synthetic oligonucleotides able to bind with high affinity and selectivity to their target) can also form G-quadruplexes.

Stabilizing G-quadruplexes in gene promoters and in telomeric regions is a proposed strategy to fight cancer. Indeed, downregulation of oncogene expression is one emerging elegant way to fight cancer because it would act directly at the source, suppressing the expression of a protein. In addition, inhibiting telomerase binding to telomeric DNA may be a strategy to reduce the proliferation of cancer cells. In both cases (oncogene promoters and telomeres), the G-quadruplexes are repressors: their formation would inhibit the binding of the transcription machinery or of the telomerase. Therefore, several groups synthesize highly affine and specific ligands, potential drugs, to stabilize the G-quadruplex over the duplex.

Objectives

The design of G-quadruplex ligands is not so simple. Indeed, G-quadruplexes may adopt different topologies, depending on the complex molecularity, strand orientation in the G-quadruplex core, the stacking of guanines, the type of loops... This polymorphism is illustrated by the human telomeric sequence for which at least 10 structures were reported.

To target G-quadruplexes, ligands are usually designed to have large aromatic moieties, to π - π stack on external G-quartets. However, if this was the only binding mode, all ligands would have similar affinities for all G-quadruplexes. Other binding modes are theoretically possible such as intercalation, binding in the grooves or to the loops... Nevertheless, no structure presenting binding modes different from π - π stacking is reported in the literature so far.

The thesis objectives consist in determining which complexes are present in solution, with and without ligands, and their respective quantities.

Many biophysical techniques can be used to study G-quadruplex folding and the binding of ligands. However, either the technique cannot bring information about G-quadruplex structures, or it does not allow the quantification of each of these structures. To answer these two questions (structure and quantity), we used native mass spectrometry. Indeed, mass spectrometry allows to obtain the stoichiometry of the complexes and quantify them. Ion mobility coupled to mass spectrometry brings information about the structure.

Results

The first problem was related to the analysis of nucleic acids and cation complexes. In fact, the potassium cation is the relevant cation present in the nucleus but is not directly mass spectrometry-compatible. In the literature, mass spectra of G-quadruplexes are typically obtained from ammonium acetate solutions. However, it is important that G-quadruplexes are folded in solutions containing a cation of relevant nature because the cation's nature can influence the topology. For this reason, we developed a sample preparation method allowing us to obtain mass spectra from potassium containing solutions. In the developed method, ionic strength is fixed using trimethylammonium acetate (TMAA) doped with up to 1 mM K⁺ required to fold the G-quadruplexes. This mass spectrometry-compatible sample preparation allows the folding of G-quadruplexes with topologies similar to those obtained in the literature by nuclear magnetic resonance (NMR), usually in 100 mM KCl.

Thanks to a mathematical treatment (adapted from a method described by Klassen) which was designed to take into account the nonspecific adducts, we determined the specific cation stoichiometries and quantified them. We determined equilibrium constants for the binding of the two potassium cations to human telomeric sequences. We found that at low KCl

concentration some complexes are formed with only one K⁺ instead of the expected two. We deduced that these complexes only contain two G-quartets. The shortest human telomeric sequences, containing the d((GGGTTA)₃GGG) core, are more prone to form this 1-K⁺ complex. The binding of the two K⁺ cations to longer sequences is more cooperative and the 1-K⁺ species become minor. By combining the results obtained by mass spectrometry and circular dichroism (CD), we deconvoluted the CD spectra of each of the 0, 1- and 2-K⁺ species. The 0 and 2-K⁺ spectra are typical of unfolded and hybrid G-quadruplexes, respectively. The 2- quartet 1-K⁺ complex is typical of antiparallel G-quadruplex. These results were also obtained for the synthetic sequences 222 et 222T, d((GGGTT)₃GGG) et d(T(GGGTT)₃GGGT).

To study the folding pathway of G-quadruplexes, K⁺ binding was also monitored as a function of time. We found that six kinetic ensembles of species were needed to fit the data: one species with no K⁺, two with 1-K⁺ and three with 2-K⁺. The best models presented a 1-K⁺ antiparallel G-quadruplex that acts as a kinetic trap. Ion mobility spectrometry, coupled to mass spectrometry confirmed the presence of at least three species with 2-K⁺. Interestingly, this branched folding pathway, obtained for the human telomeric sequences, could be generalized to other sequences. Indeed, we showed that c-myc promoter also fold according to the same mechanism. These results suggest that the folding of antiparallel 1-K⁺ structures as kinetic traps is a general tenet of G-quadruplex folding.

Finally, we investigated the binding of ligands to G-quadruplex forming sequences. First, through a screening of over 20 DNA sequences, we showed that two Cu-*ttpy* ligands bind with high affinity and cooperativity to the human telomeric G-quadruplexes. The 2:1 (L:DNA) complex is a major species whereas the 1:1 complex is almost inexistent. This extreme cooperativity is attributed to conformational changes in the human telomeric structure. Thanks to circular dichroism, we showed that the conformations of human telomeric G-quadruplexes are affected: when two ligands are bound, the G-quadruplex topologies are antiparallel. The K⁺ stoichiometry, measured by mass spectrometry, is also affected: the ligands bind to 2-K⁺ structures.

Second, we studied the binding of some of the most affine and selective G-quadruplex ligands (Phen-DC3, 360A and PDS). These ligands bind with high affinity to the human telomeric G-quadruplexes with the ejection of a K⁺, forming a 1:1:1 (DNA:L:K⁺) complex. According to

circular dichroism, this complex is antiparallel and presents the same characteristics as the complex formed at low KCl concentration or at short time scale after KCl addition. Because they provide information on the binding mode, these results are important for ligand design. Indeed, a possible way to obtain ligands specific for one G-quadruplex sequence could be to target folding intermediates.

In conclusion, these results demonstrate the importance of characterizing the stoichiometry of complexes in order to properly determine equilibrium constants and characterize the structures of complexes. In several cases, we found that structural changes were accompanied by a change in the number of K⁺ cation bound. Native mass spectrometry is a key technique to characterize these complexes. It allows the determination and quantification of different species formed in solution.

II. INTRODUCTION

II.1. NUCLEIC ACIDS

DNA is the hereditary material that serves as the storage of all the genetic information needed to make a living being. The genetic information is transmitted through generations. Living cells use nucleic acids to synthetize the proteins they need. Additionally, nucleic acids are implied in other roles, such as regulation and catalysis of chemical reactions in the cell metabolism.

Nucleic acids are present inside the nucleus of eukaryotic cells as deoxyribonucleic acids (DNA) but also in the cytoplasm as ribonucleic acids (RNA). Nucleic acids are biopolymers made of four types of subunits, the nucleotides.¹ A nucleotide is made of a phosphate group, a sugar (that can be, either a ribose for RNA or a deoxyribose for DNA), and a base (Figure 1A). Among the bases, the natural and most abundant ones are also classified in two categories (Figure 1B): the purines, made of two aromatic rings (G, guanine and A, adenine) and the pyrimidines, made of one aromatic ring (C, cytosine and T, thymine, for DNA. In RNA, the thymine is replaced by a U, uracil).

Usually a DNA or RNA sequence is named using the one letter code of its sequence, going from the 5' to the 3' end. A small *d* or *r* at the beginning of the sequence is used to indicate the sugar nature, deoxyribose or ribose, respectively. The 5' carbon of the sugar on the ith nucleotide binds to the 3' carbon of the (i-1)th nucleotide through the phosphate group to form the phosphodiester linkage (Figure 1C). The nucleotides are therefore connected covalently in a chain through the sugars and phosphates. The alternating sugar-phosphate-sugar-phosphate moieties form the backbone.

Depending on their sequences and experimental conditions (solvent, ionic strength, presence/absence of salt,...) DNA and RNA can adopt different types of structures ranging from the unfolded single stranded form to the tetrastranded G-quadruplex, discussed in the next paragraphs.



Figure 1. Nucleic acid bases. (A) Definition of a nucleotide. (B) Natural DNA and RNA bases. (C) A nucleic acid strand. The bases are color-coded through all this work: G is gold, C pink, A blue, T and U are green.

II.1.1 TRIDIMENSIONAL STRUCTURES

B-form duplex. The most common structure of DNA is the B-helix, first described by J. Watson and F. Crick in 1953.² It is a double helix that consists in two antiparallel polynucleotides chains held together by hydrogen bonding between the bases of the strands: G with C and A with T. Furthermore, the paired bases stack one upon another, stabilizing the structure. The helix is right-handed and ten base pairs are needed to make a complete turn of the helix. It possesses a major and a minor groove and the negatively charged phosphates point towards the surface of the structure.

A-form duplex. In some conditions, complementary nucleic acid sequences can adopt other structures such as the A-helix (represented in Figure 2). One major difference comes from the placement of the base pairs within the duplex axis. In the B-DNA, the base-pairs are centered on the main axis while in the A-form they are displaced from the center to be closer to the major groove. This A-form is preferentially adopted by RNA, or by DNA in a dehydrated medium.

Z-form duplex. A third type of helix can also be formed when the sequence has alternating purine-pyrimidine bases. This leads to the formation of a Z-form helix that is left-handed.



Figure 2. The B-, A- and Z- double helices. Left, the B-form structure from PDB ID 424D.³ Middle, the A-form structure from PDB ID 406D.⁴ Right, the Z-form structure from PDB ID 2DCG.⁵

Duplexes are not the only structure that DNA or RNA adopt. Other structures such as the triplex,⁶ the i-motif⁷ or the guanine-quadruplex (G-quadruplex) can be formed.

Triplex. A triple helix can be formed by the interaction of a third single strand inside the major groove of a double helix. Two strands form a double helix thanks to the Watson-Crick base pairs and the third strand interacts via hydrogen bonds between its bases and the Hoogsteen face of the bases of the duplex part (Figure 3).

i-motifs. Four strands interacting together two by two via C⁺-C base pairs (two cytosines sharing a proton) can form the i-motif structure. They are stable in acidic conditions.



Figure 3. The triplex structure. Left, the triplex structure from the PDB ID 1BWG⁸ and the hydrogen network formed for a T-A-T triplet when a triplex structure is folded. Right, the structure of i-motif from PDB ID 1YBR⁹ and the representation of a C⁺-C base pair formed in i-motifs.

II.2. G-QUADRUPLEXES

II.2.1 WHAT IS A G-QUADRUPLEX?



Figure 4. (A) A G-quartet, (B) the stacking of G-quartets gives a G-quadruplex and (C) top and lateral view of the parallel tetramolecular [d(TG₄T)]₄ G-quadruplex obtained by X-ray crystallography (PDB ID 352D).¹⁰

Guanine-rich DNA or RNA strands may form G-quadruplex structures. Bang was the first to observe G-quadruplex formation: guanylic acid in solution forms a gel.¹¹ However, it took half a century of research to discover its tetra-stranded structure.¹² G-quadruplexes are made of two or more stacked G-quartets, which are hydrogen-bonded macrocycles formed by four guanines (Figure 4A). The H on the N1 and NH₂ on the C2 of one guanine are involved in hydrogen bonds, respectively, with the O6 and N7 of the neighboring guanine. Eight hydrogen bonds per G-quartet are formed. Those hydrogen bonds are called Hoogsteen hydrogen bonds because the Hoogsteen face of one guanine is involved in the hydrogen network.¹³ Four G-tracts (runs of G) are needed to form a G-quadruplex that is furthermore stabilized by cations, coordinated between the O6 of guanines.¹³ For example on Figure 4B, the G-tracts are made of four guanines, leading to a four G-quartets G-quadruplex with three cations inbetween. The interactions stabilizing G-quadruplexes are electrostatic (hydrogen bonds and

dipole-ion interactions), π - π stacking and hydrophobic (water molecules are released upon folding).¹⁴

G-QUADRUPLEX FORMALISM

There are many ways to form G-quadruplexes, depending on the number of strands, their orientation in the G-quadruplex core, the way the loops connect the G-tracts, etc.... Those characteristics describe the topology of a G-quadruplex, and all of them are interconnected. Because many G-quadruplex topologies are possible, and to simplify their description, Webba da Silva has described a recommended geometric formalism and associated nomenclature,¹⁵ which is discussed hereafter.

Molecularity. The first characteristic to consider is the molecularity of the complex. Most of the G-quadruplexes studied here were intramolecular, made of one single DNA or ARN strand. However, they can also be made of two or four strands, called respectively bi- and tetramolecular. Although only unimolecular G-quadruplexes may be relevant *in vivo*, bi- and tetramolecular G-quadruplexes can be formed *in vitro*.



Figure 5. (A) Unimolecular, (B) bimolecular and (C) tetramolecular G-quadruplexes.

Glycosidic torsion angle. The free rotation of the bases around the glycosidic bond leads to two guanosine conformers: if the base and the sugar are on the same side of the glycosidic bound, the conformation is *syn* (Figure 6, color-coded in red); if they are on opposite sides of the bond, the conformation is *anti* (color-coded in light grey). Because each guanine has two possible configurations, there are 16 (4²) different types of G-quartets, all of them represented in Figure 6B. Depending on the secondary structure adopted, the glycosidic bond is locked in one or the other conformation, resulting in the grooves sites of variable size: wide, medium or narrow.



Figure 6. (A) anti/syn configuration of the glycosidic bond: grey = anti and red = syn; this color-code is adopted through all this work. (B) All 16 possible arrangements of the G-quartets, fixing the position of the bottom right sugar (doted sugar). Illustration adapted from Webba da Silva.¹⁵

Loop types. The next point is the type of the loops connecting the G-tracts. In a tetramolecular G-quadruplex, such as $[TG_4T]_4$ (Figure 4C), all G-tracts are independent and no loop is needed to connect the strands. However, when the G-quadruplex is bi- or unimolecular, loops are needed. Three types of loops exist: a lateral loop connecting two antiparallel adjacent strands; a propeller loop connecting two parallel adjacent strands; and a diagonal loop connecting antiparallel opposite strands (Figure 7).



Figure 7. (A) Lateral, (B) propeller and (C) diagonal loops.

Strand orientations. The strand orientation is defined by the orientation of the G-tracts involved in the G-quartets following the backbone from the 5' to the 3' end. In the literature, this is the most commonly used way to characterize the topology of a G-quadruplex. If all strands point in the same direction, the G-quadruplex is parallel and belongs to group-I. If

three strands point in one direction and the last one in the opposite, the G-quadruplex is hybrid and belongs to group-II. If the strands have opposite direction two by two, the Gquadruplex is antiparallel and belongs to group-III.



Figure 8. (A) Parallel, (B) hybrid, (C) and (D) antiparallel G-quadruplexes.

All those characteristics are interconnected, for example, changing one strand orientation implies changing the loop type, and the glycosidic torsion angle... For those reasons, the interconversion between different structures are far from being simple mechanistically. Those numerous characteristics contribute to the high polymorphism of G-quadruplexes. For example, there are in theory 26 possible unimolecular G-quadruplexes topologies of three loops with three G-quartets.¹⁵

Additional sequence/structure-specific features, such as base pairs between bases in the loops or bulges in the G-tract, also contribute to this polymorphism as illustrated with a few structures in Figure 9.



Figure 9. Illustration of the polymorphism of G-quadruplexes. Structures obtained by NMR or X-ray crystallography of different G-quadruplexes. The PDB ID is indicated below the structure: antiparallel 2-quartet G-quadruplex, 1RDE (d(GGTTGGTGGTGGGTGGG)),¹⁶ parallel 3-quartet G-quadruplex with one G-tract made of 2 guanines, 2A5P (d(TGAGGGTGGIGAGGGTGGGGAAGG), I stands for inosine),¹⁷ parallel 3-quartet G-quadruplex with one long loop, 2LPW (d(AA<u>GGGTGGGG</u>TGTAAGTGT<u>GGGTGGG</u>T)),¹⁸ and bimolecular antiparallel G-quadruplex with mini-hairpins in the loops, 2KAZ (2x d(<u>GGG</u>ACGTAGT<u>GGG</u>)).¹⁹

II.2.2 WHERE ARE G-QUADRUPLEXES FOUND IN BIOLOGICAL SYSTEMS?

Structural and biophysical studies evidenced that many G-rich sequences can fold into this non-classical type of structures. Initially, it was not known whether the G-quadruplex was a biologically relevant structure. There is now more and more evidence that G-quadruplexes are relevant DNA and RNA secondary structures *in vivo*.^{20,21} One of the most convincing evidence was brought by Biffi *et al*,²² who developed a G-quadruplex structure-specific antibody that was used to visualize quantitatively DNA G-quadruplexes in human cells.

The quest for G-quadruplex localization in the genome was started by bioinformatics studies. Two works screened for G-quadruplex motifs in the human genome.^{23,24} The first definition used to find putative G-quadruplex forming sequences was the following. Intramolecular Gquadruplexes follow the generalized sequence $G_aX_nG_bX_0G_cX_pG_d$ where G_{a-d} represent the guanine tracts and X_{n-p} represent the loops regions that can be made of any nucleotide combination (G included). The number of guanines in the G-tracts was restricted to 3-5 and the loops were between 1 and 7 nucleotides. This definition led to more than 300 000 putative occurrences. However, the limitation to seven nucleotides maximum in the loop regions is probably too stringent, given that several structures were recently found to have longer loops (examples are shown in Figure 9, PDB ID: 2KAZ, 2LPW).^{18,19,25} More generally, if the external loops are short, a long middle loop can be tolerated and could even stabilize the structure if they form secondary structures (such as base pairs^{26,27}). More recently, Phan showed that Gquadruplexes can be made of 4n-1 guanines: one G-tract can be one guanine shorter, leading to a G-quadruplex made of *n*-1 G-quartets and a triplet of guanines.²⁸ For these reasons, Mergny developed a new algorithm to determine G-richness and skewness of regions of the genome.²⁹ They found a much higher number of G4 forming sequences in the human genome (around 1.5 million). In agreement, experimentally, using a high-resolution sequencing approach, Balasubramanian identified more than 700 000 sequences forming G4 structures.³⁰

In human, yeast and bacterial genomes, G-quadruplex forming sequences are overrepresented in certain regions such as the promoters and the telomeres.^{23,24,29,31,32} Several indirect evidences for the presence of G-quadruplexes structures come from the DNA binding proteins in bacteria.³³ Some proteins are able to induce the folding of G-quadruplexes³⁴ and others to unwind them.³⁵ The evolutionary conservation of the localization and composition

of G4 motifs suggests that G4 have a role in the cells,^{36,37} even though the exact role of those structure is not yet fully understood. Some of their potential roles are presented in the next few pages.

Gene promoters. Regions close to the translation start site (TSS) in the promoters of genes (within 1000 nucleotides upstream of the TSS) are enriched in G-quadruplex-forming sequences.^{23,29,32} For example, a very well-studied G-quadruplex is present in the promoter of the transcription factor MYC in mammalians.

Telomeres. By definition, the telomere is the sequence at the end of the chromosomes and the proteins that protect it.³⁸ The primary role of the telomere is to protect the chromosome against degradation or end-to-end fusion. They are made of a double stranded region and an extended 3' single stranded overhang. Usually, the telomeric 3' end sequence consists in the repetition of a G-rich motif, for example, in vertebrates, (TTAGGG)_n,³⁹ but other G-rich sequences have been identified for bacteria, yeast or plants.⁴⁰⁻⁴² Several studies have shown that this G-rich sequence is able to form G-quadruplexes *in vivo*.^{22,30,43} In human cultured cells, a small fluorescent probe⁴⁴ and an antibody²² suggested the presence of G-quadruplexes in telomeric regions *in vivo*. However, researchers are still uncertain about their presence as their folding may be induced by those probes.

During DNA Replication. During its replication, the double stranded DNA is separated into two strands by helicases (proteins responsible for unwinding the DNA structures). First, a preformed G-quadruplex could slow down the helicase unwinding.⁴⁵ Secondly, during replication, the leading strand is continuously duplicated and is therefore continuously reforming a duplex. However, the lagging strand's replication is discontinuous and some portions are free to form secondary structures, such as G-quadruplexes, which could therefore act as transcriptional regulators.³⁸ G-quadruplexes could also be formed in the double stranded mother-part and prevent or slow down the unwinding by the helicases. A high degree of mutations within the genome of a cell lineage can be caused by most of the human helicases that are able to unwind the G-quadruplexes *in vitro*.⁴⁵⁻⁴⁷

Outside the nucleus. G-rich RNA sequences are also able to fold into G-quadruplexes, suggesting that RNA G-quadruplexes may have a translation regulation role.^{48–50} Recently, fluorescent G-quadruplex probes have been used in living cells and in isolated mitochondria.

In this work, the authors suggest their presence also in mitochondria.⁵¹ Other potential roles of the G-quadruplexes in the genome are linked with the epigenetic regulation of gene expression, origin of replication, meiosis, recombination, etc....³⁸

Viruses. G-quadruplex-forming sequences are also found in viruses. Recently, a G-rich sequence has been identified in the promoter of HIV-1 and the formation *in vitro* of several G-quadruplexes has been confirmed by NMR spectroscopy.^{26,52} Another work suggests that G-quadruplexes can play a role on hepatitis C virus replication.⁵³

G-quadruplexes as therapeutic agents. G-quadruplexes can also be used as therapeutic agents as aptamers, synthetic oligonucleotides able to bind with high affinity and selectivity to their target. Indeed, some aptamers are able to fold into a G-quadruplex structure.¹⁶ Short G-quadruplexes have been evaluated as potential anti-HIV candidates because they bind to HIV-related proteins.^{54–56} Some G-quadruplexes are also tuned with chemical modifications to improve their binding affinities towards their target.⁵⁷

II.2.3 WHY IS IT IMPORTANT TO STUDY G-QUADRUPLEXES?

ALTERING GENE EXPRESSION

Because many genes have G-quadruplex-forming sequences in their promoter, the transcription could theoretically be influenced by G-quadruplex stability. Surprisingly, there is a discrepancy within gene promoters: promoters of oncogenes (genes that have the potential to cause cancer) are even more enriched in G-quadruplex forming sequences⁵⁸ whereas promoters of tumor repressor genes are poorer in those motifs.⁵⁹ Displacing the equilibrium towards the formation of the G-quadruplex is a proposed strategy to fight cancer cells (Figure 10).

Stabilizing G-quadruplexes in or near the promoter regions may have positive or negative aspects. If present in the template strand, the G-quadruplex could affect the binding of the polymerase and downregulate the gene expression which would be positive in case of an oncogene. However, G-quadruplexes in human cells tend to be more often found on the other strand.⁶⁰ They would then have a negative effect as they could facilitate the transcription. However, they may serve as recruiters to stimulate the transcription... or as blockers by recruiting gene repressors...⁶¹ Their exact role is actually still under intense investigation.



Figure 10. Schematic representation of a G-quadruplex-forming sequence in the promoter of a gene. The formation of the G-quadruplex prevent the gene expression.

Some results suggest that the formation of the G-quadruplex would downregulate the gene expression of MYC,⁶² a transcription factor in mammalians which is associated with cell proliferation and is overexpressed in 80% of human cancers.⁶³ Displacing the equilibrium would therefore be positive for the organism.

Downregulation of oncogenes is one emerging elegant way to fight cancer because it would act directly at the source, suppressing the expression of a protein.

PREVENT TELOMERASE BINDING

At each cell cycle, the very few end-bases of the chromosome cannot replicate. This leads to telomere attrition, a decreasing of the length of the telomere with cell replication. Once the telomere is too short, the cell dies normally, and this process is called cell senescence.

Telomerase, a ribonucleoprotein responsible for the regeneration of the telomere, is inactive in normal cells but upregulated in 85-90% of malignant cells. This enzyme gives the cancer cells the ability to divide indefinitely.⁶⁴ Immortality and limitless replicative potential are two of the hallmarks of a cancer cell.⁶⁵



Figure 11. Schematic representation of the telomere overhang. The formation of the G-quadruplex prevent the binding of the telomerase and elongation of the telomere. Figure adapted from Wang *et al.*⁶⁶

The substrate for this enzyme is the single stranded DNA overhang. If that overhang is folded, the telomerase cannot bind anymore, which prevents telomere elongation (Figure 11).⁶⁷ One research objective is therefore to stabilize the G-quadruplexes formed at the telomeres in

order to provide obstacles for the binding of the telomerase, limiting the elongation of the telomere. Inhibiting the telomerase binding may indeed be a strategy to reduce the proliferation of cancer cells.^{68,69}

In addition, because the telomeres are implicated in cell divisions, they may have a role in cell ageing. Indeed, telomere attrition correlates with age-related pathologies.⁷⁰
II.2.4 HOW TO STABILIZE G-QUADRUPLEXES?

Due to their presence in key regions of the genome, it has been suggested that they have the ability to regulate biological events. To affect those biological events using the G-quadruplex structure, a strategy consists in influencing the G-quadruplex stability. Stabilizing a G-quadruplex means displacing the equilibrium represented in Figure 12 towards the folding of the G-quadruplex.



Figure 12. Factors influencing the G-quadruplex folding equilibrium.

Note that in the genome, DNA G-quadruplex formation competes with duplex (not represented in Figure 12). To the contrary, RNA does not have this constrain and the formation of G-quadruplexes is therefore even more likely to occur. Many parameters have been described to stabilize (or destabilize) G-quadruplexes. The main ones are briefly described hereafter.

Cations. The cation may displace more or less the equilibrium towards the formation of the G-quadruplex depending on its concentration, ionic radius and hydration energy (desolvation and binding of the cation is part of the equilibrium).⁷¹ The general trend for quadruplex-stabilizing cations is the following: $Sr^{2+} > Ba^{2+} > K^+ > Ca^{2+} > Na^+$, NH_4^+ , $Rb^+ > Mg^{2+} > Li^+ > Cs^+$.⁷² The nature of the cation also influences the preferred topology of the G-quadruplex as discussed later. In cells, mostly K⁺ and Na⁺ are present and their concentrations are regulated thanks to the Na⁺/K⁺ pump. In the nucleus, potassium is more concentrated than sodium⁷³ and K⁺ is therefore the most biologically relevant cation for *in vitro* assays.

lonic strength. To overcome the electrostatic repulsion between the negative charges of the backbone, a high ionic strength is needed.⁷⁴ Increasing the ionic strength is, however, not a convenient strategy to act on the equilibrium *in vivo*.

Crowding or dehydration agents. Crowding agents such as polyethylene glycol have been used to mimic cell conditions. These agents were shown to be a major determinant of the G-quadruplex stability.⁷⁵⁻⁷⁷ In a previous work, we also showed that organic co-solvents such as ethanol induced the formation of G-quadruplexes.⁷⁸ The commonly accepted reason is that water molecules are released upon folding of the G-quadruplex. Therefore, adding dehydration agents displaces the equilibrium towards the folding of G-quadruplexes.

Temperature. The Gibbs free energy (ΔG_1) of the reaction depends on the temperature: lower temperatures favor the folding of G-quadruplexes. Again, it is not possible to decrease the cell temperature of a living being. This parameter is however used to determine the "G-quadruplex stability", indicated by the melting temperature (temperature at which 50% of the G-quadruplexes in solution are folded). If the melting temperature is increased/decreased in certain conditions, it means the G-quadruplex is stabilized/destabilized by those conditions. Measuring the melting temperature in different conditions is a widely used method to study G-quadruplex stabilizing/destabilizing conditions. This particular point will be discussed in more details in the next sections.

Ligands. Because none of the above-mentioned parameters can be used to influence Gquadruplexes formation *in vivo*, researchers had to find a different way to affect the equilibrium. This can be achieved using small molecules (also called ligands) to stabilize the G-quadruplex structures. Genome-wide approaches (not targeting one specific gene) using Gquadruplex ligands have been conducted to evidence the potential use of G-quadruplexes as transcriptional regulators.^{62,79}

The equilibrium in Figure 12 becomes more complicated when a ligand binds to the different structures. The equilibrium in Figure 13 represents some of the equilibria that affect the concentrations of the different species in solution. If the goal is to displace the equilibria towards the G-quadruplex formation, it is therefore important to synthetize ligands affine and selective for G-quadruplexes over the duplex or unfolded species. A review of selected ligands is presented in a specific section (II.2.6), later.



Figure 13. Schematic representation of the solution equilibria at stake when a ligand bind to a Gquadruplex-forming sequence. The green balls are cations and the black disk is a ligand.

II.2.5 WHICH IS THE TARGET?

Targeting G-quadruplexes with ligands seems to be a promising route to fight cancer for the reasons presented above. However, once injected inside the cell, a ligand could bind to all G-quadruplexes present in the genome and therefore could alter many different processes. Additionally, G-quadruplex ligands could also theoretically induce the folding of G-quadruplexes in unwanted regions, which may promote unwanted biological effect.

Different G-rich sequences may form G-quadruplexes of different topologies, which could in principle be targeted by different ligands. However, it is not so simple because the same sequence can also fold into different types of G-quadruplexes, ensuing uncertainties about which G-quadruplex structure should be targeted.

The first step to design affine and selective ligands for the G-quadruplex structure is to know the target's structure. Because most of this work is focused on the human telomeric sequences, the main published human telomeric structures are presented in this section.

POLYMORPHISM OF THE HUMAN TELOMERIC SEQUENCES

The human telomeric sequence made of d(TTAGGG) repeats is one of the most studied biologically relevant G-quadruplex forming sequences. Over the years, several structures have come out and this sequence therefore appears to be extremely polymorphic.⁸⁰ Monomolecular G-quadruplexes are formed by one sequence of four d(TTAGGG) repeats. Bimolecular G-quadruplexes are formed either by two sequences of two repeats or by one sequence of one repeat and one sequence of three repeats. Tetramolecular human telomeric G-quadruplexes are formed by four strands, each made of one d(TTAGGG) repeat. Therefore several G-quadruplexes with different molecularities are published in the literature.

Tetra- and bi-molecular human telomeric G-quadruplexes. Historically, the first topology obtained for a human telomeric sequence is a tetramolecular parallel G-quadruplex.⁸¹ Wang and Patel published, a year later, the first G-quadruplex structure of a derivative of the *Tetrahymena* telomeric sequence (d(TTGGGGT)).⁸² Tetramolecular G-quadruplexes do not have loop constraints and, systematically, the adopted structure is parallel with all glycosidic angles in *anti*-conformation, as also observed for the human telomeric G-quadruplex

[d(TTAGGGT)]₄ (Figure 14A).⁸³ A bimolecular human telomeric G-quadruplex structure was also obtained in K⁺ conditions by X-ray crystallography.⁸⁴ The G-quadruplex is parallel with two propeller loops (Figure 14B). In solution, Phan showed that this structure is in equilibrium with an antiparallel structure (Figure 14C).⁸⁵ Those structures have grooves of different sizes. In the parallel form, all grooves are of medium size whereas in the antiparallel one, two grooves are wide and two are narrow.

Another particular structure is made of a three-repeat sequence combining with a one-repeat sequence, forming a hybrid G-quadurplex (Figure 14D).⁸⁶ Those bimolecular G-quadruplexes may be relevant, for example, in the case of the t-loop formation.



Figure 14. Structures of tetra- and bimolecular G-quadruplexes formed by the human telomeric sequence obtained in K⁺ (green balls in the schematic view) or Na⁺ (blue balls in the schematic view) solutions from NMR or X-ray. (A) Parallel tetrameric G-quadruplex [d(TTAGGGT)]₄, PDB ID: 1NP9.⁸³ (B) Parallel bimolecular G-quadruplex [d(TAGGGT)₂]₂, PDB ID: 1K8P.⁸⁴ (C) Antiparallel G-quadruplex [d(TAGGGT)₂]₂.⁸⁵ (D) Asymmetric bimolecular G-quadruplex observed for d(GGGTTA)₂GGGT + d(TAGGGT), PDB ID: 2AQY.⁸⁶

Intramolecular human telomeric G-quadruplexes. However, more relevant structures may be intramolecular, made of four G-tracts (Figure 15). Over the past 15 years, at least six different folds have been published. The final topology depends on many parameters: the nature of the monovalent cation, the concentrations, the ionic strength, the pH, the presence of crowding agents and the exact sequence. Indeed, there are 16 different ways of delimiting the d(TTAGGG)_n sequence, all containing at least the 21 bases central core d(GGGTTA)₃GGG and a maximum of 4 G-tracts.



Figure 15. Structures of the intramolecular G-quadruplexes formed by the human telomeric sequences obtained in K⁺ (green balls in the schematic view) or Na⁺ (blue balls in the schematic view) solutions from NMR or X-ray. (A) Parallel G-quadruplex formed by 22AG, PDB ID: 1KF1.⁸⁴ (B) Antiparallel G-quadruplex formed by 22AG, PDB ID: 143D.⁸⁷ (C) Hybrid-1 G-quadruplex formed by 24TTG, PDB ID: 2GKU.⁸⁸ (D) Hybrid-2 structure formed by 26TTA, PDB ID: 2JPZ.⁸⁹ (E) 2-quartet antiparallel G-quadruplex formed by 22GT, PDB ID: 2KF8.⁹⁰ (F) 2-quartet antiparallel G-quadruplex formed by 22CTA, PDB ID: 2KF8.⁹¹

By X-ray diffraction of crystals obtained from K⁺ solutions, Parkinson *et al.* obtained a parallel structure made of three G-quartets for 22AG, (d(A(GGGTTA)₃GGG) (Figure 15A).⁸⁴ However, studies in diluted solutions suggest that the crystal packing may bias the resulting structure,⁹² even though parallel G-quadruplexes could be relevant in molecular crowded conditions (and therefore in the nucleus).^{75,76,93} The NMR of 22AG in diluted solution containing 100 mM Na⁺ cations indicated that the formed G-quadruplex is antiparallel (Figure 15B).⁸⁷ In potassium solutions by NMR spectroscopy, the 22AG structure has not been elucidated but several structures for closely related sequences were obtained. The mutated sequence 24TTG (d(TT(GGGTTA)₃GGGA) folds into a hybrid-1 structure⁸⁸ as well as 23TAG, (d(TA(GGGTTA)₃GGG) (PDB ID: 2|SM)⁹⁴ and 25TAG, (d(TA(GGGTTA)₃GGGTT) (PDB ID: 2|SL) (Figure 15C).⁹⁴ Another hybrid type structure, the hybrid-2 has been published for the 26TTA sequence, (d(TTA(GGGTTA)₃GGGTT) (Figure 15D).⁸⁹ The difference between hybrid-1 and -2 comes from the order of the loops: hybrid-1 loops are respectively 5'-propeller-lateral-lateral-3' and the hybrid-2 loops are 5'-lateral-lateral-propeller-3'. A third structure, called hybrid-3, is a 2-quartet antiparallel G-quadruplex with a triplet of guanines (Figure 15E). About 70% of 22GT, (d((GGGTTA)₃GGGT) (PDB ID: 2KF8), fold into this conformation⁹⁰ and 23AG, (d(A(GGGTTA)₃GGGT), is also able to adopt this topology.⁹⁵ Finally, the modified 22CTA sequence (d(A(GGGCTA)₃GGG, the TTA loops are replaced by CTA loops) also form a 2-quartet antiparallel structure⁹¹ additionally stabilized by a G:C:G:C platform.

In the cellular context, there are uncertainties about which structures are really present. An NMR study suggest the coexistence of the 2-quartet (hybrid-3) and the hybrid-2 type G-quadruplexes for sequences with more than four repeats that form multiple G-quadruplexes.⁹⁶ Similarly, interactions between two adjacent G-quadruplexes in longer sequences have been proposed and could bring additional stability to the G-quadruplex structures formed in the telomeric regions and influence the topology.⁹⁷

II.2.6 HOW DO LIGANDS BIND TO G-QUADRUPLEXES?

A good G-quadruplex ligand has to present a good affinity and a good selectivity versus other secondary structures. In this section, we will discuss the types of ligand scaffolds synthetized and tested in the literature and their potential binding modes.

Theoretically, there are four different potential binding sites on a G-quadruplex structure: the ligand could interact with the bases in the loops and bind *via* loop recognition, it could intercalate between the G-quartets, bind in the grooves (electrostatic interactions) or stack on external G-quartets (π - π stacking) (Figure 16). The rational design mainly targets external G-quartets because it is the common feature to all G-quadruplexes. Indeed, over the years, different molecular scaffolds have appeared in the literature but one common point consists in large polyaromatic moieties, prone to π - π stack.^{98,99} Several structures of ligands stacked on external G-quartets of different G-quadruplexes were obtained by NMR spectroscopy and X-ray crystallography,¹⁰⁰⁻¹⁰⁶ therefore, the usually accepted binding mode for those molecules is π - π stacking.¹³ Note that the structure of a peptide interacting with the two external G-quartets of a parallel G-quadruplex has been published, this also supports that those binding sites are interesting G-quadruplex recognition motifs.¹⁰⁷



Figure 16. Potential interacting sites on a G-quadruplex structure.

However, other binding sites such as groove binding, loop recognition or intercalation may be as important targets but only a few structures presenting such binding modes are published in the literature. In theory, taking into consideration particular features (loops, grooves...), that are topology-dependent, may theoretically increase the selectivity for G-quadruplexes over other structures and eventually lead to inter-G-quadruplex selectivity. Only one complex implying groove binding was obtained^{108,109} and some attempts have been made to target the loops¹¹⁰ but those binding modes are as of today not the most common ones.

The most common G-quadruplex ligands usually have a very large aromatic plane, larger than those of traditional duplex intercalators.¹¹¹ This leads to increased selectivity for the G-quadruplex compared to the unfolded and/or duplex forms. Additionally, ligands have been improved to bind with higher affinity to the G-quadruplex structure *via* electrostatic interactions. Indeed, most of the ligands have positively charged groups that can interact with the negative nucleic acids backbone. However, this interaction could also bring nonspecific binding and therefore decrease the selectivity for the G-quadruplex.

In the next sections, we will briefly describe a non-exhaustive list of reported G-quadruplex ligands. They are classified into four categories according to their general molecular shape and will discuss three of each category: porphyrins and ligands with large aromatic planes, metallo-complexes, bisquinoline derivatives and macrocycles.

For a more complete list of G-quadruplex ligands, a database was published¹¹² and several reviews discuss some of the most recent ligands.^{98,99,113}

PORPHYRINS

The aromatic core of porphyrins makes their conception as G-quadruplex ligands almost straightforward. Derivatization of the core with chains of various nature, including or not positive charges, was a common way to improve the affinity of those molecules.¹¹⁴

The tetramethylpyridinium porphyrin, TMPyP4, one of the most studied G-quadruplex ligand, increases the melting temperature of some G-quadruplexes by values as high as 17 °C¹¹⁵ and inhibits telomerase.¹¹⁶ Hurley also showed that TMPyP4 was able to regulate the expression of the c-myc oncogene.¹¹⁷ In addition, a surprising structure with one of the two binding ligands interacting with an external loop has been obtained.¹¹⁸ However, this could come from a nonspecific binding. Indeed, a series of studies evidenced TMPyP4's lack of selectivity.^{115,119-}

¹²¹ It was even shown that this ligand unfolds very stable RNA G-quadruplexes, displacing the equilibrium towards the unfolded species.¹²²



Figure 17. Examples of porphyrin G-quadruplex ligands.

The development of Mn(III)-porphyrins by Pratviel's group improved the selectivity for Gquadruplexes. Surface plasmon resonance (SPR) evaluation of the dissociation constants indicated sub-micromolar constants for the G-quadruplex, compared to values higher than tens of micromolar for duplexes.¹²³ This increased selectivity comes supposedly from the steric hindrances of the side chains.¹²⁴

Those two previously described porphyrins bare positive charges that probably increase the interaction between the ligands and the DNA. A negatively charged porphyrin, N-methyl-mesoporphyrin IX (NMM), possesses two negative charges at physiological pH. Its affinity is affected by the presence of those charges but it presents a stronger selectivity for G-quadruplexes versus duplexes.¹²⁵ It was also shown that NMM has a preferential binding to parallel G-quadruplexes whose external G-quartets are particularly exposed. The exalted fluorescence when bound to a G-quadruplex and its selectivity toward parallel G-quadruplexes permitted its use as a strand orientation reporter for parallel G-quadruplexes.¹²⁶

METALLO-COMPLEXES



Figure 18. Examples of metal complexes G-quadruplex ligands.

Some of the porphyrins can complex a metal cation, changing the distribution of the charges on the ligand and potentially influencing the binding affinity.¹²⁷ Researchers also developed metal complexes to target G-quadruplexes because they are easy to synthetize and could lock an organic moiety in a specific conformation. The ideal metallo-complex binds the Gquadruplex with its metal positioned over the cation channel of the G-quadruplex and its organic moiety(ies) stacked on the quartet.¹²⁸ The Ni(II)-salophen complex was designed this way and stabilizes the G-quadruplex form by up to 33 °C, which made this ligand one of the best temperature stabilizers.¹²⁸ It is also a telomerase inhibitor.¹²⁹ Further optimization is possible by modifying the chemical groups on the central phenyl moiety and the metal nature.^{130,131}

Terpyridines in interaction with various metal cations have been employed as G-quadruplex ligands.¹³² The platinum- and palladium-tolylterpyridine (Pt- and Pd-*ttpy* complexes) were analyzed by mass spectrometry and were shown to bind covalently to the DNA.¹³³ Teulade-Fichou showed that the Pt-*ttpy* ligand is covalently bound to the adenines located in the loop regions of the human telomeric G-quadruplex.¹³⁴ Another study revealed the importance of the metal geometry of those ligands: to obtain the best selectivity over duplexes, the trigonal bipyramidal shape obtained for the Copper-tolyterpyridine, Cu-*ttpy*, ligand is the best because it impedes duplex intercalation.¹³² Derivatization strategies around the terpyridine moiety has been explored.^{135,136} This ligand complex with a G-quadruplex has also been used, for completely different purposes, as a metalloenzyme for Diels-Alder reactions catalysis.¹³⁶

Other types of metallo-complexes have been employed to target DNA G-quadruplexes such as ruthenium and di-ruthenium complexes.^{137,138} The rationale was the use of a large aromatic plane in between the metal cores that bare the charges. The dinuclear $\{Ru(phen)_2\}_2$ -tpphz binds

diastereoselectively to antiparallel human telomeric G-quadruplexes (only one of the diastereoisomers binds to the G-quadruplexes).¹³⁹ These *Ru*-complexes have been proposed as light up probes for *in vivo* direct imaging studies because they incorporates well inside the cell.¹⁴⁰

BISQUINOLINE DERIVATIVES



Figure 19. Examples of bisquinoline G-quadruplex ligands.

Another family of very affine ligands is based on derivatives of the 2,6-pyrido-dicarboxamide (PDC) core.^{141,142} These ligands were shown to have a very good telomerase inhibition activity, correlated with an increase of more than 20 °C of the melting temperature of G-quadruplexes. The tritiated ligand 360A has been used to probe the presence of G-quadruplexes in cells: the ligand localizes preferentially at the human chromosome ends, which is another evidence for the existence of G-quadruplexes *in vivo*.¹⁴³ It has however been shown that 360A could induce the folding of G-quadruplexes¹⁴⁴ and therefore the conclusion whether the G-quadruplexes are present *in vivo* remains unclear. Interestingly, the combination of 360A with Cu(II) has been tested but the Cu-360A complex had a very poor affinity for G-quadruplexes.¹⁴⁵ The explanation lies in the conformational change of the 360A induced by the copper: the crescent shape of the ligand is lost and its configuration becomes more linear (the oxygens of the amide groups and the nitrogen of the pyridine bind to the Cu center).

Knowing that the crescent shape scaffold is required, Teulade-Fichou's group extended the family to phenanthroline derivatives such as Phen-DC3.¹⁴⁶ This ligand stabilizes the G-quadruplex even in the presence of high amount of duplex competitor, demonstrating its selectivity. Because the size of the G-quartet and of the ligand are comparable, it was initially suggested that Phen-DC3 stacks upon an external G-quartet. Recently, this binding mode was confirmed by NMR on a parallel c-myc promoter G-quadruplex.¹⁴⁷ Due to its good affinity and selectivity, Phen-DC3 has become one of the benchmark tests for ligand binding assay

development.^{115,148} Biological assays have shown that this ligand could lead to genomic instability.^{149,150}

Similar stabilization effects and selectivities were found for the pyridostatin ligand (abbreviated PDS).¹⁵¹ Thanks to its high affinity (~500 nM range),¹⁵² this ligand was used for determination of G-quadruplex forming sequences *in vivo* using a high-throughput sequencing approach³⁰ and a fishing approach.¹⁵³ Both methods suggest that G-quadruplexes are present *in vivo*, notably in the telomere regions. However, it has not yet been documented if PDS induces the folding of G-quadruplexes, as well as for 360A.¹⁴⁴

MACROCYCLES



Figure 20. Examples of macrocycles G-quadruplex ligands.

Another completely different scaffold, initially discovered in 2001 with telomestatin,¹⁵⁴ consists in macrocycles. The macrocyclic shape of telomestatin is perfect for π - π interaction with an external G-quartet, and it also makes the telomestatin a bad duplex ligand. Molecular dynamics simulations also suggest that telomestatin and macrocycle analogs recruit metal ions upon binding on an G-quartet,¹⁵⁵ which was also observed experimentally by mass spectrometry.¹⁵⁶ The good affinity of this macrocycle for G-quadruplexes is suggested by the increase of melting temperature induced for G-quadruplexes ($\Delta T_m = 24 \text{ °C}$)⁹⁸ and by mass spectrometry that showed the delayed unwinding of the G-quadruplex in presence of the ligand.¹⁵⁷ Additionally, the selectivity for the human telomeric G-quadruplex d(TTAGGG)₄ in potassium (hybrid G-quadruplexes) over the sodium conditions (antiparallel G-quadruplexes) also suggests a structural preference for hybrid conformations compared to antiparallel ones.¹⁵⁸ It is also possible that the binding site between the ligand and the G-quadruplex is cation-dependent, with a preference for potassium, which could explain the preferential binding in potassium conditions. The biological effect of telomestatin was shown notably by

Riou and his team. They found that the telomeres of human tumor cells treated by the ligand eroded faster than normal cells, which was explained by the uncapping of the proteins on the chromosomal ends.¹⁵⁹

Similarly, an oxazole based nonmacrocyclic ligand, TOxaPy, has been developed.^{160,161} It exhibited a strong preference for G-quadruplexes over duplexes and more interestingly, stabilizes exclusively the human telomeric sequence folded in sodium conditions compared to potassium, suggesting a better affinity for antiparallel G-quadruplexes. To this day, it is still one of the most inter-G-quadruplex selective ligand. Its binding mode may not be as straightforward as telomestatin (considered as a π - π stacker). Indeed, TOxaPy can rearrange to linear folds and potentially fit in the G-quadruplexes grooves.¹⁶⁰

Additional coulombic interactions on macrocyclic ligands were brought by the amino side chains of the L2H2 derivatives.¹⁶² This ligand has a size and shape similar to telomestatin except that, unbound, it is not planar. When bound, NMR studies showed that L2H2 was able to become flatter and accommodate to the G-quartet, without influencing the overall human telomeric G-quadruplex folding.¹⁶³ This study also strongly suggested the cation-mediated binding of the ligand. Derivatives of this ligand were recently synthesized and found to induce the folding of human telomeric G-quadruplexes without added salt.¹⁶⁴ A dimeric derivative (two L2H2 ligands linked covalently by an alkyl chain) has also been synthesized and shown to bind to long repeats of the human telomeric sequence supposedly by π - π stacking, each L2H2 part of the ligand to each end G-quartet.¹⁶⁵

In conclusion, several families have been synthesized with sometimes very different chemical scaffolds.^{98,99} The common point is the aromatic region presumably used to stack on top of G-quartets. However, when no atomic structure is published one can only guess the binding mode for those ligands and simple stacking is sometimes not sufficient to explain the selectivity observed for different G-quadruplexes. More assays have to be developed to study not only the affinity of a ligand for a G-quadruplex but also to assess its binding mode. This question is treated in the Chapter VI.

II.2.7 HOW TO STUDY G-QUADRUPLEX FOLDING AND LIGAND BINDING MODES?

In this work, we want to answer several questions. What is the structure of the G-quadruplex? Can we elucidate G-quadruplex folding mechanism? What is the affinity of a ligand to a particular G-quadruplex? How does the ligand bind to G-quadruplexes? How many ligands? Are there conformational changes upon ligand binding?

Plenty of different biophysical methods are available, each of the techniques has pros and cons. Some can be used to determine binding affinities, others to study structural changes, others to elucidate atomic structures... This section focuses on the introduction of the most commonly used biophysical techniques to study G-quadruplex structures and G-quadruplex-ligand interactions.

STRUCTURAL CHARACTERIZATION

To characterize the structures of the G-quadruplexes and ligand induced-structural changes, several techniques can be used, depending whether or not the structure has to be characterized with atomic resolution.

Circular dichroism (CD). Circular dichroism is a chiroptical technique developed to assign the configuration of chiral molecules.¹⁶⁶ It has been popularized over the last fifty years to become one of the most widely used techniques to characterize conformational changes and supramolecular interactions. The origin of CD is due to the presence of absorbing chromophores in a chiral environment.¹⁶⁷ Depending on the chiral orientation, the chromophore absorbs differently the light of a given polarization. The difference in absorbance between left and right circularly polarized light gives the CD signal.

In the case of G-quadruplexes, a typical approximation considers the guanines as the only contributing chromophores (the absorbance of the other bases is considered negligible).¹⁶⁷ Guanines have two absorption bands at 248 and 279 nm.¹⁶⁸ In a G-quadruplex, the molecular orbitals involved in electronic transitions are delocalized on several guanines and G-quartets.¹⁶⁹ G-quartets are stacked and rotated one upon each other (Figure 21).^{14,170} Because of the usual way to represent schematically the structures (as stacked squared boxes),¹⁵ the rotation can be easily forgotten but is the key to the understanding of the origin of the CD

signal. Indeed, the position of the band is dependent on the stacking type (*syn-anti, anti-syn, anti-anti* or *syn-syn*) of the guanines and on this rotation. For those reasons, markedly different CD spectra can be obtained for different families of G-quadruplexes.¹⁷⁰⁻¹⁷² More details on how to assign the topology of G-quadruplexes are provided in the experimental approaches section, in Chapter III.



Figure 21. Transition dipole moments at 248 and 279 nm in green and red, respectively, for stacked guanines. Illustration reproduced from Webba da Silva.¹⁷⁰

Nuclear magnetic resonance spectroscopy (NMR). To obtain atomic details on a structure, NMR is also commonly used.¹⁷³ A 1D ¹H NMR experiment can be run to determine whether or not the chosen sequence is folded into a G-quadruplex. Peaks between 10.5 and 12 ppm are typical of guanine imino-protons embedded in a G-quadruplex structure (the proton on the N1 of the guanines). If the structure is not symmetrical, the number of peaks in this region can give access to the number of G-quartets (4 peaks per G-quartet). The guanine's imino-protons are quickly exchanged when they are in contact with the solvent and are therefore invisible in the spectrum. 2D and 3D experiments allow the determination of the atomic structure.

The main drawback of the technique is the difficulty to interpret the data when several conformers coexist in the solution (*a priori* one does not know from which structure a peak comes from). Very often, NMR specialists mutate the sequences to drive the equilibria towards one well defined conformation.¹⁷⁴ Moreover, highly concentrated solutions are needed and increase the cost of the experiment.

X-ray diffraction. X-ray diffraction of crystals allows the atomic resolution of structures. Unfortunately, this technique is limited by the likelihood of complexes to crystalize. Additionally, not all G-quadruplex structures crystalize well: parallel G-quadruplexes tend to crystalize more easily.⁹² Indeed, the external G-quartets of parallel G-quadruplexes are free, which allows the crystallization due to a better stacking of G-quadruplexes in the crystallographic lattice. So, even though X-ray diffraction provides atomic details on a complex, it does not necessarily reflect the possible heterogeneity of the complexes in solution.

CHARACTERIZATION OF REACTION EQUILIBRIA

To get insight into the binding affinity or folding mechanism of G-quadruplexes (Figure 22), several techniques can be employed. The requirement is that the technique measures, as a function of concentration, time or temperature, a signal that is proportional to the concentration of the complex formed. For example, spectroscopic techniques such as CD, UV absorbance, fluorescence... are commonly used to determine equilibrium constants or to monitor kinetics. In addition to the structure, NMR spectroscopy can also assess kinetics and dynamics of G-quadruplexes, determine molecular interactions or melting temperature...¹⁷⁴

Thermal denaturation experiments. A simple method to determine the stability of Gquadruplexes and compare ligand-induced stabilization is to measure the melting temperature, the temperature at which 50% of the complexes in solution are folded/unfolded. The temperature-induced unfolding of G-quadruplexes is typically monitored by UV-vis absorbance^{175,176} but other methods measuring any property sensitive to folded/unfolded ratio, such as circular dichroism (CD), fluorescence (using Förster Resonance Energy Transfer, FRET)¹⁷⁷ or nuclear magnetic resonance (NMR), can also be used.¹⁷⁸



Figure 22. Schematic representation of the solution equilibria at stake when a ligand bind to a Gquadruplex-forming sequence. The green balls are cations and the black disk is a ligand.

In practice, determination of the melting temperature of a G-quadruplex can give access to the folding equilibrium constant (linked to ΔG_1 , in Figure 22), without the possibility to obtain individual K⁺ binding constants. Comparison of the melting temperatures of a G-quadruplex with a variety of ligands is a classical ligand screening method. A common interpretation is that the higher the melting temperature stabilization (ΔT_m), the better the ligand, because the ligand displaces more the equilibrium towards the formation of the G-quadruplex.^{146,178,179} However, in this case, only two ensembles are defined: the unfolded species and the Gquadruplexes. Only an apparent equilibrium constant can be determined with those techniques: no information on the equilibria of ligand binding to the G-quadruplex (linked to ΔG_{2i} in Figure 22) and of ligand binding to unfolded species (linked to ΔG_{3i} in Figure 22) can be accessed. The equilibria can be even more complicated if multiple G-quadruplex conformations coexist in solution (such as for the human telomeric sequence). To characterize more deeply those equilibria, Chaires and coworkers have deconvoluted spectroscopic signals (CD, UV absorbance, Fluorescence) using singular value decomposition (SVD) to enumerate the number of significant species and intermediates required to properly analyze thermal denaturation experiments.¹⁸⁰⁻¹⁸²

Surface plasmon resonance (SPR). Quantification of the affinity constants and binding kinetics can be provided by surface plasmon resonance. Briefly, a target is bound to a surface and the analyte solution is flushed on the surface. The SPR machine detects the change in amount of ligand bound to the surface using the total internal reflection of a light beam on the surface. When the ligand binds to the target, a change in the refraction index close to the surface occurs and the angle at which there is plasmon resonance changes. The system is sensitive to the changes in refraction index and in the mass of the binding ligand. This property

can be used to monitor the interaction of ligands with G-quadruplexes according to the time to obtain steady state and kinetic parameters.^{183,184} Because SPR is sensitive to the mass of the ligand, it is possible to deduce the number of bound ligands.

However, because a change in mass is theoretically needed to monitor reactions using SPR, the study of G-quadruplex folding is not possible (the mass of two K⁺ cations is not sufficient). Another drawback is that one of the partners has to be immobilized on a surface and this immobilization could affect the structure of the G-quadruplex and/or the interaction with ligands.

Förster Resonance Energy Transfer (FRET) and Single-molecule FRET (smFRET). FRET is a mechanism of energy transfer between two chromophores. One of the chromophores, the donor, is irradiated and becomes electronically excited. The donor is relaxed and the second chromophore, the acceptor, is excited by resonance energy transfer. The energy transfer efficiency is inversely proportional to the sixth power of the distance between the donor and the acceptor. The experiment consists in measuring the acceptor relaxation by fluorescence. The distance between the two chromophores, measured by FRET, could in principle give information on the structure of the complex. However, if performed in bulk solution, the signal is averaged on the ensemble of structures.

To get access to binding constants and folding landscapes, smFRET can be employed. The same principle is used but one of the two binding partners is immobilized on a surface and observed under the microscope. Signals from individual pairs of chromophores are recorded as a function of time. The assay consists in measuring changes in distances between the dyes, and correlating these distance changes to the dynamics between different structures.^{185,186} Folding landscape of G-quadruplexes were obtained by smFRET.¹⁸⁷ smFRET can provide structural details because the dyes may be positioned at different distances in different G-quadruplex conformations. This can provide information on the interconversion between different structures. Ligand-induced structural changes can possibly be recorded providing that the dyes positions are affected by the binding.^{186,188,189}

The smFRET technique has two drawbacks. First, the binding constants could be influenced by the fact that one of the partners has to be immobilized on a surface. Second, the molecule is

tagged with chromophores, which could influence the binding site and/or the conformations of the G-quadruplexes.

Fluorescent Intercalator Displacement (FID). Another fluorescence-based method to screen for ligands is the fluorescent intercalator displacement assay, FID.^{190,191} The technique consists in displacing a molecule, which is fluorescent when bound to the G-quadruplex, with the target ligand and measuring the decrease in fluorescence. As opposed to FRET, the G-quadruplex does not need to be labelled. However, the ability to displace the fluorescent molecule is probably binding mode-dependent: both the fluorescent molecule and the ligand have to compete for the same binding site. The reliability of the assay has been confirmed by FRET-melting and MS experiments for some ligands,¹⁹² and assumed to be valid for any ligand.

II.3. MASS SPECTROMETRY

The previously described techniques are mainly good at determining either the binding constants, or the structures. In this section, mass spectrometry will be described focusing on its potential in addressing the two types of questions (structural and quantitative).

Mass spectrometry allows the identification of the chemical constituents of a solution by separating them according to their mass to charge m/z ratio. This allows qualitative and quantitative analysis. To achieve this, the molecules are ionized and desolvated. In the gas phase, the ions are guided inside the mass spectrometer by electric fields applied between electrodes. Ions are then separated according to their m/z ratio by the analyzer, and detected. Different types of ionization sources and mass analyzers exist. Additionally, some devices can be coupled with the mass analyzers to increase the information obtained using mass spectrometry, notably about structures.

II.3.1 NATIVE MASS SPECTROMETRY

Mass spectrometry is a research field that has grown a lot since its invention 110 years ago.¹⁹³ The mass spectrometry paternity is usually attributed to J. J. Thomson who was one of the first to separate particles of different m/z ratio. For this work, he obtained the Nobel Prize in Physics in 1906.¹⁹⁴ In 1953, Stephens used time dispersion to separate ions according to their m/z.¹⁹⁵ He invented the first time-of-flight (TOF) mass spectrometer, whose principle is described in the next section. In the same decade, several mass analyzers were developed such as the ion cyclotron resonance (ICR) MS¹⁹⁶ or the quadrupole (Q) MS.¹⁹⁷

Before the 1980s, the electron impact ionization was the most common way to desolvate and ionize molecules. It was impossible to transfer intact molecules from the solution to the gas phase and therefore it was impossible to analyze large biomolecules and biomolecule complexes. Indeed, this technique was not a soft ionization method as most chemicals were fragmented. A significantly softer ionization technique was invented first by Dole *et al.* in 1968¹⁹⁸ and adapted for the use of macrobiomolecules in 1984 by Fenn and called electrospray ionization (ESI).¹⁹⁹

Over the years, electrospray ionization has become the most widely spread ionization method to gently desolvate large molecules. John B. Fenn was awarded the Nobel Prize in Chemistry in 2002 for his contribution to the field.²⁰⁰ In 1984 in his first paper about electrospray ionization,¹⁹⁹ Fenn described ESI as a way to desolvate and analyze bigger complexes, too fragile for ionization with other methods and "non-volatile". In his 1989 Science paper,²⁰¹ he describes the use of ESI to obtain MS spectra of biopolymers of masses up to 130 kDa. Nowadays, noncovalent complexes with masses higher than the MDa²⁰² including whole ribosomes^{203,204} or viruses,²⁰⁵ can be electrosprayed. Thanks to this ionization source mass spectrometry can be used for studying very large noncovalent complexes.²⁰⁶

Native mass spectrometry allows the investigation of noncovalent complexes interactions. Note that the "native" term might be excessive because the mass spectrometry experiments are carried out on ions in the gas phase and structural rearrangement may occur when the ions go from the polar water solution (relative permittivity $\varepsilon_r = 78$) to the very apolar gas phase ($\varepsilon_r = 1$). This specific point is addressed in a dedicated section, in Chapter II.3.3.

The main advantage of native mass spectrometry is its ability to unambiguously detect simultaneously several species in a heterogeneous mixture of complexes. Unlike NMR spectroscopy and X-ray diffraction of the studied structures, mass spectrometry cannot provide atomic details and is a destructive method. However, it requires small amount of sample (< nmol for classical electrospray of nucleic acids) and provides key information on the binding partners: nature, stoichiometries, and insight into the quaternary structures.²⁰⁷ Compared to NMR spectroscopy, mass spectrometry does not require expensive labeling. In addition, there is no mass limitation and very labile complexes can be analyzed.²⁰⁸

The main disadvantage of mass spectrometry interaction studies is the expensive instrumentation required. In addition, depending on the mass spectrometer specifications different types of information can be obtained (information about the shape, the sequence...). Yet, native mass spectra can be obtained even on old cheap second-hand mass spectrometers.²⁰⁹

Native MS is an extremely powerful biophysical technique for equilibrium characterization. The information provided by native MS is: Does a ligand bind? What is the ligand stoichiometry? The cation stoichiometry? Is it possible to obtain information on the affinity

constants and binding kinetics? Being able to distinguish the number of ligands on a target, the stoichiometry, and to quantify the amount of each of those stoichiometries is very important for biophysical studies.

II.3.2 ION MOBILITY SPECTROMETRY (IMS)

On the structural point of view, the higher resolution techniques are the x-ray crystallography and the NMR spectroscopy as described previously. Those techniques require to invest a lot of time and sample. Furthermore, mixtures of products are difficult to analyze: the samples have to be purified and concentrated, which might affect their structure. The NMR-compatible complexes are additionally limited to a range of a few tens of kiloDaltons. For X-ray characterization of the complexes, the crystallogenesis might be reluctant depending on the lability of the complexes.

Several techniques were developed in order to obtain structural information from mass spectrometry experiments. Indeed, techniques such as chemical cross linking (CX-MS)²¹⁰ and hydrogen-deuterium exchange²¹¹ (HDX) techniques gave to the mass spectrometry researchers tools to reach up to 20-30 Å² resolutions.²⁰⁷ Those are techniques developed and used for proteins and protein complexes, not directly applicable to nucleic acids. They could be adapted but with significant efforts.

The recent development of ion mobility coupled to mass spectrometry (IMS-MS) gave mass spectrometry the lacking structural information.²¹² In addition with the *m/z* separation, ions are separated according to their shape.^{212,213} Because this additional separation is due to an intrinsic property of the ion, IMS-MS is directly amenable to study nucleic acids. There are several families of IMS instruments: field-asymmetric ion mobility spectrometry (FAIMS),²¹⁴ travelling-wave ion mobility spectrometry (TWIMS),²¹⁵ transversal modulation ion mobility spectrometry (TMIMS),²¹⁶ trapped ion mobility spectrometry (TIMS),^{217,218} drift tube ion mobility spectrometry (DTIMS).²¹⁹

Because a DTIMS instrument is available in our laboratory, a fundamental theory and functioning of the DTIMS (Figure 23) are described hereafter. DTIMS is a temporal dispersive method in which the more compact ions reach the detector before the more extended ones. Therefore, ions with the same *m*/*z* ratio can now be separated according to their shape, which is of great interest to separate two conformers and to study structural changes.

One of the greatest advantages of DTIMS compared to the other IMS devices is that it allows to obtain collision cross sections values (a physical parameter that is linked to the size and

shape of the ion) without calibration of the mobility cell.²²⁰ In DTIMS, under the effect of a uniform linear electric field, ions pass through a gas-filled cell. In this cell, the ions are accelerated until their velocity is constant (v_d), as the friction forces counterbalance the electric force:

$$v_d = K \times E$$

Equation 1

Where *K* is the mobility and *E* the electric field. The velocity (and hence the mobility) depends on the gas pressure and the temperature. The mobility value is related to the orientationally averaged momentum transfer collision integral, also called the collision cross section or CCS. It is a physical parameter corresponding to the surface area that makes collisions with the gas in the cell. The larger the value, the more extended the conformation of the ion.



Figure 23. Illustration of the principle of ion mobility spectrometry.

The CCS value does not directly give access to the structure. To obtain structures, molecular dynamics simulations are made and the theoretical CCS value is calculated. The best molecular structures that fit to the experimental CCS are used as potential atomic models. Several groups correlate ion mobility results and CCS with structural models either for small molecules,²²¹ large protein complexes²²² or nucleic acids.²²³

In summary, native mass spectrometry can be used to determine stoichiometries and ion mobility spectrometry coupled to mass spectrometry can provide an additional refinement in characterization of the structures that are present.

II.3.3 MASS SPECTROMETRY AS TOOL TO STUDY SOLUTION PHASE REACTIONS

Although mass spectrometry experiments are conducted in the gas phase, if conducted carefully, they can reflect solution equilibria. To study equilibria or kinetics of solution reactions by using a gas phase technique, four main conditions have to be respected (Figure 24).²²⁴



Figure 24. Schematic representation of the Hess cycle of the binding of a small molecule (black and yellow disk) to a duplex DNA.

Desolvation. To go from the solution to the gas phase, electrospray ionization (ESI) was employed and is described hereafter (Figure 25). The solution containing the analyte is injected inside a capillary. A voltage is set between the extremity of the capillary and the entrance of the mass spectrometer. The solution forms a Taylor cone that emits solution droplets. This region of the mass spectrometer is at atmospheric pressure. The droplets evaporate thanks to a coaxial gas flux (nitrogen), and the droplets of the desired polarity are attracted to the mass spectrometer.

From those droplets, two ion production mechanisms are discussed in the literature:²²⁵ the ion evaporation model (Figure 25B)²²⁶ and the charged residue model (Figure 25C).¹⁹⁸ In the ion evaporation model, as the droplets shrink, the charge density increases and at some point

the molecule with charges is ejected from the droplet. In the charged residue model, the large droplets emit daughter droplets containing only one ion. Those droplets are emitted when the mother droplet's size reaches the Raleigh limit. This is the size limit at which the Coulomb repulsion is too high for further evaporation, and leads to the explosion of the droplet.

While the ion evaporation model is commonly accepted for small molecules, the charged residue model is commonly accepted for large molecules. Indeed, it involves a last step of declustering, during which the removal of exceeding solvent and buffer molecules from the analyte is achieved. This specific step can be observed experimentally (see Chapter III). This last step is important to obtain mass spectra reflecting stoichiometries of complexes in solution. Furthermore, if structural information is desired, it is important not to denature the ions in the gas phase in order to keep their native-like folding.



Figure 25. Principle of electrospray ionization. (A) Schematic representation of the source working in negative ion mode, (B) the ion evaporation model and (C) the charge residue model.

Because the volume of the droplet decreases, one can expect to shift the equilibria toward the formation of the complexes (the concentrations increase). This is, however, the case only if the rate of binding is much faster than the shrinkage of the droplets. In 2004, Kebarle and coworkers showed that the droplets emit daughter droplets in the microsecond time scale and that the amount of complexes formed is negligible for reactions as fast as $25 \,\mu s$.²²⁴ Thanks to the very rapid evaporation, this contribution on the error on the quantifications of the species is very likely negligible.

Transfer from the solution to the gas phase. Secondly, the ionization efficiency may be different for the different species. Indeed, the number of ions produced is not necessarily directly proportional to the number of ions in solution. Therefore, the gas phase ion intensities are not necessarily directly proportional to the concentrations in solution. However, a mathematical correction involving response factors, the proportionality factors between intensities and concentrations, has been previously developed.²²⁷ Thanks to internal standards it is possible to calculate the relative response factors for each species in a titration or a kinetic experiment. To know whether the correction is needed, the mass balance upon titration (or as a function of time) is checked thanks to the internal standards. This specific point is described in more details in Chapter III.

Equilibrium in the gas phase. Finally, one could think that the equilibrium studied is related to the equilibrium in the gas phase (ΔG^G_{diss} , top equilibrium in Figure 24), the dissociation of the complex in the gas phase. However, because of the small amount of ions sprayed, once in the gas phase the partners will never meet and the association does not happen. On the other hand, the instrument is parametrized at best to conserve the noncovalent complexes. A soft tune allows to keep intact noncovalent complexes (instrument tuning is described in Chapter III).

In addition, for the binding of oppositely charged partners (such as positive small molecules binding to negatively charged nucleic acids), the interaction is enhanced in the gas phase. Indeed, in the gas phase, coulombic interactions are reinforced: repulsion of charges of the same sign and attraction of particles of opposite charges are strengthened. Therefore, the dissociation does not occur for such ion pairs.²²⁸ However, it is possible that the binding of some neutral ligands is too weak in the gas phase and therefore the method is not suited. To ensure that we are not in such a case, the experiments are usually run at several activation voltages to check its influence on the proportions of the different species.

In conclusion, even if the mass spectrometry implies the desolvation of mixtures and the quantification of the species in the gas phase, the nature of species and their abundances are related to those in solution. Therefore, the intensities of the species detected can be linked with their concentrations in solution and the study of reactions happening in solution.

Gas phase structural rearrangement. If the goal is to obtain structural information, an additional point has to be considered: structural rearrangements may occur. Even when doing everything to keep as close as possible the structures from the solution, the structures of the complexes may be affected by the desolvation. In the gas phase, we want to kinetically trap the complexes in metastable states, reflecting their initial states in solution. To study the conformations of biocomplexes in the gas phase (with IMS, for example) and trace back their structures in solution, one has to understand the desolvation process.

III. EXPERIMENTAL APPROACHES

III.1. MASS SPECTROMETRY

The aim of this section is to discuss how to prepare the solutions, choose the mass spectrometer, set up the instrumental parameters and process and analyze the data. Technical details are found in the respective articles.

III.1.1 SAMPLE PREPARATION

To study noncovalent complexes interactions, the first condition is to find solution conditions in which the complexes may fold. Usually, a solvent is MS-compatible if it is volatile: its boiling point is not too high (~100 °C). Here, we studied interactions in biologically relevant aqueous conditions. The solvent is pure water. Mixtures of solvent are also possible (for example to increase the signal). Note that for G-quadruplexes, addition of alcohols may favor the folding but also alter the topology of the G-quadruplexes.⁷⁸

Additionally, for nucleic acids folding, a high ionic strength (~100 mM anions and cations)⁷⁴ is required in order to decrease the repulsion of charges of the same sign. In the nucleic acids field, the ionic strength is commonly fixed to 100-150 mM monovalent cations and anions which is also relevant monovalent salt concentration in the nucleus.⁷³ In mass spectrometry, the most commonly used electrolyte is ammonium acetate (NH₄OAc) at pH 7. (Note that in this pH range, the ammonium acetate does not act as a pH buffer. To buffer at pH 7, imidazole can be used). However, because our complexes often interact with other cations than ammonium, we developed a sample preparation method thanks to which other cations, such as potassium or strontium, up to the 1 mM concentration can be added. This particular method is described and discussed in more details in Chapter IV.

In addition, we systematically add dT_6 (a hexamer of deoxythymines) and use it as internal standard. The purpose of this internal standard is described in the response factors section (Section III.1.5). It is also used as control for the IMS cell.

III.1.2 Which mass spectrometer should be used?

To perform mass spectrometry experiments in native conditions, there is no particular instrument required. In this paragraph, pros and cons of using each of four mass spectrometers, among those available in the institute, are discussed.

WATERS LCT PREMIER



Figure 26. The Waters LCT Premier mass spectrometer.

This instrument is composed of a source (in our case an ESI or nanoESI source) and a time-offlight (TOF) analyzer connected with ion guides. The TOF measures the time taken by the ions to go in the vacuum from the pusher, where ions of the same charge receive the same amount of kinetic energy (thanks to a voltage push), to the detector (a multichannel plate). Because the ions with lower m/z ratios go faster than the bigger ones, they reach the detector sooner. A calibration curve is typically done with Na-Cs-I clusters. From the curve, the m/z ratio of the ions of interest can be retrieved.

We modified this instrument in order to be able to increase the pressure in the source up to 45 mbar. This high pressure coupled with relatively low voltages allowed to preserve intact

the noncovalent complexes from the solution to the gas phase and to the analyzer. In-source collision induced dissociation can be conducted by increasing the voltages applied at the entrance of the mass spectrometer (Sample Cone and Ion Guide 1 Voltages).

This instrument was used because we were able to tune it to be extremely soft. It is easy to handle and a large number of experiments can be done in a small amount of time. The extraction of the data for further analysis is also very simple thanks to the MassLynx software. However, it has a low resolution because of its short TOF tube and does not possess an ion-mobility device. It also cannot perform MS/MS experiment on mass or mobility selected ions.

THERMO EXACTIVE



Figure 27. The Thermo Exactive mass spectrometer.

This instrument is almost as simple as the LCT Premier. An ESI source is connected with transfer optics to the analyzer, an orbitrap. The orbitrap technology allows to reach very high mass resolutions. Indeed, the analysis is done on ions oscillating between electrodes. The measurement of the image current and the Fourrier transform of the signal give the mass spectrum.²²⁹ Due to the high coherence of the ions inside the orbitrap, which is possible thanks to the extremely low pressure ($\approx 10^{-10}$ mbar), very high resolutions are obtained.

In addition with the in-source collision induced dissociation (CID), it is possible to send the ions in a cell where the pressure is a little bit higher (HCD cell). The ions are trapped in a potential energy well that can be adjusted to perform ion fragmentation. However, no mass selection is possible.

BRUKER AMAZON SL



Figure 28. The Bruker Amazon SL mass spectrometer.

This mass spectrometer is a quadrupole ion trap used with an ESI source. It allows mass selection and fragmentation of selected ions as well as MS^n if needed. First, the ions will be stored inside the trap using potential energy on the electrodes that allow the trapping of ions from 400 to 2200 *m/z*. Later, in the scanning mode, the voltage on the trap electrodes makes the ion trajectories become less and less stable, ejecting the ions according to their *m/z*. It is also possible to destabilize the trajectories of the ions except those of one defined *m/z*. Accelerating the ions in the Helium filled trap allows the fragmentation. Further selection of fragments and activation is called MS^n .

In our laboratory this instrument is modified allowing the coupling of a laser to perform laserbased MS/MS which leads to other types of fragmentation.²³⁰ The trap is therefore a great advantage. However, the quantification is less robust because the number of each ions kept in the trap is biased: all the ions cannot be in the center of the trap and being trapped with the same efficiency. Nevertheless, we managed to perform quantitative experiments on this mass spectrometer.

AGILENT 6560 IMS-Q-TOF



Figure 29. The Agilent 6560 IMS-Q-TOF mass spectrometer.

This instrument's TOF is much longer than the LCT's, which increase the instrumental *m/z* resolution. It also has a quadrupole to select ions followed by a collision cell to perform MS/MS of mass selected ions. However, it is not possible to fragment beyond MS/MS, unlike with the ion trap.

This mass spectrometer also has a drift tube mobility cell (DTIMS) which allows the separation of ions based on their shape and develops the mass spectra on a supplementary dimension: the time spent by the ions to go through the cell. Mobility-selected fragmentation is therefore possible. In addition with proper tuning of the voltages to desolvate and guide the ions gently to the analyzer, it is also mandatory to monitor and regulate the pressure inside the mobility cell with a very high precision and therefore, more precaution is needed.²²² The extraction and treatment of the heavier data makes using this mass spectrometer more time consuming. Therefore, this instrument was chosen only when IMS was potentially useful.
In summary, those instruments were used for different parts of the work depending on their advantages: native mass spectra were obtained either with the LCT Premier or the Agilent 6560 IMS-Q-TOF, depending if the IMS dimension was needed. The Amazon XL was used for MS/MS experiments of the complexes and the Exactive was used to measure accurate masses of synthesized compounds.

III.1.3 MS INSTRUMENTAL SETTINGS

To work in native-like conditions, the mass spectrometer tuning is important. When the ions are accelerated in the mass spectrometer, they receive energy upon collisions with the gas. This energy could cause structural rearrangement or lead to fragmentation. Those effects have to be avoided or minimized. Unfortunately, there is no clear-cut recipe of how to tune a mass spectrometer in soft conditions. To set up the instrument, a reference compound can be used.

The characteristics of a good reference compound are: bieng of close nature and size compared to the complexes of interest and being sensitive to the tuning of the instrument. Our reference compound for tuning is a dimer of dG₄T₄G₄. In 100 mM NH₄OAc, this DNA sequence folds into a bimolecular G-quadruplex that contains three ammonium cations between the G-quartets and two more labile cations on the top of each terminal G-quartet.²³¹ The ammonium cations are easily released (as neutral ammonia) upon activation. The softness of a set of instrumental parameters can be evaluated by the number of cations retained in the G-quadruplex in the gas phase (as illustrated for the LCT Premier on Figure 30).

Several charge states are usually observed. The instrumental conditions depend on the charge state of the ions (because the acceleration of the ions is proportional to their charge). The instrument is tuned for one charge state, generally the most intense one. In our case, nucleic acids of about 20 to 26 bases show a major peak intensity for the 5- charge state (which is also the charge state shown for most of the figures in this document). Zooms on the 5- charge state are also represented on the right hand side of Figure 30. The source voltage clearly affects the preservation of ammonium ions in the gas phase. On the right, a schematic representation of the gas-phase unfolding of the complex shows the loss of ammonium (red and green balls) as a function of the voltage. Further activation leads to fragmentation into two monomers.



Figure 30. Effect of source voltages on the ammonium cation preservation in the gas-phase. ESI-MS of 10 μ M dG₄T₄G₄ in 100 mM NH₄OAc from A to D increasing sample cone voltage (30, 50, 70 and 90 V). On the left, the full scale spectrum is represented, and on the right, the zooms on the dimer D⁵⁻ peak.

Once the mass spectrometer is tuned in the softest conditions possible as displayed in Figure 30A, the voltages and pressures can be used for similar complexes. The mass spectra are typically recorded in at least two voltage conditions to check for any activation-related effects.^{228,232}

III.1.4 ION MOBILITY SPECTROMETRY (IMS)

IMS allows the separation of the ions according to their shape. How to obtain experimentally the CCS value is explained in this section. The CCS values of biomolecular complexes are determined in Chapter VI.

Under the effect of a uniform linear electric field, ions will pass through a gas-filled cell. In this cell, the ions are accelerated until their velocity is constant (v_d), when the friction forces counterbalance the electric force:

$$v_d = K \times E$$

Equation 2

Where *K* is the mobility and *E* the electric field.

The electric field is the applied voltage *V* divided by the distance, here, the length of the tube *L*:

$$E = \frac{V}{L}$$

Equation 3

The reported parameter is the reduced mobility K_0 , the mobility normalized at $P_0 = 760 Torr$ and $T_0 = 273.15 K$:

$$K_0 = K \frac{P}{P_0} \frac{T_0}{T}$$

Equation 4

In those conditions, the drift time t_d depends on its mobility K and the length of the cell L:

$$t_d = \frac{L}{v_d} = \frac{L}{K \times E} = \frac{L \times P \times T_0}{K_0 \times P_0 \times T \times E} = \frac{L^2 \times P \times T_0}{K_0 \times P_0 \times T \times V}$$

Equation 5

Rearranging this equation gives the reduced mobility, K_0 :

$$K_0 = \frac{L^2}{V} \frac{1}{t_d} \frac{P}{760} \frac{273.15}{T}$$

Equation 6

The reduced mobility of an ion drifting through a buffer gas can be related to its collision cross section using the gas kinetic theory. This is the Mason-Schamp^{233,234} equation:

$$K_0 = \frac{3}{16} \sqrt{\frac{2\pi}{k_B T}} \sqrt{\frac{1}{M} + \frac{1}{m}} \times \frac{ze}{N_0} \times \frac{1}{\Omega_{avg}}$$

Equation 7

This equation involves k_B the Boltzmann constant, T the temperature, the masses of the analyte M and of the gas m, the charge in Coulomb ze, the gas number density at P_0 and T_0 ($N_0 = 2.687 \times 10^{25} m^{-3}$) and Ω_{avg} , the orientationally averaged momentum transfer collision integral, also called the collision cross section, CCS.

In practice, the arrival time t_a is measured instead of the drift time. The arrival time is the sum of the drift time t_a and the time spend by the ions elsewhere in the mass spectrometer (after the drift tube), t_0 .

$$t_a = t_d + t_0$$

Equation 8

Combining Equations 6 and 8 Equation 8 gives the following relation.

$$t_a = \frac{L^2}{V} \frac{1}{K_0} \frac{P}{760} \frac{273.15}{T} + t_0$$

Equation 9

To obtain the CCS, the IMS experiment is recorded at several voltages *V* and plotting t_a as a function of either $\frac{1}{V}$ or $\frac{P}{V}$ giving access to K_0 (via the slope) and therefore to Ω_{avg} . In our group, the drift tube of the instrument is 78.1 cm and we work with helium pressures of 3.89 Torrs at 293 K. Usually the regression using five different voltages between 390 and 790V is used.

III.1.5 DATA TREATMENT

NONSPECIFIC ADDUCTS

In order to distinguish potassium ions specifically bound to the G-quadruplex structure, we applied a method described by Klassen,²³⁵ graphically summarized in Figure 31. Briefly, in a set of conditions (ionic strength and KCl concentration) the nonspecific adducts distribution is given by the distribution obtained in the same conditions for a non-G-quadruplex forming sequence that contains the same number of bases and of guanines. Generally, an increasing amount and number of adducts is observed with decreasing charge state. 1, 2 and 3 nonspecific adducts are taken into account for the 6, 5 and 4- charge states, respectively.

Using the equations below (example for 2 nonspecific adducts, 5- charge state) the specific K⁺ distribution for each sequence in each condition is recalculated:

$$\begin{split} I_{0K} &= i0K * (1 + x + y) & \text{Equation 10} \\ I_{1K} &= (i1K - i0K * x) * (1 + x + y) & \text{Equation 11} \\ I_{2K} &= (i2K - i0K * y - (i1K - i0K * x) * x) * (1 + x + y) & \text{Equation 12} \\ I_{3K} &= (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * (1 + x + y) & \text{Equation 13} \\ I_{4K} &= (i4K - (i2K - i0K * y - (i1K - i0K * x) * x) * y & - (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * (1 + x + y) & \text{Equation 14} \\ I_{5K} &= (i5K - (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * y & - (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * y \\ &- (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * y & - (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y + (i2K - i0K * y - (i1K - i0K * x) * x) * x) \\ &+ (1 + x + y) & (11K - i0K * x) * y + (11K - i0K * x) * y + (11K - i0K * x) * y + (11K - i0K * x) * x) * x) \\ &+ (11K - i0K * x) &+ (11K - i0K * x) * y + (11K - i0K * x) * x) &+ (11K - i0K * x) * x) \\ &+ (11K - i0K * x) &+ (11K - i0K * x) * y + (11K - i0K * x) * x) &+ (11K - i0K * x) * x) &+ (11K - i0K * x) &+ (11K$$

Equation 15

Where I_{0K} , I_{1K} , I_{2K} , I_{3K} , I_{4K} and I_{5K} reflect the specific K⁺ intensity distribution. *i*0K, *i*1K, *i*2K, *i*3K, *i*4K and *i*5K are the integrals directly obtained from the MS spectrum for the 0, 1, 2, 3, 4 and 5 K⁺ stoichiometries respectively.

x and y are the values of the normalized integrals of the 1 and 2 nonspecific K^+ :

$$x = \frac{Ref_{1K+}}{Ref_0} \qquad \& \qquad y = \frac{Ref_{2K+}}{Ref_0}$$

Equations 16





The subtraction of the contribution of nonspecific adducts gives comparable results from different charge states. Consequently, each charge state can be treated separately, then the results of recalculated concentrations of each specific K⁺ stoichiometry can be combined, and the error bar can be calculated as the standard deviation of the average value.

In practice, the data treatment is relatively simple (and recently automated in our group). The main challenge is to find the right reference sequence. For G-quadruplexes and potassium, any sequence similar in composition compared with the sequence of interest can be used. We

systematically use reference sequences with the same number of guanines and bases as the sample.

RESPONSE FACTORS

The desolvation and ionization occurring in electrospray do not necessarily occur with the same efficiency for each species in solution: some of them will have a better response in electrospray than others. Before converting the relative intensities into relative concentrations, we also considered the possibility that all complexes may not be electrosprayed with the same response. To verify that the response factors are not influenced by the addition of a compound in the solutions we used an internal standard (dT₆) as reference.²²⁷ The response factor is the proportionality factor between the concentration and the mass spectrometry intensity:

 $I_A = R_A[A]$

Equation 17

 I_A is the intensity of the species *A*. It is linked to its concentration [*A*] thanks to its response factor R_A . If the mass balance as a function of a variable (KCl concentration, ligand concentration, time...) do not deviate systematically, the response factors are equal and the detected signals are directly proportional to the concentrations. In practice, the relative intensity of the detected complexes and dT₆ is used as indicator. Most of the time, this treatment was not necessary, presumably because the masses of the electrosprayed complexes were only weakly affected (either by the binding of K⁺ or the ligand). However, sometimes it might be very important to consider that all the species might not fly equivalently, mostly when multimers are formed from monomers.⁷⁸

$K_{\rm D}$ determination, kinetic model evaluation and signal deconvolution

After the previously described corrections (nonspecific adducts and response factors), it is possible to obtain the concentration of the different species as a function of the concentration (ligand or K⁺) or time. Those data can, then, be evaluated using the DynaFit software (4.05.103, BioKin Ltd Watertown, MA).²³⁶

Equilibrium constants determination. The input is the concentration in each complex as a function of the ligand (or K⁺) concentration. By defining the equilibrium constants using chemical equations, the software fits the data by iteration and the output is the value of the equilibrium constant(s) and the fit(s). An example of the script is given hereafter:

```
[task]
   data = equilibrium
   task = fit
[mechanism]
   DNA + L <==> DNAL1 : K1 dissoc
DNAL1 + L <==> DNAL2 : K2 dissoc
[constants]
   K1 = 1 ?
   K2 = 1 ?
[concentrations]
   DNA = 10
[data]
   variable L
   file ..\0.txt | response DNA = 1
file ..\0+1L.txt | response DNAL1 = 1
   file ..\0+2L.txt | response DNAL2 = 1
[output]
   directory ..\Output-Kd
[end]
```

In the task section, the types of data (equilibrium data) are indicated. The mechanism section is the place where the equilibrium constants are defined. Writing the equilibrium as chemical equations allows the software to define the equilibrium constants. In the constants section, are written the initial parameters of the constants (starting point of the iteration process). A question mark after the value means that this value will be adjusted by the software. In concentrations, the starting total concentration in DNA is indicated. In data, the variable (in this case, the ligand concentration) and the data are indicated. The data files contain two columns, one with the total ligand concentration and another with the concentration in the species defined right after the "response" word. Note that the responses are equal to 1 because the intensities obtained from the MS data were previously converted into concentrations.

Kinetics experiments. In this case, the data are as a function of time instead of a concentration.

```
[task]
  data = progress
  task = fit
[mechanism]
  DNA + K <==> DNA.K1 : ka kb
  DNA + K + K <==> DNA.K2 : kx ky
[constants]
   ka = 0.1 ?, kb = 0.1 ?
  kx = 0.1 ?, ky = 0.1 ?
[concentrations]
  DNA = 10
[progress]
  global
  file ..\OK-K200.txt | conc K = 200 | response DNA = 1
   file ...\1K-K200.txt | conc K = 200 | response DNA.K1 = 1
   file ... \2K-K200.txt | conc K = 200 | response DNA.K2 = 1
   file ...\OK-K500.txt | conc K = 500 | response DNA = 1
   file ...\1K-K500.txt | conc K = 500 | response DNA.K1 = 1
   file ...\2K-K500.txt | conc K = 500 | response DNA.K2 = 1
   file ...\OK-K1000.txt | conc K = 1000 | response DNA = 1
   file ...\1K-K1000.txt | conc K = 1000 | response DNA.K1 = 1
   file ...\2K-K1000.txt | conc K = 1000 | response DNA.K2 = 1
[output]
   directory ... \Output
[end]
```

Inside the task section, the keyword "progress" is written instead of "equilibrium". The chosen mechanism is written, in this example, the formation of the 1-K⁺ and 2-K⁺ complexes are on different pathways (parallel reactions). The differential kinetic equations for this particular case are the following:

$$\frac{d[DNA]}{dt} = -k_a[DNA][K] + k_b[DNA.K1] - k_x[DNA][K]^2 + k_y[DNA.K2]$$
Equation 18
$$\frac{d[DNA.K1]}{dt} = k_a[DNA][K] - k_b[DNA.K1]$$
Equation 19
$$\frac{d[DNA.K2]}{dt} = k_x[DNA][K]^2 - k_y[DNA.K2]$$

Equation 20

Those equations are used by the software to produce the fit. In this case, we fitted together the kinetics experiments done at different K^+ concentrations (200, 500 and 1000 μ M). The

same equations are used and the software fit the data to obtain the same kinetics parameters. The value of the total potassium concentration is taken into consideration in each case for the global fit.

In this case, the responses are equal to 1 because the intensities obtained from the MS data were previously converted into concentrations.

Several models can be evaluated at the same time and the goodness of the fit is compared automatically by calculating the BIC, Bayesian Information Criterion and the AIC, Akaike Information Criterion. Those criteria penalize models as a function of their number of parameters: the goal is to obtain the best fit with the lowest number of parameters.

Signal deconvolution. From the known concentrations of each species, it is possible to deconvolute signals coming from other techniques (for example spectroscopic techniques). From the measured total signal intensity, the response of each species is calculated. The data are fitted according to the following equation:

$$\begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ \cdots \\ S_n \end{pmatrix} = R_x \times \begin{pmatrix} C_0^x \\ C_1^x \\ C_2^x \\ \cdots \\ C_n^x \end{pmatrix} + R_y \times \begin{pmatrix} C_0^y \\ C_1^y \\ C_2^y \\ \cdots \\ C_n^y \end{pmatrix} + \cdots$$

Equation 21

Where S_0 to S_n is the recorded signal intensity as a function of ligand (or K⁺) concentration or of time with *n* the total number of points in the titration or kinetics. R_x , R_y ,... are the responses of the species *x*, *y*,..., and C_0^x to C_n^x , C_0^y to C_n^y ,... are vector matrices of the concentrations of the species *x*, *y*,... as a function of concentration or time.

For spectroscopic data, calculating the responses as a function of the wavelength gives the deconvoluted spectra, which can be reconstructed by linking the dots with a spline curve. In this work, this procedure was applied to CD data (titrations and kinetics).

In the script, the question mark next to the equilibrium constant value is removed meaning this value is known. The unknowns are in the responses section. In the data file, two columns are present: the total concentration of the variable component (here K⁺).

```
[task]
  data = equilibrium
  task = fit
[mechanism]
  DNA + K <==> DNAK1 : K1 dissoc
  DNAK1 + K <==> DNAK2 : K2 dissoc
[constants]
  K1 = 271
  K2 = 6500
[concentrations]
  DNA = 10
[responses]
  DNA = 1 ?, DNAK1 = 1 ?, DNAK2 = 1 ?
[data]
  variable K
  file ..\Titr-Signal.txt
[output]
  directory ..\Output
[end]
```

The main difference between this deconvolution method and singular value decomposition (SVD)¹⁸⁰ is that our approach determines first the number of species and their relative concentrations experimentally thanks to the mass spectrometry signals, and then the response of each species with DynaFit, whereas SVD determines the number of species and their response altogether.

III.2. CIRCULAR DICHROISM

G-quadruplexes are commonly classified into three groups according to their strand orientations and base stacking, which present specific CD signatures:

Group-I: Parallel G-quadruplexes. Parallel G-quadruplexes have all strands pointing in the same direction. All guanine configurations are *anti* according to their glycosidic bonds, leading purely to *anti-anti* stacking. The CD spectra of group-I G-quadruplexes show a negative peak at 240 nm and a positive one at 260 nm.

Group-II: Hybrid G-quadruplexes. Hybrid G-quadruplexes have three strands pointing in one direction and the fourth one pointing in the opposite direction. Linked to this strand orientation, the guanines present both types of stacks: homo- (*anti-anti* and *syn-syn*) and hetero-stacking (*anti-syn* and *syn-anti*). The CD signature of such G-quadruplexes presents a negative peak at 240 nm and two positive peaks at 260 and 290 nm.

Group-III: Antiparallel G-quadruplexes. Antiparallel G-quadruplexes have two strands pointing one direction and two in the opposite one. Those G-quadruplexes have guanines configurations alternating between *anti* and *syn*, leading to *anti-syn* and *syn-anti* stacking. For antiparallel G-quadruplexes, two positive CD peaks are found at 240 and 290 nm and one negative at 260 nm. In this group-III, some differences can be observed in the CD signature for particular structures containing only two G-quartets, for example, the *Giardia* telomeric G-quadruplex.^{170,237}

Typical CD spectra of the three groups are represented in Figure 32. However, particularities, such as bases pairs, additional base stacking, G-triplet... may influence the recorded spectrum.



Figure 32. Spectra recorded in house of typical G-quadruplexes of the three groups in conditions in which their folding is known. In green, the parallel Pu24, d(TGAGGGTGGIGAGGGTGGGGAAGG), in 100 mM K⁺ (PDB ID: 2A5P)¹⁷; in red, the hybrid-1 human telomeric 24TTG, d(TT(GGGTTA)₃GGGA), in 100 mM K⁺ (PDB ID: 2GKU)⁸⁸ and in blue, the antiparallel human telomeric 22AG, d(A(GGGTTA)₃GGG), in 100 mM Na⁺ (PDB ID: 143D).⁸⁷

In practice, in solution, the sample is irradiated with linearly polarized light that can be seen as the linear combination of opposite circularly polarized light of equal amplitude and phase (Figure 33A). If the sample absorbs more the left or right polarized light, the detected light that passed through the sample is not linearly polarized anymore but elliptically polarized (Figure 33B). The θ angle is measured in millidegrees and corresponds to the ellipticity. It can be transformed into molar circular-dichroic absorption $\Delta \varepsilon$ using the following equation:

$$\Delta \varepsilon = \theta / (32980 \times C \times l)$$

Equation 22

Where *C* and *l* are, respectively, the compound concentration in the solution and the optical path length. In this work, θ was systematically converted to $\Delta \varepsilon$ in $(cm^2 mmol^{-1})$. Scanning as a function of the wavelength gives the CD spectrum.



Figure 33. A, the polarization of the light before the sample absorbs and B, after partial absorption of the right polarized light. In practice, the θ angle is measured. This angle is different from zero when the light is elliptically polarized.

III.3. THERMAL DENATURATION EXPERIMENTS

Thermal denaturation experiments are monitored by any technique that gives a change in signal proportionally to the folded/unfolded ratio. A commonly used technique is UV-absorbance, typically at 295 nm which shows the greatest change with the temperature, as illustrated in Figure 34A.

For 2-state 1-step reactions (G-quadruplex versus unfolded species), the temperature-induced unfolding curve is a sigmoid (Figure 34B). From this sigmoid, the position of the inflexion point correspond to the melting point, 50% of the species in solution are folded/unfolded. It is possible to extrapolate the equilibrium constant at a given temperature because the fraction folded/unfolded is known as a function of temperature.¹⁷⁵ The thermodynamic parameters ΔH and ΔS can be obtained from the steepness of the sigmoid curve around the inflexion point.¹⁷⁶



Figure 34. UV-melting experiment of 24TTG in 100 mM TMAA and 1 mM KCl. (A) The absorbance spectra at 4 and 85 °C are slightly different and (B) the absorbance at 295 nm which shows the greatest change as a function of temperature. The melting temperature (41 °C) is indicated with a red dot.

IV. NATIVE MS FROM POTASSIUM SOLUTIONS

IV.1. INTRODUCTION

The first mass spectrum of a G-quadruplex was published in 1993 by Richard D. Smith's group²³⁸, using an electrospray ionization source in negative ion mode. When 10 mM sodium phosphate buffer and 0.1 mM ethylenediaminetetraacetic acid solution were used, the mass of ions corresponding to a complex made of four strands of d(CGCG₄GCG) was detected. This peak was attributed to a G-quadruplex. Because solutions containing high concentrations of sodium were injected, the complexes are detected with many adducts: in their case, up to 16 Na⁺ adducts as compared to the three expected inter-G-quartet cations. Reputedly MS-incompatible buffers were used in this study. Indeed, the sodium phosphate buffer tends to suppress the signal. To bypass this difficulty, the authors used very high DNA concentration (approximately 1 mM).

To be able to study biomolecule complexes, decreasing the electrostatic repulsion between charges of the same sign, in order to allow the folding of the biomolecule, is achieved by increasing the ionic strength. To keep a physiological ionic strength (~100-150 mM range), the ammonium acetate (NH₄OAc) buffer is commonly used. During the desolvation process the ammonium cations become neutral ammonia by losing protons and acetate anions become acetic acids capturing protons.^{231,239-241}

However, G-quadruplexes do not only require a certain ionic strength but also cations to fill their binding sites, between the G-quartets. Ammonium can bind and induce the folding of the G-quadruplexes.²³¹ However, the nature of the salt, its concentration and the experimental conditions can influence the folding.²⁴² In the cell nucleus, the salt concentrations are relatively high: about 10 mM of Na⁺ and 140 mM K⁺.⁷³ However, in 100 mM KCl, and at low DNA concentration (micromolar range), only potassium chloride clusters can be observed by MS and the peaks of interest are masked (Figure 35). This salt is not MS-compatible because it is not "volatile".



Figure 35. Mass spectrum of 10 µM 23TAG, d(TA(GGGTTA)₃GGG), in 100 mM KCl. No ion of interest is detected, only KCl clusters.

Obtaining mass spectra of nucleic acids (in particular G-quadruplexes) in the presence of relevant salt concentration and nature is of great interest but is still a challenge. Several groups worked on this problem. One of the easiest ways to obtain mass spectra of G-quadruplexes folded in potassium conditions is to prepare the solution in presence of high KCI concentration before precipitation and re-dissolution in pure water.^{243,244} However, the method offers poor control on the final salt concentration and is therefore not suitable for biophysics quantitative analysis. Others have used gas additive to selectively eliminate the nonspecific adducts already desolvated in the gas phase.²⁴⁵ This technique was applied on DNA and RNA complexes in a hexapole just before the mass analyzer.²⁴⁶ Unfortunately, this requires a specific instrumentation and is probably not applicable on a daily basis.

In this work, we combined two main ideas. The first one is the use of trimethylammonium acetate (TMAA). Indeed, alkyl-ammonium and -amine such as TMAA, triethylammonium acetate (TEAA) or triethylamine (TEA) were used by several groups of mass spectrometry researchers to remove nonspecific Na⁺ or K⁺ adducts.^{247–249} Longer alkyl chains on the ammonium allow one to obtain mass spectra with fewer adducts on the peaks of interest.

Therefore, for this purpose, the ammonium acetate is the less efficient buffer, increasing the length of the alkyl chain is the first step to get cleaner spectra. The second idea comes from a study of the potassium binding to G-quadruplexes,²⁵⁰ in which the authors used the tetrabutylammonium phosphate buffer to fix the ionic strength. The tetrabutylammonium cation is too bulky to fit between the quartets and to induce the folding of G-quadruplexes. In this work we showed that TMA⁺ is also too big to induce the folding of G-quadruplexes but a 100 mM TMAA solution supplemented with up to 1 mM KCl gave good signal of the G-quadruplexes containing potassium cations.

IV.2. ARTICLE 1

Native Electrospray Mass Spectrometry of DNA G-Quadruplexes in Potassium Solution

A. Marchand & V. Gabelica

Journal of the American Society for Mass Spectrometry **2014**, 25, 1146–1154.

For supplementary information, see Annex I.



FOCUS: MASS SPECTROMETRY AND DNA DAMAGE: RESEARCH ARTICLE

Native Electrospray Mass Spectrometry of DNA G-Quadruplexes in Potassium Solution

Adrien Marchand,^{1,2} Valerie Gabelica^{1,2}

¹University Bordeaux, IECB, ARNA Laboratory, F-33600, Pessac, France ²INSERM, U869, ARNA Laboratory, F-33000, Bordeaux, France



Abstract. A commonly used electrolyte in electrospray mass spectrometry (ESI-MS) of biomolecules is ammonium acetate (NH₄OAc). Although some nucleic acid structures such as duplexes require only proper physiological ionic strength (whatever the monovalent ions) to be properly folded in ESI-MS conditions, the folding of some other nucleic acid structures such as DNA G-quadruplexes also depends on direct binding of specific cations. Here, we developed ESI-MS compatible conditions that allow one to observe DNA G-quaduplexes with \vec{K} ions specifically bound between G-quartets. NH₄OAc was replaced with trimethylammonium acetate (TMAA), at concentrations up to 150 mM to provide physiological ionic strength, and the solution was doped with KCI at concentra-

tions up to 1 mM. The trimethylammonium ion is too large to coordinate between G-quartets, where only K ⁺ ions bind. Compared with the equivalent NH₄OAc/KCI mixtures, the TMAA/KCI mixtures provide cleaner spectra by suppressing the nonspecific adducts, and favor the formation of similar stacking arrangements as in 100 mM KCI (physiologically relevant cation) for the polymorphic human telomeric DNA G-quadruplexes. This new sample preparation method can be exploited to determine the number of potassium binding sites in new sequences, to screen ligand binding to the structures favored in potassium, and to transfer potassium-bound G-quadruplexes to the mass spectrometer for gas-phase structural probing, as illustrated herein with ion mobility spectrometry experiments.

Key words: DNA structures, G-quadruplexes, Native, Mass spectrometry, Ion mobility, Potassium, Trimethylammonium, Adducts, Cation binding

Received: 28 January 2014/Revised: 11 March 2014/Accepted: 17 March 2014

Introduction

G-quadruplexes are non-classical DNA or RNA structures consisting of stacked guanine quartets (G-quartets) made by Hoogsteen hydrogen-bonding between guanines (Figure 1a). Potassium, sodium, or ammonium cations also stabilize the structure by coordinating to the O6 oxygen of guanines. Owing to their size, ammonium, and potassium ions are located in-between two G-quartets and, therefore, coordinate with eight O6 atoms, whereas smaller sodium ions can also access the center of G-quartets. Gquadruplexes can be found in vitro in the promoter regions of oncogenes [1–4] or in the human telomeric region [5]. Some recent work from Balasubramanian's group showed that G-quadruplex DNA structures are present in human cells [6]. Because of their supposed role in cancer, G-quadruplexes are often studied as targets for anti-cancer drugs [7–10].

The most studied G-quadruplex forming sequence is probably the human telomeric sequence, consisting of $(TAGGGT)_n$ repeats. Human telomeric G-quadruplexes present a high level of polymorphism in vitro depending on their core sequence, flanking bases, salt conditions, the nature of the sugar (DNA versus RNA),... [11]. For example, parallel G-quadruplexes have been reported by X-ray crystallography studies for $d(A(GGGTTA)_3GGG)$ in K+conditions (Figure 1b) [12]. In contrast, by NMR in solution, an antiparallel basket-type form with three Gquartets was observed in Na⁺ conditions for $d(A(GGGTTA)_3GGG)$ (Figure 1c) [13], and one with only two G-quartets in K⁺ conditions for $d((GGGTTA)_3GGGT)$ [14]. In K⁺ solution, a hybrid form can also be adopted by sequences $d(TT(GGGTTA)_3GGGA)$ (Figure 1d) [15],

Electronic supplementary material The online version of this article (doi:10.1007/s13361-014-0890-3) contains supplementary material, which is available to authorized users.

Correspondence to: Valerie Gabelica; e-mail: v.gabelica@iecb.u-bordeaux.fr



Figure 1. (a) A guanine quartet (G-quartet). (b) X-ray crystal structure of the parallel G-quadruplex form of $d(A(GGGTTA)_3GGG)$ in K⁺ conditions [12] (PDB ID: 1KF1). (c) NMR structure of the antiparallel G-quadruplex form of $d(A(GGGTTA)_3GGG)$ in Na⁺ solution [13] (PDB ID: 143D). (d) NMR structure of the hybrid G-quadruplex form of $d(TT(GGGTTA)_3GGGA)$ in K⁺ solution [15] (PDB ID: 2GKU). The schemes underline strand directionality, and base-sugar conformations (light grey for *anti*, dark grey for *syn*)

d(TA(GGGTTA)₃GGG) [16], or d(TA(GGGTTA)₃GGGTT) [17].

Taking this high level of polymorphism into account is important to interpret experimental results obtained by different techniques, which sometimes impose using different solution conditions. G-quadruplexes are indeed often studied in different buffer conditions depending on the assay. Phosphate and Tris/HCl buffers are often used to obtain neutral pH, and are supplemented by sodium chloride or potassium chloride, usually at a concentration around 100 mM. When screening ligands towards Gquadruplexes, a sensible approach to acknowledge this polymorphism is to test ligand binding to different sequences and under a range of solution conditions. For example, ligand binding assays are often compared in Na^+ and in K^+ to probe whether a ligand has any preference for antiparallel forms (in Na⁺) or more parallel or hybrid forms (in K⁺). In indirect ligand binding assays such as changes in melting temperature, sometimes the Kconcentration is decreased to 10 or even 1 mM in order to destabilize the G-quadruplex, so as to be able to measure the effect of the ligand on the stability [18].

Mass spectrometry (MS)-based G-quadruplex drug binding assays offer several advantages compared with other traditional methods [19] because it allows direct stoichiometry observation and quantification of each peak [20]. Nevertheless, none of the buffers mentioned above are compatible with native electrospray ionization because they are nonvolatile. To bypass this limitation, ammonium acetate (NH4OAc) is used to fix the ionic strength. When Gquadruplex forming sequences are electrosprayed in NH₄OAc, specific ammonium cations are retained in the complexes. They are able to coordinate between the G-quartets [21, 22]. G-quadruplex structures made in ammonium acetate solution are, however, less stable and more polymorphic than those formed in potassium chloride solution [20].

Spraying nucleic acids in the presence of physiological cations is an important challenge that has been addressed by several groups, either for G-quadruplexes (K^+) or for RNA structures (Mg^{2+}) . Previously, Evans et al. [23] and the team of Brodbelt [24] reported tetramolecular G-quadruplexes electrosprayed with potassium adducts. In their respective works, the Gquadruplexes are preformed into potassium conditions before being partially desalted by ethanol precipitation or filtering, and resuspended in ammonium acetate for ESI-MS analysis. However, although this worked well for tetramolecular G-quadruplexes that are parallel-stranded whatever the cation, for polymorphic G-quadruplexes, ammonium acetate replacement will also displace the conformational equilibria. McLuckey's group exploited ionmolecule reactions, and used vaporized acetic or formic acid inside the desolvation cell to displace metal counter-ions that stick to the phosphate groups [25]. This method allowed keeping native species in solution because the interaction with the acid is made in the gas phase and not in the solution phase. However, gas-phase structures may be disrupted. Fabris' group exploited ion-ion reactions in FTICR-MS to clean up sodium and magnesium adducts from diverse DNA and RNA structures [26]. However these approaches are not practical on an everyday basis and require specialized instrumentation.

Here, we show that trimethylammonium acetate (TMAA) can advantageously replace ammonium acetate (NH₄OAc) to carry out ESI-MS experiments on Gquadruplexes in the presence of K⁺. The key idea is that the trimethylammonium cation is too bulky to coordinate between quartets, and that only K⁺ can fill the inter-quartet sites. A similar approach was used by Gray and Chaires for solution spectroscopy-monitored G-quadruplex titrations by KCl in the presence of tetrabutylammonium phosphate [27]. Trialkylamines or trialkylammonium buffers are commonly used in ESI-MS. In 1994, Bischoff's group compared the efficiency of triethylamine (TEA), trimethylamine (TMA), and NH₄OAc to remove sodium adducts on oligomers from 12- to 132-mer oligonucleotides [28]. They showed that ammonium acetate is the least efficient in reducing the amount of adducts. Ostrander's group [29] proposed the triethylammonium acetate (TEAA) additive to clean up their mass spectra from sodium and potassium adducts. In their case, they fixed the pH to 7 thanks to imidazole. Lemaire et al. [30] also reported that triethylammonium bicarbonate is useful to study proteins and proteins complexes under non-denaturing conditions. They showed a charge state displacement to lower charges induced by this additive, and were able to distinguish NADH adducts on the dimer made by two alcohol dehydrogenase. The present study shows for the first time the advantage of using TMAA for studying nucleic acids by electrospray mass spectrometry in native conditions (neutral pH, physiological ionic strength, and with sufficient amounts of KCl to form presumably native G-quadruplex structures).

Experimental

Materials

Oligonucleotides were purchased lyophilized and RP-cartridge purified from Eurogentec (Seraing, Belgium). In this work, the following nomenclature will be used: the hybrid G-quadruplex forming sequence $d(TT(GGGTTA)_3GGGA) =$ 24TTG; the hybrid G-quadruplex forming sequence d(TA(GGGTTA)₃GGG)=23TAG; a 22-mer control oligonucleotide with twelve guanines but incapable of forming a G-quadruplex: d(GGG-ATG-CGA-CAG-AGA-GGA-CGG-G)=22non-G4; a tetramolecular parallel G-quadruplex made by four d(TGGGGT) strands=TG₄T; and a bimolecular Gquadruplex made by two d(GGGGTTTTGGGGG) strands= G₄T₄G₄. Stock solutions were first prepared at 0.5 to 1 mM in water. The stock concentrations were determined by UV absorbance at 260 nm, with extinction coefficients calculated using Cavaluzzi-Borer Correction [31], and measured on an Agilent Cary 100 UV spectrophotometer. Water was nuclease-free grade from Ambion (Applied Biosystems, Lennik, Belgium). Ammonium acetate (NH4OAc, Ultra for Molecular Biology, Fluka), trimethylammonium acetate (TMAA, Ultra for UPLC, Fluka), potassium (KCl, **9**99.999%, Sigma) and sodium chloride (NaCl, **9**99.999%, Sigma) were purchased from Sigma-Aldrich (Saint-Quentin

Fallavier, France). The solutions were prepared at room temperature (no annealing), 2 to 3 h before the circular dichroism (CD) or MS measurements were performed. Some measurements were repeated after 1 wk, and no change in CD or MS results was observed. The pH of TMAA and DNA-containing TMAA solutions was in the 6.7–6.8 range. TMAA solution was therefore always used as received without complementary pH-adjustment.

Mass Spectrometry (MS)

All mass spectra were obtained in negative ion mode with injection concentrations of 5 µM in G-quadruplexes and 5 μ M internal standard dT₆. Native electrospray mass spectra of oligonucleotides were obtained using several instruments, with similar results. (1) On an LCT Premier mass spectrometer (Waters, Manchester, UK), the ESI source voltage was set to 2.2 kV with a desolvation temperature of 60°C. Unless otherwise mentioned, the ion guide 1 voltage is 50 V. The source pressure was increased to 35 mbar. This pressure is measured using a Center Two probe (Oerlikon Leybold Vacuum, Cologne, Germany). The syringe injection flow was fixed to 200 µL/h. In all experiment 5 µM dT₆ (monoisotopic mass 1762.318 Da) was used as internal standard. (2) An Exactive ESI-Orbitrap mass spectrometer (Thermo Scientific, Bremen, Germany) was also used to monitor the titration of TG₄T by KCl. In this case, ESI spray voltage and capillary voltage are 2.75 kV and -20 V, respectively. Capillary temperature is set to 150°C. Tube lens and skimmer voltage are fixed to -180 V and -10 V. In order to help the desolvation process to occur, the HCD cell voltage is set to 10 eV and the cell pressure is 2.5×10^{-5} mbar.

Ion Mobility Spectrometry (IMS)

The collision cross sections of NH₄⁺-containing and K⁺containing G-quadruplexes were determined using a Synapt G2S mass spectrometer (Waters, Manchester, UK). The electrospray capillary was at 2.04 kV, the source and desolvation temperatures were 30°C and 40°C, respectively, and the sampling cone was at 28 V. The trap and transfer collision energies were 2 and 4 V, respectively. All these parameters ensured minimal fragmentation outside the IMS cell. The helium gas flow in the pre-IMS cell was 200 mL/min, and the nitrogen gas flow in the IMS cell was 90 mL/min. The IMS T-wave was operated at 40 V and 600 m/s. The CCS was calibrated following the procedure of Ruotolo et al. [32] using dT_6^{2-} (306 Å²), dT_6^{3-} (333 Å²), $[dTG_4T]_4^{4-}$ (730 Å²), $[dTG_4T]_4^{5-}$ (775 Å²) [33], and $[dTG_4T]_4^{6-}$ (795 Å²). These five collision cross sections were measured in a helium drift tube mass spectrometer [34]. For the telomeric Gquadruplexes, collision activation upon entrance in the IMS cell is promoted by changing the trap DC bias from 18 to 30 V. The helium cell DC is 30.2 V and the IMS DC entrance is 7.3 V.

Circular Dichroism (CD)

CD experiments were performed to obtain information on the DNA strands orientation. Experiments were run on a Jasco J-815 spectrophotometer using a quartz cell of 2 mm path length. All spectra shown here are the sum of three scans, acquired at 20°C with a scan speed of 50 nm/min and a 2 nm bandwidth. The CD data were normalized to molar circular-dichroic absorption ($\Delta \varepsilon$) based on DNA concentration using the following Equation,

$$\Delta \epsilon \frac{1}{32980 \text{ X } c \text{ X } l}$$
 õlþ

where θ is the CD ellipticity in milidegrees, *c* is DNA concentration in mol L⁻¹ (5×10⁻⁶ mol.L⁻¹), and *l* is the pathlength in cm (0.2 cm).

Results and Discussions

G-Quadruplexes from TMAA/KCl Solutions Show a Specific Number of K^+ Adducts Corresponding to Coordination Between the G-Quartets

The mass spectrometer is tuned so that when the ion guide 1 is 50 V, the native structure is kept in the gas phase, as tested with the sensitive G-quadruplex [dG₄T₄G₄]₂ in 100 mM NH₄OAc (Supporting information S1) [22]. Figure 2a shows a mass spectrum of the human telomeric sequence 23TAG in 100 mM NH₄OAc in native conditions (low ion guide 1 voltage: 50 V) (Figure 2a, top). G-quadruplexes chelating two ammonium cations constitute the most abundant species, but a distribution of adducts is nevertheless observed (from 0 up to 5 or 6 ammonium cations). If the structure was unknown, it would actually be difficult to conclude on the preferred ammonium binding stoichiometry, as there is no clear bias in the ammonium ion distribution at any voltage. When the ion guide 1 voltage is increased from 50 to 70 V (Figure 1a, bottom) the complex is disrupted, the ammonium ions are released as NH3 and the most abundant species becomes the free oligonucleotide.

In order to examine the effect of adding potassium to the solution, we doped the ammonium acetate solution by 1 mM potassium chloride (Figure 2b). In these conditions, the resulting mass spectra are noisy at low m/z range (not shown) due to clusters formation, and there are numerous adducts on the G-quadruplexes peaks. As only two specifically bound K⁺ ions are anticipated, additional adducts are therefore nonspecific external adducts. Therefore, in ammonium acetate solutions doped with KCl, it is not possible to deduce the number of specifically bound potassium ions in the G-quadruplexes. Besides, two series of peaks are observed at low source voltage, indicating that each inter-quartet binding site may be occupied either by NH₄⁺ or by K⁺.

To drive K^+ ions to occupy inter-quartet sites specifically, we used trimethylammonium acetate (TMAA) as volatile



Figure 2. ESI-MS of 5 μ M 23TAG into (a) 100 mM NH₄OAc, (b) 100 mM NH₄OAc+1 mM KCI, (c) 100 mM TMAA, and (d) 100 mM TMAA+1 mM KCI. The top mass spectra were recorded with an ion guide 1 voltage of 50 V and the bottom ones with 70 V. The annotations represent the number of adducts on the 5- charge state. The peak series with stars corresponds to mixed NH₄⁺/K⁺ adducts

buffer. ESI-MS recorded in TMAA alone show predominantly no TMA adduct (Figure 2c), in line with the fact that the trimethylammonium cation is too bulky to fit between Gquartets. At low ion guide 1 voltage (50 V, top) some TMA adducts can be observed but the shape of the distribution is typical of nonspecific adducts. This was confirmed with a control single strand incapable of forming a Gquadruplex in solution (Supporting Information S2). When the ion guide 1 voltage is set to 70 V (bottom) the only resulting peak is the DNA alone.

Figure 2d shows 23TAG electrosprayed in TMAA doped by 1 mM KCl. The most abundant species is the Gquadruplex chelating 2 potassium cations. This stoichiometry was expected because the structure adopted by 23TAG contains three G-quartets, and because K^+ specific binding sites are located between quartets, there are two potassium binding sites per G-quadruplex. When the voltage is increased to 70 V, the remaining trimethylammonium (TMA) adducts are expelled, but the K⁺ adduct distribution remains identical, and clearly biased towards the 2-K⁺ adduct. The monovalent ion binding stoichiometry can be obtained with much more confidence in the TMAA/KCl conditions, and evidence of intramolecular G-quadruplex formation can be concluded with more confidence than in pure NH₄OAc.

TMAA Suppresses Nonspecific Alkali Adducts More Efficiently than NH₄OAc, Including in Native Conditions

The comparison between Figure 2b and d illustrates that using TMAA instead of NH₄OAc as electrolyte at pH=7 allows the elimination of undesirable nonspecific K^+ adducts. This may be due to the fact that the trimethylammonium ion has greater affinity than ammonium ion for the external nonspecific binding sites, so that it is more efficient in competing with alkali cations. To test that hypothesis, we investigated the effect of the TMAA concentration on potassium and sodium adducts in the 24TTG human telomeric sequence (Figure 3). For all tested TMAA concentration (50 mM, Figure 3a; 100 mM, Figure 3b, and 150 mM, Figure 3c) in the case of potassium the most abundant species is the G-quadruplex with two potassium cations. What changes is the number of additional, nonspecific K^+ adducts. When TMAA concentration is increased, nonspecific adducts are increasingly suppressed to give almost exclusively the G-quadruplex with 2 potassium ions.

In the case of sodium (Figure 3d, e, and f), the same effect was observed. However the final distribution is biased towards zero Na⁺ adduct, suggesting that no G-quadruplex is formed. This was confirmed by CD spectroscopy (Figure 4 and section below). The nonspecific adduct suppression effect was also confirmed with other DNA structures (see Supporting Information S3). The maximum KCl or NaCl concentration amenable to electrospray in 100 mM TMAA is about 1 mM. Above this concentration the signal/noise ratio decreases significantly and the formation of clusters at lower m/z becomes predominant. This concentration is sufficient for potassium, but not sodium, to induce the formation of the G-quadruplexes studied here. The lesser affinity of the G-quadruplexes for sodium than for potassium is well documented [35]. The remaining portion of the present paper focuses on G-quadruplexes formed in potassium conditions.

Besides displacing nonspecific adducts, increasing the TMAA concentration has also another effect: the charge states are displaced to lower values (Supporting Information S4). The lower charge states presented more numerous nonspecific adducts than the 5- presented here. Keeping the



Figure 3. ESI-MS of 5 μ M 24TTG into increasing amount of TMAA. Left: 1 mM KCl with (a) 50 mM, (b) 100 mM, and (c) 150 mM TMA, respectively. Right: 1 mM NaCl with (d) 50 mM, (e) 100 mM, and (f) 150 mM TMAA, respectively. Mass spectra were recorded with an ion guide 1 voltage of 70 V. The annotations represent the potassium (a)-(c) or sodium (d)-(f) adducts on the 5- charge state



Figure 4. CD spectra of 5 μ M 24TTG (a) and 23TAG (b) in different solution conditions (see also annotation in panel (b): 20 mM KH₂PO₄/K₂HPO₄+70 mM KCI at pH 7 (purple), in 100 mM NH₄OAc (blue), in 100 mM NH₄OAc+1 mM KCI (green), in 100 mM TMAA (grey), in 100 mM TMAA+1 mM KCI (red), or in 100 mM TMAA+1 mM NaCI (black)

TMAA concentration at 100 mM allowed a good compromise to determine stoichiometry thanks to 5- charge state without losing too much signal. Triethylammonium acetate (TEAA) was also tested because it was reported to displace alkali ions even better than TMAA [28]. However, TEAA displaced totally the charge state 5- to 4-, and the adduct suppression effect of TEAA was less efficient on the 4charge state than TMAA on the 5- charge state (data not shown). TMAA is the best compromise to obtain clean ESI mass spectra at low ion internal energies, which is important to preserve the gas-phase structure (see below).

The G-Quadruplex Structures Formed in 100 mM TMAA+1 mM KCl Are Similar to Those Formed in 100 mM KCl/Phosphate Buffer

To check whether the structures of the G-quadruplexes formed in TMAA/KCl are closer to those formed in pure NH4OAc or in pure KCl, circular dichroism (CD) spectroscopy was used. This technique is sensitive to the orientation of the strands in G-quadruplexes studies [36]: a maximum at 260 nm and a minimum at 240 nm indicate the presence of parallel G-quadruplex with homo-stacking between all-*anti* guanine quartets. A maximum at 290 nm and a minimum at 260 nm indicate the presence of antiparallel G-quadruplex with alternate *syn-anti* guanine stacking motifs. Hybrid Gquadruplexes will show two maximum peaks at 260 and 290 nm, as they possess both types of stacking. A mixture of structures in solution will result in a CD spectrum containing the weighted average of each contribution.

In Figure 4, we compare the CD spectra of the human telomeric DNA sequences in ammonium acetate, in phosphate buffer conditions (20 mM KH₂PO₄/K₂HPO₄+70 mM KCl, the buffer in which the NMR structures were determined), and in the different solution compositions tested here by ESI-MS (24TTG, Figure 4a and 23TAG, Figure 4b). The fact that the structures differ in KCl and in NH₄OAc (Figure 4, purple and blue, respectively) was already known [20]. The structure of G-quadruplexes formed in ammonium acetate are generally intermediate between the ones formed in KCl (predominant in the nucleus) [37] and in NaCl [20].

In TMAA without KCl (Figure 4, grey), both DNA sequences adopt the same kind of CD spectra, characteristic of unfolded DNA. When 1 mM NaCl is added to TMAA (Figure 4, black), the species remains unfolded, confirming the anticipation from the absence of specific 2-Na⁺ adduct observed in the MS experiments in the same conditions (Figure 3e). However, when 1 mM KCl is added in the TMAA solution (Figure 4, red) the G-quadruplexes formed adopt almost exactly the same CD spectra as the ones in the phosphate buffer conditions. KCl-doped ammonium acetate solutions (Figure 4, green) showed CD spectra closer to the purely KCl form, but not as close as the KCl-doped TMAA conditions. In conclusion, the 100 mM TMAA/1 mM KCl conditions are compatible with electrospray mass spectrom-

etry and favor the formation of structures that possess almost identical proportion of homo-stacking versus alternate stacking as in 100 mM K⁺ solutions. Further studies, for example by NMR in deuterated TMAA, will be needed to determine if the structures are indeed identical.

Ion Mobility Spectrometry Experiments Indicate a Similar Shape for Telomeric G-Quadruplexes Preserving Two K^+ or NH_4^+

In MS experiments on G-quadruplex in NH4OAc, ammonium ion preservation in the gas phase structure is believed to ensure the preservation of the solution-phase structure [22, 38]. Is that the same in the case of K^+ ions? To probe the gas-phase structure of the G-quadruplexes, we used ion mobility spectrometry coupled to mass spectrometry (IMS-MS). This experiment links the arrival time of an ion inside the mobility cell to a shape factor called collision cross section (CCS). Figure 5 shows the results: two conditions are compared, 100 mM TMAA+1 mM KCl (red) and 100 mM NH₄OAc (brown and blue) for the sequence 24TTG. Very similar results were obtained for 23TAG (Supporting Information S5). The top graphs show the results for the 6- charge state [TMAA/KCl in (A), NH₄OAc in (B)] and the bottom graphs are for the 5- ions. The bias voltage right before the mobility cell entrance was varied (18 V, 25 V, or 30 V). The NH4⁺-binding stoichiometries vary with the bias voltage, in a similar way as they vary with source voltages (see Figure 2 and discussion thereof). For that reason, we represented the integrated CCS of all adduct species (0, 1, and 2 NH_4^+ adducts summed, in brown, and the CCS integrated only over the mass spectral peak corresponding to two NH4⁺ ions bound (blue). CCS distributions were normalized to their total area in order to reflect the proportion of folded or unfolded G-quadruplexes with two ammonium ions.

At low bias voltage (18 V) the CCSs were similar in both conditions: one peak around 800 Å² for 24TTG for the 6ions (Figure 5a and b). For the 5- ions (Figure 5c and d) the CCS peak appears at smaller values; which is typical in ion mobility spectrometry because the overall structure is less charged and then less dilated. In native conditions, the structures are very close in term of CCS in NH₄⁺ (light brown) or K⁺ (light red). Interestingly, the Gaussian curve that fits the data is a little bit broader in ammonium ion conditions. This can be due to a greater floppiness or to a greater polymorphism in ammonium ion solution. When the bias voltage increases (from light to dark colors) the peaks at the initial CCS value decrease and at the expense of peaks at higher CCS (for 6- charge state) or lower CCS (for 5- charge state). Expansion of high charge states and compaction of lower charge states upon energy increase was noted for other G-quadruplexes in ammonium conditions [38].

In the case of ammonium ions, the gas-phase structure is changing when the voltage is increased, and ammonia loss observed in the mass spectra gives a hint that denaturation



Figure 5. Collision cross section distribution of 5 μ M 24TTG into 100 mM TMAA+1 mM KCl (a) and (c) or 100 mM NH₄OAc (b) and (d). (a) and (b) represent the 6- ions and (c) and (d) the 5- ions. Bias voltage increases from light to dark colors (18, 25, and 30 V, respectively). Brown colors represent the collision cross section reconstructed from the sum of the G-quadruplexes with zero, one, and two ammonium cations and blue ones are the collision cross section of the G-quadruplexes with two ammonium ions exclusively. Red colors are the G-quadruplexes with two potassium ions

happens in the gas phase. When two potassium ions are trapped inside the G-quadruplex formed from TMAA/KCl, the collisional activation has the same unfolding effect, although the potassium ion adducts are not lost. Therefore, although the voltage at which unfolding occurs is higher in the presence of potassium than ammonium, gas-phase unfolding still occurs. Importantly, these changes in ATD are not accompanied by visible changes in the mass spectra for the TMAA/KCl conditions: the K⁺ ions cannot escape from the multiply charged anions.

The KCl Concentration Required for G-Quadruplex Formation in TMAA Depends on the Sequence

Above, we showed results with the maximum concentration we could electrospray while keeping clean mass spectra (1 mM). Here we describe KCl titration experiments that probe the minimum KCl concentration required to fold the G-quadruplexes. Figure 6 shows two KCl titrations in 100 mM TMAA for the intramolecular telomeric sequence 23TAG (Figure 6a) and for the tetramolecular G-quadruplex [dTG₄T]₄ (Figure 6b). At 100 μ M KCl, 23TAG Gquadruplexes are not formed: only a very small amount of potassium adducts are observed. When the KCl concentration is increased, the adduct distribution is shifted towards two potassium ions bound. Mass spectra at 200 and 300 μ M unambiguously demonstrate cooperativity in K⁺ binding: the cation adducts distribution is non-statistical and depleted in 1-K⁺ adduct. This is very important to understand the folding mechanism of G-quadruplexes, and mass spectrometry experiments allow to unambiguously visualize the concerted incorporation of two potassium ions.

The second example is the titration of the single strand dTG₄T, which forms a tetramolecular G-quadruplex with three coordinated monovalent ions. The titration is made on a lower concentrations range, as the G-quadruplex formation is already complete at 80 μ M in KCl. Despite the fact that ammonium ions were not added intentionally, some ammonium adducts appear here at low KCl concentrations, and mixed NH₄⁺/K⁺ adducts appear at intermediate KCl concentrations (the sum of adducts always being three cations, i.e., one between each G-quartet). Ammonium traces come from the oligonucleotide synthesis and, therefore, traces of ammonium-ion bound tetramers are preformed in the lyophilized oligonucleotides. However the single strand is prevalent (see full scale mass spectra in the Supporting Information S6).



Figure 6. KCI titration of (a) 5 μ M 23TAG and (b) 20 μ M TG₄T single strand into 100 mM TMAA. The annotations indicate the number of adducts on the 5- charge state. In titration (a), the single strand was allowed to react for 4 h at room temperature before the mass spectra were recorded using the LCT mass spectrometer. In titration (b), the solutions were prepared with 100 μ M in TG₄T single strand for 24 h before being diluted to 20 μ M, and the mass spectra were recorded on the Exactive mass spectrometer

The mechanism of formation of the tetramolecular Gquadruplex formation is very different from the formation of the intramolecular quadruplex. Adding a very small amount of potassium into the solution induced the cooperative tetramerization of TG₄T. To be stabilized, the tetramer must always contain total of three cations in its central channel. Then, because potassium has a higher affinity for Gquadruplexes than ammonium, the equilibrium is displaced specifically towards the 3-K⁺ adduct by addition of higher KCl amounts. This example further illustrates the correlation between the observation of *n* K⁺ adducts and the presence of *n*+*I*G-quartets [21].

Conclusions

We reported here for the first time that by using trimethylammonium acetate (TMAA) as electrolyte, electrospray mass spectra of G-quadruplexes can be obtained in the presence of the likely physiologically relevant potassium cation. Moreover, circular dichroism experiments indicate that the guanine stacking motifs and, therefore, relative strand orientations are very similar in 100 mM TMAA/KCl and 100 mM KCl, so the presence of TMAA in the ESI-MS compatible solution does not affect the overall structure of the DNA G-quadruplexes studied here. This is particularly important for polymorphic G-quadruplexes such as those formed by the human telomeric sequence. Also, our conditions make it easier to determine the number of potassium ions specifically coordinated to the structure and, hence, to ascertain the presence of a G-quadruplex in solution and how many G-quartets are involved.

The perspectives open by the present work are numerous. First, as in TMAA/KCl solutions the K⁺-binding stoichiometry is determined easily. In NH₄OAc conditions, the ESI-MS determination of cation binding stoichiometry was clear only for parallel G-quadruplexes, which robustly preserved their ammonium ions in the gas phase, whereas the antiparallel structures did not. The new method can be used to probe the number of specific binding sites in other sequences, for example issued from bioinformatics studies. Second, as TMAA allows spraying potassium-bound Gquadruplexes in soft conditions, this opens the way to probing preserved structures in the gas phase, as shown by our preliminary experiments with ion mobility spectrometry. Third, the TMAA/KCl conditions are interesting for ligand screening, in order to test the binding stoichiometry and affinity for potassium-bound structures. ESI-MS can also be used to understand the G-quadruplex folding mechanism and, more specifically, the interplay between cation binding and folding. Detailed investigation of the folding of various human telomeric sequences upon K⁺ or ligand titration is currently under way in our laboratory. Finally, it would be interesting to test whether the use of TMAA can be extended to other systems than DNA G-quadruplexes, for example to RNA-magnesium complexes, or to protein complexes.

Acknowledgments

The authors thank Frédéric Rosu and Miguel Garavis for fruitful discussion, the Structural Biophysico-chemistry Platform of the IECB (CNRS UMS 3033 and Inserm US 001) for access to the MS instruments, Kevin Giles and Jonathan Williams (Waters, Manchester, UK) for assistance in performing the IMS experiments on the Synapt G2S, and Christian Bleiholder and Mike Bowers (UC Santa Barbara) the opportunity to measure the drift tube CCS of the new calibrants. This work was funded by the Inserm (ATIP-Avenir), the EU (Marie Curie CIG), and the Conseil Régional Aquitaine (Crédits d'installation). A.M. has a Ph.D. fellowship from Inserm and Conseil Régional Aquitaine.

Open Access

This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S., Neidle, S.: Putative DNA quadruplex formation within the human c-kit oncogene. J. Am. Chem. Soc. 127(6), 10584–10589 (2005)
- Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., Hurley, L.H.: 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. 99(18), 11593–11608 (2002)
- Cogoi, S., Xodo, L.E.: G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. 34(9), 2536–2549 (2006)
- Sun, D., Guo, K., Rusche, J.J., Hurley, L.H.: Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. Nucleic Acids Res. 33(18), 6070– 6780 (2005)
- 5. Wang, Y., Patel, D.J.: Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. Biochemistry 31(35), 8112-8119 (1992)
- Biffi, G., Tannahill, D., McCafferty, J., Balasubramanian, S.: Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. 5(3), 182–186 (2013)
- Neidle, S.: Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. FEBS J. 277(5), 1118–1125 (2010)
- Rodriguez, R., Muller, S., Yeoman, J.A., Trentesaux, C., Riou, J.F., Balasubramanian, S.: A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. J. Am. Chem. Soc. 130(47), 15758–15759 (2008)
- Gomez, D., Wenner, T., Brassart, B., Douarre, C., O'Donohue, M.F., El Khoury, V., Shin-Ya, K., Morjani, H., Trentesaux, C., Riou, J.F.: Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. J. Biol. Chem. 281(50), 38721–38729 (2006)
- Phatak, P., Cookson, J.C., Dai, F., Smith, V., Gartenhaus, R.B., Stevens, M.F., Burger, A.M.: Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. Br. J. Cancer 96(8), 1223–1233 (2007)
- Phan, A.T.: Human telomeric G-quadruplex: structures of DNA and RNA sequences. FEBS J. 277(5), 1107–1117 (2010)

- Parkinson, G.N., Lee, M.P., Neidle, S.: Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417(6891), 876–880 (2002)
- Wang, Y., Patel, D.J.: Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. Structure 1(4), 263–282 (1993)
- Lim, K.W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D.J., Luu, K.N., Phan, A.T.: 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(12), 4301–4309 (2009)
- Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L., Patel, D.J.: Structure of the human telomere in K+solution: an intramolecular (3+1) Gquadruplex scaffold. J. Am. Chem. Soc. 128(30), 9963–9970 (2006)
- Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A., Yang, D.: Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic Acids Res. 34(9), 2723–2735 (2006)
- Phan, A.T., Luu, K.N., Patel, D.J.: Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+solution. Nucleic Acids Res. 34(19), 5715-5719 (2006)
- De Cian, A., Guittat, L., Kaiser, M., Sacca, B., Amrane, S., Bourdoncle, A., Alberti, P., Teulade-Fichou, M.P., Lacroix, L., Mergny, J.L.: Fluorescence-based melting assays for studying quadruplex ligands. Methods 42(2), 183–195 (2007)
- Yuan, G., Zhang, Q., Zhou, J., Li, H.: Mass spectrometry of Gquadruplex DNA: formation, recognition, property, conversion, and conformation. Mass Spectrom. Rev. 30(6), 1121–1142 (2011)
- Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E.S., Bowers, M.T., De Pauw, E., Gabelica, V.: G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. J. Am. Chem. Soc. 130(31), 10208-10216 (2008)
- Gros, J., Rosu, F., Amrane, S., De Cian, A., Gabelica, V., Lacroix, L., Mergny, J.L.: Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. Nucleic Acids Res. 35(9), 3064–3075 (2007)
- Balthasart, F., Plavec, J., Gabelica, V.: Ammonium ion binding to DNA G-quadruplexes: do electrospray mass spectra faithfully reflect the solution-phase species? J. Am. Soc. Mass Spectrom. 24(1), 1–8 (2013)
- Evans, S.E., Mendez, M.A., Turner, K.B., Keating, L.R., Grimes, R.T., Melchoir, S., Szalai, V.A.: End-stacking of copper cationic porphyrins on parallel-stranded guanine quadruplexes. J. Biol. Inorg. Chem. 12(8), 1235–1249 (2007)
- Pierce, S.E., Wang, J., Jayawickramarajah, J., Hamilton, A.D., Brodbelt, J.S.: Examination of the effect of the annealing cation on higher order structures containing guanine or isoguanine repeats. Chem. Eur. J. 15(42), 11244–11255 (2009)
- 25. Kharlamova, A., Prentice, B.M., Huang, T.-Y., McLuckey, S.A.: Electrospray droplet exposure to gaseous acids for reduction of metal

counter-ions in nucleic acid ions. Int. J. Mass Spectrom. 300(2/3), 158-166 (2011)

- Turner, K.B., Monti, S.A., Fabris, D.: Like polarity ion/ion reactions enable the investigation of specific metal interactions in nucleic acids and their noncovalent assemblies. J. Am. Chem. Soc. 130(40), 13353– 13363 (2008)
- Gray, R.D., Chaires, J.B.: Isothermal folding of G-quadruplexes. Methods 57(1), 47–55 (2012)
- Potier, N., Van Dorsselaer, A., Cordier, Y., Roch, O., Bischoff, R.: Negative electrospray ionization mass spectrometry of synthetic and chemically modified oligonucleotides. Nucleic Acids Res. 22(19), 3895–3903 (1994)
- Deguchi, K., Ishikawa, M., Yokokura, T., Ogata, I., Ito, S., Mimura, T., Ostrander, C.: Enhanced mass detection of oligonucleotides using reverse-phase high-performance liquid chromatography/electrospray ionization ion-trap mass spectrometry. Rapid Commun. Mass Spectrom. 16(9), 2133–2141 (2002)
- Lemaire, D., Marie, G., Serani, L., Laprévote, O.: Stabilization of gasphase noncovalent macromolecular complexes in electrospray mass spectrometry using aqueous triethylammonium bicarbonate buffer. Anal. Chem. 73(8), 1699–1706 (2001)
- Cavaluzzi, M.J., Borer, P.N.: Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. Nucleic Acids Res. 32(1), e13 (2004)
- Ruotolo, B.T., Benesch, J.L., Sandercock, A.M., Hyung, S.J., Robinson, C.V.: Ion mobility-mass spectrometry analysis of large protein complexes. Nat. Protoc. 3(7), 1139–1152 (2008)
- Gidden, J., Baker, E.S., Ferzoco, A., Bowers, M.T.: Structural motifs of DNA complexes in the gas phase. Int. J. Mass Spectrom. 240(3), 183–193 (2005)
- Wyttenbach, T., Kemper, P.R., Bowers, M.T.: Design of a new electrospray ion mobility mass spectrometer. Int. J. Mass Spectrom. 212(1), 13-23 (2001)
- Chaires, J.B.: Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. FEBS J. 277(5), 1098– 1106 (2010)
- Karsisiotis, A.I., Hessari, N.M., Novellino, E., Spada, G.P., Randazzo, A., Webba Dasilva, M.: Topological characterization of nucleic acid Gquadruplexes by UV absorption and circular dichroism. Angew. Chem. Int. Ed. Engl. 50(45), 10645–10648 (2011)
- Dick, D.A.: The distribution of sodium, potassium, and chloride in the nucleus and cytoplasm of Bufo bufo oocytes measured by electron microprobe analysis. J. Physiol. 284, 37–53 (1978)
- Gabelica, V., Baker, E.S., Teulade-Fichou, M.P., De Pauw, E., Bowers, M.T.: 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. 129(4), 895–904 (2007)

IV.3. ADDITIONAL WORK

IV.3.1 Additional structural and quantitative information from MS

While testing this sample preparation on several other G-quadruplex forming sequences, we systematically found that the main stoichiometry observed corresponded to the expected number of G-quartets in the G-quadruplex, minus one. This had already been observed and used in previous studies in ammonium acetates conditions.^{78,242,251–255} The Figure 36 illustrates this principle: *n* cations = (*n*+1) quartets.



Figure 36. *n* cations = (n + 1) quartets. Illustration of the principle with examples. Mass spectra, zoomed on the charges states 4 or 5-, of 10 µM DNA G-quadruplex in 100 mM TMAA + 1 mM KCl corresponding to the structures obtained by NMR spectroscopy, indicated at the bottom (2-quartet G-quadruplex, d(<u>GGTTGGGTGGGGTGGGGAAGG</u>), PDB ID: 1RDE,¹⁶ 3-quartet G-quadruplex, d(TGA<u>GGGTGGGGGGGGGGGGGGAAGG</u>), PDB ID: 2MGN,¹⁴⁷ 4-quartet G-quadruplex, 4x (dT<u>GGGGT</u>), PDB ID: 352D,¹⁰ and 4-quartet G-quadruplex, 2x d(<u>GGGTTTTGGGG</u>), PDB ID: 156D.²⁵⁶ The expected cation binding sites are highlighted in red and correspond to the main stoichiometry detected by MS (indicated in red on the spectra, the additional adducts are nonspecific).

The main stoichiometry corresponds to the one expected if the cations are trapped between the G-quartets. However, additional cation stoichiometries were observed and correspond to nonspecific cations, not G-quadruplex specific. Indeed, a potassium ion distribution on each DNA peak is observed by MS when a KCI-containing solution is injected. The detected potassium binding distribution is not the same for each charge state. Therefore, necessarily, a part of the binding detected is nonspecific and happens along the ESI process. It is also possible that a fraction of these adducts comes from K⁺ binding to external phosphate groups, and these are also not G-quadruplex specific. The goal of subtraction of nonspecific contribution is to deduce the fraction of K⁺ coming from specific binding, in-between Gquartets. In a set of conditions, the nonspecific distribution is given by the distribution obtained for a non-G-quadruplex forming sequence (Figure 37, for technical details see section III.1.5). The mass spectrometry is therefore a technique giving very important structural and quantitative information that could be complementary to those obtained with other techniques.





On the other hand, a limitation of this sample preparation method was observed for dimers of G-quadruplexes. Indeed, the dimerization of G-quadruplexes by end-stacking was observed for parallel G-quadruplexes by mass spectrometry in ammonium acetate²⁴¹ but was not occurring in TMAA conditions. One explanation for the suppression of this interaction is that the dimerization of G-quadruplexes occurs thanks to the presence of more ammonium cations, presumably trapped between the two G-quadruplex units or shielding the negative phosphates. In TMAA + KCl conditions, the K⁺ concentration is too low to populate the cation binding site between the two G-quadruplex units, probably of lower affinity, preventing dimerization. This effect could be wanted if the monomeric G-quadruplex is of interest.

IV.3.2 HIGHER SALT CONCENTRATIONS

In this work, the highest K⁺ concentration used is typically 1 mM. This is not a strict limit. Indeed, the signal drops continuously with increasing KCl concentration until the signal is completely lost among the clusters peaks. This means that higher KCl concentrations than 1 mM could, in principle, be reached by paying the price for the signal-to-noise ratio.

At concentrations as high as 10 mM KCl, signals corresponding to the DNA were still detectable, despite being masked by the clusters peaks (Figure 38A). Zooming on the region of interest allowed to distinguish the peaks of the DNA from the clusters but the mass spectrum is noisy and the identification of the peaks might be difficult and the quantification impossible (Figure 38C).

Thanks to ion mobility spectrometry coupled with mass spectrometry, it was possible to separate G-quadruplexes from the clusters (Figure 38B). Reconstruction of the mass spectrum corresponding to the IMS region of the G-quadruplexes (Filtered out area in Figure 38B) allows to obtain mass spectra of G-quadruplexes in up to 10 mM KCl (Figure 38D). However, for instrument preservation reasons, the limit was usually fixed to 1 mM. In addition, higher concentrations increase the amount and number of nonspecific adducts, which could lead to more difficult stoichiometry determination, eventually without providing more information. Therefore, it is useful to detect what is the first peak of the distribution (here 2-K⁺) but not practical to quantitate precisely.


Figure 38. Mass spectra of 10 μ M 22AG, d(A(GGGTTA)₃GGG) in 100 mM TMAA + <u>10 mM KCl</u>. (A) Full scale mass spectrum, (B) the 2D heatmap of the arrival time in function of the m/z, (C) zoom on the region of interest of the mass spectrum and (D) the mass spectrum reconstructed from the area indicated in B.

IV.3.3 OTHER CATIONS THAN K⁺

In principle, any cation can be used with this methodology. However, the upper limit concentration could be different for each cation. Indeed, each salt could make clusters more or less easily when electrosprayed.

Another intensively used cation in the G-quadruplex field is the sodium. Indeed, Na⁺ induces the folding of G-quadruplexes with sometimes a different structure (for example, the human telomeric sequence⁸⁰). Mass and CD spectra of G-quadruplexes were obtained in TMAA + 1 mM NaCl (See Article 1, Figure 3 & 4). However, in those conditions, G-quadruplexes presented the characteristics of unfolded structures: the main detected stoichiometry is 0 and the CD is typical of unfolded. Usually, G-quadruplexes are less stable in presence of sodium than potassium.⁷² 1 mM NaCl is not sufficient to fold these G-quadruplexes and a higher concentration, not attainable using MS, is required. It was therefore not possible to study sodium binding to G-quadruplexes using the TMAA conditions.

Finally, TMAA doped with strontium dichloride salt was also tested. The Sr²⁺ cation is a good G-quadruplex stabilizer.⁷² Indeed, only 50 μ M Sr²⁺ sufficed to fold 10 μ M of G-quadruplex (Figure 39A). Interestingly, we found differences in terms of cation stoichiometry from the potassium conditions (Figure 39B & C). Indeed, for this sequence, G-quadruplexes are formed with two preferential potassium cations but only one strontium. Unfortunately, we could not subtract the nonspecific contribution with the method previously described. Indeed, we found that the strontium cation binds with a very good affinity to the guanines-rich single stranded reference.²⁵⁷ This is an example where there are more nonspecific adducts on the reference than on the G-quadruplex (Figure 39B & D) and therefore, only qualitative analysis could be performed.



Figure 39. Mass spectra of 10 μ M 222T, d(T(GGGTT)₃GGGT), in 100 mM TMAA supplemented with, (A) 50 μ M SrCl₂, (B) 1 mM SrCl₂, (C) 1 mM KCl. (D) mass spectrum of the reference shuf1, d(TGTGGTGTGTGTGTGTGGG), in 100 mM TMAA with 1 mM SrCl₂.

IV.4. CONCLUSIONS

In this chapter, we have shown that mass spectrometry of potassium containing Gquadruplexes can be easily obtained using trimethylammonium acetate as electrolyte. Up to a millimolar concentration of potassium cation was co-injected with G-quadruplex forming sequences. This new sample preparation method offers several advantages:

- 1) The potassium cation is more physiologically relevant than ammonium. Because some polymorphic sequences can fold into different structures in different salt conditions, the ability to obtain mass spectra of G-quadruplexes in presence of potassium is of greater biological interest. It is now possible to correlate the mass spectrum of detected species to their respective structures, obtained in potassium conditions by high resolution techniques such as NMR spectroscopy or x-ray diffraction.
- 2) The potassium cation interacts very strongly with DNA in the gas phase thanks to Coulomb attractions. This provides an advantage compared to the ammonium conditions in that it is impossible to dissociate K⁺ from the complex in the gas phase (whereas ammonium could be released as ammonia by proton transfer). Therefore, the tuning of the mass spectrometer is a little bit more forgiving.
- 3) The potassium stoichiometry can be used to deduce structural details: n cations = (n+1)G-quartets. The determination of the G-quartet number thanks to the cation stoichiometry was also valid for ammonium but the ability to subtract the contribution of nonspecific adducts allows extraction of quantitative information. The determination of the number of G-quartets and their stacking types given by CD combine very well for topology characterization of G-quadruplexes.
- 4) The polymorphism of some sequences can be studied using mass spectrometry, allowing one to separate different potassium binding stoichiometries and quantify them. Ion mobility coupled to mass spectrometry can also give information about the polymorphism of one defined stoichiometry.

V. POTASSIUM BINDING TO G-QUADRUPLEXES

V.1. INTRODUCTION

Even though cations are required for folding G-quadruplexes, few studies take into account the number of ions bound, probably because no technique (other than MS) can determine directly this stoichiometry. Measuring the free K⁺ concentration allowed previously to determine indirectly the number of cations bound: Chaires found that up to ten cations binds,²⁵⁰ some being possibly nonspecific.²⁵⁸



Figure 40. Proposed folding landscape for the human telomeric G-quadruplexes. Illustration reproduced from Chaires *et al.*²⁵⁹ and Mashimo *et al.*²⁶⁰

The folding pathway of G-quadruplexes (and, in particular, of the human telomeric sequences) is currently studied by several groups. One proposed folding pathway of the human telomeric sequence (Figure 40) involves first the collapse of the unfolded species in hairpin-like structures, involving G:G base pairs.^{260,261} Those species then fold into antiparallel G-quadruplexes^{259,262} before folding into the hybrid structure, one of the possible intermediates being a triplex species.²⁶³⁻²⁶⁵

Mass spectrometry has previously been used to study the folding pathway of tetramolecular G-quadruplexes, in ammonium acetate conditions without and with addition of co-solvents.²⁵³ We previously (Master 2 thesis) showed that the formation of bimolecular human telomeric G-quadruplexes could be induced by the addition of co-solvent.^{78,255} We partitioned the complexes formed thanks to their ammonium distribution and were able to distinguish four ensembles: one unfolded, one misfolded with one NH4⁺ and two G-quadruplexes with 2-NH4⁺, one antiparallel and one parallel. Interestingly, the data suggested that those two species were on different branches in the mechanism.

In the present work, we focused on the folding of intramolecular human telomeric Gquadruplexes formed in KCl containing solutions. We performed equilibrium experiments to determine the equilibrium binding constants of K⁺ to different human telomeric sequences (differing only by a few flanking bases). We found that the shortest sequences (with few or no flanking bases) were prone to fold into a 1-K⁺ antiparallel G-quadruplex whereas longer sequences bound K⁺ in a more cooperative manner. Kinetic experiments showed that the folding occurs through a multiple pathway mechanism and that the structures that fold first are antiparallel. Those complexes are on a dead-end in the reaction mechanism. To fold into the thermodynamically most stable hybrid G-quadruplexes, these antiparallel Gquadruplexes first have to unfold. It is then possible to generalize this mechanism to other Gquadruplexes as we also show in this work that antiparallel species were folded at short time scales for the parallel Pu24 sequence.

V.2. ARTICLE 2

Folding and Misfolding Pathways of G-quadruplex DNA

A. Marchand & V. Gabelica

Nucleic Acids Research, 2016, doi: 10.1093/nar/gkw970.

For supplementary information, see Annex II.

Nucleic Acids Research, 2016 1 doi: 10.1093/nar/gkw970

Folding and misfolding pathways of G-quadruplex DNA

Adrien Marchand and Valérie Gabelica^{*}

INSERM, CNRS, Univ. Bordeaux, U1212 / UMR5320 - Acides Nucléiques: Régulations Naturelle et Artificielle, IECB, 2 rue Robert Escarpit, 33607 Pessac, France

Received August 10, 2016; Revised October 07, 2016; Editorial Decision October 10, 2016; Accepted October 15, 2016

ABSTRACT

G-quadruplexes adopt various folding topologies, but information on their folding pathways remains scarce. Here, we used electrospray mass spectrometry to detect and quantify the specifically bound potassium ions, and circular dichroism to characterize the stacking topology of each ensemble. For human telomeric (hTel) sequences containing the d((GGGTTA)₃GGG) core, K⁺ binding affinity and cooperativity strongly depends on the chosen construct. The shortest sequences bind only one K⁺ at low KCI concentration, and this 2-quartet Gquadruplex is antiparallel. Flanking bases increase the K⁺ binding cooperativity. To decipher the folding pathways, we investigated the kinetics of K+ binding to telomeric (hybrid) and c-myc (parallel) G-quadruplexes. G-quadruplexes fold via branched pathways with multiple parallel reactions. Up to six states (one ensemble without K⁺, two ensembles with 1-K⁺ and three ensembles with 2-K⁺) are separated based on their formation rates and ion mobility spectrometry. All G-quadruplexes first form longlived misfolded structures (off-pathway compared to the most stable structures) containing one K⁺ and two quartets in an antiparallel stacking arrangement. The results highlight the particular ruggedness of G-quadruplex nucleic acid folding landscapes. Misfolded structures can play important roles for designing artificial G-quadruplex based structures, and for conformational selection by ligands or proteins in a biological context.

INTRODUCTION

G-quadruplex nucleic acid structures can form in important regions of the human genome such as telomeres, genes promoters or transcription start sites (1), and are involved in several cellular regulation pathways such as regulation of gene expression (2-8) and telomere maintenance (911). Recently G-quadruplexes have been revealed in human cells using an antibody (12). G-quadruplexes are tetrahelical structures formed by guanine-rich sequences (13). Four guanines, interacting via Hoogsteen hydrogen bonds, form a G-quartet. Structures with at least two stacked Gquartets are called G-quadruplexes. G-quadruplex stabilization is ensured by monovalent cations such as Na⁺, K⁺ or NH_4^+ (14,15), which they intercalate between consecutive G-quartets and are octa-coordinated by the O6 of the guanines (16). G-quadruplexes can adopt different topologies (Figure 1) (17), and this structural polymorphism is exploited in applications such as DNA functionalization and nanodevices (18-20). The topology is defined by which bases are engaged in G-quartets, the orientation of the strands in the stem, the position of the loops connecting the four parts of the stem, and the stacking mode of the bases in the stem (all four characteristics being inter-related) (17). Many parameters influence the adopted topology: the nature of the monovalent cation, the concentration, the ionic strength, the pH, the presence of crowding or dehydration agents, and the sequence.

The human telomeric sequence, with the repetitive 3^1 -end sequence $d(TTAGGG)_n$, is one of the most studied biologically relevant G-quadruplexes. Incidentally, this is also one of the most polymorphic G-quadruplex forming sequences (21). Here, we investigated the intramolecular folding of telomeric sequences with four tracts of guanines, in the presence of K⁺ cations. All sequences contain the central $d((GGGTTA)_3GGG)$ core, but the final topology depends on the flanking bases.

In 100 mM K⁺ (NMR conditions), a hybrid-1 type G-quadruplex structure was observed for the sequence $d(TT(GGGTTA)_3GGGA)$ (abbreviated 24TTG, PDB ID: 2GKU (22), Figure 1A, see Table 1 for abbreviations) and for $d(TA(GGGTTA)_3GGG)$ (23TAG, PDB ID: 2JSM (23)), the latter being slightly more polymorphic. The se-

quence d(TTA(GGGTTA)₃GGGTT) (26TTA, PDB ID: 2JPZ (24), Figure 1B) folds into a hybrid-2 type Gquadruplex. The hybrid topologies differ by the or- der of the loops: in the hybrid-1 the loops are respec- tively 5^1 -propeller-lateral-lateral- 3^1 type. For the hybrid-2 the loops are 5^1 -lateral-lateral-propeller- 3^1 . The sequence

Θ The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.

^{*}To whom correspondence should be addressed. Tel: +33 5 4000 2940; Email: valerie.gabelica@inserm.fr

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

d((GGGTTA)₃GGGT)) (22GT, PDB ID: 2KF8 (25), Figure 1C) folds predominantly (\sim 70%) into another topology, called the hybrid-3. Here, only two quartets are formed. The structure is further stabilized by a triplet of guanines, and the circular dichroism spectrum therefore resembles that of the other hybrid structures (26). 22AG (d(A(GGGTTA)₃GGG)), a sequence very frequently used for drug binding assays, folds in an antiparallel structure in sodium (Figure 1D) (27) but is polymorphic in KCl (28-30). The structure of 21G (d((GGGTTA)₃GGG)) has not been solved by NMR, most likely because it is also polymorphic. It is the shortest possible human telomeric sequence containing four tracks of guanines, and it is frequently used in FRET (Förster Resonance Energy Transfer) melting assays with the FAM and TAMRA dyes at the 5^1 and 3^1 ends (31). 21G, with and without the fluorescent tags, have also been included in the present study, as well as the FRET homologue of 26TTA. The modified human telomeric DNA 22CTA (d(A(GGGCTA)₃GGG), PDB ID: 2KM3 (32), Figure 1E) was also included. This G-quadruplex structure contains two G-quartets plus a G:C:G:C platform.

Telomeric sequences form intramolecular parallel folds only in the crystal state or in molecular crowing conditions, but this species is not significantly populated in dilute aqueous solution (33,34). To include a parallel quadruplex in our panel, we also studied the sequence Pu24 (d(TGAG₃TG₄AG₃TG₄A₂G₂), PDB ID: 2A5P (35), Figure 1F), a variant of the c-myc oncogene promoter nuclease hypersensitivity element.

The structural polymorphism of G-quadruplexes complicates their biophysical studies. The thermodynamically stable folds of telomeric G-quadruplex structures have been well characterized (21) but there is yet no consensus on the folding pathway (36-38). Spectroscopic techniques such as UV-absorbance or UV-melting (39), fluorescence or FRETmelting (40), and circular dichroism (CD) are commonly used. Gibbs free energy of folding can be obtained by assuming a 1-step (2-state) folding (41). However, singular value decomposition (SVD) analysis revealed that thermal folding and unfolding of G-quadruplexes was not 2-state (29,42), and that the isothermal cation-induced folding involves at least three states (37, 43). The nature of these states is however still debated. The folding pathway may include a collapse of the unfolded species into hairpin-like structures with G:G base pairs (44-47), which then fold into an antiparallel G-quadruplex (37,48), then interconvert progressively into triplexes intermediates (49–51) before folding properly into a G-quadruplex (37). This model is sequential, and therefore does not consider side-reactions in the mechanism, i.e. 'parallel' or 'branched' pathways (note that for avoiding confusion between 'parallel' G-quadruplex structures and 'parallel' reactions in chemical kinetics, we will use the term 'branched' instead of 'parallel' when talking about mechanisms). Others have proposed branched mechanisms where both the unfolded species, an intermediate (triplex or other) and the G-quadruplex are in equilibrium, with (52) or without intermediates (38,53). Molecular dynamics simulations also suggest that the folding of the human telomeric sequences proceeds through a complex branched mechanism (49,54,55).

A major hurdle in studying the folding mechanisms experimentally is the difficulty in defining the different states (or ensembles) contributing to the mechanism. In the studies cited above, the ensembles were defined based on bulk spectroscopic properties or from single-molecule measurements. In the present study, we used mass spectrometry to separate and quantify the DNA complexes in a different way: with no cation, one potassium ion or two potassium ions specifically bound (56). We studied the equilibrium dissociation constants and the kinetics (on the minute to hour time scale) of K⁺ binding to several variants of the human telomeric sequences and one c-myc promoter Gquadruplex, at up to 1 mM final KCl concentrations (57). Each ensemble defined based on the potassium binding stoichiometry was then sub-divided based on kinetic behavior, and confirmed by ion mobility spectrometry separation of the mass spectral peaks (IMS-MS). The topology of the ensembles defined are also characterized by CD spectroscopy.

Although we are blind to the details of the early folding events, our experiments allowed to uncover important slow events. While a 1-K⁺ structure with antiparallel base stacking always forms quickly, hybrid conformations take much longer to form. Branched folding pathways prevail, and the folding energy landscapes are so rugged that misfolded states can linger for several minutes. These results highlight that 2-quartet topologies are worthy of consideration among the biologically relevant ones.

MATERIALS AND METHODS

Materials

Oligonucleotides were purchased in lyophilized and RPcartridge purified form from Eurogentec (Seraing, Belgium). They were solubilized in nuclease-free grade water from Ambion (Fisher Scientific, Illkirch, France) at approximately 1 mM. The stock concentrations were determined using the Beer-Lambert law, using absorbance at 260 nm on a Uvikon XS, and molar extinction coefficients calculated using the IDT website using the Cavaluzzi-Borer Correction (58). The stock solutions were then diluted to exactly 200 μ M. Trimethylammonium acetate (TMAA, Ultra for UPLC, Fluka) and potassium chloride (KCl, >99.999%) were purchased from Sigma-Aldrich (Saint-Quentin Fallavier, France). The pH of the final solutions is 6.9. Table 1 lists the sequences and the abbreviations used in this work.

Electrospray mass spectrometry (ESI-MS)

All equilibrium (ESI-MS) titration experiments were performed in negative ion mode on a LCT Premier mass spectrometer (Waters, Manchester, UK). The titration experiment on the 21G sequence was also repeated on a Bruker Amazon SL mass spectrometer (Bruker, Bremen, Germany). All tuning parameters are given in Supporting Information. Each titration point was obtained from an independent solution (no sequential dilution). The solutions at 10 μ M DNA, 100 mM TMAA and the proper amount of KCl were annealed at 85°C for 5 min and cooled at room temperature in the dark for at least one night before measurement in order to avoid any kinetic effect.

Table 1. K_D values at 20°C of first and second K⁺ binding to G-quadruplex forming sequences (results of the fitting of the ESI-MS titration curves using DynaFit, with the errors from the fits)

		G4 structure obtained by NMR in 100				
Abbreviation	Sequence	mM K ⁺	$K_{\rm D1}~(imes 10^{-6}~{\rm M})$	$K_{\rm D2}~(\times 10^{-6}~{\rm M})$	$K_{\rm D1}/K_{\rm D2}$	$K_{\text{D-Glob}} (\times 10^{-9} \text{ M}^2)$
21G	(GGG TTA)3 GGG	Mixture	18 ± 1	1660 ± 40	0.01	30 ± 2
22GT	(GGG TTA)3 GGG T	Hybrid-3 (2-quartet antiparallel) (25)	22 ± 1	1050 ± 30	0.02	23 ± 2
22CTA	A (GGG CTA)3 GGG	2-quartet antiparallel (32)	271 ± 6	6500 ± 600	0.04	1750 ± 200
22AG	A (GGG TTA)3 GGG	Mixture	60 ± 4	1170 ± 70	0.05	70 ± 9
F21T	FAM-(GGG TTA)3 GGG-TAMRA	n.d. ^a	177 ± 4	700 ± 10	0.25	123 ± 5
23TAG	TA (GGG TTA)3 GGG	Hybrid-1 (23)	310 ± 30	170 ± 15	2	50 ± 10
26TTA	TTA (GGG TTA)3 GGG TT	Hybrid-2 (24)	1500 ± 100	190 ± 20	8	280 ± 50
F26T	FAM-TTA (GGG TTA)3 GGG TT-TAMRA	n.d. ^a	6500 ± 600	560 ± 75	12	3700 ± 800
24TTG	TT (GGG TTA) ₃ GGG A	Hybrid-1 (22)	270 ± 40	10.4 ± 1.6	26	2.7 ± 0.8
Pu24	TGAG ₃ TG ₄ AG ₃ TG ₄ A ₂ G ₂	Parallel (35)	550 ± 100	3.8 ± 0.9	145	2.2 ± 0.9

^aG-quadruplex structure, topology not documented.

 K_{D1} and K_{D2} are defined in Equations (2) and (3). The sequences are ranked with increasing K_{D1}/K_{D2} ratio.



Figure 1. Structures and their respective schematic topologies of (A) the hybrid-1 24TTG, (22) (B) the hybrid-2 26TTA, (24) (C) the antiparallel 2-quartet 22GT, (25) (D) the antiparallel in sodium 22AG, (27) (E) the antiparallel 2-quartet 22CTA, (32) and (F) the parallel Pu24.(35) These structures were obtained by NMR in 100 mM K⁺ except for 22AG whose structure was obtained in 100 mM Na⁺.

The kinetics experiments were carried out on an Agilent 6560 IMS-Q-TOF (Agilent Technologies, Santa Clara, CA, USA) with a dual ESI source in negative ion mode (see supporting information for the tuning parameters). This instrument allows one to monitor the drift time distributions of each m/z range, and this further helps discerning the existence of several ensembles in species of a given stoichiometry. One MS spectrum is accumulated every 1.1 s. The first two minutes are missed or discarded because it is the time lapse to manually mix the solutions and obtain a stable ESI-MS signal and flow rate. The solutions were prepared by mixing the proper volumes of stock solutions (previously annealed in pure water) to reach 10 µM DNA, 100 mM TMAA and the desired KCl concentration. dT₆ was also added at 2.5 µM for the mass spectrometry experiments and used as an internal standard to validate that the relative intensities reflect the relative concentrations in solution (see Supporting Information for more detail) (59). The start time of the experiment is the KCl addition (t_0) . Identical results were obtained whether we annealed the oligonucleotide or not in the TMAA buffer prior to KCl addition.

$K_{\rm D}$ determination and kinetic model evaluation

The concentrations of the species detected by mass spectrometry were obtained after integration of the peaks and subtraction of the nonspecific adducts. We also verified that all complexes have the same response factors. The detailed procedure is reported in the supplementary materials and methods (Supplementary Figures S1–S10). Both K_D determination and kinetic model evaluation were carried out using the DynaFit software (4.05.103 BioKin Ltd Watertown, MA, USA) (60). The concentrations of specific complexes of each stoichiometry were fitted as function of the potassium concentration with the chemical Equations (1b) and (2b), correlated to the chemical equilibria (1a) and (2a), respectively.

$$\mathsf{DNA} + \mathsf{K}^+ \rightleftharpoons \mathsf{DNA} \bullet \mathsf{K} \tag{1a}$$

$$K_{\text{D1}} = [\text{DNA}][\text{K}^+]/[\text{DNA} \bullet \text{K}]$$
(1b)

$$\mathsf{DNA} \bullet \mathsf{K} + \mathsf{K}^+ \rightleftharpoons \mathsf{DNA} \bullet \mathsf{2}\mathsf{K} \tag{2a}$$

$$K_{D2} = [DNA \bullet K][K^+]/[DNA \bullet 2K]$$
(2b)

The DynaFit software works by iteration on the two K_D values in order to fit all the concentrations of the three sto-

ichiometries as a function of total KCl added. The K_{D-Glob} is defined by the product of K_{D1} by K_{D2} .

$$K_{\text{D-glob}} = K_{\text{D1}} \times K_{\text{D2}} = [\text{DNA}][\text{K}^+]^2 / [\text{DNA} \cdot 2\text{K}]$$
 (3)

For the kinetics experiments, we fitted altogether the kinetic data obtained for the 0-K⁺, specific 1-K⁺ and 2-K⁺ ensembles at four different KCl concentrations (100, 200, 500 and 1000 μ M), i.e. a total of 12 curves. The chemical reaction models are an input for Dynafit, and the software derives the kinetic differential equations linked with each chemical model, and fits the data. Over 20 different reaction models were tested and ranked using the Akaike information criterion (AIC) (61). The AIC is applicable when the experiments depend on a multitude of parameters, and one is searching for the most useful model to describe the experimental set (62). Several starting initial parameters were tested, and converged to the same results. We also examined manually the fit quality, the errors on the rate constants and their correlation (see results and discussion).

Circular dichroism (CD) spectroscopy

CD is commonly used to characterize the stacking of the guanines in a G-quadruplex, and therefore indirectly gives information on the strand orientations. Parallel Gquadruplexes have only *anti*-guanines, so *anti/anti* stacking motifs exist in the resulting topology, and the CD shows a maximum at 260 nm. Antiparallel G-quadruplexes have alternated stacking of the guanines (*anti/syn* and *syn/anti*) leading to a minimum in the CD spectrum at 260 nm and a maximum at 290 nm. Hybrid G-quadruplexes have both types of stacking in the same topology, and present two positive peaks at 270 and 290 nm (26). Note that in some cases, 2-quartet structures presenting antiparallel strand orientation can display hybrid-type CD spectra, due to stacking of additional bases on the main two quartets (26).

CD experiments were carried out on solutions prepared exactly as for ESI-MS. Experiments were run on a Jasco J-815 spectrophotometer at 20°C (unless otherwise mentioned)—see supporting information for the instrumental details. CD data were normalized to molar circulardichroic absoption ($\Delta \epsilon$) based on DNA concentration using Equation (4),

where θ is the CD ellipticity in milidegrees, *C* is the DNA concentration in mol/l (in this work: 10^{-5} mol/l) and *l* is the pathlength in cm (0.2 cm). The concentrations used for CD are the same as those for MS, to allow direct comparison of the results.

The DynaFit software was used to obtain the molar ellipticity coefficient of each ensemble defined based on ESI-MS, as a function of the wavelength (here, each 5 nm). The software uses the equilibrium or rate constants to calculate the concentration of each ensemble, and then determines the molar response (here the molar ellipticity coefficient) of each ensemble (see supporting information for the data processing details).

RESULTS

K⁺ binding equilibria

Mass spectrometry titrations reveal a sequence-dependent K⁺ binding cooperativity. Thanks to the mass resolution, an independent signal is detected for each potassium binding stoichiometry, allowing non-ambiguous stoichiometry determination. Figure 2A-C shows representative mass spectra (zoom on the 5- charge state) obtained upon KCl titration of three human telomeric sequences, 24TTG, 23TAG and 22GT. Full titration experiments, full-scale mass spectra, zooms on other charge states, and results on other sequences are shown in Supplementary Figures S11-S20. After subtracting the nonspecific adducts contribution (which account for $\sim 15\%$, $\sim 30\%$ or $\sim 70\%$ of the total signal for charge states 6-, 5- and 4-, respectively, see Supplementary Figure S7), we quantify the concentration of specific K⁺ complexes of each stoichiometry for each point in the titration (Figure 2D-F, for the other sequences, see Supporting Supplementary Figure S21). After this subtraction, only 1-K⁺ and 2-K⁺ specific complexes remain. Interestingly, a 1-K⁺ complex predominates for the shortest sequences (21G, 22GT and 22AG). Although this was expected for 22GT (\sim 70% is folded into a 2-quartet form), this was more surprising for 22AG and 21G, usually represented as 3-quartet hybrid G-quadruplexes in K⁺. The 22CTA sequence presents almost exclusively the 1-K⁺ stoichiometry, as anticipated based on its 2-quartet structure (32).

Figure 2D-F and Supporting Supplementary Figure S21 show the best fits and the experimental data for each sequence, and Table 1 summarizes the $K_{\rm D}$ values. The experiments were successfully reproduced on another mass spectrometer (Supporting Supplementary Figure S22). K_{D1} is smallest for the 21-mer and 22-mer sequences, and the opposite is found for K_{D2} . This indicates that, although these short sequences contain tracts of three guanines, they predominantly adopt a 1-K⁺ binding stoichiometry in sub-mM KCl concentrations. The K_{D-Glob} values allow to compare the sequences in terms of their propensity to form high amounts of 2-K⁺ species at equilibrium, whereas the ratio K_{D1}/K_{D2} reflects the K⁺ binding cooperativity. As described previously (63), the 2:1 binding can be considered as positively cooperative if $K_{D1}/K_{D2} > 1/4$. Interestingly, all sequences containing more than 22 bases (i.e. more than one flanking base) showed positive cooperativity in K⁺ binding. 24TTG and Pu24 are the most cooperative. Note that 24TTG is a mutated telomeric sequence, which is particularly reputed for being monomorphic (22). The parallel Pu24 G-quadruplex has an even higher K⁺ binding cooperativity and in equilibrium conditions almost no 1-K⁺ complex is formed, whatever the KCl concentration.

The effect of the FRET dyes (FAM and TAMRA) on the K_D values was also investigated. For 21G, adding the dyes increases K_{D1} 10-fold and decreases K_{D2} 2.4-fold. Adding the dyes therefore increases the cooperativity by a factor of ~25. For 26TTA, the same effect is observed, but it is less pronounced as this sequence binds K^+ already in a more cooperative manner than 21G. The K_{D-Glob} is also affected by the addition of dyes, reflecting that the affinity for potassium (and hence the relative stability of the folded



Figure 2. (A–C) Mass spectrometry titrations (20°C) of 10 μ M (A) 24TTG, (B) 23TAG and (C) 22GT, in 100 mM TMAA with increasing amounts of KCl (zooms on the 5- charge state, see Supplementary Figures S11-S20 for the full scale mass spectra). (D–F) Concentration of the complexes with 0, 1 or 2 specific potassium ions bound in green, blue and orange, respectively, as a function of total KCl concentration, for (D) 24TTG, (E) 23TAG and (F) 22GT. The error bars are the standard errors on the treatment of the three charge states. The solid lines represent the best fits of the models (Equations 2 and 3).

form) of the labeled oligonucleotides decreases compared to the unlabeled analogues. These results highlight that Gquadruplex folding thermodynamics is modified by fluorophore attachment.

The 1-K⁺ complexes formed at equilibrium are antiparallel. To deduce the predominant topology of each ensemble distinguished by ESI-MS (the specific 0-K⁺, 1-K⁺ and 2-K⁺ complexes), we monitored the same titration experiments by circular dichroism spectroscopy (Figure 3 and Supporting Supplementary Figure S23). Without KCl in solution the DNA sequences are not folded into G-quadruplexes (green), except for 22GT and 21G whose bands at 295 nm could be indicative of a small fraction of folded G-quadruplex. Because of the high ionic strength (100 mM monovalent cation and anion), it is possible that sequences starting with a 5¹-guanine tract form secondary structures not requiring K⁺ binding, similar to those reported by the

Plavec group (44). Here we observed the same CD spectra with and without pre-annealing of the solution.

The CD spectra change upon KCl addition to the oligonucleotides, indicating a structural change induced by K⁺ binding. For 24TTG, the folding into a hybrid Gquadruplex occurs at low KCl concentration, as anticipated based on the MS results. The isoelliptic point in the CD titration is in line with the cooperativity observed by ESI-MS: the 1-K⁺ species not being significantly populated, the transition is 2-state. For 23TAG and 26TTA, the CD spectrum of hybrid structure(s) is progressively formed, but there is no isoelliptic point, and this indicates the presence of more than two states. This is in line with the ESI-MS titrations that revealed three states, the 1-K⁺ population being significantly populated at intermediate KCl concentrations (Figure 2).

Interestingly, for 21- and 22-mer sequences (21G, 22GT, 22AG and F21T), the amplitude of the CD signals at 260



Figure 3. Circular dichroism titrations (20°C) of 10 μ M (A) 24TTG, (B) 23TAG and (C) 22GT in 100 mM TMAA with increasing amount of KCl (0 in yellow, 25, 50, 75, 100, 200, 500 in gray and 1000 μ M KCl in red). (D–F), CD spectra reconstructed for the 0-K⁺ (green), 1-K⁺ (blue) and 2-K⁺ (orange) complexes, for sequences (D) 24TTG (the 1-K⁺ complex was not abundant enough to reconstruct a CD spectrum), (E) 23TAG and (F) 22GT.

nm and 295 nm reaches a maximum at about 0.1 mM KCl (Figure 3C and Supporting Supplementary Figure S23C, D and G) and decreases at higher KCl concentrations. In order to obtain structural information for each ensemble detected and quantified by MS, we deconvoluted the CD spectra into those of the 0-K⁺, 1-K⁺ and 2-K⁺ ensembles, using the abundances deduced from ESI-MS. The generated CD spectra are shown in Figure 3D–F and Supporting Supplementary Figure S24.

For 24TTG and Pu24, the CD spectra for the 1-K⁺ structures were not deconvoluted because of the low population of 1-K⁺ ensemble (see for example Figure 3A for 24TTG). Also, given that the titration involves only two main states, the CD titration data can be fitted using Equation (3). The $K_{\text{D-Glob}}$ obtained by fitting the CD titration agrees with the one obtained by fitting the ESI-MS titration (2.69 × 10⁻⁹ M² versus 2.70 × 10⁻⁹ M², respectively) (Table 1), and this also validates the quantitative mathematical treatment of the ESI-MS data.

For all the human telomeric sequences in which it is significantly populated, the 1-K⁺ ensemble has clearly an antiparallel-type CD spectrum. For 22GT, based on the published structure (PDB ID: 2KF8 (25)), we inter- pret the results as follows. The final structure contains two G-quartets and a triplet of guanines, so the two ex- pected cation binding sites are therefore not equivalent. The highest-affinity binding site is most likely the one inbetween the two G-quartets, while the lower-affinity binding site is the one between the G-quartet and the G-triplet. One cation suffices to stabilize a 2-quartet G-quadruplex structure, which is antiparallel. Upon addition of higher amounts of KCl and population of the second binding site, the negative band at 260 nm becomes less negative, due to the triplet stacking, resulting in a hybrid-like CD spectrum (26). Mechanistically, given that no dramatic topology rearrangements is required between the 1-K⁺ and the 2-K⁺ structure, the K⁺ binding events could be sequential.

The 21G and 22AG sequences have similar bind- ing characteristics as 22GT, suggesting potentially similar folding pathways. Interestingly, the sequence d(A(GGGTTA)₃GGGT), a close derivative of both 22GT and 22AG, was also shown to form an antiparallel G-quadruplex with two quartets in potassium (64). So, although the 22AG sequence folds into an antiparallel 3-quartet G-quadruplex in NaCl (27), it may form a 2-quartet, antiparallel structure at low KCl concentrations. A very similar CD spectrum has previously been deconvoluted by Bončina et al. for 22AG in presence of a ligand (65). The 22CTA sequence behaves differently from 21G, 22AG and 22GT, in that the 1-K⁺ stoichiometry is the only one significantly populated (with an antiparallel signature in CD, see Supporting Supplementary Figure S24H). The 22CTA NMR structure does not involve G-triplets, and this could prevent the creation of additional K⁺ binding sites.

In summary, an antiparallel 1-K⁺ complex, presumably containing two G-quartets, exists at equilibrium, particularly for human telomeric sequences with no or few flanking bases. In such case, a second K⁺ cation can subsequently bind with lower affinity and without structural change, and a sequential binding pathway is plausible. For sequences containing longer overhangs, the 1-K⁺ complex is formed only at low KCl concentration and sometimes with low abundance. For those sequences we therefore wanted to determine whether the 1-K⁺ structure was transiently populated along the folding pathway, and if yes, whether the folding pathways are sequential or branched. To elucidate the pathways, we studied the kinetics of potassium binding to those sequences.

Kinetics of K⁺ binding

The 1-K⁺ stoichiometry is transiently populated, even for sequences that bind 2 K⁺ very cooperatively. On the minute to the hour time scale, we followed the potassium binding stoichiometry by ESI-MS at room temperature (20°C). The recording is started when KCl is added to the solution (t_0). After subtraction of the nonspecific adducts, we distinguish and quantify three different stoichiometries (0-K⁺, 1-K⁺ and 2-K⁺) as a function of reaction time.

The least cooperative sequences (21G to F21T in Table 1) form both $1-K^+$ and $2-K^+$ ensembles faster than the dead time of our experiment (Supporting Supplementary Figures S25-S29). In contrast, K⁺ binding to 26TTA, 23TAG and 24TTG was slower (on 20-min to the hour time scale, 26TTA being the fastest, Supporting Supplementary Figures S30-S32). We investigated more in depth the sequences 24TTG, 23TAG and Pu24, because they gave clearer variations of the signals in the KCl concentration range accessible here. Figure 4 shows the temporal evolution of the K⁺ binding stoichiometries for 24TTG (500 µM KCl) as well as the quantification of the three stoichiometries as a function of time. Those experiments have been quantitatively reproduced on two different MS instruments (Supplementary Figure S33). The transient 1-K⁺ ensemble is clearly detected even for the very cooperative K⁺-binding sequence 24TTG (kinetics at other KCl concentrations are shown in supplementary Supplementary Figure S31). This ensemble also appeared for 23TAG and Pu24 (Supplementary Figure S34).

G-quadruplexes fold via multiply branched pathways. The abundances of 0-K⁺, 1-K⁺ and 2-K⁺ ensembles, measured at four different KCl concentrations (i.e. a total of 12 curves), were fitted globally using over 20 binding models (listed in supporting information Scheme S1). We included some that had been already suggested in the literature. The models, ranked based on the Akaike information criteria—AIC—calculated by DynaFit) are shown in supporting Supplementary Tables S1–S3 for sequences 24TTG, 23TAG and Pu24, respectively. We also visually examined the residuals, and the coefficients of variation (CV) on the fitted rate constants.

The 0-1-2 model, proposed based on UV spectroscopy (66), shows good CV on the rate constants, but very bad residuals when used to fit our data (see Figure 4B and Supplementary Figure S35 for 24TTG), indicating that this model does not properly account for the experimental data. The branched reaction model $0 \times (1|2|2)$ with two different 2-K⁺ complexes is derived from a recent paper from the Schwalbe group (67). The authors demonstrated by NMR that, for 24TTG and 26TTA, the hybrid-2 is formed first before unfolding to form the hybrid-1. We added a 1-K⁺ ensemble (not mentioned in their study) as another branched reaction. The residuals are lower, but still clearly deviate from zero (Figure 4C and Supplementary Figure S36). We also fitted our data with the model proposed by Chaires (named here $0 \times ((2-1-2)|(2-1-2)|)$ (37), involving first the folding into antiparallel G-quadruplexes, which convert to G-triplexes, before proper folding into hybrid Gquadruplexes. The antiparallel structure was proposed to

have three quartets (2-K⁺), and the triplex ensemble was proposed to bind 1 K⁺. This model works better, but does convincingly fits of our data even if it contains more species and more adjustable parameters (Supplementary Figure S37). However, if we assume that the antiparallel structures of Chaires' mechanism have only one K⁺ ion bound, we reduce the number of species in the mechanism and the fitting becomes more satisfactory (Figure 4D and Supplementary Figure S38, $0 \times ((1-2))(1-2)$) model). The residuals are good, but the CV values are large for most rate constants.

Besides, this is not the only folding mechanism that might fit. Here we have to evaluate models in a framework where the data sets may not cover the whole parameter space of the model: our experiments cover a limited concentration range and time scale, and we separate only three ensembles based on stoichiometry. In the Akaike framework, all tested models are to some extent wrong, but the AIC helps finding the model(s) that maximize predictive accuracy (62). After exploring over 20 binding models, we found that all mechanisms giving satisfactory AIC shared the following characteristics: they have are multiple branches, the two main ones leading respectively to a significantly populated 1-K⁺ ensemble, and to the final 2-K⁺ structures. For 24TTG, three ensembles with 2-K⁺ stoichiometries can be distinguished kinetically: two form slowly on the time scale of the experiment, and one is in fast equilibrium (hence the CVs of its $k_{\rm on}$ and $k_{\rm off}$ are high, and actually only the ratio $k_{\rm on}/k_{\rm off}$ can be reliably determined). The best-ranked model for 24TTG is shown in Figure 4E and is compared with the other best ones in Supplementary Figures S39 and S40.

We performed the same pathway analysis for the kinetics of K⁺ binding to 23TAG (Supplementary Table S2). The same set of mechanisms with multiple branched reactions were found to fit the best. The most satisfactory fit is obtained with one fewer 2-K⁺ ensemble (Supplementary Figure S41), using the $0 \times ((1-2)|(1-2)|)$ model. This model is a simplification of the best $0 \times ((1-2)|(1\times(2|2)))$ model, and corresponds to Chaires' model, as discussed above. The coefficients of variations were high for several rate constants, revealing that two species (1b and 2b) form fast on the time scale of our experiments. The fits using other satisfactory models are shown in supplementary Figures S42–S44.

Finally, in order to test the general applicability of the highly branched pathways, we also studied the folding of the c-myc Pu24 sequence (Supplementary Table S3 and Supplementary Figure S45). The folding of Pu24 is even slower than that of 24TTG (>20 min at 1 mM KCl). Again, we observed clearly a transient $1-K^+$ complex. We found that the same family of branched mechanisms fits the data. Strikingly, the best-ranked mechanism for the c-myc Pu24 is the same as for the telomeric sequence 24TTG. The values of the rate constants tell us that the 1b and 2b ensembles are in fast equilibrium with the 0 and 1a ensembles, respectively. In summary, highly branched folding pathways with multiple branched reactions is valid for G-quadruplexes of parallel topology as well as for the telomeric sequences, and therefore potentially constitute a tenet of G-quadruplex folding.

Ion mobility spectrometry confirms the presence of up to three distinct ensembles with two potassium cations bound. Separation methods such as size-exclusion chromatogra-



Figure 4. Kinetics of K⁺-binding (20°C) monitored by mass spectrometry of 10 μ M 24TTG in 100 mM TMAA after adding KCl to reach a final concentration of 500 μ M. (**A**) Mass spectra after addition of KCl at different time (integrated over 1 min accumulation). (**B**–**E**) Concentration of the complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of time after K⁺ addition. The data were fitted using the models described above each graph.

phy (SEC) are able to separate multimers but not Gquadruplexes of different topologies (68). For this reason, we used ion mobility spectrometry (56,69–71) to separate in the gas phase ions of a given mass-to-charge ratio according to a second dimension linked to their shape. For a given mass and charge, the time t_D needed for the ions to drift through the helium-filled cell, in which a uniform electric field is applied, will be influenced by the ion shape: a more extended conformer will be subjected to more friction and will travel more slowly. For G-quadruplex structures, the Gstack core is preserved in the gas phase upon gentle ionization and ion transfer conditions (72). If several solution ensembles give rise to m/z ensembles of significantly different collision cross sections, these will be separated based on their drift time.

A full scale IMS-MS spectrum is shown in the supplementary Figure S46. For each potassium binding stoichiometry, we integrated the ion mobility drift time distribution, at different reaction time intervals following KCI addition (Figure 5A and B for $(24TTG)^{4-}$ and $(23TAG)^{4-}$ respectively at different time points after addition of 500 µM KCI; for the other KCl concentrations and for charge state 5-, see supplementary Figures S47–S49). The drift time distribution reflects the distribution of shapes for each K⁺ binding stoichiometry. For the 2-K⁺ complex of 24TTG, we observe very clearly three peaks at short reaction time scale, indicating three conformational ensembles for the 2-K⁺ stoichiometry. This observation validates a mechanism with three different species having 2-K⁺, and discards the $0 \times ((1-2))(1-2))$ and $0 \times (111 \times (212))$ models.

Using the kinetic parameters obtained from the fitting we simulated the fraction of each of the ensembles as a function of time (see Figure 5C for the best model

 $0 \times ((1-2)|(1 \times (2|2)))$ on 24TTG, and Supplementary Figures S38–S40 for the other models). The three complexes with 2-K⁺ appear at different time scales. One is formed fastest then disappears (2c), one forms slowest (2a) and the last one forms with intermediate speed (2b). Consistent results between simulated relative abundances and ion mobility peak abundances are obtained at the other KCl concentrations for the mechanisms.

For 23TAG, the simpler $0 \times ((1-2)|(1-2)|)$ model is compatible with the IMS data: only two ensembles are distinguished by ion mobility on the 2-K⁺ stoichiometry (Figure 5B Supplementary Figure and S49). For the 1-K⁺ stoichiometry, we observe only one broad mobility peak (Supplementary Figures S47–S49). This is also consistent with the abundances predicted by the best mechanisms: indeed, only one of the two 1-K⁺ species (1b) is significantly populated. For Pu24, ion mobility separation gives only one 2-K⁺ ensemble (not shown). Note however that, although several ion mobility peaks indicate several populations initially in solution, the reverse is not true: several populations can end up in structures of very similar collision cross sections, not separable by ion mobility spectrometry.

Circular dichroism characterization of the transient species.

To obtain more structural information on the subensembles, we investigated their stacking arrangements by CD spectroscopy. The binding/folding kinetics were monitored by CD and correlated with the abundances predicted from ESI-MS (Supplementary Figures S50-58). Using a global fitting with the best models, we generated the CD spectra for each kinetically distinguishable species that is significantly populated (i.e. species 0, 1b, 2a, 2b and 2c for 24TTG, see Figure 5E and Supplementary Figure S38-40,



Figure 5. (A and B) Arrival time distribution in function of the time for the 4- charge state for the $2-K^+$ stoichiometry at 500 μ M KCl from green (3 min) to red (24h) through gray, respectively for 10 μ M (A) 24TTG and (B) 23TAG. On each panel, arrival time distributions were area-normalized for visualization of the relative abundances. (C and D) Simulation of the concentrations of the species obtained using the best models at 500 μ M KCl: (C) $0 \times ((1-2)|(1 \times (212)))$ for 24TTG and (D) $0 \times ((1-2)|(1-2)|)$ for 23TAG. (E and F) Deconvoluted CD spectra for the ensembles defined by the kinetics model, respectively for (E) 24TTG and (F) 23TAG.

bottom panel). In practice, the lowly populated 1a species was discarded by setting its $\Delta\epsilon$ to zero at all wavelengths. Similarly, we extracted the CD spectra of four species for 23TAG (Figure 5F and Supplementary Figure S41), and five for Pu24 (Supplementary Figure S45).

For all telomeric sequences, the $0-K^+$ species' CD is the one at the starting point of the kinetic. Two species with $2-K^+$ (2a and 2b) present CD spectra typical of hybrid Gquadruplex stacking (maxima at both 295 and 265 nm). They however differ around 245 nm. There is currently no interpretation of the 245-nm CD signal in terms of structure, and this subject may be worthy of further investigation. For 24TTG, the third $2-K^+$ species (2c), is never very abundant over time, so the molar ellipticity coefficients are entailed with larger errors and the reconstruction of its CD spectrum gave different results depending on the kinetic

model. Using the model $0 \times ((1-2)!(1 \times (2!2)))$, a negative band at around 270–280 nm is obtained, which is remi-

niscent of the recently published CD spectrum of a left handed G-quadruplex (73) (even though less intense in our case). Using the model $0 \times (1|2|(1 \times (2|2)))$ (Supplementary Figure S40), however, two positive peaks between 250 and 290 nm are obtained, reminiscent of the CD spectrum reported for the *Giardia* telomeric sequence d(TAGGG)4, which folds into an antiparallel 2-quartet form (26,74). Both structural hypotheses for these off-pathway 2-K⁺ folds are worthy of future investigation by higher-resolution techniques or by molecular modeling.

The other major off-pathway structure of 24TTG, i.e. the 1-K⁺ ensemble '1b', robustly presents an antiparallel-type CD spectrum, independently of the fitting model. For 23TAG as well, species '1b' shows a negative band around 260–270 nm and a positive band around 290 nm (Supplementary Figure S41). This highlights again the prevalence of $1-K^+/2$ -quartet antiparallel structures in the folding mechanism, which is off-pathway as compared to the

most stable 2-K⁺ species. The deconvolution of the CD for Pu24 gave intriguing results as well (Supplementary Figure S45). Both off-pathway species (1b and 2c) display positive CD signals at 260 and 290 nm. The spectrum of 1b is again very similar to the CD spectrum reported for the 2-quartet antiparallel *Giardia* telomeric sequence d(TAGGG)₄. The CD spectrum of 2c, which forms sequentially from 1b, is similar to that of 1b but with a higher intensity for the positive peak at 260 nm, and this could be due to an additional homo-stacking on the 'Giardia-like' antiparallel 2-quartet structure.

Finally, to check if the proposed mechanisms also hold for higher KCl concentrations and temperatures, we monitored the folding of 24TTG and Pu24 by CD at 37°C with 1 and 25 mM KCl (Supplementary Figures S55 and S58). For 24TTG, the folding remained slow even at 25 mM KCl, as also shown by Chaires' group (37). The spectrum recorded <0.8 min after the addition of KCl (highlighted in black in Supplementary Figure S55) presents a significant peak at 295 nm, typical of antiparallel arrangements. At 37°C, the folding of Pu24 is faster (Supplementary Figure S58), yet the short time scale spectrum presents a band at 298 nm as also monitored in the previous conditions. The folding kinetics monitored at 298 nm showed that this peak persists for about 5 min. These experiments confirm that offpathway antiparallel structures linger for minutes, even at higher KCl concentration and physiological temperature.

DISCUSSION

In summary, our results suggest a general G-quadruplex folding landscape in which 1-K⁺ complexes, having presumably antiparallel 2-quartet structures, form fast, but are off-pathway as compared to the final states. These antiparallel 2-quartet structures can therefore be considered as 'misfolded' structures.

The 1-K⁺ antiparallel G-quadruplex structures

Potassium ions can bind with very different cooperativity to different telomeric sequences. The shortest sequences have the lowest K⁺ binding cooperativity, and the 1-K⁺ structure predominates at low KCl concentration. This was unexpected for sequences such as 21G and 22AG, but for sequences 22GT and 22CTA this observation is in line their 2-quartet structure (hence presumably one preferred inter-quartet binding site). The Chaires group mea- sured indirectly the K^+ /oligonucleotide stoichiometry by measuring the kfree concentration with a fluorescent probe, and deduced that a wide range of stoichiometries can be formed (75). By MS, we observe directly the complexes with only one or two specific potassium ions, the rest being nonspecific adducts that also form in non-Gquadruplex sequences. The Chaires group also observed a complete folding at 2-3 mM KCl, concentration at which the *ip*-inding stoichiometry is 2:1 (K:DNA) (43), and this fully agrees with our results. Higher stoichiometries described previously could therefore come from nonspecific interaction with the negatively charged backbone (14). Single molecule FRET imaging also revealed two folded ensembles for 21G (d((GGGTTA)₃GGG)) (52), or 23TAG

(d(TA(GGGTTA)₃)) in which a low FRET ensemble is mainly present at low KCl concentration (76). In both studies, those ensembles were associated to parallel and antiparallel/hybrid structures. Based on our results, we would propose the 'antiparallel' species to be the $1-K^+/2$ -quartet antiparallel G-quadruplex and the 'parallel' ones to be the folded hybrid G-quadruplexes.

Using differential scanning calorimetry (DSC), the Vesnaver group proposed a folding mechanism for 22AG in which both a folded, an unfolded and an intermediate are in equilibrium (53). The authors also estimated the difference in number of cations bound to these three states, and found 1.5 \pm 0.1 cations for the intermediate and 3.2 \pm 0.2 for the final folded structure. The numbers of cations bound, extracted from fitting, do not match the integer numbers detected directly by mass spectrometry, but the difference could again reflect the extra nonspecific adducts. The intermediate is often considered in the literature as a triplex, but in light of our results we rather propose the major intermediate ensemble to be antiparallel structures with one K⁺ and two G-quartets, and the final form to be hybrid structures with $2-K^+$ and, depending on the sequences, either three Gquartets or two G-quartets and one G-triplet.

Increasing the number of flanking bases decreases the population of the 1-K⁺ complex to the profit of the 2-K⁺ one. Vorlìčková's group reported that if flanking bases are added the CD signature becomes less antiparallel-like and more hybrid-like (77). Here, we bring the crucial information that the antiparallel structures actually contain only one specifically bound potassium ion.

The sequence-dependent K⁺ binding cooperativity

Past reports based on DSC also showed that the f,G° of folding of human telomeric G-quadruplexes could vary by up to ~7 kcal/mol from one human telomeric sequence to another (78). The large differences can actually be correlated to the influence of a single nucleotide on the K⁺ binding constants and binding cooperativity. Here, we show that the K⁺ binding constants (and therefore the f,G° values) are highly influenced even by a single flanking base or by flanking fluorophores.

In traditional spectroscopic methods, the binding cooperativity can be assessed by the Hill coefficient (79). For the short variant 21G (d((GGGTTA)₃GGG)) (66), a Hill coefficient of 0.9 \pm 0.1 was obtained by the Balasubramanian group. This indicates a non-cooperative binding of K⁺, in agreement with our results on the shortest sequences. The authors also reported a positive cooperativity with a Hill coefficient of 1.8 \pm 0.1 for TERRA, a parallel RNA Gquadruplex. Our experiments on Pu24 show that the K⁺ binding to this parallel G-quadruplex is also very cooperative. The Chaires group also commented on potassium (and sodium) binding cooperativity, and showed that adding nucleotides increases the binding cooperativity (80). Our results are in line with previous studies, but the ability to quantitatively determine the individual $K_{\rm D}$ values for the first and second potassium binding event (Equations 2 and 3, Table 1) refines the interpretation of equilibrium binding data obtained from calorimetric or spectroscopic studies. Similar effects have been reported when terminal thymine

residues are added to synthetic sequences (81), or when fluorescent probes increase the K^+ binding cooperativity and modify the folding thermodynamics (31).

A generalized G-quadruplex folding pathway

Several G-quadruplex folding pathways have been proposed previously, including intermediates such as hairpins and triplexes. Thanks to the ability to separate populations based on the K⁺ binding stoichiometry and on distinct kinetic ensembles, we can propose the generalized folding landscape shown in Figure 6. The kinetic data revealed a complex mechanism with several branched reactions resulting in several 'branches' in the folding landscape. At least five groups of structures are involved for 23TAG and six for 24TTG: one unfolded ensemble, two ensembles with only one potassium cation and two or three ensembles with two cations.

The sequences fold into the 1-K⁺ antiparallel Gquadruplex (1b), then grab one more cation (2c) to form a 22GT-like structure (or a Giardia-like structure). The reason this pathway is the fastest may lie in probability: there are many more ways to form a misaligned (or 'misfolded') 2-quartet quadruplex than ways to form a perfectly aligned 3-quartet structure, so although the latter may be enthalpically favorable, the former may be entropically favorable. Future temperature-dependent experiments or calorimetric experiments will help in validating this hypothesis. Importantly, the 1-K⁺ antiparallel ensemble is significantly populated at the beginning of the reaction and persists on time scales up to 20 min at 20°C and 1 mM KCl (even though it is not a thermodynamically stable final state for all sequences longer than 22-nt). It is still detected at 37°C and 25 mM KCl. These structures later rearrange into the proper hybrid-1 (2a) and/or -2 (2b), but to do so they must first completely turn back to unfolded forms without specifically bound cations. This turning back is complete for sequences 23TAG and 24TTG. Previous kinetic studies indicated that the folding of 24TTG is about five times slower than for 22AG when initiated with 25 mM KCl addition (80) and we observe the same trend. The turning back may however not always be complete, and some sequences may permanently have part of their population sampling the 'misfolded' pathway. Chaires also observed that the folding of 22AG is faster than sequences with flanking bases (37). It is possible that the lack of flanking bases lowers the steric hindrance, stabilizing the '2c' complex and therefore stopping the folding in this state. The folding landscape proposed therefore also holds for those short sequences, which preferentially take the fastest branch (0-1b-2c).

Previous works have been devoted to elucidating the folding pathways of G-quadruplexes (82), and multi-pathway mechanisms were also predicted structurally by the extensive modeling work of the Sponer group (49,55). Hairpins and triplexes are commonly invoked reaction intermediates. Where would these intermediates fit within our classification according to the number of specific cations bound to each state? Mashimo *et al.* (83) proposes a collapse from the unfolded species to a hairpin ensemble, and the Plavec group has demonstrated cation-independent formation of GG base pairs (44). Hairpins would therefore belong to our 0-K⁺ ensemble. The G-triplex is also a commonly invoked intermediate (37,38,53). Our experiments could neither directly confirm nor dismiss any triplex structure. However, taking into account that molecular modeling shows significant cation affinity for the triplets (49), and that the control sequences used to subtract nonspecific adducts contained fewer than three G-tracts, a G-triplex structure is a good candidate for our low-abundance 1-K⁺ intermediate 1a. Based on existing data (67), we attributed the firstly folded 2-K⁺ hybrid G-quadruplex to the hybrid-2, whereas the hybrid-1 is formed at longer time scales. The rate constants then suggest that intermediate 1a is structurally closer to the hybrid-2 than to the hybrid-1 structure. Molecular modeling shows that hairpins and G-triplexes are stable only when their strands are antiparallel (49,54). We therefore propose that for telomeric sequences, the 1a intermediate would have its 5¹-end folded in an antiparallel manner. The fact that 1a is not abundantly populated means that if its formation is the rate-limiting step en route to folding into perfect Gquadruplex structures.

Interestingly, the folding landscape proposed here holds also for the Pu24 sequence (the only difference being that the final 2-K⁺ structures are parallel). A branched mechanism is needed to fit the data, and an antiparallel 2-quartet structure is consistently a prominent off-pathway species. The mechanism is also valid for the recently studied synthetic sequences $d((G_3T_2)_3G_3)$ and $d(T(G_3T_2)_3G_3T)$ (81), which were presented as artificial structural switches that respond to cation presence and cation concentration.

CONCLUSION

The detection and quantification of the $1-K^+$ and $2-K^+$ ensembles, made possible by ESI-MS, provide novel insight into the folding pathway of DNA G-quadruplexes. Unlike the sequential folding pathway that is typically considered, we demonstrated that misfolded species may be formed, and significantly influence the global folding kinetics. In that respect, nucleic acids (and especially G-quadruplex folding) differs from proteins in that their folding energy landscape is roughened by numerous hydrogen bonds and ionic interactions, which create high barriers between ensembles. As a result, misfolded structure are not just a mere frustration along the folding pathway, but can persist on the second to the minute time scale.

The importance of 2-quartet misfolded structures should therefore not be ignored. For example, 2-quartet structures have been found in the context of longer, more biologicallyrelevant sequences (84). In the oligonucleotide sequences studied here, often used for biophysical studies, the $1-K^+$ structures are not negligible. Our results also highlight the critical effect of choosing the exact sequence (terminal bases or mutations) and of labeling the sequences, as those modifications can drive the folding into different pathways. This aspect should be taken into account for future studies, and in comparing published results obtained on different sequence variants.

Admittedly the potassium concentration and temperature of our experimental studies are not the physiologically relevant ones. Changing the concentrations will change the relative abundances of the different ensembles in the pop-



Figure 6. Proposed folding pathway. The number of cations is shown by the number of green circles, and the quartet guanine relative orientations are shown by arrows. Example topologies compatible with each ensemble are shown in gray (*anti* bases in light gray, *syn* bases in dark gray).

ulation, yet the folding pathway is an inherent property of the system. Besides, even if 2-quartet 'misfolded' quadruplexes are a minor part of the population, conformational selection processes may still fish them out. The implications of the proposed folding pathway are therefore also important in the field of G-quadruplex ligands design. Indeed, the characterization of ligand binding to G-quadruplexes should not be limited to binding affinity, but should also consider ligand-induced changes in the target conformation. It was recently shown that the binding of some of the most potent ligands known to date (360A, Phen-DC3 and pyridostatin) was accompanied by the ejection of one cation (85, 86), suggesting that the ligands preferentially bind to 2quartet antiparallel structures. Species that are off-pathway in the folding of model oligonucleotides in vitro can become the predominant species in another environment. One should therefore not overlook that ligands and/or proteins may displace the equilibrium toward the 'misfolded' 1-K⁺ antiparallel structures. Better understanding of the folding pathways, including minor branches, may help revealing new structural targets and new strategies for inter-Gquadruplex selectivity.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors acknowledge Miho Takematsu for technical assistance and Frédéric Rosu for technical support and reading the manuscript. We thank the Structural Biophysical Chemistry platform of the Institut Européen de Chimie et Biologie (CNRS UMS 3033/Inserm US001/Univ. Bordeaux) for access to the mass spectrometers.

FUNDING

Inserm [ATIP-Avenir Grant no. R12086GS]; Conseil Régional Aquitaine [20121304005]; EU Seventh Framework Programme [FP7-PEOPLE-2012-CIG-333611]. Funding for open access charge: FP7 Marie Curie Career Integration Grant 'BIOPHYMS'.

Conflict of interest statement. None declared.

REFERENCES

- 1. Maizels,N. and Gray,L.T. (2013) The G4 Genome. *PLoS Genet.*, **9**, e1003468.
- 2. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) 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.*, **99**, 11593–11598.
- Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, 3, 218–221.
- Cahoon,L.A. and Seifert,H.S. (2009) An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science*, 325, 764–767.
- 5. Paeschke,K., Capra,J.A. and Zakian,V.A. (2011) DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*, **145**, 678–691.
- Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*40, 4727–4741.
- Cogoi,S. and Xodo,L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, 34, 2536–2549.

- Sun,D., Guo,K., Rusche,J.J. and Hurley,L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, 33, 6070–6080.
- 9. Wang,Y. and Patel,D.J. (1992) Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry*, **31**, 8112–8119.
- Mergny, J.-L., Riou, J.-F., Mailliet, P., Teulade-Fichou, M.-P. and Gilson, E. (2002) Natural and pharmacological regulation of telomerase. *Nucleic Acids Res.*, 30, 839–865.
- Rizzo, A., Salvati, E., Porru, M., D'Angelo, C., Stevens, M.F., D'Incalci, M., Leonetti, C., Gilson, E., Zupi, G. and Biroccio, A. (2009) Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic Acids Res.*, 37, 5353–5364.
- Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, 5, 182–186.
- Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, 334, 364–366.
- Wong,A. and Wu,G. (2003) 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., 125, 13895–13905.
- Largy, E., Mergny, J.-L. and Gabelica, V. (2016) Role of alkali metal ions in G-quadruplex nucleic acid structure and stability. In: Sigel, A, Sigel, H and Sigel, RKO (eds). *The Alkali Metal Ions: Their Role for Life*. Springer, pp. 203–258.
- Parkinson,G.N. (2006) Fundamentals of quadruplex structures. In: Neidle,S and Balasubramanian,S (eds). *Quadruplex Nucleic Acids*. Royal Society of Chemistry, Cambridge, pp. 1–30.
- 17. Webba da Silva, M. (2007) Geometric formalism for DNA quadruplex folding. *Chemistry*, **13**, 9738–9745.
- Davis, J.T. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, 43, 668–698.
- Tintoré, M., Eritja, R. and Fábrega, C. (2014) DNA nanoarchitectures: steps towards biological applications. *ChemBioChem*, 15, 1374–1390.
- Zhou, J., Amrane, S., Korkut, D.N., Bourdoncle, A., He, H.-Z., Ma, D.-L. and Mergny, J.-L. (2013) Combination of i-motif and G-quadruplex structures within the same strand: formation and application. *Angew. Chemie Int. Ed.*, **52**, 7742–7746.
- Phan,A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.*, 277, 1107–1117.
- 22. 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.
- 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.
- 24. 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.
- 25. 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.
- Karsisiotis, A.I., Hessari, N.M., Novellino, E., Spada, G.P., Randazzo, A. and Webba da Silva, M. (2011) Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed. Engl.*, 50, 10645–10648.
- 27. Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d(AG3(T2AG3)3) G-tetraplex. *Structure*, **1**, 263–282.
- Gray,R.D., Petraccone,L., Trent,J.O. and Chaires,J.B. (2010) Characterization of a K+-induced conformational switch in a human telomeric DNA oligonucleotide using 2-aminopurine fluorescence. *Biochemistry*, 49, 179–194.
- Buscaglia, R., Gray, R.D. and Chaires, J.B. (2013) Thermodynamic characterization of human telomere quadruplex unfolding. *Biopolymers*, 99, 1006–1018.

- Miller, M.C., Buscaglia, R., Chaires, J.B., Lane, A.N. and Trent, J.O. (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d(AG3(TTAG3)3). J. Am. Chem. Soc., 132, 17105–17107.
- De Rache, A. and Mergny, J.-L. (2015) Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie*, 115, 194–202.
- 32. Lim,K.W., Alberti,P., Guédin,A., Lacroix,L., Riou,J.F., Royle,N.J., Mergny,J.L. and Phan,A.T. (2009) Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G·C·G·C tetrad. *Nucleic Acids Res.*, **37**, 6239–6248.
- Hansel, R., Lohr, F., Foldynova-Trantirkova, S., Bamberg, E., Trantirek, L. and Dotsch, V. (2011) The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.*, 39, 5768–5775.
- 34. Li,J., Correia,J.J., Wang,L., Trent,J.O. and Chaires,J.B. (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res.*, 33, 4649–4659.
- Phan,A.T., Kuryavyi,V., Gaw,H.Y. and Patel,D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.*, 1, 167–173.
- Gabelica, V. (2014) A pilgrim's guide to G-quadruplex nucleic acid folding. *Biochimie*, 105, 1–3.
- Gray, R.D., Trent, J.O. and Chaires, J.B. (2014) Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.*, 426, 1629–50.
- Koirala,D., Ghimire,C., Bohrer,C., Sannohe,Y., Sugiyama,H. and Mao,H. (2013) Long-loop G-quadruplexes are misfolded population minorities with fast transition kinetics in human telomeric sequences. *J. Am. Chem. Soc.*, **135**, 2235–2241.
- Mergny, J.-L. and Lacroix, L. (2009) UV Melting of G-Quadruplexes. In: *Current Protocols in Nucleic Acid Chemistry*. John Wiley & Sons, Inc., Hoboken, pp. 17.1.1–17.1.15.
- De Cian,A., Guittat,L., Kaiser,M., Saccà,B., Amrane,S., Bourdoncle,A., Alberti,P., Teulade-Fichou,M.-P., Lacroix,L. and Mergny,J.-L. (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods*, 42, 183–195.
- Mergny, J. and Lacroix, L. (2003) Analysis of thermal melting curves. Oligonucleotides, 537, 515–537.
- 42. Gray, R.D., Buscaglia, R. and Chaires, J.B. (2012) Populated intermediates in the thermal unfolding of the human telomeric quadruplex. J. Am. Chem. Soc., 134, 16834–16844.
- 43. Gray, R.D. and Chaires, J.B. (2012) Isothermal folding of G-quadruplexes. *Methods*, **57**, 47–55.
- Ceru,S., Sket,P., Prislan,I., Lah,J. and Plavec,J. (2014) A new pathway of DNA G-quadruplex formation. *Angew. Chem. Int. Ed. Engl.*, 53, 4881–4884.
- 45. Mashimo, T. and Sugiyama, H. (2007) Folding pathways of human telomeric hybrid G-quadruplex structure. *Nucleic Acids Symp. Ser.* (*Oxf*)., doi:10.1093/nass/nrm120.
- Mashimo, T., Sannohe, Y., Yagi, H. and Sugiyama, H. (2008) Folding pathways of hybrid-1 and hybrid-2 G-quadruplex structures. *Nucleic Acids Symp. Ser. (Oxf).*, doi:10.1093/nass/nrn208.
- 47. Rajendran, A., Endo, M., Hidaka, K. and Sugiyama, H. (2014) Direct and single-molecule visualization of the solution-state structures of G-hairpin and G-triplex intermediates. *Angew. Chem. Int. Ed.*, **53**, 4107–4112.
- Su,D.G.T., Fang,H., Gross,M.L. and Taylor,J.-S.A. (2009) Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc. Natl. Acad. Sci. U.S.A.*, 106, 12861–12866.
- 49. Stadlbauer, P., Trantírek, L., Cheatham, T.E., Koča, J. and Sponer, J. (2014) Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie*, **105**, 22–35.
- Limongelli, V., De Tito, S., Cerofolini, L., Fragai, M., Pagano, B., Trotta, R., Cosconati, S., Marinelli, L., Novellino, E., Bertini, I. *et al.* (2013) The G-triplex DNA. *Angew. Chem. Int. Ed. Engl.*, **52**, 2269–2273.
- 51. Koirala,D., Mashimo,T., Sannohe,Y., Yu,Z., Mao,H. and Sugiyama,H. (2012) Intramolecular folding in three tandem guanine

repeats of human telomeric DNA. Chem. Commun. (Camb)., 48, 2006–2008.

- Lee, J.Y., Okumus, B., Kim, D.S. and Ha, T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 18938–18943.
- Bončina, M., Lah, J., Prislan, I. and Vesnaver, G. (2012) Energetic basis of human telomeric DNA folding into G-quadruplex structures. J. Am. Chem. Soc., 134, 9657–9663.
- 54. Stadlbauer,P., Kührová,P., Banáš,P., Koča,J., Bussi,G., Trantírek,L., Otyepka,M. and Šponer,J. (2015) Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res.*, **43**, 9626–9644.
- 55. Stadlbauer, P., Krepl, M., Cheatham, T.E., Koca, J. and Sponer, J. (2013) Structural dynamics of possible late-stage intermediates in folding of quadruplex DNA studied by molecular simulations. *Nucleic Acids Res.*, 41, 7128–7143.
- 56. Ferreira, R., Marchand, A. and Gabelica, V. (2012) Mass spectrometry and ion mobility spectrometry of G-quadruplexes. A study of solvent effects on dimer formation and structural transitions in the telomeric DNA sequence d(TAGGGTTAGGGT). *Methods*, 57, 56–63.
- 57. Marchand, A. and Gabelica, V. (2014) Native electrospray mass spectrometry of DNA G-quadruplexes in potassium solution. *J. Am. Soc. Mass Spectrom.*, **25**, 1146–1154.
- Cavaluzzi, M.J. and Borer, P.N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, 32, e13.
- Gabelica, V., Rosu, F. and De Pauw, E. (2009) A simple method to determine electrospray response factors of noncovalent complexes. *Anal. Chem.*, 81, 6708–6715.
- Kuzmic,P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.*, 237, 260–273.
- Akaike,H. (1973) Information theory and an extension of the maximum likelihood principle. *Int. Symp. Inf. Theory*, doi:10.1016/j.econlet.2011.12.027.
- Aho,K., Derryberry,D. and Peterson,T. (2014) Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*, 95, 631–636.
- Rosu,F., De Pauw,E. and Gabelica,V. (2008) Electrospray mass spectrometry to study drug-nucleic acids interactions. *Biochimie*, 90, 1074–1087.
- 64. Zhang,Z., Dai,J., Veliath,E., Jones,R.A. and Yang,D. (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.*, 38, 1009–1021.
- 65. Bončina, M., Hamon, F., Islam, B., Teulade-Fichou, M.-P., Vesnaver, G., Haider, S. and Lah, J. (2015) Dominant driving forces in human telomere quadruplex binding-induced structural alterations. *Biophys. J.*, **108**, 2903–2911.
- Zhang,A.Y.Q. and Balasubramanian,S. (2012) The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. J. Am. Chem. Soc., 134, 19297–19308.
- Bessi, I., Jonker, H.R.A., Richter, C. and Schwalbe, H. (2015) Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex. *Angew. Chem.*, **127**, 8564–8568.
- Largy, E. and Mergny, J.-L. (2014) Shape matters: size-exclusion HPLC for the study of nucleic acid structural polymorphism. *Nucleic Acids Res.*, 42, e149.
- Bohrer,B.C., Merenbloom,S.I., Koeniger,S.L., Hilderbrand,A.E. and Clemmer,D.E. (2008) Biomolecule analysis by ion mobility spectrometry. *Annu. Rev. Anal. Chem.*, 1, 293–327.

- Gabelica, V., Baker, E.S., Teulade-Fichou, M.-P., De Pauw, E. and Bowers, M.T. (2007) 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., **129**, 895–904.
- Shi,L., Holliday,A.E., Shi,H., Zhu,F., Ewing,M.A., Russell,D.H. and Clemmer,D.E. (2014) Characterizing intermediates along the transition from polyproline I to polyproline II using ion mobility spectrometry-mass spectrometry. *J. Am. Chem. Soc.*, **136**, 12702– 12711.
- Balthasart, F., Plavec, J. and Gabelica, V. (2013) Ammonium ion binding to DNA G-quadruplexes: do electrospray mass spectra faithfully reflect the solution-phase species? J. Am. Soc. Mass Spectrom., 24, 1–8.
- Chung,W.J., Heddi,B., Schmitt,E., Lim,K.W., Mechulam,Y. and Phan,A.T. (2015) Structure of a left-handed DNA G-quadruplex. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 2729–2733.
- 74. 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.
- Gray,R.D. and Chaires,J.B. (2011) Linkage of cation binding and folding in human telomeric quadruplex DNA. *Biophys. Chem.*, 159, 205–209.
- Long,X. and Stone,M.D. (2013) Kinetic partitioning modulates human telomere DNA G-quadruplex structural polymorphism. *PLoS One*, 8, e83420.
- Renciuk, D., Kejnovská, I., Skoláková, P., Bednárová, K., Motlová, J. and Vorlícková, M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K+ solutions. *Nucleic Acids Res.*, 37, 6625–6634.
- Lane, A.N., Chaires, J.B., Gray, R.D. and Trent, J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, 36, 5482–5515.
- Weiss, J.N. (1997) The Hill equation revisited: uses and misuses. FASEB J., 11, 835–841.
- Gray, R.D. and Chaires, J.B. (2008) Kinetics and mechanism of K+and Na+-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res.*, 36, 4191–4203.
- Largy, E., Marchand, A., Amrane, S., Gabelica, V. and Mergny, J.-L. (2016) Quadruplex turncoats: cation-dependent folding and stability of quadruplex-DNA double switches. *J. Am. Chem. Soc.*, **138**, 2780–2792.
- Chaires, J.B. (2010) Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. *FEBS J.*, 277, 1098–1106.
- Mashimo,T., Yagi,H., Sannohe,Y., Rajendran,A. and Sugiyama,H. (2010) Folding pathways of human telomeric type-1 and type-2 G-Quadruplex structures. *J. Am. Chem. Soc.*, **132**, 14910–14918.
- Hänsel, R., Löhr, F., Trantirek, L. and Dötsch, V. (2013) High-resolution insight into G-overhang architecture. J. Am. Chem. Soc., 135, 2816–2824.
- Marchand,A., Granzhan,A., Iida,K., Tsushima,Y., Ma,Y., Nagasawa,K., Teulade-Fichou,M.-P. and Gabelica,V. (2015) Ligand-induced conformational changes with cation ejection upon binding to human telomeric DNA G-quadruplexes. J. Am. Chem. Soc., 137, 750–756.
- Bončina, M., Podlipnik, Č., Piantanida, I., Eilmes, J., Teulade-Fichou, M.-P., Vesnaver, G. and Lah, J. (2015) Thermodynamic fingerprints of ligand binding to human telomeric G-quadruplexes. *Nucleic Acids Res.*, doi:10.1093/nar/gkv1167.

V.3. ADDITIONAL WORK

V.3.1. K⁺ BINDING TO SYNTHETIC SEQUENCES

In parallel with the work on the telomeric sequences, we have worked on K⁺ binding to synthetic sequences in collaboration with the group of Dr. Jean-Louis Mergny.²⁶⁶ Using potassium titrations, the presence of a 1-K⁺ complex at low concentration was evidenced also for the synthetic 222T (d(T(GGGTT)₃GGGT)) and 222 (d((GGGTT)₃GGG)) sequences (Figure 41), which suggest that the formation of a 1-K⁺ complex is not human telomeric G-quadruplex specific. Additionally, also for those sequences, the addition of flanking bases increases the K⁺ binding cooperativity. Indeed, the 1-K⁺ complex is more prominent at low KCl concentrations for 222 compared with 222T. The CD results also indicate a similar conclusion as for the human telomeric sequences: the 1-K⁺ complex is antiparallel. These discoveries expand the results obtained for the previously studied human telomeric sequences to other types of G-quadruplexes forming sequences: the K⁺ binding is not cooperative for sequences without flanking bases whereas it is cooperative as soon as 3' or 5' bases are added and the 1-K⁺ complex is antiparallel.



Figure 41. ESI-MS KCl titrations of 10 μ M 222 (left) and 222T (right), zooms on the 4- charge state. The quantification of the species as a function of KCl concentration are also reported. The fitting and K_D values were obtained using DynaFit.

In this recently published work,²⁶⁶ we discuss the use of G-quadruplexes as double switches (meaning that the conformations of the G-quadruplex change with a given condition). The first

switch is the interconversion from an antiparallel 2-quartet G-quadruplex at low KCl concentration to a parallel structure with 2-K⁺ at higher KCl concentration. Mass spectrometry was a key technique to reveal the states involved. The second exploited condition is the cation nature structural dependency of G-quadruplexes: the selected sequences 222 and 222T interconvert from an antiparallel conformation in sodium conditions to a parallel conformer in potassium.

V.4. CONCLUSIONS

In this chapter, we have used advantageously the method developed in Chapter IV. We quantified the intensities of each of the detected complexes, containing different numbers of K⁺ ions, subtracting mathematically the nonspecific adducts contribution.

- a) We evaluated the binding constants of potassium to a variety of sequences including several human telomeric sequences previously described in the literature, a c-myc promoter forming a parallel G-quadruplex and two artificial sequences. Systematically, we found that the sequences with flanking bases bind potassium cooperatively whereas the 1-K⁺ complex is prominent at low KCl concentration for the shortest sequences. This new behavior could be of interest in nanotechnology as described in the double switch paper²⁶⁶ but it also gives new information about the folding of such sequences. Indeed, the potential to form G-quadruplexes containing only 1-K⁺ is general. This 1-K⁺ complex could be an interesting target for ligands. Model sequences such as 222 and 222T could be of interest to study G-quadruplex folding landscape or interactions with small molecules.
- b) To determine whether the 1-K⁺ 2-quartet species is an intermediate on the folding pathway towards the final 2-K⁺ structure, we studied K⁺ binding kinetics by ESI-MS, and used CD to obtain hints on the structures of the 1-K⁺ species. The kinetics of formation of the 2-K⁺ complex for the cooperatively binding sequences are much slower than for the non-cooperative families. We found that up to 6 ensembles are needed to account for all the kinetic data. All best kinetic models involve a "misfolding" pathway, in which the species are quickly folded into antiparallel G-quadruplexes with only two G-quartets. The proposed model seems a general tenet of G-quadruplex folding. The 2-quartet structures could play important roles in biological pathways, and could be the preferred target of some ligands or proteins.

V.5. PERSPECTIVES

To decompose the Gibbs free energy into enthalpy and entropy contributions and to characterize the energy barriers and build a complete folding landscape, temperature-dependent experiments are needed. Additionally, it would be interesting to measure the melting temperature based on the potassium stoichiometry by mass spectrometry and identify potential unfolding intermediates.^{180,182} Unfortunately, in mass spectrometry the temperature of the experiment is room temperature (in our case 20 ± 1 °C) and is not variable.

During a collaboration with Bob Hommersom from Dr. Ron Heeren's group, we could test the nanoelectrospray ionization (nESI) source they had developed. The solution is heated (or cooled): the glass needle is held between metallic clamps, connected to a Peltier which regulates the temperature. The solution is slowly pushed in the thermostated glass needle and sprayed as in classical nESI source. This source was adapted on our Exactive mass spectrometer.

In Figure 42A, the distribution of potassium cations is shown at 25 °C and at 70 °C. At 25 °C, the human telomeric G-quadruplex (24TTG, d(TT(GGGTTA)₃GGGA)) is folded with two K⁺ (with additional nonspecific adducts) as obtained with a conventional ESI source. At 70 °C, the distribution is clearly shifted towards 0-K⁺ (free DNA), suggesting that it is unfolded. A temperature ramp of 1 °C/min from 25 to 70 °C was used to record the intensities of each stoichiometries as a function of temperature (Figure 42B). The transition from the 2, 3 and 4-K⁺ stoichiometries to the 0 and 1-K⁺ occurs at about 50 °C. We also compared the melting temperature value obtained by CD in the same conditions (Figure 42C & D). The T_m value obtained by CD (43 °C) is not perfectly the same as the one obtained by MS but they are in good agreement. To get more precise and quantitative information from the MS data, further work on the contribution of the nonspecific adducts has to be done.



Figure 42. A, MS spectra of 10 µM 24TTG in 100 mM TMAA and 1 mM KCl, with 10% iPrOH added just before injection, at 25 or 70 °C. B, Quantification of the different stoichiometries as a function of the temperature. The temperature ramp was 1 °C/min. C, CD-melting experiment in the same conditions as for the MS, from 25 °C (in blue) to 64 °C (in red). D, Evolution of the CD signal as a function of temperature at the indicated wavelengths (290, 270 and 240 nm).

Temperature-dependent mass spectrometry of G-quadruplexes was done for the first time and the (preliminary) results are promising: it is possible to use a thermostated ESI source to study the thermal denaturation of G-quadruplexes, investigating the number of bound cations. The perspectives linked with the temperature-dependent experiments are numerous. It could become possible to characterize the folding landscape of G-quadruplexes as a function of temperature and get access to the energy barriers or to investigate the unfolding intermediates of G-quadruplexes with and without ligand. The device could also be coupled with ion mobility spectrometry and study the conformational changes as a function of temperature.

VI. G-QUADRUPLEX LIGAND BINDING MODES

VI.1. INTRODUCTION

In Chapter II.2.6, plausible binding modes were presented (π - π stacking, groove binding, loop recognition or intercalation) and even if all these possibilities are theoretically possible, the most widely spread binding mode is the π - π stacking on external G-quartets. However, if this was the only binding mode, the ligands would bind with similar affinities to all G-quadruplexes (providing that the external G-quartets are exposed). G-quadruplex-selective ligands however exist.^{125,126,267} This highlights the needs of understanding binding modes of ligands to G-quadruplexes.

In addition, ligand-induced conformational changes had previously been evidenced, for example, for the human telomeric G-quadruplexes and the ligand TMPyP4. The G-quadruplex structures were influenced to form more hybrid G-quadruplexes.²⁶⁸ Similarly, NMM binds preferentially to parallel structures, which could drive the equilibrium to more parallel structures.¹²⁵ Also, Platinum-tolylterpyridine complexes induce a shift from hybrid to antiparallel G-quadruplexes.²⁶⁹ It is now clear that ligands can influence G-quadruplex structures.

To accurately characterize the binding of a ligand, we first need the stoichiometry of the complexes. Mass spectrometry is powerful to determine stoichiometries and has been used to evidence unexpected behavior for ligands binding to DNA. For example, Rosu *et al.* have shown that some ligands could bind with high cooperativity to a DNA duplex.²⁷⁰ Multiple stoichiometries are common, also for ligands binding to G-quadruplexes. Up to four Thiflavin T ligands were found to bind to human telomeric G-quadruplexes.²⁷¹ ESI-MS demonstrated the importance of considering the stoichiometry for equilibrium constants measurement.

In the first part of this chapter, copper-terpyridine (such as Cu-tolylterpyridine, Cu-*ttpy*) Gquadruplex ligands were studied. Those metal complexes have been previously described as good binders, based on the increased melting temperature they induced for Gquadruplexes.¹³² Using mass spectrometry, we were able to simultaneously detect the stoichiometries in term of number of DNA strands, ligands and cations. We evidenced a highly

133

cooperative binding to (and only to) the human telomeric G-quadruplexes. We also observed two potassium ion bound, suggesting a preferential binding to three G-quartets structures. Circular dichroism also indicated a conformational change to antiparallel G-quadruplexes.

In the second part of this chapter, we investigated the binding of Phen-DC3, 360A and PDS (described in the introduction) to human telomeric G-quadruplexes. These ligands bind with high affinity to those G-quadruplexes as previously described.^{143,144,146,152} In addition, we found that these ligands were also able to induce the formation of an antiparallel G-quadruplex with only one cation trapped and therefore presumably with only two G-quartets. This was the first description of cation released upon binding of a ligand.

This whole chapter presents mass spectrometry as a useful tool to detect structural changes through a series of unexpected binding modes.

VI.2. ARTICLE 3

Selective and Cooperative Ligand Binding to Antiparallel Human Telomeric DNA G-quadruplexes

A. Marchand, D. Strzelecka & V. Gabelica

Chemistry – A European Journal **2016**, 22, 9551-9555

For supplementary information, see Annex III.

Selective and Cooperative Ligand Binding to Antiparallel Human Telomeric DNA G-quadruplexes

A. Marchand^[a,b,c], D. Strzelecka^[a,b,d], V. Gabelica^[a,b,c]*

Abstract: The quest for ligands that bind specifically to particular Gquadruplex nucleic acid structures is particularly important to conceive molecules with specific effects on gene expression or telomere maintenance, or conceive structure-specific molecular probes. Here, using electrospray mass spectrometry in native conditions, we reveal a highly cooperative and selective 2:1 binding of Cu(II)-tolylterpyridine complexes to human telomeric G-quadruplexes. Circular dichroism and comparisons of affinities for different sequences reveal a marked preference for antiparallel structures with diagonal loops and/or widemedium-narrow-medium groove width order. The cooperativity is attributed to conformational changes in the polymorphic telomeric Gquadruplex sequences, which convert preferably to an antiparallel, 3quartet topology upon binding two ligands.

G-quadruplex structures formed by guanine-rich DNA and RNA sequences are stabilized by Hoogsteen hydrogen bonding between four guanines (forming a G-quartet), by cations coordination between the quartets, and by base stacking. Gquadruplexes in key regions of the genome can influence the DNA replication processes^[1] or the telomere structure.^[2] G-quadruplex binding ligands could influence oncogene expression and telomere maintenance, and are therefore considered as potential anti-cancer therapeutics.^[3]

Several families of ligands were reported to date.^[4] Metal complexes of planar geometry can target G-quadruplexes: the positive core interacts favourably with the external quartets and the organic moiety stabilizes the G-quadruplex by π - π stacking.^[5] Good selectivity for G-quadruplexes versus other DNA structures was reported, and planar complexes using copper, platinum and palladium show good affinity.^[6] Copper is a good G-quadruplex stabilizer when bound to a salphen^[5c] or tolylterpyridine *ttpy* (Figure 1).^[5a, 7] Metal cations were also used in porphyrin complexes, which showed telomerase^[8] or oncogene expression inhibition.^[9]

Although specificity for quadruplexes vs. other DNA structures has been achieved by several ligand families, a good selectivity between G-quadruplex families has been much more rarely reported. We studied here the binding of the Cu(II)-4'-(*p*-tolyl)-2,2':6',2"-terpyridine (Cu-*ttpy*) ligand (Figure 1) and of some

[a]	A. Marchand, D. Strzelecka, Dr. V. Gabelica*			
	Univ. Bordeaux, IECB, ARNA Laboratory			
	Pessac, France			
	E-mail:valerie.gabelica@inserm.fr			
[b]	A. Marchand, D. Strzelecka, Dr. V. Gabelica*			
	Inserm, U1212, ARNA Laboratory			
	Bordeaux, France			
[c]	A. Marchand, Dr. V. Gabelica*			
	CNRS, UMR 5320, ARNA Laboratory			
	Talence, France			
[d]	Current affiliation: Division of Biophysics, Institute of Expe			

[d] Current affiliation: Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw.

Supporting information for this article is given via a link at the end of the document.

derivatives to several oligonucleotide sequences (listed in ESI Table S1). The Cu, Zn, Pt and Pd complexes were synthesized in-house by mixing the ligand and the metal salt (Cl⁻ or NO₃·)^{I5a, 7]} (ESI Section 2 and Figures S1-S10). We report herein that the binding of Cu-*ttpy* derivatives is very specific for telomeric DNA G-quadruplexes, and that the ligand binding to telomeric sequences is accompanied by a striking 2:1 (ligand:DNA) binding cooperativity, which is correlated to changes of the telomeric structure to an antiparallel conformation with 3 G-quartets.



Figure 1. Structure of the ligand Cu-ttpy (dinitrate salt).

Electrospray mass spectrometry (ESI-MS) in nondenaturing conditions was key to study the stoichiometries of the complexes and determine their individual $K_{\rm D}$ values. Previous studies by CD and UV-vis spectroscopy had indeed failed to detect the cooperative binding. The G-quadruplexes were prepared here in ESI-compatible 1-mM potassium conditions using trimethylammonium acetate (TMAA).^[10] To determine the $K_{\rm D}$ values (Eq 1-2), we used the peak areas of the 5- charge state of free DNA and DNA bound with 1 or 2 ligands, assuming that the response factors are equal for all the species (the verifications that validate this hypothesis are presented in Figure S11). Although this sounds surprising, equal response seems the rule rather than the exception in DNA-ligand complexes: the electrospray response depends more on the total charge than on how the individual charges are distributed over the molecule.^[11]

$$K_{D1} = [DNA][L]/[DNA•L]$$
(Eq. 1)

$$\mathcal{K}_{D2} = [DNA \bullet L][L]/[DNA \bullet L_2]$$
(Eq. 2)

Figure 2 shows the ESI-MS titrations of two G-quadruplexes by the Cu-*ttpy* ligand. The curves of DNA concentration vs total ligand concentration were fitted using the DynaFit software^[12] to obtain K_{D1} and K_{D2} . The DynaFit software works by iteration on the two K_D values in order to fit all the concentrations of the three stoichiometries as a function of total ligand added. The inputs are the concentrations of each species obtained by MS (0:1, 1:1 and 2:1 L:DNA stoichiometries) and the output consist in the values of the K_D and the fits. We observed ligand binding, without nitrate

COMMUNICATION

counter-ions, to both sequences. We quantified the complexes of each stoichiometry and fitted the titration data for the human telomeric G-quadruplex 22GT (dGGG(TTAGGG)₃T),^[13] for the parallel G-quadruplex from the c-myc promoter Pu24,^[14] (Figure 2), and for the bimolecular antiparallel G-quadruplex [dG₄T₄G₄]₂^[15] (Figure S12 for full titration). For the other sequences, the *K*₀ values were estimated using single-concentration measurements.^[11a] In such case, the ligand free concentration is deduced using the total ligand concentration and the concentration of all complexes.



Figure 2. Top, MS titrations of 10 μ M 22GT or Pu24 by Cu-*ttpy* in 100 mM TMAA + 1 mM KCI. Zooms on the 5- charge state. Bottom, quantification and fitting.

Equilibrium dissociation constants for 21 sequences including 8 human telomeric sequences, 11 other Gquadruplexes (both parallel and antiparallel), one hairpin and one single strand (see ESI Table S1 for the full list of sequences) are reported in Figure 3 (see ESI Table S3 for the values). The product $K_{D1} \times K_{D2}$ represents the global constant for the binding of two ligands. The total amount of ligand bound was also calculated, and the percentage ligand bound in the 10µM:10µM (ligand:DNA) mixtures is shown in Figure 3. We observed a very good binding of Cu-ttpy to G-quadruplexes, with a high affinity for all the human telomeric ones. Remarkably, the amount of bound Cu-ttpy on telomeric sequences (>90%) is similar to what is obtained for some of the most potent G-quadruplex ligands reported to date (360A, Phen-DC3 or pyridostatin, see ESI Figure S15). Cu-ttpy binds with low affinity to the single stranded 24nonG4 and the hairpin ds26. This result confirms the selectivity in favour of G-

quadruplexes, and a particular selectivity for all telomeric sequence derivatives.

Indeed, even not all pre-formed G-quadruplexes are targets for this ligand. For example, the sequence ckit1 binds the ligand exclusively without trapping any potassium ion (the fraction unfolded in solution), but the quadruplex form (which contains 2 potassium ions) does not bind the ligand (Figure S16A). The HIV-PRO1, a 2-quartet antiparallel G-quadruplex, is not folded in our MS conditions (the 0-K⁺ form predominates). When one ligand bind to the sequence, although the amount of 1-K⁺ complex is a little increased, the ligand binds mostly to the 0-K⁺ species (Figure S16B). Cu-*ttpy* did not bind either to the thrombin binding aptamer TBA (Figure S16C), an antiparallel G-quadruplex with only two Gquartets.



Figure 3. Comparison of the log K_{D1} (orange) and log K_{D2} (green) and the percentage of ligand bound at 1 equivalent (10 μ M) for the 21 sequences for Cu-*ttpy*. The sequences are ranked according to their K_{D1}/K_{D2} ratio (Table S3).

For the telomeric sequences only, the distribution of ligand binding stoichiometries is strongly biased towards the 2:1 complex. Upon titration the 1:1 (Ligand:DNA) complex is almost not populated, in contrast with the 2:1 complex (see Figure 2). The ligand binding is therefore highly cooperative. We also tested other electrolyte conditions (100 mM NH₄OAc) to probe whether the low potassium conditions were responsible for the cooperativity, and the results were very similar to those obtained in TMAA + KCI (Figure S17).

We ranked the sequences according to their ratio K_{D1}/K_{D2} , which indicates cooperativity (Table S3),^[11a] and found a very high cooperativity ($K_{D1}/K_{D2} > 5$) for almost all human telomeric sequences (22GT, 25TAG, 23TAG, 24TTG and 26TTA) (Figure 3). The highest cooperativity is obtained for 22GT, K_{D1} being almost 2 orders of magnitude larger than K_{D2} . Only the two human telomeric sequences beginning with an adenine (22AG and 23AG) were binding Cu-*ttpy* less cooperatively ($5 > K_{D1}/K_{D2} > 0.7$). No cooperativity is observed for any of the other sequences, for example all derivatives forming parallel G-quadruplexes (c-myc promoter sequences Pu24, CMA and CMT), the [dTG₄T]₄ tetramolecular parallel G-quadruplex and 26CEB, a parallel G-quadruplex with one very long loop).

To probe which ligand properties are essential to the cooperativity, we synthetized two derivatives of the Cu-complexes (Scheme S1): Cu-tpy (lacking the tolyl group) and Cu-Clptp (Cu-Chlorophenyterpyridine, the methyl group being replaced by a chlorine, see ESI section 2). Cu-tpy was previously described as a poor binder^[5a] and Cu-Clptp is reported here for the first time. ESI-MS confirmed the poor binding affinity of the Cu-tpy complex (Figure S14), and the small amount of ligand that binds does so non-cooperatively (Figure S18 and Table S3). In contrast, Cu-Clptp behaves exactly as Cu-ttpy, although being a little bit less affine for G-quadruplexes (see Figure S13 and Table S3 for the $K_{\rm D}$ values). The presence of the fourth aromatic moiety therefore seems essential for both the high affinity and high cooperativity of ligand binding. Next, to test the effect of the metal centre, we replaced copper by Zn, Pd or Pt (Figure S18). As known already from past studies,^[5a] Zn-ttpy is not a good ligand (it binds with poor affinity to all DNA sequences), due to the trigonal bipyramidal geometry adopted by Zn(II), which leads to steric hindrance. Ptand Pd-complexes bind to the G-quadruplex^[7] but without any cooperativity: only a 1:1 complex is forming. For the Pt derivatives, the ligand first binds non-covalently, then loses chlorine and binds covalently to the DNA, as described previously.^[7, 16] Changing the metal core of the ligand abolished the cooperative behaviour. We also synthesized and tested the chloride salt of the ligands. Like for NO3⁻, the counter ion was not observed directly bound (the ligand binds as CuL²⁺). The nature of the anion did not influence the binding affinity or cooperativity of those ligands. In conclusion, both the extended aromatic moiety and the preferred square pyramidal penta-coordination mode enabled by Cu(II) are essential both for the high binding affinity^[6a] and for the binding cooperativity.

In order to get information on possible conformational changes induced by the ligand, we examined the number of potassium ions bound, as detected using ESI-MS, on the complexes with and without ligands (Figure 4A). In our conditions, the free G-quadruplex of 22GT is formed with both 1 and 2 potassium ions. The minor 1:1 complex presents a K⁺ distribution close to that of the free G-quadruplex. In contrast, the 2:1 complex is different: the 2-K⁺ stoichiometry is preferred. We previously reported potassium ions ejection from the human telomeric Gquadruplexes upon 1:1 complex formation with ligands 360A, Phen-DC3 or pyridostatin.^[17] Here with Cu-ttpy we observe the complete opposite. For all human telomeric sequences, the 2:1 complex is binding 2-K⁺ even when the free sequence did not contain purely 2-K+: 23AG, 22AG and 22CTA (ESI Figure S19). Interestingly, even if 22CTA^[18] binds the ligand in a noncooperative manner on the population level (total 2:1 complex vs. total 1:1 complex), the second ligand binding event also induces the capture of one more potassium ion. Those hints of structural changes were observed only for the human telomeric sequence.

Circular dichroism (CD) was also performed on the 22GT and 24TTG sequences to study the ligand-induced changes in preferential base-stacking (Figure 4C-D). The CD spectra indeed

display different signatures according to the stacking of the guanines involved in G-quartets: a positive peak at 260 nm is indicative of parallel G-quadruplexes, a negative peak at 260 nm and a positive at 290 nm of antiparallel G-quadruplexes, and two positive peaks at 270 and 290 nm indicate hybrid G-quadruplexes. For 22GT, the starting structure has a CD spectrum close to the typical antiparallel one. Addition of Cu-ttpy influences the CD signature, making the negative band at 260 nm more negative. The presence of isodrichroic points at 254 and 286 nm indicates a single transition between two species, i.e. the free G-quadruplex and the 2:1 complex. Following the same trend, the CD titration of 24TTG shows a transition from the hybrid structure to a more antiparallel one. On the other hand, the CD spectra of Pu24, a parallel G4, and [dG₄T₄G₄]₂, an antiparallel one, are not influenced by the binding of Cu-ttpy (Figure S20). Indeed, no change in the potassium stoichiometry was observed. Despite a small intensity decrease for Pu24's CD signal, the fact that there is no change in the CD spectrum upon binding also rules out possible ligandinduced CD signal that could affect the CD signature.

The structural preference for the antiparallel human telomeric G-quadruplex has been observed before for triarylpyridine-platinum complexes^[16,19] or a dinuclear Rucomplex.^[20] Other ligands were previously shown to alter the human telomeric G-quadruplexes, as for example TMPyP4, inducing the formation of more hybrid G-quadruplexes.^[21] Another example is the N-methyl mesoporphyrin IX that binds preferentially to parallel G-quadruplexes.^[22] Cu-*ttpy* is however the first example of a ligand for which this structural preference is accompanied by the cooperative binding of two ligands.



Figure 4. Topology characterization of the binding of Cu-*ttpy*. MS and determination of the number of potassium in the complexes for one Cu-*ttpy* equivalent A, for 22GT and B, for 24TTG. CD of C, 22GT and D, 24TTG with increasing amount of ligand.

In order assess whether the ligands are bound to proximal or distant sites, we prepared an equimolar mixture of Cu-*ttpy* and Cu-*Clptp* ligands, which was added to the 22GT G-quadruplex (Figure S21). We found a maintained cooperativity, but a purely statistical distribution of each ligand, suggesting that the nature of one ligand bound does not condition the nature of the second ligand bound. The result of this competition experiment therefore

COMMUNICATION

suggests that the ligands do not interact with the quadruplex as pre-formed dimers.

A rather common ligand binding mode in G-quadruplexes, compatible with the 2:1 stoichiometry, is stacking of one ligand to each terminal G-quartet. To test that structural hypothesis easily with mass spectrometry, we used the bimolecular G-quadruplexes $[dG_4T_4G_4]_2$ and $[G_4T_3G_4]_2$. These quadruplexes present identical environments on each terminal G-quartets, and the 2:1 Cu-*ttpy* complexes are formed with no marked cooperativity (Figure 3). The complexes were prepared in NH₄⁺, and the charge state 6- was selected so as to facilitate strand separation upon MS/MS. Upon collisional activation, the dimer⁶⁻ breaks into its two strands (Figures S22-S23), each containing exclusively either one ligand or one Cu (neutral *ttpy* being ejected from the complex). This strongly suggests that the two ligands are bound at distal positions, so the ligands presumably bind to each of the two external G-quartets.



Figure 5. A) The various topologies that the human telomeric sequence can adopt in solution. B) One of the possible preferred topology of the 2:1 complex with Cu-*ttpy*: an antiparallel structure with 3 G-quartets.

In summary, for the intramolecular telomeric Gquadruplexes, the preferred binding of Cu-*ttpy* is therefore to an antiparallel, three-quartet structure, presumably on each terminal G-quartet (Figure 5B). The ligand is also selective among several possible antiparallel structures. Indeed, the binding to the antiparallel guadruplexes HIV-PRO1 and TBA is extremely poor. These G-quadruplexes are characterized by lateral-lateral-lateral loop arrangement and a wide-narrow-wide-narrow groove width order. This is also in line with the higher binding affinity for $[dG_4T_4G_4]_2$ and $dG_4(T_4G_4)_3$ as compared to $[dG_4T_3G_4]_2$. Among antiparallel structures, the ligand Cu-ttpy apparently prefers structures with diagonal loops and/or wide-medium-narrowmedium groove width order. In the human telomeric sequences, the preferred quadruplex fold would therefore be the antiparallel with lateral-diagonal-lateral loops, i.e. the same fold as formed in Na⁺ by 22AG. Interestingly, past results showed a higher FRET-

 ΔT_m when 5 eq. Cu-*ttpy* are added to the human telomeric sequence F21T in NaCl (+15.3°C)^[5a] compared to in KCl (+6°C).^[7]

The binding of the two ligands to the human telomeric Gquadruplexes is accompanied by a conformational change. This conformational change could occur after the binding of the ligands: this mechanism is called induced fit.^[23] For ligands prone to π - π sta ck on external G-quartets, this scenario is very unlikely because ligand binding should then be followed by an unfolding of the G-quadruplex. Indeed, to form the antiparallel, 3-quartet Gquadruplex starting from any of the hybrid topologies, all Gquartets must first be broken to release at least one strand,^[24] and if starting from hybrid I or II, to allow the groove widths to rearrange in order to be compatible with the antiparallel strand arrangement. We therefore favor a conformational selection mechanism (Figure 5), in which the antiparallel, 3-guartet structure, although minor in the ligand-free form, becomes predominant upon binding of two ligands which stack on the top and bottom quartet.

In conclusion, Cu(II)-tolylterpyridinium complex derivatives bind all the human telomeric G-quadruplex variants with high affinity, high specificity, and high cooperativity to form 2:1 complexes. The cooperative ligand binding is accompanied by a conformational change to an antiparallel structure with three Gquartets. The cooperativity is due to a combined effect of (1) allosteric changes in the G-quadruplex upon ligand binding, (2) conformational preference of the ligand, and (3) the fact that one ligand is not sufficient to promote the conformational equilibrium displacement. This ligand binding mechanism points to a structural selectivity for one particular fold, and the ability to switch the conformational equilibria of all telomeric sequence variants to this fold.

Acknowledgements

This work was financially supported by the Inserm (ATIP-Avenir Grant no. R12086GS), the Conseil Régional Aquitaine (Grant no. 20121304005), and the EU (FP7-PEOPLE-2012-CIG-333611 and ERC-2013-CoG-616551). D.S. benefited from an Erasmus+ traineeship from the University of Warsaw to the University of Bordeaux. The authors thank Dr. Eric Largy and Mélanie Lenfant (M1 student, U. Bordeaux) for contributing preliminary experiments, Dr. Céline Douat (CBMN laboratory) for her assistance with the synthesis of the Zn, Pt and Pd complexes and Dr. Stéphane Massip (Structural biophysicochemistry platform, IECB) for the crystal structures of the ligands.

Keywords: mass spectrometry • G-quadruplex • ligand • cooperativity • affinity

- a) L. A. Cahoon and H. S. Seifert, *Science* 2009, *325*, 764-767; b) S. Kumari,
 A. Bugaut, J. L. Huppert and S. Balasubramanian, *Nature Chem. Biol.* 2007, *3*, 218-221; c) K. Paeschke, J. A. Capra and V. A. Zakian, *Cell* 2011, *145*, 678-691; d) A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L.
 H. Hurley, *Proc. Natl. Acad. Sci. USA* 2002, *99*, 11593-11598.
- [2] a) J.-L. Mergny, J.-F. Riou, P. Mailliet, M.-P. Teulade-Fichou and E. Gilson, *Nucleic Acids Res.* 2002, *30*, 839-865; b) A. Rizzo, E. Salvati, M. Porru, C. D'Angelo, M. F. Stevens, M. D'Incalci, C. Leonetti, E. Gilson, G. Zupi
and A. Biroccio, *Nucleic Acids Res.* **2009**, *37*, 5353-5364; c) Y. Wang and D. J. Patel, *Biochemistry* **1992**, *31*, 8112-8119.

- [3] T.-M. Ou, Y.-J. Lu, J.-H. Tan, Z.-S. Huang, K.-Y. Wong and L.-Q. Gu, *ChemMedChem* **2008**, *3*, 690-713.
- [4] D. Monchaud and M.-P. Teulade-Fichou, Org. Biomol. Chem. 2008, 6, 627-636.
- [5] a) H. Bertrand, D. Monchaud, A. De Cian, R. Guillot, J.-L. Mergny and M.-P. Teulade-Fichou, *Org. Biomol. Chem.* 2007, *5*, 2555-2555; b) N. H. Campbell, N. H. A. Karim, G. N. Parkinson, M. Gunaratnam, V. Petrucci, A. K. Todd, R. Vilar and S. Neidle, *J. Med. Chem.* 2012, *55*, 209-222; c) J. E. Reed, A. A. Arnal, S. Neidle and R. Vilar, *J. Am. Chem. Soc.* 2006, *128*, 5992-5993.
- [6] a) A. Arola-Arnal, J. Benet-Buchholz, S. Neidle and R. N. Vilar, *Inorg. Chem.*2008, 47, 11910-11919; b) S. E. Evans, M. A. Mendez, K. B. Turner, L.
 R. Keating, R. T. Grimes, S. Melchoir and V. A. Szalai, *J. Biol. Inorg. Chem.*2007, *12*, 1235-1249; c) J. Pan and S. Zhang, *J. Biol. Inorg. Chem.*2009, *14*, 401-407.
- [7] E. Largy, F. Hamon, F. Rosu, V. Gabelica, E. De Pauw, A. Guédin, J.-L. Mergny and M.-P. Teulade-Fichou, *Chem. Eur. J.* 2011, 17, 13274-13283.
- [8] a) E. Izbicka, R. T. Wheelhouse, E. Raymond, K. K. Davidson, R. A. Lawrence, D. Sun, B. E. Windle, L. H. Hurley and D. D. Von Hoff, *Cancer Res.* **1999**, *59*, 639-644; b) D.-F. Shi, R. T. Wheelhouse, D. Sun and L. H. Hurley, *J. Med. Chem.* **2001**, *44*, 4509-4523.
- [9] a) J. Alzeer, B. Vummidi, P. Roth and N. W. Luedtke, *Angew. Chem. Int. Edit.* 2009, 48, 9362-9365; b) A. Membrino, M. Paramasivam, S. Cogoi, J. Alzeer, N. W. Luedtke and L. E. Xodo, *Chem. Commun.* 2010, 46, 625-627.
- [10] A. Marchand and V. Gabelica, J. Am. Soc. Mass Spectr. 2014, 25, 1146-1154.

- [11] a) F. Rosu, E. De Pauw and V. Gabelica, *Biochimie* 2008, *90*, 1074-1087;
 b) V. Gabelica, F. Rosu and E. De Pauw, *Anal. Chem.* 2009, *81*, 6708-6715.
- [12] P. Kuzmic, Anal. Biochem. 1996, 237, 260-273.
- [13] K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu and A. T. Phan, J. Am. Chem. Soc. 2009, 131, 4301-4309.
- [14] A. T. Phan, V. Kuryavyi, H. Y. Gaw and D. J. Patel, Nat Chem Biol 2005, 1, 167-173.
- [15] P. Schultze, F. W. Smith and J. Feigon, Structure 1994, 2, 221-233.
- [16] H. Bertrand, S. Bombard, D. Monchaud, E. Talbot, A. Guédin, J.-L. Mergny, R. Grünert, P. J. Bednarski and M.-P. Teulade-Fichou, Org. Biomol. Chem. 2009, 7, 2864-2864.
- [17] A. Marchand, A. Granzhan, K. lida, Y. Tsushima, Y. Ma, K. Nagasawa, M.-P. Teulade-Fichou and V. Gabelica, *J. Am. Chem. Soc.* 2015, *137*, 750-756.
- [18] K. W. Lim, P. Alberti, A. Guédin, L. Lacroix, J. F. Riou, N. J. Royle, J. L. Mergny and A. T. Phan, *Nucleic Acids Res.* 2009, 37, 6239-6248.
- [19] J.-T. Wang, Y. Li, J.-H. Tan, L.-N. Ji and Z.-W. Mao, *Dalton Trans.* 2011, 40, 564-566.
- [20] T. Wilson, P. J. Costa, V. Félix, M. P. Williamson and J. A. Thomas, J. Med. Chem. 2013, 56, 8674-8683.
- [21] R. D. Gray, J. Li and J. B. Chaires, *J. Phys. Chem. B* **2009**, *113*, 2676-2683.
- [22] J. M. Nicoludis, S. P. Barrett, J.-L. L. Mergny and L. A. Yatsunyk, *Nucleic Acids Res.* 2012, 40, 5432-5447.
- [23] S. Gianni, J. Dogan and P. Jemth, Biophys Chem 2014, 189, 33-39.
- [24] P. Stadlbauer, M. Krepl, T. E. Cheatham, 3rd, J. Koca and J. Sponer, *Nucleic Acids Res* 2013, 41, 7128-7143.

VI.3. ARTICLE 4

Ligand-Induced Conformational Changes with Cation Ejection upon Binding to Human Telomeric DNA G-Quadruplexes

A. Marchand, A. Granzhan, K. Iida, Y. Tsushima, Y. Ma, K. Nagasawa, M-P. Teulade-Fichou & V. Gabelica

Journal of the American Chemical Society **2015**, 137, 750-756.

For supplementary information, see Annex IV.

Article





Ligand-Induced Conformational Changes with Cation Ejection upon Binding to Human Telomeric DNA G-Quadruplexes

Adrien Marchand,^{†,‡} Anton Granzhan,[§] Keisuke Iida,^{||} Yamato Tsushima,^{||} Yue Ma,^{||} Kazuo Nagasawa,^{||} Marie-Paule Teulade-Fichou,[§] and Valerie Gabelica^{*,†,‡}

[†]IECB, ARNA Laboratory, University of Bordeaux, 33600 Pessac, France

[‡]Inserm, U869, ARNA Laboratory, Institut National de la Santé et de la Recherche Medicale, 33000 Bordeaux, France [§]Institut Curie, Centre de Recherche, CNRS UMR176, Centre Universitaire Paris Sud, Bat. 110, 91405 Orsay, France ^{II}Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho. 184-8588 Koganei-shi, Tokyo Japan

* Supporting Information

ABSTRACT: The rational design of ligands targeting human telomeric DNA G-quadruplexes is a complex problem due to the structural polymorphism that these sequences can adopt in physiological conditions. Moreover, the ability of ligands to switch conformational equilibria between different G-quadruplex structures is often overlooked in docking approaches. Here, we demonstrate that three of the most potent G-quadruplex ligands (360A, Phen-DC3, and pyridostatin) induce conformational changes of telomeric DNA G-quadruplexes to an antiparallel structure (as determined by circular dichroism) containing only one specifically coordinated K^+ (as determined by electrospray mass spectrometry) and, hence, presumably only two consecutive G-quartets. Control ligands TrisQ, known to bind preferentially to hybrid than to antiparallel structures, and L2H2-6M(2)OTD, known not to disrupt the



hybrid-1 structure, did not show such K^+ removal. Instead, binding of the cyclic oxazole L2H2-6M(2)OTD was accompanied by the uptake of one additional K^+ . Also contrasting with telomeric G-quadruplexes, the parallel-stranded Pu24-myc G-quadruplex, to which Phen-DC3 is known to bind by end-stacking, did not undergo cation removal upon ligand binding. Our study therefore evidences that very affine ligands can induce conformational switching of the human telomeric G-quadruplexes to an antiparallel structure and that this conformational change is accompanied by removal of one interquartet cation.

INTRODUCTION

There is now strong evidence that guanine-rich sequences can form G-quadruplex (G4) structures in human cells,¹ including in the regions of telomeres that comprise a 3'-overhang consisting of $d(TTAGGG)_n$ repeats. Ligands targeting telomeric G-quadruplexes are developed as anticancer therapeutics, and some were shown to perturb telomere structure.^{2–5} Current screening assays usually probe the ligand ability to bind to prefolded G-quadruplexes with both high affinity and high selectivity as compared to double stranded DNA.

The rational design of telomeric G4 ligands is however further complicated by the polymorphism that these sequences adopt *in vitro*⁶ and the ensuing uncertainties about which structures are relevant *in vivo*.^{7–10} For example, four-repeat sequences containing the central core d(GGG(TTAGGG)₃) have been shown to be able to form three different Gquadruplex structures in 100 mM K⁺: Hybrid-1 formed by sequences 24TTG¹¹ (dTTGGG(TTAGGG)₃A, Figure 1A) or 23TAG (dTAGGG(TTAGGG)₃), Hybrid-2 formed by 26TTA¹² (d(TTAGGG)₄TT, Figure 1B), and the two-quartet Hybrid-3 structure¹³ formed by 22GT (dGGG(TTAGGG)₃T, Figure 1C).

Some ligands were previously reported to bind better to the Na⁺structure or to the K⁺structure, as deduced from thermal denaturation and fluorescence displacement assays,¹⁴ but these studies did not discuss possible ligand-induced changes in G-quadruplex conformation. Other groups documented ligand-induced formation of particular G-quadruplex structures,^{15–18} but with few exceptions¹⁹ most of these studies were carried out in the absence of salt, so the relevance to physiologically relevant conditions remains unclear.

We studied here the interaction of ligands Phen-DC3,²⁰ 360A,^{21,22} pyridostatin²³ (abbreviated PDS), TrisQ,²⁴ L2H2-6M(2)OTD²⁵ (abbreviated L2H2), and TMPyP4 (ligand structures in Chart 1, molar extinction coefficients in Supporting Information S1) with the different human telomeric sequences 26TTA, 24TTG, 23TAG, and 22GT. Some of these ligands were previously shown to be very potent G4 binders and therefore have been used as G4 probes in a variety of

Received: September 26, 2014



Figure 1. (Top) Structures of human telomeric G-quadruplexes obtained by NMR in 70 mMK⁺and 20 mM K₂HPO₄/KH₂PO₄. (Bottom) Their respective schematic view with expected cation binding sites in-between successive G-quartets. (A) Hybrid-1 formed by 24TTG, PDB ID: 2GKU¹ (B) Hybrid-2 formed by 26TTA, PDB ID: 2JPZ.¹²(C) Hybrid-3 formed by 22GT, PDB ID: 2KF8.³ For full sequences, see main text.

Chart 1. Ligand Structures



biological systems and in cell-based assays. For example, both 360A and PDS promote G4-specific DNA damage and induce growth arrest in human cancer cells.^{26,27} In addition functional derivatives of both compounds have been used for G4 imaging,²⁶ pull down assays,²⁸ *in vitro* selection of G4 DNA²⁹ and more recently to photo-cross-link G-quadruplexes.³⁰ Finally, transcriptomic studies showed that both Phen-DC3 and 360A affect the expression of genes with G4 motif in their transcription start site.³¹ More specifically, Phen-DC3 has been

used to trigger quadruplex-dependent genetic instability in yeast^{32,33} and subsequently was shown to inhibit the unwinding activity of FANCJ helicase.³⁴

Despite their wide use, the exact binding mode of these ligands to telomeric G-quadruplex is not known. Indeed, only a few structures of the human telomeric sequence in the presence of ligands have been released. Except for L2H2 for which NMR revealed binding to the 24TTG sequence without changing the hybrid-1 G-quadruplex topology,²⁵ all reported structures of the human telomeric DNA interacting with ligands are parallel G-quadruplexes obtained by X-ray crystallography.^{35–41} Unfortunately, crystallization is known to bias the structural population by selecting the parallel topology.⁴²

There is therefore a strong need for in-solution assays able to quickly assess ligand binding affinity and binding mode in polymorphic G-quadruplexes such as the human telomeric sequence. Here we used native electrospray mass spectrometry (ESI-MS) to assay ligand ability to bind to G-quadruplexes. Thanks to the mass separation, single cation binding can be distinguished in the free and bound quadruplexes. Our experiments revealed ligand-induced cation ejection upon binding, thereby providing novel insight into the ligand binding mode.

RESULTS AND DISCUSSION

Ligand and Cation Binding, Simultaneously Probed by ESI-MS, Reveals K⁺ Ejection upon Binding of 360A, Phen-DC3, or PDS to Telomeric DNA G-Quadruplexes. We recently reported a sample preparation method in 100 mM trimethylammonium acetate (TMAA) and up to 1 mM KCl, which allows recording ESI-MS spectra in conditions where the solution fold is very similar as in 100 mM K⁺, as indicated by circular dichroism (CD) spectroscopy,43 although the fraction folded at room temperature is somewhat lesser in 1 mM KCl/ 100 mM TMAA than in 100 mM K⁺ (see melting experiments in Figure S2 and separate discussion of 22GT below). This is due to the K⁺ concentration effect on the folding equilibrium. Comparison between LiCl/KCl and TMAA/KCl mixtures with different K⁺ concentrations confirm that TMAA acts merely as a spectator electrolyte (Table S1 and Figure S3). In TMAA/ KCl conditions, ESI-MS allows the direct simultaneous measurement of ligand and cation binding stoichiometries from KCl solutions. On the contrary to the commonly used ammonium acetate conditions,44 the determination of potassium ion binding stoichiometry by ESI-MS from TMAA/KCl preparations is robust and independent of instrument choice or tune.43

Figure 2 shows the ESI-MS spectra of 5 µM 24TTG in 100 mM TMAA and 1 mM KCl without ligand (Figure 2A), a reference single-strand 24nonG4 (dTGGGATGCGACAGA-GAGGACGGGA) without ligand (Figure 2B), and 5 µM 24TTG incubated for 3 days at room temperature in the presence of 5 μ M ligand (Figure 2C-H) (1 day to 1 month, or annealing did not make any difference). Figure 2 shows the mass spectral region corresponding to free G4 and the 1:1 complex in the 5-charge state. Full-scale spectra in Figure S4 illustrate that the 2:1 complex abundance is very minor. Ligand relative affinities are estimated from the relative intensities of the peaks without and with ligands. The relative abundance of 1:1 complex relative to free G4 ranked: Phen-DC3 \sim TrisQ \sim 360A > PDS ~ L2H2 > TMPyP4. Similar ranking was obtained with 26TTA, 23TAG, and 22GT (Figures S5–S7, respectively). Tests with a reference hairpin duplex (ds26) allowed us to



Figure 2. Electrospray mass spectra (all in 100 mM TMAA + 1 mM KCl) of (A) 5 μ M 24TTG alone, (B) 24nonG4 alone, (C–H) 5 μ M 24TTG incubated 3 days with 5 μ M of each ligand: (C) 360A, (D) TrisQ, (E) Phen-DC3, (F) L2H2, (G) PDS, and (H) TMPyP4.

assess ligand binding selectivity (Figure S8). TMPyP4 binds very strongly to the hairpin duplex, in line with its known total lack of selectivity. TrisQ binds to the duplex, but less well as to the G-quadruplexes. The other four ligands do not bind to the hairpin and are confirmed as very selective G-quadruplex binders.

Labels in Figure 2 also show the major specific K^+ binding stoichiometries in each complex. Without ligand, 24TTG specifically binds two K⁺ ions (Figure 2A), in line with its known three-quartet NMR structure (Figure 1A), and with the known interquartet octahedraK⁺coordination mode. The additional K⁺ binding (third and fourth adduct) detected on the right-hand side of the peak shows the same distribution as for 24nonG4 (first and second adduct) and are therefore nonspecifically bound adducts. In order to distinguish and quantify quadruplex-specific K⁺ adducts, we used the following method described by Klassen's group.45 Briefly, a control molecule, as similar as possible to the analyzed molecule but lacking the specific K^+ binding sites, is used as a reference. The intensity ratios of the K⁺-bound versus unbound reference are taken as the signature of the nonspecific adducts distribution. This intensity distribution is then used to subtract nonspecific adducts contribution for other analytes. Using non-G4 forming sequences (as confirmed by CD, see Figure S9) with same guanine content and length as each quadruplex, we reconstructed the specific K⁺ distributions for each complex in the form of bar graphs (see supporting text and Figure S10 for detailed method and Figure S11 for results), which reflect what is annotated in the raw mass spectra of Figure 2.

Strikingly, the K⁺-binding stoichiometry is affected by ligand binding, in a ligand-dependent manner, and the same behavior is observed for all four variants of the human telomeric sequence (24TTG, 23TAG, 26TTA, and 22GT; see Figure S11). For 360A, Phen-DC3, and PDS, a single K⁺ ion is bound to the 1:1 complex, while two are bound to the free G4. In contrast, TrisQ only very partially displaced one cation.

Article

Because Ks⁺ supposedly coordinated in-between consecutive G-quartets, a 1:1:1 (DNA:ligand:K⁺) stoichiometry suggests that the ligand-bound structure contains only two consecutive G-quartets. It is particularly interesting that the 24TTG sequence, one of the most monomorphic variants of the human telomeric sequence in KCl, as shown by NMR,¹¹ is capable of such ligand-induced rearrangement.

In order to prove that the observed stoichiometries are indeed the ones present in solution and are not due to some kind of mass spectrometry artifact, we repeated the experiment with L2H2, which was previously shown by NMR to bind to 24TTG without changing its topology.²⁵ The two interquartet K⁺ binding sites should therefore be preserved. Moreover, cyclic oxazoles might even capture one extra K⁺ ion by coordination, in the same way as telomestatin.^{46,47} This is exactly what we observe in our mass spectra: the 1:1 DNA:ligand complex was detected with either two or three K⁺ ions bound (Figure 2F).

As a second control we used the Pu24 c-myc sequence (dTGAG₃TG₄AG₃TG₄AAG₂). Indeed, the recently published structure of Pu24 interacting with Phen-DC3 revealed that this ligand did not disrupt the parallel-stranded three-quartet structure.⁴⁸ When analyzed by ESI-MS in the TMAA/KCl preparation (Figure S12), the Pu24 sequence retains the expected two K⁺ ions, both without ligand and with all ligands including Phen-DC3. Coordination of a thirk⁺ ion with L2H2 is observed as well. The CD spectra were not affected upon ligand binding to Pu24 (Figure S13). These controls highlight that cation removal upon binding of 360A, Phen-DC3, and PDS is a binding mode that is peculiar to the human telomeric G-quadruplexes.

CD Spectroscopy Indicates That the G-Quartets in the $1\text{-}K^{+}$ Form Have Alternated Stacking. In order to obtain information about conformational changes occurring upon ligand binding and cation ejection, we performed CD experiments on the same solutions. The CD spectra give information on the stacking mode of consecutive guanines. Type-I spectra (positive peak at 265 nm, negative peak at 240 nm) indicate guanine stacking all in the same orientation (e.g., anti–anti), type-II spectra (positive peaks at 265 and 295 nm, negative peak at 240 nm) indicate guanine stacking partly in the same, and partly in alternating orientation (e.g., anti-syn or syn-anti), and type-III spectra (positive peaks at 295 and 240 nm, negative peak at 260 nm) indicate guanine stacking in exclusively alternating orientations.⁴⁹ If mixtures are present, the CD spectrum will be the weighted average spectrum of all conformers present in solution and may resemble Type-II spectra.

Figure 3A illustrates the CD spectral changes observed when adding ligands to 24TTG (hybrid-1 structure with CD spectrum of Type II). In the case of Phen-DC3, 360A, and PDS, the CD spectra are shifting to Type-III, indicating exclusively alternated stacking. In the corresponding ESI-MS spectra, almost all 24TTG is bound to 360A or Phen-DC3 in a 1:1:1 (DNA:ligand:K⁺) stoichiometry. The 1-K⁺ and presumably 2-quartet, ligand-bound structure has therefore the alternate G-quartet stacking typically encountered in antiparallel folds. For PDS, ESI-MS tells us that some free 24TTG remains, explaining the lesser shift of the CD spectrum compared to 360A and Phen-DC3, but the ligand-bound



Figure 3. CD spectra of 5 μ M 24TTG in (A) 100 mM TMAA + 1 mM KCl and in (B) 20 mM KH₂PO₄/K₂HPO₄ + 70 mM KCl. CD spectra were recorded after 3 days without ligand (black) and in the presence of Phen-DC3 (yellow), 360A (red), PDS (green), L2H2 (orange), and TrisQ (blue).

structure has the same characterisctics. In the cases of TrisQ and L2H2, the CD spectra of 24TTG are almost unaffected. The same results are obtained for 26TTA, 23TAG, and 22GT (Figures S14–S16, respectively).

Figure 3B shows the CD spectra recorded in 100 mM K⁺. In 100 mM TMAA + 1 mM KCl conditions, the CD spectrum of 24TTG alone is the same as in 100 mM K⁺.⁴³ The ligand ability to induce a shift to the Type-III spectrum is still observed, although to a lesser extent than in TMAA/KCl, probably due to the higher stability of the hybrid-1 structure in high KCl concentration. Phen-DC3 is clearly the most prone to change the G-quadruplex structure upon binding, even in 100 mM K⁺.

Some of the Most Affine and Selective G-Quadruplex Ligands Display a New Binding Mode with Ligand-Induced Cation Ejection. The ligand-bound G-quadruplex structure containing only two quartets (as deduced from the number of K⁺ ions coordinated) displays alternated stacking (as deduced from CD). Conversely, the ligands that do not displace cations do not lead to such antiparallel-type CD spectra. There are mainly three hypotheses on what this ligandbound G-quadruplex structure could be (Figure 4): (1) a structure with ligand intercalation (the ligand would take the place of one K⁺), (2) a hybrid-1 or hybrid-2 type structure with the quartet in homostacking disrupted, or (3) a hybrid-3 type



Figure 4. Ligand binding modes compatible with our data: (A) an intercalated ligand on a hybrid-2 G-quadruplex, (B) a ligand disrupting a quartet of a hybrid-2 G4, and (C) a ligand taking the place of the triplet of guanines in a hybrid-3 G4.

structure with 2 quartets and with the triplet stacking hampered by ligand end-stacking.

To our knowledge, only one study to date has reported ligand intercalation in G-quadruplexes, with TMPyP4 intercalation in long parallel G-wires.⁵⁰ However, intercalation in the G-wires only occurred in absence of salt. Here, true intercalation (Figure 4A) would probably involve a strong structural rearrangement with first a G-quartet opening and K⁺ displacement before closing again the G-quartet by sandwiching the ligand. G-quartet opening an K+displacement are of course unfavorable thermodynamically, but in turn, four guanines will gain conformational degrees of freedom, and that a solvated K⁺ will be released from the complex, so the penalty might be compensated by the free energy gained upon ligand binding. However, it seems difficult to assume that reclosing of the G-quartet on the other side of the ligand would be thermodynamically favorable, because the small enthalpically favorable contribution due to extra stacking would be compensated entropically by reducing the degrees of freedom. The binding mode of Figure 4B is therefore more likely than that of Figure 4A.

The second hypothesis, namely the disruption of G-quartets (Figure 4B), was proposed previously for an anthrathiophene-dione on telomeric G4,¹⁹ for TMPyP4 on an RNA G4,⁵¹ and for a triarylpyridine ligand on c-kit1 and c-kit2.52 However, in all reports these ligands were promoting total unfolding of the G-quadruplexes. In our case the melting temperature of the Gquadruplexes with the ligands is higher than without ligand (see Figures S17 and S18), meaning that the ligand-bound 2-quartet structure is more thermally stable than the starting structure. Also the broad melting transitions of 360A and Phen-DC3 complexes are typical of multistep denaturation that may indicate conformational rearrangements.53 However, the thermal denaturation transition is narrower and at lower temperature for PDS, indicating that ligand binding mode and thermal stability are not necessarily linked: these are two different tenets of ligand binding that must both be characterized. For example, the relative abundance of 1:1 complex relative to free G4 at room temperature (see Figures S3-S6) correlates well with the T_m only for the three ligand sharing the newly described binding mode: Phen-DC3 > 360A > PDS. Ligand L2H2, although having a different binding mode without disturbing the structure and with an additional potassium ion, also fit in this trend, with an affinity in the same range as that of PDS. Ligand TrisQ, with its propensity to induce more parallel structures either at room temperature (with 26TTA, see Figure S17) or at higher temperature (with 24TTG, see Figure S16), does however follow different rules. In general, studies of the ligand binding mode and ligandinduced conformational changes should also consider the temperature dependence of these phenomena.

The last hypothesis is the ligand-induced conformational selection of a hybrid-3 structure such as that formed predominantly by 22GT. This structure contains only 2 quartets at 100 mM K⁺, plus a triplet of guanines stacked on one terminal quartet.¹³ Notably, this 2-quartet structure is as thermally stable as the other 3-quartet structures¹³ (Figure S2), and therefore extra stabilization by a ligand sounds reasonable. To further explore the correlation between the potassium binding stoichiometry and the hybrid-3 structure, we titrated 22GT by KCl, from 0 to 1 mM using ESI-MS and CD and up to 100 mM using CD (Figure S19). In pure TMAA and no KCl, the DNA remains unfolded.⁴³ Interestingly, at 200 µM KCl concentration, the main stoichiometry observed is 1 K^+ per DNA strand, and the corresponding CD spectrum is purely antiparallel, just like in the presence of ligands 360A, Phen-DC3, and PDS. We therefore assign the first K^+ binding event to the binding site between the two full G-quartets of hybrid-3. At 1 mM KCl concentration, a second potassium ion binding site is detected, which we attribute to the filling of a second binding site between the G-triplet and one G-quartet. In the same solution, the CD spectra show that the hump at 270 nm, attributed to the stacking of the triplet,49 is increasing accordingly. Ligand binding to the hybrid-3 by end-stacking on the top quartet would disrupt triplet stacking, and this binding mode is therefore compatible with all our observations.

CONCLUSIONS

In summary, we evidenced here that ligands can induce conformational changes in human telomeric G-quadruplexes at physiological ionic strength and in KCl. Three ligands (Phen-DC3, 360A, and PDS), among the most potent described to date were found to favor G4 structures with only one K⁺, hence two consecutive G-quartets, and heterostacking between these G-quartets. This ligand-induced K⁺ ejection could be due to a quartet disruption of hybrid structures or to the conformational selection of the hybrid-3 structure.

Interestingly, only the three most flexible⁵⁴ of the ligands tested here, all based on the dicarboxamide scaffold (Phen-DC3, 360A, and PDC), are capable of ejecting K⁺ cation upon binding, while stabilizing the antiparallel 2-quartet G-quadruplex structure. It would be interesting in the future to explore the enthalpy and entropy contributions to ligand binding in those particular cases and to partition the intrinsic folding thermodynamic parameters from the polyelectrolyte and water capture/release contributions. Although high-resolution atomistic methods would be required to determine the exact binding mode, our work opens new ideas on possible G-quadruplex ligand binding modes. It also highlights that, thanks to ESI-MS monitoring of cation binding, hints on the ligand binding mode can be obtained with small amounts of sample and even in the presence of polymorphic structures. The technique is amenable to higher throughput screening of ligands.

Our results also have important implications for the design of ligands targeting polymorphic G-quadruplexes such as the human telomeric sequence. The availability of high-resolution structures makes it tempting to develop drug design programs based on the lock-and-key model. In contrast, we show here that several ligands with high G-quadruplex affinity and specificity induce structural changes in telomeric G-quadruplexes and that induced fit or conformational selection must therefore be taken into account for the drug design.

EXPERIMENTAL SECTION

Materials. Oligonucleotides were purchased lyophilized and RPcartridge purified from Eurogentec (Seraing, Belgium) and used without further purification. We used the following nomenclature for the DNA sequences: $26TTA = d((TTAGGG)_4TT); 24TTG =$ $d(TT(GGGTTA)_3GGGA); 23TAG = d(TA(GGGTTA)_3GGG);$ $22GT = d((GGGTTA)_3GGGT); Pu24 c-myc =$ d-(TGAGGGTGGGGAGGGTGGGGAAGG), 24nonG4 = d-(TGGGATGCGACAGAGAGAGGACGGGA), 22nonG4 = (GGGATGCGACAGAGAGAGGACGGG), 23nonG4 = dd-(TGGGATGCGACAGAGAGGGACGGG), 26nonG4 = d(TTGGGATGCGACAGAGAGGAGGACGGGAT), and ds26 =d-(CA2TCG2ATCGA2T2CGATC2GATC2GAT2G). Analyzed solutions contained 5 µM DNA diluted from 200 µM stock solution. Water was nuclease-free from Ambion (Applied Biosystems, Lennik, Belgium). The concentrations of the initial stocks solutions were measured by UV absorbance at 260 nm on a Uvikon XS. Molar absorption coefficients were obtained on the IDT Web site and corrected using Cavaluzzi–Borer correction.⁵⁵ 100 mM trimethylammonium acetate (TMAA, Ultra for HPLC, Fluka analytical) was used to fix the ionic strength, and the solutions were doped with up to 1 mM KCl (>99.999%, Sigma) as described previously. Ligands were either synthesized as described elsewhere (360A iodide salt, Phen-DC3 methylsulfonate salt,20 TrisQ bromide salt,24 L2H2 trifluoroacetate salt),56 or purchased from Sigma-Aldrich (PDS trifluoroacetate salt and TMPyP4 *p*-toluenesulfonate salt). When documented the molar absorption coefficients were used to measure concentration using the UV absorbance. When not known, the powders were weighted, and the coefficient was determined (Figure S1 summarizes all molar extinction coefficients). 1:1 DNA:ligand molar ratios were used to acquire the ESI-MS and the CD data. TMAA and KCl were first added to the water solution, and then the proper DNA solution volume is added. Ligands were added 1 h after, to ensure starting from preformed G-quadruplex. Solutions were stored at room temperature in the dark before analysis. Annealing experiments were also performed on freshly prepared solutions, with 95 °C heating for 5 min and then leaving the sample for 4 h at room temperature before further analysis (results did not change any further when checked after one month left at room temperature, indicating that equilibrium was reached).

Electrospray Mass Spectrometry. Native ESI-MS spectra were obtained using a LCT Premier mass spectrometer (Waters, Manchester, U.K.). The source voltage is set to 2200 V. The desolvation temperature is 60 °C, and the gas temperature is 40 °C. The source pressure is increased to 45 mbar (measured by a Center Two probe, Oerlikon Leybold Vacuum, Cologne, Germany). The sample cone voltage is 200 V. At this voltage TMA adducts on the mass spectra are totally avoided, but the K⁺ stoichiometry is not affected. The syringe injection rate is 200 µL/h. Shown MS spectra are the sum of 3 min accumulations (1 scan per 1.1 s). Savitzky–Golay smoothing has been performed with a smooth window of 5 points. Baselines were also subtracted using curves with a polynomial order of 40.

Circular Dichroism. CD experiments were run on a Jasco J-815 spectrophotometer using a quartz cell of 2 mm path length. Shown spectra are the sum of 3 accumulations at 20 °C with a scan speed of 50 nm/min and an integration time of 0.5 s. Data were normalized to molar circular-dichroic absorption $\Delta \epsilon$ based on DNA concentration: $\Delta \epsilon = \theta/(32980 \times c \times l)$, with θ the CD ellipticity in milidegrees, *c* is the DNA concentration in M (5 μ M), and *l* is the cell length (2 mm). Baselines were subtracted using either 100 mM TMAA in water solution or 20 mM KH₂PO₄/K₂HPO₄ + 70 mM KCl depending on the experimental conditions. A second linear baseline was removed from the raw data using the mean value obtained for each experiment between 320–350 nm, as no peak was observed in this area. A smoothing was then performed using locally weighted scatterplot smoothing method (LOWESS). Reported CD spectra display raw data points and the smoothed curve.

Melting Monitored by CD. Melting experiments were performed on the same Jasco J-815 spectrophotometer. The temperature ramp was set to $0.4 \,^{\circ}$ C/min from 4 to 90 °C. One spectrum per minute was acquired at 100 nm/min scan speed with 2 accumulations. The temperature ramp is stopped during each recording. A linear baseline was subtracted from the spectra using the mean value for each experiment between 320 and 350 nm.

Melting Monitored by UV Absorbance. Melting experiments on the G-quadruplexes without ligands were also performed on a SAFAS UV mc² spectrophotometer (Montecarlo, Monaco) equipped with a Peltier temperature controller. The temperature ramp was 0.2 °C/min from 4 to 90 °C. The absorbance was monitored at 295 nm.

ASSOCIATED CONTENT

* Supporting Information

Supplementary mass spectra, CD spectra, and CD melting data as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

valerie.gabelica@inserm.fr

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Inserm (ATIP-Avenir grant no. R12086GS to V.G.), the Conseil Regional Aquitaine (grant no. 20121304005 to V.G.), and the EU (FP7-PEOPLE-2012-CIG-333611 to V.G.). A.M. benefited from an Inserm/CRA Ph.D. fellowship.

REFERENCES

(1) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Nat. Chem. 2013, 5, 182.

(2) Neidle, S. FEBS J. 2010, 277, 1118.

(3) Rodriguez, R.; Muller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. J. Am. Chem. Soc. 2008, 130, 15758.

- (4) Gomez, D.; Wenner, T.; Brassart, B.; Douarre, C.; O'Donohue, M.-F.; El Khoury, V.; Shin-Ya, K.; Morjani, H.; Trentesaux, C.; Riou,
- J.-F. J. Biol. Chem. 2006, 281, 38721.
 (5) Phatak, P.; Cookson, J. C.; Dai, F.; Smith, V.; Gartenhaus, R. B.; Stevens, M. F. G.; Burger, A. M. Br. J. Cancer 2007, 96, 1223.

(6) Tua, A.; Phan, A. T. FEBS J. 2010, 277, 1107.

(7) Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. *J. Am. Chem. Soc.* 2010, *132*, 17105.

(8) Hansel, R.; Löhr, F.; Foldynova-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. Nucleic Acids Res. 2011, 39, 5768.

(9) Kan, Z.-Y.; Yao, Y.; Wang, P.; Li, X.-H.; Hao, Y.-H.; Tan, Z. Angew. Chem., Int. Ed. Engl. 2006, 45, 1629.

(10) Chaires, J. B. FEBS J. 2010, 277, 1098.

- (11) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. J. Am. Chem. Soc. 2006, 128, 9963.
- (12) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Nucleic Acids Res. 2007, 35, 4927.
- (13) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. *J. Am. Chem. Soc.* 2009, 131, 4301.
- (14) Chen, M.; Song, G.; Wang, C.; Hu, D.; Ren, J.; Qu, X. *Biophys. J.* 2009, 97, 2014.
- (15) Nakatani, K.; Hagihara, S.; Sando, S.; Sakamoto, S.; Yamaguchi, K.; Maesawa, C.; Saito, I. J. Am. Chem. Soc. 2003, 125, 662.

(16) Rahman, K. M.; Reszka, A. P.; Gunaratnam, M.; Haider, S. M.; Howard, P. W.; Fox, K. R.; Neidle, S.; Thurston, D. E. *Chem. Commun.* (*Cambridge, U. K.*) 2009, 4097.

- (17) Xu, L.; Xu, Z.; Shang, Y.; Feng, S.; Zhou, X. *Bioorg. Med. Chem. Lett.* 2012, 22, 2988.
- (18) Shi, S.; Geng, X.; Zhao, J.; Yao, T.; Wang, C.; Yang, D.; Zheng, L.; Ji, L. *Biochimie* 2010, 92, 370.

(20) De Cian, A.; Delemos, E.; Mergny, J.-L.; Teulade-Fichou, M.-P.; Monchaud, D. *J. Am. Chem. Soc.* 2007, 129, 1856.

(21) De Cian, A.; Mergny, J.-L. Nucleic Acids Res. 2007, 35, 2483.

(22) Granotier, C.; Pennarun, G.; Riou, L.; Hoffschir, F.; Gauthier, L. R.; De Cian, A.; Gomez, D.; Mandine, E.; Riou, J.-F.; Mergny, J.-L.; Mailliet, P.; Dutrillaux, B.; Boussin, F. D. *Nucleic Acids Res.* 2005, 33, 4182.

(23) Koirala, D.; Dhakal, S.; Ashbridge, B.; Sannohe, Y.; Rodriguez, R.; Sugiyama, H.; Balasubramanian, S.; Mao, H. *Nat. Chem.* 2011, *3*, 782.

(24) Bertrand, H.; Granzhan, A.; Monchaud, D.; Saettel, N.; Guillot, R.; Clifford, S.; Guedin, A.; Mergny, J.-L.; Teulade-Fichou, M.-P. *Chemistry* 2011, 17, 4529.

(25) Chung, W. J.; Heddi, B.; Tera, M.; Iida, K.; Nagasawa, K.; Phan, A. T. J. Am. Chem. Soc. 2013, 135, 13495.

(26) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. *Nat. Chem. Biol.* 2012, *8*, 301.

(27) Pennarun, G.; Granotier, C.; Hoffschir, F.; Mandine, E.; Biard, D.; Gauthier, L. R.; Boussin, F. D. *Nucleic Acids Res.* 2008, 36, 1741.

(28) Lam, E. Y. N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. Nat. Commun. 2013, 4, 1796.

(29) Renaud de la Faverie, A.; Hamon, F.; Di Primo, C.; Largy, E.; Dausse, E.; Delauriere, L.; Landras-Guetta, C.; Toulme, J.-J.; Teulade-Fichou, M.-P.; Mergny, J.-L. *Biochimie* 2011, 93, 1357.

(30) Verga, D.; Hamon, F.; Poyer, F.; Bombard, S.; Teulade-Fichou, M.-P. Angew. Chem., Int. Ed. Engl. 2014, 53, 994.

(31) Halder, R.; Riou, J.-F.; Teulade-Fichou, M.-P.; Frickey, T.; Hartig, J. S. BMC Res. Notes 2012, 5, 138.

(32) Piazza, A.; Boulé, J.-B.; Lopes, J.; Mingo, K.; Largy, E.; Teulade-Fichou, M.-P.; Nicolas, A. *Nucleic Acids Res.* 2010, *38*, 4337.

(33) Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M.-P.; Foiani, M.; Nicolas, A. *EMBO J.* 2011, *30*, 4033.

(34) Bharti, S. K.; Sommers, J. a; George, F.; Kuper, J.; Hamon, F.; Shin-ya, K.; Teulade-Fichou, M.-P.; Kisker, C.; Brosh, R. M. *J. Biol. Chem.* 2013, 288, 28217.

(35) Parkinson, G. N.; Ghosh, R.; Neidle, S. *Biochemistry* 2007, 46, 2390.

(36) Bazzicalupi, C.; Ferraroni, M.; Bilia, A. R.; Scheggi, F.; Gratteri, P. *Nucleic Acids Res.* 2013, 41, 632.

(37) Campbell, N. H.; Karim, N. H. A.; Parkinson, G. N.; Gunaratnam, M.; Petrucci, V.; Todd, A. K.; Vilar, R.; Neidle, S. J. *Med. Chem.* 2012, 55, 209.

(38) Micco, M.; Collie, G. W.; Dale, A. G.; Ohnmacht, S. a; Pazitna, I.; Gunaratnam, M.; Reszka, A. P.; Neidle, S. *J. Med. Chem.* 2013, 56, 2959.

(39) Collie, G. W.; Promontorio, R.; Hampel, S. M.; Micco, M.; Neidle, S.; Parkinson, G. N. J. Am. Chem. Soc. 2012, 134, 2723.

(40) Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. J. Am. Chem. Soc. 2008, 130, 6722.

- (41) Parkinson, G. N.; Cuenca, F.; Neidle, S. J. Mol. Biol. 2008, 381,
- 1145.

(42) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* 2005, 33, 4649.

(43) Marchand, A.; Gabelica, V. J. Am. Soc. Mass Spectrom. 2014, 25,

1146.

(44) Balthasart, F.; Plavec, J.; Gabelica, V. J. Am. Soc. Mass Spectrom.

2013, 24, 1.

(45) Sun, J.; Kitova, E. N.; Wang, W.; Klassen, J. S. *Anal. Chem.* 2006,

78, 3010.

(46)Rosu, F.; Gabelica, V.; Smargiasso, N.; Mazzucchelli, G.; Shin-Ya, K.; De Pauw, E. J. Nucleic Acids 2010, 2010, 1. (48) Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M.-P.; Phan, A. T. Angew. Chem., Int. Ed. Engl. 2013, 1.

(49) Karsisiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. Angew. Chem., Int. Ed. Engl. 2011, 50, 10645.

(50) Lubitz, I.; Borovok, N.; Kotlyar, A. *Biochemistry* 2007, 46, 12925.
(51) Morris, M. J.; Wingate, K. L.; Silwal, J.; Leeper, T. C.; Basu, S. Nucleic Acids Res. 2012, 40, 4137.

(52) Waller, Z. A. E.; Sewitz, S. A.; Hsu, S.-T. D.; Balasubramanian, S. J. Am. Chem. Soc. 2009, 131, 12628.

(53) Buscaglia, R.; Gray, R. D.; Chaires, J. B. Biopolymers 2013, 99, 1006.

(54) Monchaud, D.; Yang, P.; Lacroix, L.; Teulade-Fichou, M.-P.; Mergny, J.-L. Angew. Chem., Int. Ed. Engl. 2008, 47, 4858.

(55) Cavaluzzi, M. J.; Borer, P. N. Nucleic Acids Res. 2004, 32, e13. (56) Tera, M.; Ishizuka, H.; Takagi, M.; Suganuma, M.; Shin-ya, K.; Nagasawa, K. Angew. Chem., Int. Ed. Engl. 2008, 47, 5557.

VI.4. ADDITIONAL WORK

VI.4.1 COVALENT LIGANDS

In these articles, we demonstrated that G-quadruplex ligand binding modes may be more complicated than simple π - π stacking on external G-quartets and that ligand binding can lead to conformational changes. To understand more deeply the binding mode of those ligands, more structural information is needed. The most straightforward way to understand the binding mode is to obtain 3D structures of the complex. However, due to the polymorphism of those G-quadruplexes, it is not easy to obtain NMR structures.

For this reason, we used another MS-based approach. The objective is to fragment the DNA in the mass spectrometer.²⁷² Because the ligand is positively charged, it is not easily ejected upon collisions inside the mass spectrometer and remains on DNA fragments. The principle consists in determining on which DNA base the ligand remains after fragmentation in order to determine the binding site. However, collision-induced dissociation (CID) may lead to rearrangement of the complex in the gas phase.^{273,274}



Figure 43. Structure of the ligands 360A the chlorambucyl derivative.

To maintain the ligand in the binding site, chemical modifications were needed in order to covalently bind the ligand to the DNA. We used this approach to characterize the binding mode of the ligands able to eject a potassium cation upon binding. Such derivatizations of these ligands were previously done with a biotin group,²⁷⁵ or with an alkylating agent.²⁷⁶ The group of Teulade-Fichou derivatized the 360A G-quadruplex ligand (Figure 43) with a chlorambucyl function which allows the ligand to bind covalently to the DNA bases (preferentially to the N7 of adenines), with a HCl loss.



Figure 44. A and B, MS of 10 μ M 23TAG in 100 mM TMAA and 1 mM KCl with, respectively, 10 μ M 360A and 10 μ M PDC-C4-C. C, Zoom on the main peak of the 23TAG+PDC-C4C distribution (m/z=1617) and attribution of the stoichiometry using the isotopic distribution. D, CD spectra of 10 μ M 23TAG and 10 μ M PDC-C4-C in 100 mM TMAA + 1 mM KCl.

First, we checked that the binding of those derivatives also induces the potassium ejection and the transition to antiparallel G-quadruplexes. The potassium and chloride stoichiometries of the 1:1 (DNA:L) complex were investigated using high resolution mass spectrometry (Figure 44A-C). The potassium adducts distribution of the G-quadruplex bound with one ligand is represented and the isotopic distribution of the main peak (*m*/*z*=1617) matches the theoretical mass corresponding of the 1:1:1 DNA:L:K⁺ complex with the loss of a HCl molecule. This is the expected stoichiometry if the ligand binds covalently to the DNA with the same binding mode as 360A. In addition, the CD spectrum (Figure 44D) of the G-quadruplex bound to the covalent ligand also becomes more antiparallel, although not to the same extent as with 360A. This observation can be explained by a lower affinity of the chlorambucyl derivative for this G- quadruplex compared with 360A (Figure 44A & B). We concluded that the covalent ligand binds to the G-quadruplex with the same binding mode and that, therefore, the chlorambucyl derivatives could be used for further MS/MS experiments.

MS/MS fragmentation indicated that the ligand is covalently bound in the central loop of the G-quadruplex. Unfortunately, it was not possible to obtain only one structure compatible with these results. Still, we could discard some of the compatible structures. This MS/MS investigation is the object of an article, currently in preparation.²⁷⁷

VI.5. CONCLUSIONS

In this chapter, we have shown that ligands binding to G-quadruplexes are more complicated than the usually assumed π - π stacking on external G-quartets. Ligand-induced conformational changes are indeed more common than expected (See Figure 45 for a summary of the discoveries). MS is a quick method to reveal binding modes involving changes in stoichiometries.



Figure 45. Schematic representation of the G-quadruplex conformations for the human telomeric sequence and summary of the binding modes of the ligands studied in this work.

- 1) Ligand stoichiometry. In Article 3, we evidenced the cooperative binding of two Cu*-ttpy* ligands to human telomeric G-quadruplexes. Other methods failed to detect this highly cooperative binding. We found that the binding was accompanied by the preferential uptake of two K⁺ ions and the transition to antiparallel structures of the human telomeric G-quadruplexes.
- 2) Cation stoichiometry. Detecting the number of cations bound can provide hints on ligand-induced structural changes. In Article 4, a new binding mode involving the ejection of one K⁺ for some of the most affine G-quadruplex ligands (Phen-DC3, 360A and PDS), for the human telomeric G-quadruplexes. Because K⁺ binding sites are between G-quartets, we deduced that the ligand-bound G-quadruplexes have only two

G-quartets. CD indicated an antiparallel stacking. This complex present similar characteristics as the complex observed either transiently or at low KCl concentration in Chapter V. Other works support the cation ejection binding mode via interpretation of the entropic and enthalpic contributions for the binding of those ligands to human telomeric G-quadruplexes.^{278,279} We also showed that some ligands (L2H2) can bind with uptake of one additional cation (see Article 4).

Atomic resolution structures of those complexes would provide more information on the binding modes. However, no NMR or X-ray diffraction structure of a human telomeric G-quadruplex with one of those ligand is published, presumably because of the structural heterogeneity of the complexes. The ensemble of structures may be too complicated to be resolved by NMR, and this emphasizes the needs of developing new techniques to obtain structural information.

Our results demonstrate the power of MS: determining the stoichiometry of a complex is of prime importance. The additional information given by the IMS could potentially be very useful for binding mode determination.

VI.6. PERSPECTIVES

In this chapter, we have used mainly mass spectrometry to characterize G-quadruplexes and G-quadruplex/ligand complexes. In this section, we will discuss the use of ion mobility spectrometry coupled with mass spectrometry. We selected a simple well-characterized, tetramolecular parallel 4-quartet/3-cation G-quadruplex [d(TG₄T)]₄ (Figure 4)¹⁰ and discuss the interaction of this G-quadruplex with the macrocyclic L2H2 ligand.

For this example, ammonium acetate, instead of TMAA + KCl, was used. To the contrary of the human telomeric G-quadruplexes, for $[d(TG_4T)]_4$, a higher cation concentration than 1 mM is needed to favor the binding mode involving the uptake of an additional cation. In addition, the ion mobility spectra of the complexes with nonspecific potassium adducts could not be distinguished from those with additional specific cations. Finally, $[d(TG_4T)]_4$ has also been extensively studied in ammonium acetate by molecular dynamics and ion mobility in our group.²²³ The theoretical ^{DT}CCS_{He} of the parallel G-quadruplex containing three ammonium cations matches very well the experimental value (^{DT}CCS_{He} = 785Å², Figure 46). Additionally, the ion mobility peaks of this G-quadruplex are sharp (resolving power of ~60 in helium), suggesting that only one conformer is present.

When one ligand binds with no uptake of cation, the distribution is made of two peaks (Figure 46B, dark blue). When the binding is accompanied by the uptake of one more NH_4^+ (total of 4 NH_4^+), only one peak, corresponding to the most extended structure, remains (light blue). This peak was attributed (thanks to the NMR structure¹⁶³) to the ligand bound to the G-quadruplex by π - π stacking.

The binding of two L2H2 ligands is accompanied by the binding of one more cation (total of 4 NH₄⁺). The arrival time distribution of the 1:2:4 (G4:L:cation) complex presents two peaks (Figure 46C, red). Interestingly, the larger 1:2:4 complex ($^{DT}CCS_{He} = 869 \text{ Å}^2$) is 43 Å² larger than the 1:1:4 complex ($^{DT}CCS_{He} = 826 \text{ Å}^2$), itself 41 Å² larger than de free G-quadruplex ($^{DT}CCS_{He} = 785 \text{ Å}^2$). This peak can potentially be attributed to a structure of the G-quadruplex with two π - π stacked ligands, one at each external G-quartet.



Figure 46. A, MS spectrum of 40 μ M d(TG₄T) in 100 mM NH₄OAc with 10 μ M L2H2, charge state 5-. B and C, Normalized arrival time distribution of the complexes without (green) and with one (blue) and two ligands (red). All peaks correspond to the peak of [d(TG₄T)]₄⁵⁻ with additional cation(s) and ligand(s), indicated in corresponding color.

Thanks to the structure obtained by NMR, we could correlate the changes in CCS and the cation uptake with a π - π stacking binding mode for the L2H2 ligand. However, ion mobility data indicate that other binding modes inducing smaller changes in CCS exist as well. Complementary experiments, with ligands whose binding modes are known, are needed in order to better characterize ligand binding modes using IMS-MS. The goal is to furthermore separate the ensembles of ligand complexes based on their CCS (in addition to their K⁺ binding stoichiometry, Chapter V).

VII. GENERAL DISCUSSION AND CONCLUSION

VII.1. SUMMARY OF THE RESULTS

In Chapter IV, we describe a new sample preparation that allows using K^+ to fold the Gquadruplexes in a MS-compatible solution. The ionic strength is ensured by trimethylammonium acetate (TMAA) and up to 1 mM KCl could be added. It is well known that G-quadruplexes can fold into different topologies, depending on the salt nature. We could show that the ensembles of conformations formed in TMAA + KCl were the same as those obtained in the NMR conditions (usually at 100 mM K⁺).

Thanks to the mathematical treatment on the nonspecific adducts, we could determine the specific stoichiometry in terms of cation bound, and quantify each specific complex. In Chapter V, individual potassium ion binding constants to human telomeric G-quadruplexes were calculated. We found that 1-K⁺ complexes were formed at low KCl concentration for the shortest human telomeric sequences. When flanking bases are added, the binding of the two potassium cations is cooperative and the 1-K⁺ complex becomes less prominent. Combining the MS data with the circular dichroism data allowed deconvoluting the CD spectra for each of the 0-, 1- and 2-K⁺ complexes. The 0- and 2-K⁺ complexes present CD spectra typical of unfolded and hybrid conformation, respectively. We evidenced that the 1-K⁺ complex is antiparallel. These results were generalized to other sequences, as demonstrated for two synthetic sequences.

The binding of K⁺ to G-quadruplexes was also monitored as a function of time in order to study the folding mechanism. We found that up to six ensembles of species are needed to properly fit the data: one species without K⁺, two with one K⁺ and three with two K⁺. The kinetics indicates a misfolding pathway involving an antiparallel 1-K⁺ (2-quartet) topology. The folding pathway of telomeric sequences is transposable to other systems, including a c-myc gene promoter, suggesting that this is a general tenet of G-quadruplex folding. The implications of the proposed folding pathway are important also in the field of G-quadruplex ligands design. Indeed, Chapter VI highlights that the characterization of ligand binding to G-quadruplexes should not be limited to binding affinity, but should also consider ligand-induced changes in the target conformation. First, through a screening over more than 20 different DNA sequences, we evidenced a highly affine and cooperative binding of two Cu-ttpy ligands to and only to the human telomeric G-quadruplexes: the 2:1 L:DNA stoichiometry prevails and the 1:1 complex is almost nonexistent upon titration. This unprecedented cooperativity is attributed to conformational changes in the polymorphic telomeric G-quadruplexes. Indeed, as indicated by CD, the structures convert to antiparallel topologies upon binding of the two ligands. The G-quadruplex-specific potassium cations bound are also affected: the ligands bind to a 2-K⁺ structure. Secondly, we investigated the binding of some of the most affine and selective G-quadruplex ligands reported to date (Phen-DC3, 360A and PDS). These ligands bind with high affinity to the human telomeric G-quadruplexes, and displace one potassium cation upon binding. The 1:1:1 (DNA:L:K⁺) complex has antiparallel-like stacking of the guanines in the G-quadruplex core. This exemplifies that species that do not necessarily prevail at equilibrium can become the predominant species in another environment, or in the presence of ligands and/or proteins which can displace the equilibrium towards other structures.

VII.2. DISCUSSION ABOUT THE MS RESULTS

VII.2.1 MASS SPECTROMETRY FOR BIOPHYSICAL STUDIES

Our results show the importance of characterizing the stoichiometry of a complex, in order to determine the equilibrium constants and obtain information on the structures of the complexes. Because native MS can easily detect the different species directly from the solution, it is a method of choice for characterizing ternary complexes. In the case of G-quadruplexes/ligands/cations systems, this capability brought important structural insight.

As distinct signals are recorded for each stoichiometry, mass spectrometry is a very powerful technique for the characterization of individual binding equilibria of complexes of multiple stoichiometries, whereas most other techniques can provide only one phenomenological equilibrium constant. In addition, very cooperative binding such as these observed for the binding of Cu*-ttpy* can be missed by more conventional techniques. For example, in a UV-melting experiment, the signal is proportional to the folded/unfolded ratio, all folded species belonging to the same ensemble.

Fitting the mass spectrometry data using DynaFit was found very powerful for obtaining equilibrium and kinetic constants for several reactions, and allowed us to build a folding landscape. The characterization of this folding mechanism was made possible by mass spectrometry because independent signals (distinct m/z) were recorded for several ensembles. Additional separation of the species was qualitatively obtained using ion mobility spectrometry. A valuable objective for future work would be to quantitatively extract the concentration of different species based both on their stoichiometry and on their shape.

VII.2.2 The importance of the buffer for MS of cation-binding biomolecules

Thanks to the development of a new sample preparation method, we were able to obtain ESI-MS spectra of nucleic acids/K⁺ complexes. We focused on G-quadruplexes in negative ionization mode. In principle, the TMAA buffer can be used for other cation-binding biomolecules, for which the ammonium and the cation of interest can compete for the same binding site. TMAA can also be beneficial in reducing undesirable nonspecific alkali cation adducts.

In our group, one perspective offered by this finding was to study the binding of magnesium cations to RNA-RNA kissing complexes.²⁸⁰ A kissing complex is made of two hairpins that interact through their loop regions, making base pairs. Both monovalent and divalent (magnesium) cations are needed for this interaction to take place. However, nothing is known to date in terms of magnesium stoichiometry and binding sites in the loop-loop region. We found that the sample preparation using TMAA as buffer actually led to a disruption of the kissing complexes. This highlights the importance of having small (NH₄⁺ rather than TMA⁺) monovalent cations in that binding motif.

Testing this method for proteins in positive ionization mode or for protein/DNA complexes could bring new information about the cation binding sides. Indeed, many biologically relevant complexes require cations to fold such as, for example, zinc finger proteins... However, we have to be careful because the TMAA buffer could also stabilize the unfolded species or destabilize the folded ones, therefore displacing the equilibrium towards the unfolding of the complexes.

VII.2.3 Sources of error in the treatment of the MS data

Before fitting the data points and obtaining equilibrium and/or kinetic constants, several steps of data treatment are needed. The first one consists in extracting the data from the MS spectra. This is usually done by integrating the area under a region of interest, directly on the MS spectra. The background is also subtracted considering the region of the theoretical peak of the "-1 stoichiometry". This step is straightforward and is not a source of error.

The second step is the response factor correction. Because not all species are desolvated equally, their concentration is not necessarily directly obtained from their intensity. To check whether the response factors correction has to be applied or not, the mass balance is made for each point in the titration or kinetic experiment. If there is significant deviation of the total signal (normalized with an internal standard), the correction is needed and the response factors can be calculated. Because a mathematical treatment of the data is needed, this could increase the error on the result. However, in our case, there was no significant deviation of the total signal as a function either of time or KCl concentration and therefore, the response factors correction was not needed.

The last step is the nonspecific adducts treatment. These nonspecific adducts presumably bind to external phosphates and are additional to the specific ones, bound in specific binding sites. We have adapted the mathematical treatment first described by Klassen. The way this correction is done can influence the result. Indeed, depending on the experimental reference, the amount of nonspecific adducts can vary. In our case, we have been very careful to select reference sequences of the same size and containing the same number of guanines and considering each charge state separately. To ensure that the nonspecific adducts were properly taken into account, a simple verification can, however, be done after treatment. It consists in checking that the concentrations of species of very high specific stoichiometries (in our case higher than 2-K⁺) are zero, meaning that their detection was only due to nonspecific adducts. In addition, a comparison between charge states can also confirm the results. Once these steps done, the data can be normalized (between 0 and the total DNA concentration) and are ready to be fitted. One has to be aware of the possible sources of error.

VII.3. DISCUSSION ABOUT THE G-QUADRUPLEX RESULTS

VII.3.1 The 1-K⁺ ANTIPARALLEL G-QUADRUPLEXES

In Chapter V, we found that a 1-K⁺ G-quadruplex was formed at short time scale and/or at low KCl concentration. This was unexpected for some complexes. Deconvolution of the CD signal allowed assigning the 1-K⁺ structure to an antiparallel G-quadruplex. The folding of G-quadruplex with the same characteristics was induced by the binding of Phen-DC3, 360A, PDS or other ligands and discussed in Chapter VI (Article 4).

These conformational changes were observed exclusively for the human telomeric Gquadruplexes, potentially because of the polymorphism of this sequence. Indeed, in our conditions, several conformations are populated at room temperature because their energy levels are close. When a ligand is added, the potential energy surface changes, which can potentially affect which is (are) the most stable structure(s).

We do not known yet whether these 1-K⁺ complexes with and without ligands have the exact same structure or not. As only 8 of the 12 guanines are involved in G-quartets, it is possible that several conformations co-exists in solution with the same characteristics. Therefore, the experimental elucidation of those structures will probably be extremely complicated to achieve with the traditional structural techniques. However, one perspective from the MS data could be using these information to displace the equilibrium towards only one species, easier to study by NMR and more probable to crystalize for X-ray diffraction.

VII.3.2 DO LIGANDS BIND TO G-QUADRUPLEXES BY INDUCED FIT OR CONFORMATIONAL SELECTION?

In Chapter VI, we found that the binding of some ligands induces a conformational change in G-quadruplexes. The ligand binding mechanism is defined by determining if the ligand binds to a preferred conformation and displaces equilibria, or instead induces conformational changes after the binding. These mechanisms are called "conformational selection" and "induced fit", respectively. To distinguish the mechanisms, it is crucial to understand what is the (ensemble of) structure(s) recognized by the ligand (the original structures or the final structures). In principle, this information can be accessed from kinetics experiments and fitting as made for the binding of K⁺ in Chapter V.^{281–283}

However, even though we did not perform such kinetics experiments in the presence of ligand, we can argue about the binding mechanism in light of our folding pathway. The induced fit should be very unlikely for G-quadruplex-selective ligands. In the case of the rearrangement of hybrid G-quadruplexes to antiparallel ones (with either 1 or 2-K⁺), one of the four strands making the G-quadruplex has to break all preformed G-quartets, unfold, and refold.²⁸⁴ Additionally, the unfolding of the G-quadruplex would transiently lead to the formation of a non-G-quadruplex structure, for which the ligands have very low affinities. The ligand is very likely to unbind this structure and, therefore, could not act through an induced fit mechanism. Hence, for very selective G-quadruplex ligands such as those described in this thesis, conformational selection is a more reasonable mechanism: the ligand binds to one of the structures present in solution with high affinity and shifts all the equilibria towards it.

VII.3.3 LIGAND BINDING MODES: STRATEGIES FOR FUTURE WORK

Because structures are not easily obtained from NMR spectroscopy or X-ray diffraction, other techniques are needed to elucidate ligand binding modes. We have shown in this work that mass spectrometry can bring information about the affinity and the selectivity of a ligand for its target, and give information about the structures. Here are some strategies to study ligand binding modes using mass spectrometry.

First, a screening for the binding of a ligand to G-quadruplexes presenting different molecularities, strand orientations, loop types, base stacking... helps sorting out to which structures the ligand binds with the highest affinity. If the ligand binds to a G-quadruplex with high affinity and without inducing structural changes, this G-quadruplex may be the preferred structure. We are currently working on a systematic screening of over 50 DNA sequences for the binding of the three ligands that eject a K⁺ upon binding.²⁸⁵ We hope to find preferential binding to particular G-quadruplex structures and provide more information about the 1-K⁺ structure obtained for the human telomeric G-quadruplexes.

Secondly, studying the collision cross section (CCS) changes upon binding of the ligand could provide structural information if linked with the atomic details brought by molecular modeling. Using ion mobility spectrometry, we are going to study the preferential binding site of a ligand on a G-quadruplex presumably unable to structurally rearrange, $[d(TG_4T)]_4$. To obtain insights on ligand binding modes, it is possible to dock the ligands in different places on the G-quadruplex structure and to use molecular dynamics to correlate the experimental CCS with these atomic models. Ion mobility spectrometry could in the future become a method for fast screening of ligand binding modes.

Another strategy is the covalent labeling of the DNA. Labeling techniques in presence and absence of ligand could give information on the binding pocket of the ligand. For example, methylation of free guanines followed by fragmentation is a potential method to determine which guanines are not involved in G-quartets. The same strategy could be used with hydrogen-deuterium exchange to determine binding sites.

165

VII.4. OUTLOOK

Mass spectrometry highlighted the interplay between folding pathways and ligand binding. We have shown that the binding of ligands often implies changes in the folding landscape of DNA G-quadruplexes and that the structural changes often go with changes in stoichiometries. Mass spectrometry is a suitable technique for distinguishing, quantifying and characterizing the different ensembles formed.

The next key step consists in studying quaternary complexes made by DNA sequences, potassium cations, ligands and proteins. Indeed, so far, we focused on the characterization of the G-quadruplex-ligand interactions. It would be very interesting to consider the protein binding to the complex and, therefore, to determine whether the G-quadruplex bound to the ligand can prevent gene expression or telomerase binding or not.

This is an area of research where mass spectrometry could also provide key insights.

VIII. PH.D. THESIS PUBLICATIONS

List of published articles related to the present thesis.

- 1. Article 1: Marchand, A., Gabelica, V. (2014). Native electrospray mass spectrometry of DNA G-quadruplexes in potassium solution. *J. Am. Soc. Mass Spectrom.*, 25(7), 1146–54.
- Article 4: Marchand, A., Granzhan, A., Iida, K., Tsushima, Y., Ma, Y., Nagasawa, K., Teulade-Fichou, M-P., Gabelica, V. (2015). Ligand-induced conformational changes with cation ejection upon binding to human telomeric DNA G-quadruplexes. *J. Am. Chem. Soc.*, 137(2), 750–6.
- Largy, E., <u>Marchand, A.</u>, Amrane, S., Gabelica, V., Mergny, J.-L. (**2016**). Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *J. Am. Chem. Soc.*, 138(8), 2780–2792.
- <u>Article 3:</u> <u>Marchand, A.</u>, Strzelecka, D., Gabelica, V. (**2016**) Selective and Cooperative Ligand Binding to Antiparallel Human Telomeric DNA G-quadruplexes. *Chem. A Eur. J.*, 22, 9551–9555.
- 5. <u>Article 2: Marchand, A.</u>, Gabelica, V. (**2016**) Folding and Misfolding Pathways of Gquadruplex DNA. *Nucleic Acids Res.*, doi: 10.1093/nar/gkw970.
- Lecours, M. J., <u>Marchand, A.</u>, Hopkins, W. S., Gabelica, V., What Soichiometries Determined by Mass Spectrometry Tell us about the Ligand Binding Mode to G-quadruplex Nucleic Acids. *BBA-General Subjects*, submitted.
- Paul, D., <u>Marchand, A.</u>, Verga, D., Teulade-Fichou, M-P., Bombard, S., Rosu, F., Gabelica, V.
 Determination of Ligand and Cation Binding Sites in Noncovalent Complexes using Mass
 Spectrometry with Electron Photodetachment Dissociation (EPD), in preparation.

IX. REFERENCES

- 1. Alberts, B., Johnson, A. & Lewis, J. in *Molecular Biology of the Cell. 4th edition.* (2002).
- Watson, J. D. & Crick, F. H. C. A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738 (1953).
- Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. S. & Shakked, Z. Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc. Natl. Acad. Sci.* 95, 15194–15199 (1998).
- Shah, S. A. & Brunger, A. T. The 1.8 å crystal structure of a statically disordered 17 base-pair RNA duplex: principles of RNA crystal packing and its effect on nucleic acid structure. *J. Mol. Biol.* 285, 1577–1588 (1999).
- 5. Wang, A. H. *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–6 (1979).
- Frank-Kamenetskii, M. D. & Mirkin, S. M. Triplex DNA Structures. *Annu. Rev. Biochem.* 64, 65– 95 (1995).
- 7. Day, H. A., Pavlou, P. & Waller, Z. A. E. i-Motif DNA: Structure, stability and targeting with ligands. *Bioorg. Med. Chem.* **22**, 4407–4418 (2014).
- 8. Asensio, J. L., Brown, T. & Lane, A. N. Solution conformation of a parallel DNA triple helix with 5' and 3' triplex-duplex junctions. *Structure* **7**, 1–11 (1999).
- 9. Esmaili, N. i-motif solution structure and dynamics of the d(AACCCC) and d(CCCCAA) tetrahymena telomeric repeats. *Nucleic Acids Res.* **33**, 213–224 (2005).
- 10. Phillips, K., Dauter, Z., Murchie, A. I. H., Lilley, D. M. J. & Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95Å resolution. *J. Mol. Biol.* **273**, 171–182 (1997).
- 11. Bang, I. Untersuchungen über die Guanylsäure. *Biochem. Zeitschrif* **26**, 293–311 (1910).
- 12. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* **48**, 2013–8 (1962).

- Parkinson, G. N. in *Quadruplex Nucleic Acids* (eds. Neidle, S. & Balasubramanian, S.) 1–30 (Royal Society of Chemistry, 2006). doi:10.1039/9781847555298-00001
- 14. Lech, C. J., Heddi, B. & Phan, A. T. Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Res.* **41**, 2034–2046 (2013).
- 15. Webba da Silva, M. Geometric formalism for DNA quadruplex folding. *Chemistry* **13**, 9738–45 (2007).
- 16. Mao, X., Marky, L. A. & Gmeiner, W. H. NMR structure of the thrombin-binding DNA aptamer stabilized by Sr2+. *J. Biomol. Struct. Dyn.* **22**, 25–33 (2004).
- 17. Phan, A. T., Kuryavyi, V., Gaw, H. Y. & Patel, D. J. Small-molecule interaction with a fiveguanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* **1**, 167–173 (2005).
- Amrane, S. *et al.* Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite. *J. Am. Chem. Soc.* **134**, 5807–5816 (2012).
- 19. Balkwill, G. D., Garner, T. P., Williams, H. E. L. & Searle, M. S. Folding Topology of a Bimolecular DNA Quadruplex Containing a Stable Mini-hairpin Motif within the Diagonal Loop. *J. Mol. Biol.* **385**, 1600–1615 (2009).
- 20. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* **19**, 414–22 (2009).
- 21. Murat, P. & Balasubramanian, S. Existence and consequences of G-quadruplex structures in DNA. *Curr. Opin. Genet. Dev.* **25**, 22–29 (2014).
- 22. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–6 (2013).
- 23. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–16 (2005).
- 24. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901–7 (2005).
- 25. Dai, J., Chen, D., Jones, R. A., Hurley, L. H. & Yang, D. NMR solution structure of the major G-

quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **34**, 5133–44 (2006).

- 26. Amrane, S. *et al.* Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J. Am. Chem. Soc.* **136**, 5249–52 (2014).
- 27. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S. & Patel, D. J. Structure of an unprecedented Gquadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.* **129**, 4386–92 (2007).
- 28. Heddi, B., Martín-Pintado, N., Serimbetov, Z., Kari, T. M. A. & Phan, A. T. G-quadruplexes with (4 n 1) guanines in the G-tetrad core: formation of a G-triad-water complex and implication for small-molecule binding. *Nucleic Acids Res.* **44**, gkv1357 (2015).
- 29. Bedrat, A., Lacroix, L. & Mergny, J.-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).
- 30. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (2015).
- 31. Rawal, P. Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Res.* **16**, 644–655 (2006).
- 32. Maizels, N. & Gray, L. T. The G4 Genome. *PLoS Genet.* 9, e1003468 (2013).
- 33. Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342**, 825–829 (1989).
- Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* 12, 847–854 (2005).
- 35. Zaug, A. J., Podell, E. R. & Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl. Acad. Sci.* **102**, 10864–10869 (2005).
- Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces cerevisiae. *PLoS Comput. Biol.* 6, e1000861 (2010).
- 37. Nakken, S., Rognes, T. & Hovig, E. The disruptive positions in human G-quadruplex motifs

are less polymorphic and more conserved than their neutral counterparts. *Nucleic Acids Res.* **37**, 5749–5756 (2009).

- 38. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
- 39. Dreesen, O., Li, B. & Cross, G. A. M. Telomere structure and shortening in telomerasedeficient Trypanosoma brucei. *Nucleic Acids Res.* **33**, 4536–43 (2005).
- Jacob, N. K., Skopp, R. & Price, C. M. G-overhang dynamics at Tetrahymena telomeres. *EMBO J.* 20, 4299–4308 (2001).
- 41. Wellinger, R. J., Wolf, A. J. & Zakian, V. A. Saccharomyces telomeres acquire single-strand TG1–3 tails late in S phase. *Cell* **72**, 51–60 (1993).
- Riha, K., McKnight, T. D., Fajkus, J., Vyskot, B. & Shippen, D. E. Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. *Plant J.* 23, 633–641 (2000).
- 43. Paeschke, K., McDonald, K. R. & Zakian, V. A. Telomeres: structures in need of unwinding. *FEBS Lett.* **584**, 3760–72 (2010).
- 44. 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.* **38**, 1022–1033 (2010).
- 45. Mendoza, O., Bourdoncle, A., Boulé, J.-B., Brosh, R. M. & Mergny, J.-L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989–2006 (2016).
- 46. London, T. B. C. *et al.* FANCJ Is a Structure-specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts. *J. Biol. Chem.* **283**, 36132–36139 (2008).
- 47. Hale, T. K., Norris, G. E., Jameson, G. B. & Filichev, V. V. Helicases, G4-DNAs, and Drug Design. *ChemMedChem* 1–5 (2014). doi:10.1002/cmdc.201402068
- Bonnal, S. *et al.* A Single Internal Ribosome Entry Site Containing a G Quartet RNA Structure Drives Fibroblast Growth Factor 2 Gene Expression at Four Alternative Translation Initiation Codons. *J. Biol. Chem.* **278**, 39330–39336 (2003).

- 49. M. Wallgren, J. B. Mohammad, K. Yan, P. P.-L. & , M. Ebrahimi, and N. S. G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures in vitro and are unwound by the Pfh1 DNA helicase. *Nucleic Acids Res.* 1–19 (2016). doi:10.1093/nar/gkw349
- 50. Huppert, J. L. Hunting G-quadruplexes. *Biochimie* **90**, 1140–8 (2008).
- 51. Huang, W.-C. *et al.* Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acids Res.* gkv1061 (2015). doi:10.1093/nar/gkv1061
- 52. De Nicola, B. *et al.* Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic Acids Res.* gkw432 (2016). doi:10.1093/nar/gkw432
- 53. Wang, S.-R. *et al.* A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci. Adv.* **2**, e1501535 (2016).
- 54. Phan, A. T. *et al.* An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 634–9 (2005).
- 55. Wyatt, J. R. *et al.* Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proc. Natl. Acad. Sci.* **91**, 1356–1360 (1994).
- Hotoda, H. *et al.* Biologically Active Oligodeoxyribonucleotides. 5. 1 5'-End-Substituted d(TGGGAG) Possesses Anti-Human Immunodeficiency Virus Type 1 Activity by Forming a G-Quadruplex Structure. *J. Med. Chem.* 41, 3655–3663 (1998).
- 57. Romanucci, V. *et al.* Hairpin oligonucleotides forming G-quadruplexes: New aptamers with anti-HIV activity. *Eur. J. Med. Chem.* **89**, 51–8 (2015).
- 58. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406–413 (2006).
- 59. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887–96 (2006).
- 60. Huppert, J. L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning
and end of UTRs. Nucleic Acids Res. 36, 6260-6268 (2008).

- 61. Qin, Y. & Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA Gquadruplexes found in eukaryotic promoter regions. *Biochimie* **90**, 1149–1171 (2008).
- 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.* **1**, 565–73 (2002).
- 63. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **277**, 3459–3469 (2010).
- 64. Campbell, P. J. Telomeres and Cancer: From Crisis to Stability to Crisis to Stability. *Cell* 148, 633–635 (2012).
- 65. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
- 66. Wang, Q. *et al.* G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res.* **39**, 6229–6237 (2011).
- 67. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by Gquartet DMA structures. *Nature* **350**, 718–720 (1991).
- Sun, D. *et al.* Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *J. Med. Chem.* 40, 2113–2116 (1997).
- 69. Oganesian, L. & Bryan, T. M. Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *BioEssays* **29**, 155–165 (2007).
- Geserick, C. & Blasco, M. A. Novel roles for telomerase in aging. *Mech. Ageing Dev.* 127, 579– 583 (2006).
- 71. Zaccaria, F., Paragi, G. & Fonseca Guerra, C. The role of alkali metal cations in the stabilization of guanine quadruplexes: why K + is the best. *Phys. Chem. Chem. Phys.* (2016). doi:10.1039/C6CP01030J
- 72. Largy, E., Mergny, J.-L. & Gabelica, V. in *The Alkali Metal lons: Their Role for Life* (eds. Sigel, A., Sigel, H. & Sigel, R. K. O.) 203–258 (Springer, 2016). doi:10.1007/978-3-319-21756-7_7

- 73. Dick, D. A. The distribution of sodium, potassium and chloride in the nucleus and cytoplasm of Bufo bufo oocytes measured by electron microprobe analysis. *J. Physiol.* **284**, 37–53 (1978).
- 74. Sastre de Vicente, M. E. The Concept of Ionic Strength Eighty Years after Its Introduction in Chemistry. *J. Chem. Educ.* **81**, 750 (2004).
- 75. Miyoshi, D., Karimata, H. & Sugimoto, N. Hydration regulates thermodynamics of Gquadruplex formation under molecular crowding conditions. *J. Am. Chem. Soc.* **128**, 7957–63 (2006).
- Miyoshi, D. & Sugimoto, N. Molecular crowding effects on structure and stability of DNA. *Biochimie* **90**, 1040–51 (2008).
- 77. Miller, M. C., Buscaglia, R., Chaires, J. B., Lane, A. N. & Trent, J. O. Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d(AG3(TTAG3)3). *J. Am. Chem. Soc.* **132**, 17105–7 (2010).
- 78. Marchand, A. *et al.* Sequence and Solvent Effects on Telomeric DNA Bimolecular G-Quadruplex Folding Kinetics. *J. Phys. Chem. B* **117**, 12391–12401 (2013).
- 79. Wei, Q. & Paterson, B. M. Regulation of MyoD function in the dividing myoblast. *FEBS Lett.***490**, 171–178 (2001).
- Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* 277, 1107–17 (2010).
- 81. Wang, Y. & Patel, D. J. Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* **31**, 8112–9 (1992).
- Wang, Y. & Patel, D. J. Solution Structure of a Parallel-stranded G-Quadruplex DNA. *J. Mol. Biol.* 234, 1171–1183 (1993).
- 83. Gavathiotis, E. & Searle, M. S. Structure of the parallel-stranded DNA quadruplex d(TTAGGGT)4 containing the human telomeric repeat: evidence for A-tetrad formation from NMR and molecular dynamics simulations. *Org. Biomol. Chem.* **1**, 1650–1656 (2003).

- 84. Parkinson, G. N., Lee, M. P. H. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–80 (2002).
- 85. Phan, A. T. & Patel, D. J. Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.* **125**, 15021–7 (2003).
- 86. Zhang, N., Phan, A. T. & Patel, D. J. (3 + 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Am. Chem. Soc.* **127**, 17277–85 (2005).
- 87. Wang, Y. & Patel, D. J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] Gtetraplex. *Structure* **1**, 263–82 (1993).
- Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L. & Patel, D. J. Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **128**, 9963– 70 (2006).
- 89. Dai, J., Carver, M., Punchihewa, C., Jones, R. A. & Yang, D. 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 (2007).
- 90. Lim, K. W. *et al.* Structure of the human telomere in K+ solution: a stable basket-type Gquadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.* **131**, 4301–9 (2009).
- 91. 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.* **37**, 6239–6248 (2009).
- Li, J., Correia, J. J., Wang, L., Trent, J. O. & Chaires, J. B. Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res.* 33, 4649–59 (2005).
- Yue, Y. *et al.* Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in
 K+ solution under molecular crowding condition. *J. Am. Chem. Soc.* **129**, 11185–91 (2007).
- 94. Phan, A. T., Kuryavyi, V., Luu, K. N. & Patel, D. J. Structure of two intramolecular Gquadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Res.* **35**, 6517–25 (2007).

- 95. Zhang, Z., Dai, J., Veliath, E., Jones, R. A. & Yang, D. Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.* 38, 1009–1021 (2010).
- 96. Hänsel, R., Löhr, F., Trantirek, L. & Dötsch, V. High-Resolution Insight into G-Overhang Architecture. *J. Am. Chem. Soc.* **135**, 2816–2824 (2013).
- 97. Renciuk, D. *et al.* Arrangements of human telomere DNA quadruplex in physiologically relevant K+ solutions. *Nucleic Acids Res.* **37**, 6625–34 (2009).
- 98. Monchaud, D. & Teulade-Fichou, M.-P. A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* **6**, 627–36 (2008).
- Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. J. Med. Chem. 59, 5987– 6011 (2016).
- 100. Gavathiotis, E., Heald, R. a., Stevens, M. F. G. & Searle, M. S. Drug Recognition and Stabilisation of the Parallel-stranded DNA Quadruplex d(TTAGGGT)4 Containing the Human Telomeric Repeat. *J. Mol. Biol.* **334**, 25–36 (2003).
- 101. Hounsou, C. *et al.* G-Quadruplex Recognition by Quinacridines: a SAR, NMR, and Biological Study. *ChemMedChem* **2**, 655–666 (2007).
- 102. Tawani, A. & Kumar, A. Structural Insight into the interaction of Flavonoids with Human Telomeric Sequence. *Sci. Rep.* **5**, 17574 (2015).
- 103. Campbell, N., Collie, G. W. & Neidle, S. in *Current Protocols in Nucleic Acid Chemistry* 17.6.1-17.6.22 (John Wiley & Sons, Inc., 2012). doi:10.1002/0471142700.nc1706s50
- 104. Collie, G. W., Campbell, N. H. & Neidle, S. Loop flexibility in human telomeric quadruplex small-molecule complexes. *Nucleic Acids Res.* **43**, 4785–4799 (2015).
- 105. Bazzicalupi, C., Ferraroni, M., Bilia, A. R., Scheggi, F. & Gratteri, P. The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Res.* **41**, 632–8 (2013).
- 106. Collie, G. W. et al. Structural basis for telomeric G-quadruplex targeting by naphthalene

diimide ligands. J. Am. Chem. Soc. 134, 2723-31 (2012).

- 107. Heddi, B., Cheong, V. V., Martadinata, H. & Phan, A. T. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide–quadruplex complex. *Proc. Natl. Acad. Sci.* **112**, 9608–9613 (2015).
- 108. Cosconati, S. *et al.* Structural and Conformational Requisites in DNA Quadruplex Groove Binding: Another Piece to the Puzzle. *J. Am. Chem. Soc.* **132**, 6425–6433 (2010).
- 109. Martino, L. *et al.* Structural and Thermodynamic Studies of the Interaction of Distamycin A with the Parallel Quadruplex Structure [d(TGGGGT)] 4. *J. Am. Chem. Soc.* **129**, 16048–16056 (2007).
- 110. Campbell, N. H. *et al.* Selectivity in ligand recognition of G-quadruplex loops. *Biochemistry*48, 1675–80 (2009).
- 111. Komor, A. C. & Barton, J. K. The path for metal complexes to a DNA target. *Chem. Commun.***49**, 3617 (2013).
- 112. Li, Q. *et al.* G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* **41**, D1115–D1123 (2013).
- 113. Zhang, S., Wu, Y. & Zhang, W. G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem* **9**, 899–911 (2014).
- 114. Romera, C. *et al.* Improvement of porphyrins for G-quadruplex DNA targeting. *Biochimie* **93**, 1310–1317 (2011).
- 115. Largy, E., Hamon, F. & Teulade-Fichou, M.-P. Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Anal. Bioanal. Chem.* **400**, 3419–3427 (2011).
- 116. Shi, D.-F., Wheelhouse, R. T., Sun, D. & Hurley, L. H. Quadruplex-Interactive Agents as Telomerase Inhibitors: Synthesis of Porphyrins and Structure–Activity Relationship for the Inhibition of Telomerase. *J. Med. Chem.* **44**, 4509–4523 (2001).
- 117. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. 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. 99, 11593–11598 (2002).

- 118. Parkinson, G. N., Ghosh, R. & Neidle, S. Structural basis for binding of porphyrin to human telomeres. *Biochemistry* **46**, 2390–7 (2007).
- 119. Ren, J. & Chaires, J. B. Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry* **38**, 16067–75 (1999).
- 120. De Rache, A. & Mergny, J.-L. Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie* **115**, 194–202 (2015).
- 121. Marchand, A. *et al.* Ligand-induced conformational changes with cation ejection upon binding to human telomeric DNA G-quadruplexes. *J. Am. Chem. Soc.* **137**, 750–756 (2015).
- 122. Morris, M. J., Wingate, K. L., Silwal, J., Leeper, T. C. & Basu, S. The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res.* **40**, 4137–45 (2012).
- 123. Dixon, I. M. *et al.* A G-Quadruplex Ligand with 10000-Fold Selectivity over Duplex DNA. *J. Am. Chem. Soc.* **129**, 1502–1503 (2007).
- 124. Romera, C., Sabater, L., Garofalo, A., M. Dixon, I. & Pratviel, G. Interaction of Cationic Nickel and Manganese Porphyrins with the Minor Groove of DNA. *Inorg. Chem.* **49**, 8558–8567 (2010).
- 125. Nicoludis, J. M., Barrett, S. P., Mergny, J.-L. L. & Yatsunyk, L. a. Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.* **40**, 5432–47 (2012).
- 126. Sabharwal, N. C. *et al.* N -methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J.* **281**, 1726–1737 (2014).
- 127. Sabharwal, N. C. *et al.* Investigation of the interactions between Pt(II) and Pd(II) derivatives of 5,10,15,20-tetrakis (N-methyl-4-pyridyl) porphyrin and G-quadruplex DNA. *JBIC J. Biol. Inorg. Chem.* (2016). doi:10.1007/s00775-015-1325-8
- 128. Reed, J. E., Arnal, A. A., Neidle, S. & Vilar, R. Stabilization of G-Quadruplex DNA and Inhibition of Telomerase Activity by Square-Planar Nickel(II) Complexes. *J. Am. Chem. Soc.* **128**, 5992– 5993 (2006).

- 129. Lecarme, L. *et al.* Efficient Inhibition of Telomerase by Nickel-Salophen Complexes. *ChemMedChem* **11**, 1133–1136 (2016).
- 130. Campbell, N. H. *et al.* Molecular basis of structure-activity relationships between salphen metal complexes and human telomeric DNA quadruplexes. *J. Med. Chem.* **55**, 209–22 (2012).
- 131. Davis, K. J. *et al.* Synthesis and characterisation of nickel Schiff base complexes containing the meso-1,2-diphenylethylenediamine moiety: selective interactions with a tetramolecular DNA quadruplex. *Dalt. Trans.* **44**, 3136–3150 (2015).
- 132. Bertrand, H. *et al.* The importance of metal geometry in the recognition of G-quadruplex-DNA by metal-terpyridine complexes. *Org. Biomol. Chem.* **5**, 2555 (2007).
- 133. Largy, E. *et al.* Tridentate N-donor palladium(II) complexes as efficient coordinating quadruplex DNA binders. *Chemistry* **17**, 13274–83 (2011).
- 134. Bertrand, H. *et al.* Exclusive platination of loop adenines in the human telomeric Gquadruplex. *Org. Biomol. Chem.* **7**, 2864 (2009).
- 135. Ghosh, S. *et al.* Assembly of Palladium(II) and Platinum(II) Metallo-Rectangles with a Guanosine-Substituted Terpyridine and Study of Their Interactions with Quadruplex DNA. *Chem. A Eur. J.* **20**, 4772–4779 (2014).
- 136. Li, Y. *et al.* Terpyridine–Cu(II) targeting human telomeric DNA to produce highly stereospecific G-quadruplex DNA metalloenzyme. *Chem. Sci.* **6**, 5578–5585 (2015).
- 137. Shi, S. *et al.* Interaction of [Ru(bpy)2(dppz)]2+ with human telomeric DNA: preferential binding to G-quadruplexes over i-motif. *Biochimie* **92**, 370–7 (2010).
- 138. Rickling, S. *et al.* A rigid dinuclear ruthenium(II) complex as an efficient photoactive agent for bridging two guanine bases of a duplex or quadruplex oligonucleotide. *Chemistry* **16**, 3951–61 (2010).
- 139. Wilson, T., Costa, P. J., Félix, V., Williamson, M. P. & Thomas, J. A. Structural studies on dinuclear ruthenium(II) complexes that bind diastereoselectively to an antiparallel folded human telomere sequence. *J. Med. Chem.* **56**, 8674–83 (2013).
- 140. Gill, M. R. et al. A ruthenium(II) polypyridyl complex for direct imaging of DNA structure in

living cells. Nat. Chem. 1, 662–7 (2009).

- 141. Lemarteleur, T. *et al.* Stabilization of the c-myc gene promoter quadruplex by specific ligands' inhibitors of telomerase. *Biochem. Biophys. Res. Commun.* **323**, 802–808 (2004).
- 142. Pennarun, G. *et al.* Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene* **24**, 2917–28 (2005).
- 143. Granotier, C. *et al.* Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.* **33**, 4182–90 (2005).
- 144. De Cian, A. & Mergny, J.-L. Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. *Nucleic Acids Res.* **35**, 2483–93 (2007).
- Monchaud, D., Yang, P., Lacroix, L., Teulade-Fichou, M.-P. & Mergny, J.-L. A metal-mediated conformational switch controls G-quadruplex binding affinity. *Angew. Chem. Int. Ed. Engl.* 47, 4858–61 (2008).
- 146. De Cian, A., Delemos, E., Mergny, J.-L., Teulade-Fichou, M.-P. & Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.* **129**, 1856–7 (2007).
- 147. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M.-P. & Phan, A. T. Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC 3. *Angew. Chemie Int. Ed.*53, 999–1002 (2014).
- 148. Mendoza, O., Gueddouda, N. M., Boule, J.-B., Bourdoncle, A. & Mergny, J.-L. A fluorescence-based helicase assay: application to the screening of G-quadruplex ligands. *Nucleic Acids Res.*43, e71 (2015).
- 149. Lopes, J. *et al.* G-quadruplex-induced instability during leading-strand replication. *EMBO J.***30**, 4033–46 (2011).
- 150. Piazza, A. *et al.* Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acids Res.* **38**, 4337–48 (2010).
- 151. Rodriguez, R. et al. A novel small molecule that alters shelterin integrity and triggers a DNA-

damage response at telomeres. J. Am. Chem. Soc. 130, 15758-9 (2008).

- 152. Koirala, D. *et al.* A single-molecule platform for investigation of interactions between Gquadruplexes and small-molecule ligands. *Nat. Chem.* **3**, 782–7 (2011).
- 153. Müller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated Gquadruplex isolation from human cells. *Nat. Chem.* **2**, 1095–1098 (2010).
- 154. Shin-ya, K. *et al.* Telomestatin, a Novel Telomerase Inhibitor from Streptomyces anulatus. *J. Am. Chem. Soc.* **123**, 1262–1263 (2001).
- 155. Linder, J., Garner, T. P., Williams, H. E. L., Searle, M. S. & Moody, C. J. Telomestatin: formal total synthesis and cation-mediated interaction of its seco-derivatives with G-quadruplexes. *J. Am. Chem. Soc.* **133**, 1044–51 (2011).
- 156. Rosu, F. *et al.* Cation involvement in telomestatin binding to g-quadruplex DNA. *J. Nucleic Acids* **2010**, 1–7 (2010).
- 157. Rosu, F., Gabelica, V., Shin-ya, K. & De Pauw, E. Telomestatin-induced stabilization of the human telomeric DNA quadruplex monitored by electrospray mass spectrometry. *Chem. Commun.* **2**, 2702 (2003).
- 158. Rezler, E. M. *et al.* Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J. Am. Chem. Soc.* **127**, 9439–47 (2005).
- 159. Gomez, D. *et al.* Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. *J. Biol. Chem.* **281**, 38721–9 (2006).
- 160. Hamon, F. *et al.* An acyclic oligoheteroaryle that discriminates strongly between diverse Gquadruplex topologies. *Angew. Chem. Int. Ed. Engl.* **50**, 8745–9 (2011).
- 161. Petenzi, M. *et al.* Cationic Pentaheteroaryls as Selective G-Quadruplex Ligands by Solvent-Free Microwave-Assisted Synthesis. *Chem. - A Eur. J.* **18**, 14487–14496 (2012).
- 162. Tera, M. *et al.* Macrocyclic hexaoxazoles as sequence- and mode-selective G-quadruplex binders. *Angew. Chem. Int. Ed. Engl.* **47**, 5557–60 (2008).
- 163. Chung, W. J. et al. Solution structure of an intramolecular (3 + 1) human telomeric G-

quadruplex bound to a telomestatin derivative. J. Am. Chem. Soc. 135, 13495–501 (2013).

- 164. Sakuma, M. *et al.* Design and synthesis of unsymmetric macrocyclic hexaoxazole compounds with an ability to induce distinct G-quadruplex topologies in telomeric DNA. *Org. Biomol. Chem.* **24**, (2016).
- 165. lida, K., Majima, S., Nakamura, T., Seimiya, H. & Nagasawa, K. Evaluation of the Interaction between Long Telomeric DNA and Macrocyclic Hexaoxazole (6OTD) Dimer of a Gquadruplex Ligand. *Molecules* **18**, 4328–4341 (2013).
- 166. Berova, N., Bari, L. Di & Pescitelli, G. Application of electronic circular dichroism in configurational and conformational analysis of organic compounds. *Chem. Soc. Rev.* **36**, 914 (2007).
- 167. Masiero, S. *et al.* A non-empirical chromophoric interpretation of CD spectra of DNA Gquadruplex structures. *Org. Biomol. Chem.* **8**, 2683–92 (2010).
- 168. Fülscher, M. P., Serrano-Andrés, L. & Roos, B. O. A Theoretical Study of the Electronic Spectra of Adenine and Guanine. *J. Am. Chem. Soc.* **119**, 6168–6176 (1997).
- 169. Rosu, F. *et al.* UV Spectroscopy of DNA Duplex and Quadruplex Structures in the Gas Phase. *J. Phys. Chem. A* **116**, 5383–5391 (2012).
- 170. Karsisiotis, A. I. *et al.* Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed. Engl.* **50**, 10645–8 (2011).
- 171. Paramasivan, S., Rujan, I. & Bolton, P. H. Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods* **43**, 324–31 (2007).
- 172. Vorlíčková, M. et al. Circular dichroism and guanine quadruplexes. *Methods* **57**, 64–75 (2012).
- 173. Webba da Silva, M. NMR methods for studying quadruplex nucleic acids. *Methods* 43, 264–277 (2007).
- 174. Adrian, M., Heddi, B. & Phan, A. T. NMR spectroscopy of G-quadruplexes. *Methods* **57**, 11–24 (2012).
- 175. Mergny, J. & Lacroix, L. Des Tms, encore des Tms, toujours des Tms *Regards sur la Biochim.*2, 36–52 (2002).

- 176. Mergny, J. & Lacroix, L. Analysis of thermal melting curves. *Oligonucleotides* **537**, 515–537 (2003).
- 177. De Cian, A. *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods* **42**, 183–95 (2007).
- 178. Rachwal, P. a. & Fox, K. R. Quadruplex melting. *Methods* 43, 291–301 (2007).
- 179. Mergny, J.-L. & Lacroix, L. in *Current Protocols in Nucleic Acid Chemistry* 17.1.1-17.1.15 (John Wiley & Sons, Inc., 2009). doi:10.1002/0471142700.nc1701s37
- 180. Gray, R. D. & Chaires, J. B. in *Current Protocols in Nucleic Acid Chemistry* 17.4.1-17.4.16 (John Wiley & Sons, Inc., 2011). doi:10.1002/0471142700.nc1704s45
- 181. Gray, R. D., Buscaglia, R. & Chaires, J. B. Populated intermediates in the thermal unfolding of the human telomeric quadruplex. *J. Am. Chem. Soc.* **134**, 16834–44 (2012).
- 182. Buscaglia, R., Gray, R. D. & Chaires, J. B. Thermodynamic characterization of human telomere quadruplex unfolding. *Biopolymers* **99**, 1006–18 (2013).
- 183. Redman, J. E. Surface plasmon resonance for probing quadruplex folding and interactions with proteins and small molecules. *Methods* **43**, 302–312 (2007).
- 184. Read, M. *et al.* Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci.* **98**, 4844–4849 (2001).
- 185. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* 5, 507–516 (2008).
- 186. Tippana, R., Xiao, W. & Myong, S. G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **42**, 8106–8114 (2014).
- 187. Noer, S. L. *et al.* Folding dynamics and conformational heterogeneity of human telomeric Gquadruplex structures in Na+ solutions by single molecule FRET microscopy. *Nucleic Acids Res.* gkv1320 (2015). doi:10.1093/nar/gkv1320
- 188. Jena, P. V. *et al.* G-Quadruplex DNA Bound by a Synthetic Ligand is Highly Dynamic. *J. Am. Chem. Soc.* **131**, 12522–12523 (2009).

- 189. Debnath, M. *et al.* Small molecule regulated dynamic structural changes of human Gquadruplexes. *Chem. Sci.* **7**, 3279–3285 (2016).
- 190. 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-quadruplex structures. *Biochimie* **93**, 1288–96 (2011).
- 191. Largy, E. Ciblage d'Acides Nucléiques G-quadruplexes. *Thesis* (2011).
- 192. Monchaud, D. *et al.* Ligands playing musical chairs with G-quadruplex DNA: a rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie* **90**, 1207–23 (2008).
- 193. Maher, S., Jjunju, F. P. M. & Taylor, S. ColloquiumD: 100 yearsPerspectives and future trends. *Rev. Mod. Phys.* 87, 113–135 (2015).
- 194. Nobelprize.org. The Nobel Prize in Physics 1906. *Nobel Media AB 2014* at http://www.nobelprize.org/nobel_prizes/physics/laureates/1906/
- 195. Wolff, M. M. & Stephens, W. E. A Pulsed Mass Spectrometer with Time Dispersion. *Rev. Sci. Instrum.* **24**, 616 (1953).
- 196. Hipple, J. A., Sommer, H. & Thomas, H. A. A Precise Method of Determining the Faraday by Magnetic Resonance. *Phys. Rev.* **76**, 1877–1878 (1949).
- 197. Paul, W. & Steinwedel, H. Notizen: Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforsch. A* **8**, (1953).
- 198. Dole, M. Molecular Beams of Macroions. J. Chem. Phys. 49, 2240 (1968).
- 199. Yamashita, M. & Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme.*J. Phys. Chem.* 88, 4451–4459 (1984).
- 200. Fenn, J. B. Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chemie Int. Ed.* **42**, 3871–3894 (2003).
- 201. Fenn, J., Mann, M., Meng, C., Wong, S. & Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-.).* **246,** 64–71 (1989).

- 202. Dyachenko, A. *et al.* Tandem Native Mass-Spectrometry on Antibody–Drug Conjugates and Submillion Da Antibody–Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal. Chem.* **87**, 6095–6102 (2015).
- 203. Benjamin, D. R., Robinson, C. V, Hendrick, J. P., Hartl, F. U. & Dobson, C. M. Mass spectrometry of ribosomes and ribosomal subunits. *Proc. Natl. Acad. Sci.* **95**, 7391–7395 (1998).
- 204. Videler, H., Ilag, L. L., McKay, A. R. C., Hanson, C. L. & Robinson, C. V. Mass spectrometry of intact ribosomes. *FEBS Lett.* **579**, 943–947 (2005).
- 205. Fuerstenau, S. D. *et al.* Mass Spectrometry of an Intact Virus. *Angew. Chemie* **113**, 559–562 (2001).
- 206. Sharon, M. & Robinson, C. V. The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. *Annu. Rev. Biochem.* **76**, 167–193 (2007).
- 207. Walzthoeni, T., Leitner, A., Stengel, F. & Aebersold, R. Mass spectrometry supported determination of protein complex structure. *Curr. Opin. Struct. Biol.* **23**, 252–60 (2013).
- 208. Beveridge, R., Chappuis, Q., Macphee, C. & Barran, P. Mass spectrometry methods for intrinsically disordered proteins. *Analyst* **138**, 32–42 (2013).
- 209. Heck, A. J. R. & van den Heuvel, R. H. H. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **23**, 368–389 (2004).
- 210. Rappsilber, J. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**, 530–540 (2011).
- 211. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
- 212. Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **6**, 281–294 (2014).
- 213. Clemmer, D. E. & Jarrold, M. F. Ion Mobility Measurements and their Applications to Clusters and Biomolecules. *J. Mass Spectrom.* **32**, 577–592 (1997).

- 214. Shvartsburg, A. *Differential Ion Mobility Spectrometry*. (CRC Press, 2008). doi:10.1201/9781420051070
- 215. Shvartsburg, A. A. & Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **80**, 9689–9699 (2008).
- Vidal-de-Miguel, G., Macía, M. & Cuevas, J. Transversal Modulation Ion Mobility Spectrometry (TM-IMS), A New Mobility Filter Overcoming Turbulence Related Limitations. *Anal. Chem.* 84, 7831–7837 (2012).
- 217. Loboda, A. Novel ion mobility setup combined with collision cell and time-of-flight mass spectrometer. *J. Am. Soc. Mass Spectrom.* **17**, 691–699 (2006).
- 218. Michelmann, K., Silveira, J. A., Ridgeway, M. E. & Park, M. A. Fundamentals of Trapped Ion Mobility Spectrometry. *J. Am. Soc. Mass Spectrom.* **26**, 14–24 (2015).
- 219. Ion Mobility Spectrometry Mass Spectrometry: Theory and Applications. (CRC Press, 2010).
- Sun, Y., Vahidi, S., Sowole, M. A. & Konermann, L. Protein Structural Studies by Traveling Wave Ion Mobility Spectrometry: A Critical Look at Electrospray Sources and Calibration Issues. *J. Am. Soc. Mass Spectrom.* 27, 31–40 (2016).
- 221. Lapthorn, C., Pullen, F. & Chowdhry, B. Z. Ion mobility spectrometry-mass spectrometry (IMS-MS) of small molecules: separating and assigning structures to ions. *Mass Spectrom. Rev.* **32**, 43–71 (2013).
- 222. Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J. & Robinson, C. V. Ion mobilitymass spectrometry analysis of large protein complexes. *Nat. Protoc.* **3**, 1139–52 (2008).
- 223. D'Atri, V., Porrini, M., Rosu, F. & Gabelica, V. Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass Spectrom.* **50**, 711–726 (2015).
- 224. Peschke, M., Verkerk, U. H. & Kebarle, P. Features of the ESI mechanism that affect the observation of multiply charged noncovalent protein complexes and the determination of the association constant by the titration method. *J. Am. Soc. Mass Spectrom.* **15**, 1424–1434 (2004).

- 225. Wilm, M. Principles of Electrospray Ionization. *Mol. Cell. Proteomics* **10**, M111.009407 (2011).
- 226. Iribarne, J. V. On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**, 2287 (1976).
- 227. Gabelica, V., Rosu, F. & De Pauw, E. A simple method to determine electrospray response factors of noncovalent complexes. *Anal. Chem.* **81**, 6708–15 (2009).
- 228. Yin, S., Xie, Y. & Loo, J. A. Mass Spectrometry of Protein-Ligand Complexes: Enhanced Gas-Phase Stability of Ribonuclease-Nucleotide Complexes. *J. Am. Soc. Mass Spectrom.* **19**, 1199– 1208 (2008).
- 229. Hardman, M. & Makarov, A. A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal. Chem.* **75**, 1699–705 (2003).
- 230. Silvestri, C. & Brodbelt, J. S. Tandem mass spectrometry for characterization of covalent adducts of DNA with anticancer therapeutics. *Mass Spectrom. Rev.* **32**, 247–66 (2013).
- 231. Balthasart, F., Plavec, J. & Gabelica, V. Ammonium ion binding to DNA G-quadruplexes: do electrospray mass spectra faithfully reflect the solution-phase species? *J. Am. Soc. Mass Spectrom.* **24**, 1–8 (2013).
- Benesch, J. L. P. Collisional Activation of Protein Complexes: Picking Up the Pieces. J. Am. Soc.
 Mass Spectrom. 20, 341–348 (2009).
- 233. Mason, E. A. & Schamp, H. W. Mobility of gaseous ions in weak electric fields. *Ann. Phys.* 4, 233–270 (1958).
- 234. Revercomb, H. E. & Mason, E. A. Theory of plasma chromatography/gaseous electrophoresis. Review. *Anal. Chem.* **47**, 970–983 (1975).
- 235. Sun, J., Kitova, E. N., Wang, W. & Klassen, J. S. Method for distinguishing specific from nonspecific protein-ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal. Chem.* **78**, 3010–8 (2006).
- 236. Kuzmic, P. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **237**, 260–273 (1996).
- 237. Hu, L., Lim, K. W., Bouaziz, S. & Phan, A. T. 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 (2009).

- 238. Goodlett, D. R., Camp, D. G., Hardin, C. C., Corregan, M. & Smith, R. D. Direct observation of a DNA quadruplex by electrospray ionization mass spectrometry. *Biol. Mass Spectrom.* 22, 181–3 (1993).
- 239. Baker, E. S., Dupuis, N. F. & Bowers, M. T. DNA Hairpin, Pseudoknot, and Cruciform Stability in a Solvent-Free Environment. *J. Phys. Chem. B* **113**, 1722–1727 (2009).
- 240. Kellersberger, K. a, Yu, E., Kruppa, G. H., Young, M. M. & Fabris, D. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Anal. Chem.* **76**, 2438–45 (2004).
- 241. 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.* **132**, 9328–34 (2010).
- 242. Smargiasso, N. *et al.* G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.* **130**, 10208–16 (2008).
- 243. Evans, S. E. *et al.* End-stacking of copper cationic porphyrins on parallel-stranded guanine quadruplexes. *J. Biol. Inorg. Chem.* **12**, 1235–49 (2007).
- 244. Pierce, S. E., Wang, J., Jayawickramarajah, J., Hamilton, A. D. & Brodbelt, J. S. Examination of the effect of the annealing cation on higher order structures containing guanine or isoguanine repeats. *Chemistry* **15**, 11244–55 (2009).
- 245. Kharlamova, A., Prentice, B. M., Huang, T. & McLuckey, S. Electrospray droplet exposure to gaseous acids for reduction of metal counter-ions in nucleic acid ions. *Int. J. Mass Spectrom.* **300**, 158–166 (2011).
- 246. Turner, K. B., Monti, S. a & Fabris, D. Like polarity ion/ion reactions enable the investigation of specific metal interactions in nucleic acids and their noncovalent assemblies. *J. Am. Chem. Soc.* **130**, 13353–63 (2008).
- 247. Potier, N., Van Dorsselaer, A., Cordier, Y., Roch, O. & Bischoff, R. Negative electrospray ionization mass spectrometry of synthetic and chemically modified oligonucleotides. *Nucleic*

Acids Res. 22, 3895–903 (1994).

- 248. Deguchi, K. *et al.* Enhanced mass detection of oligonucleotides using reverse-phase highperformance liquid chromatography/electrospray ionization ion-trap mass spectrometry. *Rapid Commun. Mass Spectrom.* **16**, 2133–41 (2002).
- 249. Lemaire, D., Marie, G., Serani, L. & Laprévote, O. Stabilization of gas-phase noncovalent macromolecular complexes in electrospray mass spectrometry using aqueous triethylammonium bicarbonate buffer. *Anal. Chem.* **73**, 1699–706 (2001).
- 250. Gray, R. D. & Chaires, J. B. Linkage of cation binding and folding in human telomeric quadruplex DNA. *Biophys. Chem.* **159**, 205–9 (2011).
- 251. Rosu, F., Gabelica, V., Houssier, C., Colson, P. & Pauw, E. De. Triplex and quadruplex DNA structures studied by electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **16**, 1729–36 (2002).
- 252. 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.* **35**, 3064–75 (2007).
- 253. Rosu, F., Gabelica, V., Poncelet, H. & De Pauw, E. Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry. *Nucleic Acids Res.* **38**, 5217–25 (2010).
- 254. Rosu, F., De Pauw, E. & Gabelica, V. Electrospray mass spectrometry to study drug-nucleic acids interactions. *Biochimie* **90**, 1074–1087 (2008).
- 255. Ferreira, R., Marchand, A. & Gabelica, V. Mass spectrometry and ion mobility spectrometry of G-quadruplexes. A study of solvent effects on dimer formation and structural transitions in the telomeric DNA sequence d(TAGGGTTAGGGT). *Methods* **57**, 56–63 (2012).
- 256. Schultze, P., Smith, F. W. & Feigon, J. Refined solution structure of the dimeric quadruplex formed from the Oxytricha telomeric oligonucleotide d(GGGGTTTTGGGGG). *Structure* **2**, 221–233 (1994).
- 257. Ennifar, E., Walter, P. & Dumas, P. A crystallographic study of the binding of 13 metal ions to two related RNA duplexes. *Nucleic Acids Res.* **31**, 2671–2682 (2003).
- 258. Wong, A. & Wu, G. 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.* **125**, 13895–13905 (2003).

- 259. Gray, R. D., Trent, J. O. & Chaires, J. B. Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.* **426**, 1629–50 (2014).
- Mashimo, T., Yagi, H., Sannohe, Y., Rajendran, A. & Sugiyama, H. Folding pathways of human telomeric type-1 and type-2 G-Quadruplex structures. *J. Am. Chem. Soc.* **132**, 14910–14918 (2010).
- 261. Ceru, S., Sket, P., Prislan, I., Lah, J. & Plavec, J. A new pathway of DNA G-quadruplex formation. *Angew. Chem. Int. Ed. Engl.* **53**, 4881–4 (2014).
- Su, D. G. T., Fang, H., Gross, M. L. & Taylor, J.-S. A. Photocrosslinking of human telomeric Gquadruplex loops by anti cyclobutane thymine dimer formation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12861–6 (2009).
- 263. Stadlbauer, P., Trantírek, L., Cheatham, T. E., Koča, J. & Sponer, J. Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie* **105**, 22–35 (2014).
- 264. Stadlbauer, P. *et al.* Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res.* **43**, 9626–44 (2015).
- 265. Limongelli, V. et al. The G-triplex DNA. Angew. Chem. Int. Ed. Engl. 52, 2269–73 (2013).
- Largy, E., Marchand, A., Amrane, S., Gabelica, V. & Mergny, J.-L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *J. Am. Chem. Soc.* **138**, 2780–2792 (2016).
- 267. Diveshkumar, K. V. *et al.* Specific Stabilization of c-MYC and c-KIT G-Quadruplex DNA Structures by Indolylmethyleneindanone Scaffolds. *Biochemistry* **55**, 3571–3585 (2016).
- 268. Gray, R. D., Li, J. & Chaires, J. B. Energetics and kinetics of a conformational switch in Gquadruplex DNA. *J. Phys. Chem. B* **113**, 2676–83 (2009).
- 269. Wang, J.-T., Li, Y., Tan, J.-H., Ji, L.-N. & Mao, Z.-W. Platinum(II)-triarylpyridines complexes with

electropositive pendants as efficient G-quadruplex binders. Dalt. Trans. 40, 564–566 (2011).

- 270. Rosu, F., Gabelica, V., De Pauw, E., Mailliet, P. & Mergny, J.-L. Cooperative 2:1 Binding of a Bisphenothiazine to Duplex DNA. *ChemBioChem* **9**, 849–852 (2008).
- 271. Gabelica, V. *et al.* Multiple and Cooperative Binding of Fluorescence Light-up Probe Thioflavin T with Human Telomere DNA G-Quadruplex. *Biochemistry* **52**, 5620–5628 (2013).
- 272. Murray, K. K. DNA sequencing by mass spectrometry. J. Mass Spectrom. 31, 1203–15 (1996).
- 273. Niu, S. & Ruotolo, B. T. Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding. *Protein Sci.* **24**, 1272–1281 (2015).
- 274. Mehmood, S. *et al.* Mass spectrometry captures off-target drug binding and provides mechanistic insights into the human metalloprotease ZMPSTE24. *Nat. Chem.* (2016). doi:10.1038/nchem.2591
- 275. Renaud de la Faverie, A. *et al.* Nucleic acids targeted to drugs: SELEX against a quadruplex ligand. *Biochimie* **93**, 1357–67 (2011).
- 276. Rajski, S. R. & Williams, R. M. DNA Cross-Linking Agents as Antitumor Drugs. *Chem. Rev.* **98**, 2723–2796 (1998).
- 277. Paul, D. *et al.* Determination of Ligand and Cation Binding Sites in Noncovalent Complexes using Mass Spectrometry with Electron Photodetachment Dissociation (EPD). In preparation.
- 278. Bončina, M. *et al.* Thermodynamic fingerprints of ligand binding to human telomeric Gquadruplexes. *Nucleic Acids Res.* gkv1167 (2015). doi:10.1093/nar/gkv1167
- 279. Bončina, M. *et al.* Dominant Driving Forces in Human Telomere Quadruplex Binding-Induced Structural Alterations. *Biophys. J.* **108**, 2903–2911 (2015).
- 280. Lebars, I. *et al.* Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res.* **36**, 7146–7156 (2008).
- 281. Gianni, S., Dogan, J. & Jemth, P. Distinguishing induced fit from conformational selection. *Biophys. Chem.* **189**, 33–39 (2014).

- 282. Sharon, M. & Horovitz, A. Probing allosteric mechanisms using native mass spectrometry. *Curr. Opin. Struct. Biol.* **34**, 7–16 (2015).
- 283. Vogt, A. D., Pozzi, N., Chen, Z. & Di Cera, E. Essential role of conformational selection in ligand binding. *Biophys. Chem.* **186**, 13–21 (2014).
- 284. Stadlbauer, P., Krepl, M., Cheatham, T. E., Koca, J. & Sponer, J. Structural dynamics of possible late-stage intermediates in folding of quadruplex DNA studied by molecular simulations. *Nucleic Acids Res.* **41**, 7128–7143 (2013).
- 285. Lecours, M. J., Marchand, A., Hopkins, W. S. & Gabelica, V. What Soichiometries Determined by Mass Spectrometry Tell us about the Ligand Binding Mode to G-quadruplex Nucleic Acids. *BBA-General Subj.* Submitted.

X. ANNEXES

SUPPORTING INFORMATION ARTICLE 1

Native Electrospray Mass Spectrometry of DNA G-Quadruplexes in Potassium Solution

A. Marchand & V. Gabelica

Journal of the American Society for Mass Spectrometry **2014**, 25, 1146–1154

Native electrospray mass spectrometry

of DNA G-quadruplexes in potassium solution

A. Marchand^{1,2}, V. Gabelica^{1,2,*}

¹ Univ. Bordeaux, IECB, ARNA Laboratory, F-33600 Pessac, France.

² Inserm, U869, ARNA Laboratory, F-33000 Bordeaux, France.

* Corresponding author email: valerie.gabelica@inserm.fr

SUPPORTING INFORMATION



Figure S1. ESI-MS of 10 μ M G₄T₄G₄ in 100 mM NH₄OAc. From A to D (30, 50, 70 and 90 V, respectively): the effect of the ion guide 1 voltage on the ammonium cation preservation in the gas-phase. Left is the full scale spectrum and on the right is the zoom on the dimer D⁵⁻ peak.



Figure S2. ESI-MS of 5 μ M 22non-G4 sequence d(GGG-ATG-CGA-CAG-AGA-GGA-CGG-G). From A to E and from F to J the ion guide 1 voltage is increased to 30, 50, 70, 90 and 110 V, respectively. (A-E): 22non-G4 in 100 mM TMAA and (F-J): 22non-G4 in 100 mM TMA + 1 mM KCl. The predominant stoichiometry is the oligonucleotide without any adducts.



Figure S3. ESI-MS of 5 μ M duplex DNA made by two d(CGC-GAA-TTC-GCG) strands = DK66 (A-D) and 5 μ M hairpin DNA d(CAA-TCG-GAT-CGA-ATT-CGA-TCC-GAT-TG) hp26 (E-H) in 100 mM NH₄OAc + 1mM NaCl (A-B and E-F) or 100 mM TMAA + 1 mM NaCl (C-D and G-H). (A, C, E and G) lon guide 1 voltage is 50 V and (B, D, F and H) 70 V. Zoom on the 5- charge state. It appears clearly that TMAA is more efficient to decrease the amount of nonspecific adducts.



Figure S4. ESI-MS of 5 μ M 23TAG; (A) and (B): electrosprayed in 100 mM NH₄OAc with ion guide 1 of 50 and 70 V, respectively; (C) and (D) in 100 mM TMAA + 1 mM KCl with ion guide 1 of 50 and 70 V, respectively. Note the 2× zoom in the 1200 – 2200 m/z region. In NH₄OAc, charge states are clearly centered on the 5- form. In TMAA + KCl at 70 V the 4- becomes the predominant charge state.



Figure S5. Collision cross section distribution of 5 μ M 23TAG in 100 mM TMAA + 1mM KCl (red) or 100 mM NH₄OAc (brown and blue). Bias voltage increase from light to dark colors (respectively 18, 25 and 30 V). Brown colors represent the collision cross section reconstructed from the sum of the G-quadruplexes with zero, one and two ammonium cations and blue ones are the collision cross section of the G-quadruplexes with two ammonium ions exclusively. Red colors are the G-quadruplexes with two ammonium ions exclusively. Red colors are the G-quadruplexes with two potassium ions.



Figure S6. Mass spectrometry titration of 20 μ M TG₄T into 100 mM TMAA doped by KCl. Full scale ESI-MS are showed. M and Q represent the monomer and quadruplex species, respectively. The solutions were prepared with 100 μ M in TG₄T single strand for 24 hours before being diluted to 20 μ M.

SUPPORTING INFORMATION ARTICLE 2

Folding and Misfolding Pathways of G-quadruplex DNA

A. Marchand & V. Gabelica

Nucleic Acids Research, 2016, doi: 10.1093/nar/gkw970.

Folding and Misfolding Pathways of G-quadruplex DNA

Adrien Marchand¹, Valérie Gabelica^{1,*}

¹Univ. Bordeaux, INSERM, CNRS, Laboratoire ARNA, IECB, 2 rue Robert Escarpit, 33607 Pessac, France.

Supporting Information Table of Contents

Materials and methods

Supplementary materials and methods	S2-6
Figure S 1. Nonspecific adducts subtraction	S4
Figure S 2-7. Mass spectrometry titrations of the nonG4 forming sequences	S7-12
Figure S 8. Non-specific adducts on nonG4 forming sequence over time	S12
Figure S 9-10. Evaluation of the need of response factors correction	S13-14
K⁺ binding equilibria	
Figure S 11-20. Mass spectrometry titrations of G4 forming sequences	S15-24
Figure S 21-22. Binding constant determination (Fitting)	S25-26
Figure S 23. Circular dichroism titrations of G4 forming sequences	S27
Figure S 24. Generated Circular dichroism spectra for the G4 forming sequences	S28
Kinetics of K ⁺ binding	
Figure S 25-33. Evolution over time of the 0, 1 and 2 -K ⁺ species	S29-31
Figure S 34. Mass spectra of 24TTG, 23TAG and Pu24 as a function of time	S32
Scheme S 1. List of kinetic mechanisms compared	S33
Table S 1. Ranking of the kinetic mechanisms for 24TTG	S34
Figure S 35-42. Kinetics mechanism evaluation for 24TTG (Fitting)	S35-40
Table S 2. Ranking of the kinetic mechanisms for 23TAG	S41
Figure S 43-47. Kinetics mechanism evaluation for 23TAG (Fitting)	S42-45
Table S 3. Ranking of the kinetic mechanisms for Pu24	S46
Figure S 48. Fitting of the best mechanism for Pu24	S47
Figure S 49. Full scale MS and IMS spectra of 24TTG	S48
Figure S 50-52. Arrival time distribution as a function of the time for 24TTG and 23TAG	S49-51
Figure S 53-58. Kinetics of K^{+} binding monitored by CD for the G4 forming sequences	S52-56
Supplementary references	S57

SUPPLEMENTARY MATERIALS AND METHODS

Electrospray mass spectrometry tuning.

Equilibrium (ESI-MS) titration experiments were performed in negative ion mode on a LCT Premier mass spectrometer (Waters, Manchester, UK). The ESI source voltage was 2.2 kV and the sample cone voltage was 200 V. The source temperature was 60 °C. The source pressure was increased to 45 mbar and measured using a Center Two probe (Oerlikon Leybold Vacuum, Cologne, Germany). The syringe injection flow rate was 200 μ L/h.

The titration experiment on the 21G sequence was also repeated on a Bruker Amazon SL mass spectrometer (Bruker, Bremen, Germany) with a spray voltage of 3 kV and an end plate offset of 500 V. The desolvation gas temperature was 180 °C. The Smart Parameters Settings (SPS) of the trap were set to 1573 m/z and the trap was filled for 20 ms. The syringe injection flow rate was 180 μ L/h. Mass spectra were accumulated for approximately 3 min.

The kinetics experiments were carried out on an Agilent 6560 IMS-Q-TOF (Agilent Technologies, Santa Clara, CA) with a dual ESI source (spray voltage 3.5 kV) in negative ion mode. This instrument allows one to monitor the drift time distributions of each m/z range, and this further helps discerning the existence of several ensembles in species of a given stoichiometry. The injection syringe flow rate was 180 µL/h. The drying gas temperature was 200 °C. The ion mobility cell was filled with helium (Alphagaz 1 grade from Air Liquide, H_2O <3ppm, C_nH_m <0.5ppm, O_2 <2ppm) and the pumping system was modified using a second Triscroll 40 pump (Agilent Technologies) (1). A pressure differential of 0.22 Torr was used in order to ensure only helium was present in the drift tube. Ion mobility spectrometry data are recorded simultaneously with the mass spectrometry data (2).

Data integration and non-specific adducts subtraction. A potassium adduct distribution on each DNA peak is observed by MS when KCI-containing solutions are injected. Because the detected potassium adduct distribution is not the same for each charge state, necessarily part of the binding detected is non-specific and happens along the ESI process. It is also possible that a fraction of these adducts come from K⁺ binding to external phosphate groups (3), and these are also not G-quadruplex specific. The goal of signal subtraction of non-specific adduct contribution is to deduce the fraction of K⁺ coming from specific binding in-between G-quartets, i.e. K⁺ specific of G4 structures.

In order to distinguish potassium ions specifically bound to the G-quadruplex structure, we applied a method described by Klassen (4), which we already applied to G-quadruplexes (5, 6). Supporting information Figure S1 shows graphically and mathematically how the adducts are accounted for. In each set of KCl concentration, oligonucleotide length and charge state, the non-specific adduct distribution is given by the distribution recorded with the same conditions for a non-G-quadruplex forming sequence containing the same number of bases and of guanines (abbreviated here 26nonG4, 24nonG4, 23nonG4, 22nonG4 and 21nonG4, see Supplementary data). The fraction of K⁺ nonspecifically bound to the G-quadruplex is low (<1%) and does not influence significantly the K⁺ concentration available for binding. The titrations done for these non-G-quadruplex forming sequences and the normalized integrals for each charge state as a function of the KCl concentration are given in Supporting Information Figures S2-S7. Using the equations described in the supplementary data, the specific K⁺ distribution for each sequence in each condition is then recalculated. For the titration experiments, the non-specific adducts were taken into account on the 3-min summed mass spectra. For the kinetics experiments, the treatment is done on each single spectrum (one per 1.1 s), by assuming that the non-specific adducts binding is instantaneous (a time

dependent experiment showing that this is the case for 21nonG4 is shown in Supporting Information Figure S8).

Before converting the relative intensities into relative concentrations, we also considered the possibility that all complexes may not have the same ESI-MS response. To verify that the response factors are not influenced by the KCl concentration and/or by the presence or not of potassium in the complexes, we used an internal standard (dT_6) as reference (7). Figure S9 shows the relative intensity of the detected complexes and dT₆. The ratios do not deviate systematically upon titration, and this indicates that the response factors are all equal, and that the detected signals are directly proportional to the concentrations. To verify that the total signal and the charge state distributions are not affected by the KCI concentration we plotted the ratio between the intensity of each charge state divided by the total DNA counts (Figure S10). The signal for each charge state remains constant, meaning that each charge state can be treated independently. For 26TTA, the change in the distribution is probably due to the larger number of non-specific adducts and this translates into larger error bars. The smallest error bars are obtained when the equilibrium is reached at low KCI concentrations because at those concentrations the non-specific adducts contribution is minor. Overall the cleaning of non-specific adducts gives reproducible recalculated abundances from different charge states. Consequently, each charge state can be treated separately, then the results of recalculated concentrations of each specific K⁺ stoichiometry obtained for each charge state can be combined and the error bars calculated as the standard deviation of the average value. The same procedure was applied to treat the time-dependent experiments, except that only the results obtained on the 5- charge state are reported here (the treatment was also done on the 4- charge state with very similar results, but larger error due to more numerous nonspecific adducts to subtract).



Figure S 1. Schematic representation of the mathematical treatment done to obtain the specific K^+ distributions (illustrated here with 2 adducts).

The values x and y depend on the potassium concentration and the strand length. We performed the experiments on control sequence for each strand length (21, 22, 23, 24 and 26 bases) at each potassium concentration. Each charge state is treated separately.

Equations for the 5- charge state (2 nonspecific adducts)

$$\begin{aligned} 0K &= i0K * (1 + x + y) \\ 1K &= (i1K - i0K * x) * (1 + x + y) \\ 2K &= (i2K - i0K * y - (i1K - i0K * x) * x) * (1 + x + y) \\ 3K &= (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * (1 + x + y) \\ 4K &= (i4K - (i2K - i0K * y - (i1K - i0K * x) * x) * y \\ &- (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * (1 + x + y) \\ 5K &= (i5K - (i3K - (i1K - i0K * x) * y - (i2K - i0K * y - (i1K - i0K * x) * x) * x) * y \\ &- (i4K - (i2K - i0K * y) + (i1K - i0K * x) * y) + (i1K - i0K * x) * x) * (1 + x + y) \end{aligned}$$

* (1 + x + y)Where 0K, 1K, 2K, 3K, 4K and 5K reflect the specific K⁺ distribution. x and y are the values of the

Where 0K, 1K, 2K, 3K, 4K and 5K reflect the specific K⁺ distribution. x and y are the values of the normalized integrals of the 1 and 2 non-specific K⁺. i0K, i1K, i2K, i3K, i4K and i5K are the initial integrals for the 0, 1, 2, 3, 4 and 5 K⁺ stoichiometries.

Data processing details for nonspecific adducts subtraction:

We took into account the following stoichiometries observed for each charge state: from 0 to 3 K for the 6- charge state, from 0 to 4 K for the 5- charge state and from 0 to 5 K for the 4- charge state for all KCl concentrations. After smoothing using Savitzky-Golay method and background subtraction with a 99th order polynomial curve directly done using the MassLynx software, the peaks were integrated from -1 to +1 m/z for the 6- charge state, from -1.2 to +1.2 for the 5- charge state and from -1.4 to +1.4 for the 4- charge state, in order to integrate the full isotope distribution of each peak. The noise was subtracted using the area under an m/z range for the "-1 K" stoichiometry for each spectrum.

Non-G4 forming sequences used as references for the nonspecific adducts subtration

26nonG4	$T_2G_3ATGCGACAGAGAG_2ACG_3AT$
24nonG4	TG3ATGCGACAGAGAG2ACG3A
23nonG4	TG ₃ ATGCGACAGAGAG ₂ ACG ₃
22nonG4	G3ATGCGACAGAGAG2ACG3
21nonG4	G ₃ ATGCGAC GAGAG ₂ ACG ₃

Circular dichroism spectrophotometer operation.

CD experiments were carried out on solutions prepared exactly as for ESI-MS. Experiments were run on a Jasco J-815 spectrophotometer at 20 °C (unless otherwise mentioned) with a 2 mm path length quartz cell (Helma). For the titrations, the scanning speed was 50 nm/min, and the bandwith was 2 nm in the range 210-350 nm. CD spectra shown are the sum of 3 accumulations. Baselines were determined in the same conditions and same cuvette, without DNA. For the kinetic experiments, 100 nm/min scan speed was chosen; these parameters allowed the acquisition of 1 scan per 0.8 min. Spectra shown herein were smoothed using the LOWESS method and a linear baseline was subtracted using the mean value for each scan between 320 and 350 nm.

The CD response deconvolution, either reconstructed from the titrations or from the kinetics experiments, is obtained using the ESI-MS equilibrium constants and rate constants, respectively. From these parameters, the proportion of each species involved in the reaction mechanism can be predicted as a function of KCI concentration or of time, respectively. From the known concentrations of each species and measured total CD intensity, the response of each species can therefore be calculated. For each wavelength, we fitted the data, adjusting the molar ellipticity coefficient of each species according to the following equation:

$$\begin{pmatrix} I_0^{WL1} \\ I_1^{WL1} \\ I_2^{WL1} \\ \dots \\ I_n^{WL1} \end{pmatrix} = \Delta \varepsilon_x \times \begin{pmatrix} C_0^x \\ C_1^x \\ C_2^x \\ \dots \\ C_n^x \end{pmatrix} + \Delta \varepsilon_y \times \begin{pmatrix} C_0^y \\ C_1^y \\ C_2^y \\ \dots \\ C_n^y \end{pmatrix} + \cdots$$
(Eq. S1)

where I_n^WL is the CD intensity as a function of KCl concentration or time at a given wavelength with n the total number of different conditions . $\Delta \varepsilon_x$, $\Delta \varepsilon_y$,... are the responses of the species x, y,..., and C_x , C_y ,... are vector matrices of the concentrations of the species x, y,... as a function of KCl concentration or time. Calculating the responses $\Delta \varepsilon_x$, $\Delta \varepsilon_y$,... as a function of the wavelength (here, at each 5 nm) gives the deconvoluted CD spectra, which is reconstructed by linking the dots with a spline curve. The main difference between our deconvolution method and singular value decomposition analysis (SVD) (8) is that our approach determines first the number of species and their relative concentrations thanks to the separate mass spectrometry signals obtained for each stoichiometry, and then the response of each species by Eq. 5, whereas SVD determines the number of species and their response altogether.


Figure S 2. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **21nonG4** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 3. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **22nonG4** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 4. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **23nonG4** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 5. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **24nonG4** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 6. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **26nonG4** in 100 mM TMAA with increasing amount of KCl. Full MS spectra and zooms on the charge states.



Figure S 7. Quantification of adducts on non-G-quadruplex forming sequences. In green, blue, orange and yellow the 0, 1, 2 and $3-K^+$ stoichiometry, respectively. Data were normalized according to the $0-K^+$ stoichiometry.



Figure S 8. Non-specific adducts on non-G-quadruplex forming sequence **21nonG4** monitored over time by MS (1 mM KCl added at t0), normalized according to the O-K⁺ stoichiometry, (A) for the 6-(B) for the 5- and (C) for the 4- charge state. Non-specific adducts binding is instantaneous. Note that the 6- charge state is noisy: the total signal for this charge state is very low. However, the mean, represented by the solid line matches very well with the value obtained from the fit (Figure S 7).



Figure S 9. Evolution of the ratio between the total signal of the G4 forming sequence and the internal standard dT_6 as a function of the KCl concentration in percentage of deviation around the mean for all the sequences. In black the experimental points and in red the mean. For all the sequences KCl concentration does not influence the response factors and the standard deviations (grey lines) around the means are of about 5% (10% for Pu24).



Figure S 10. Evolution of the charge state distribution in function of the KCl concentration. (A-E + G-J), the human telomeric sequences and (F) the c-myc Pu24 sequence. In orange the 4- charge state, in blue the 5- and in green the 6-. A darker line corresponding to the mean of the points on the studied KCl range is also displayed. For all the sequences except 26TTA there is no KCl concentration dependence on the charge state distribution.

SUPPLEMENTARY RESULTS, Part I: K⁺ binding equilibria



Figure S 11. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **24TTG** in 100 mM TMAA with increasing amount of KCl. Full MS spectra and zooms on the charge states.



Figure S 12. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **23TAG** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 13. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **22GT** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 14. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **26TTA** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 15. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **22AG** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 16. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **21G** in 100 mM TMAA with increasing amount of KCl. Full MS spectra and zooms on the charge states.



Figure S 17. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **F21T** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 18. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **F26T** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states. NB: for this sequence, the observed charge states are different from the 26nonG4 reference. The 7- charge state observed here wasn't taken into consideration for the determination of the K_D's.



Figure S 19. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **Pu24** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 20. Mass spectrometry titrations of 2.5 μ M dT6, 10 μ M **22CTA** in 100 mM TMAA with increasing amount of KCI. Full MS spectra and zooms on the charge states.



Figure S 21. Concentration of the complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, in function of KCl concentration from 0 to 1 mM, (A) for 24TTG, (B) for Pu24, (C) for 22GT, (D) for 22AG, (E) for 23TAG, (F) for 26TTA, (G) for 21G, (H) for 22CTA, (I) for F21T and (J) for F26T. The error bars reflect the errors on the treatment of the three charge states. The solid lines represent the fitting of the data using an iterative fitting procedure on the chemical equations using DynaFit.



Figure S 22. Concentration of the complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, in function of KCl concentration from 0 to 1 mM for **21G**, (A) on the **Waters LCT Premier mass spectrometer** and (B) on the **Bruker Amazon mass spectrometer**.



Figure S 23. Circular dichroism titrations of 10 μ M (A) 24TTG, (B) Pu24, (C) 22GT, (D) 22AG, (E) 23TAG, (F) for 26TTA, (G) for 21G, (H) for 22CTA, (I) for F21T and (J) for F26T. Underlined in colors: 0 μ M KCl in green, 100 μ M KCl in blue and 1 mM KCl in orange.



Figure S 24. Generated Circular dichroism spectra of 10 μ M (A) 24TTG, (B) Pu24, (C) 22GT, (D) 22AG, (E) 23TAG, (F) for 26TTA, (G) for 21G, (H) for 22CTA, (I) and for F21T in 100 mM obtained using the KD's obtained by MS for the 0, 1 and 2-K⁺ forms, respectively, in green, blue and orange. The abundancies of the 1-K⁺ complex for 24TTG and Pu24 and of the 2-K⁺ complex of 22CTA were low (Figure S 21). Therefore the response of those species were fixed to 0 in order to reconstruct properly the other species. Similarly, the F26T CD spectra were not deconvoluted because it is not significantly folded.

SUPPLEMENTARY RESULTS, Part II: Kinetics of K⁺ binding



Figure S 25. Concentration of the **22GT** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 200 μ M, (B) 500 μ M and (C) 1 mM KCl.



Figure S 26. Concentration of the **22AG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 200 μ M, (B) 500 μ M and (C) 1 mM KCl.



Figure S 27. Concentration of the **21G** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 500 μ M and (B) 1 mM KCl.



Figure S 28. Concentration of the **F21T** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 500 μ M and (B) 1 mM KCl.



Figure S 29. Concentration of the **22CTA** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 500 μ M and (B) 1 mM KCl.



Figure S 30. Concentration of the **26TTA** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 200 μ M, (B) 500 μ M and (C) 1 mM KCl.



Figure S 31. Concentration of the **24TTG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100 μ M, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl.



Figure S 32. Concentration of the **23TAG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100 μ M, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl.



Figure S 33. Concentration of the **24TTG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of 500 μ M. The data obtained on **two different MS instruments** are identical.



Figure S 34. Mass spectra of 10 μ M (A) 24TTG, (B) 23TAG and (C) Pu24 in 100 mM TMAA as a function of time after addition of 500 μ M KCl.



Scheme S 1. List of kinetic mechanisms compared, classified in number of species.

model	n _P	AIC	w ^(AIC)
0x((1-2) (1x(2 2)))	10	-37648	0.896
0x(1 2 (1x(2 2)))	10	-37644	0.104
0x(1 (1x(2 2)))	8	-37311	0
0x(1 (1x(2 2 2)))	10	-37306	0
0x(1 2- (1x(2 2)))	12	-37298	0
0x((1-2) (1-2))	8	-36905	0
0x(1 2 (1-2))	8	-36496	0
0x(1 2 (0x(1 2)))	10	-36352	0
0x(1 2 2 (1-2))	10	-33345	0
0x((2-1-2 (2-1-2))	12	-33103	0
0x(((1x(2 2)) 2))	8	-32534	0
0x(1 1 (0x(2 2)))	10	-32370	0
0x(1 2 2)	6	-32193	0
0x(1 1 2 2)	8	-32145	0
0-2-1-2	6	-31750	0
0x(2 (1-2))	6	-31750	0
0x((1-2) 2 2)	8	-31746	0
0-0x(1 2)	6	-31482	0
0x(1 2 (1-2) (1-2))	12	-30926	0
0x(1 2)	4	-28845	0
0x(1 1 2)	6	-28843	0
0-1-2	4	-26307	0

Table S1. Ranking of the reaction mechanisms for **24TTG**.



<u>0-1-2</u>

Figure S 35. Concentration of the **24TTG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0-1-2 model**.

<u>0x(1|2|2)</u>



Figure S 36. Concentration of the **24TTG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x(1/2/2) model**.

<u>0x((2-1-2|(2-1-2))</u>



 $\begin{array}{c} {\sf Pal} \\ {\sf k}_{2a} \\ {\sf k}_{2b} \\ {\sf k}_{1a} \\ {\sf k}_{2c} \\ {\sf k}_{2c} \\ {\sf k}_{2c} \\ {\sf k}_{2c} \\ {\sf k}_{2d} \\ {$

r#Set	Initial	Final	Std. Error	CV (%)	Corr mat	k_{2a}^{on}	k_{2a}^{off}	k_{2b}^{on}	k_{2b}^{off}	k_{1a}^{on}	k_{1a}^{off}	k_{1b}^{on}	k_{1b}^{off}	k_on	k_{2c}^{off}	k_{2d}^{on}	k_{2d}^{off}
on	0.001	0.0003088 (µM ⁻² min ⁻¹)	0.00054829	177.58	k_2a on	*	100	-11	-11	-89	-90	-16	-5	31	27	-4	-2
off	1000	11.6262 (min ⁻¹)	19.9071	171.23	k _{2a} off	100	*	-11	-11	-89	-89	-16	-5	31	27	-4	-2
on	100	4.19466 (μM ⁻² min ⁻¹)	74.5785	1777.94	k _{2b} ^{on}	-11	-11	*	100	18	18	-28	-56	-46	-19	-40	-6
off	1	950578 (min ⁻¹)	1.63E+07	1713.54	k _{2b} ^{off}	-11	-11	100	*	17	18	-28	-56	-46	-19	-40	-6
on	1000	11.4225 (min ⁻¹)	24.0595	210.63	k _{1a} on	-89	-89	18	17	*	100	5	-5	-53	-52	4	7
off	1	0.0806989 (µM ⁻¹ min ⁻¹)	0.178369	221.03	k _{1a} off	-90	-89	18	18	100	*	5	-5	-53	-52	4	7
on	1.00E-05	3.48E-09 (min ⁻¹)	3.60E-10	10.33	k _{1b} on	-16	-16	-28	-28	5	5	*	72	26	32	-38	-2
off	0.01	10158 (µM ⁻¹ min ⁻¹)	0.0001648	0	k _{1b} ^{off}	-5	-5	-56	-56	-5	-5	72	*	30	13	-21	9
on	1.00E-06	0.0011909 (µM ⁻¹ min ⁻¹)	2.06E-05	1.73	k _{2c} ^{on}	31	31	-46	-46	-53	-53	26	30	*	86	27	-19
off	0.01	0.0204068 (min ⁻¹)	0.00060823	2.98	k _{2c} off	27	27	-19	-19	-52	-52	32	13	86	*	5	-30
on	0.01	88.7599 (μM ⁻¹ min ⁻¹)	8.31267	9.37	k _{2d} on	-4	-4	-40	-40	4	4	-38	-21	27	5	*	0
off	1000	0.0009807 (min ⁻¹)	1747.93	1.7E+08	k _{2d} ^{off}	-2	-2	-6	-6	7	7	-2	9	-19	-30	0	*





Figure S 37. Concentration of the **24TTG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using **0x((2-1-2))(2-1-2)) model**.

<u>0x((1-2)|(1-2))</u>

Sequence 24TTG



Figure S 38. Top panel, concentration of the **24TTG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x((1-2))(1-2)) model**. Middle panel, simulation of the abundances of the species at (E) 100, (F) 200 μ M, (G) 500 μ M and (H) 1 mM KCl. Bottom panel, reconstructed CD spectra of the different species (I) with the 1a species response fixed to 0.



Figure S 39. Top panel, concentration of the 24TTG complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x((1-2))(1x(2))) model**. Middle panel, simulation of the abundances of the species at (E) 100, (F) 200 μM, (G) 500 μM and (H) 1 mM KCl. Bottom panel, reconstructed CD spectra of the different species (I) with the 1a species response fixed to 0.

<u>0x(1|2|(1x(2|2)))</u>

Sequence 24TTG



Figure S 40. Top panel, concentration of the **24TTG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x(1/2/(1x(2/2))) mode**l. Middle panel, simulation of the abundances of the species at (E) 100, (F) 200 μ M, (G) 500 μ M and (H) 1 mM KCl. Bottom panel, reconstructed CD spectra of the different species (I) with the 1a species response fixed to 0.

model	nP	AIC	w(AIC)				
0x((1-2) (1-2))	8	-41761	0.658				
0x(1 2 2 (1-2))	10	-41759	0.273				
0x((1-2) (1x(2 2)))	10	-41756	0.067				
0x(1 2 (1-2))	8	-41748	0.001				
0x(1 2 (1-2) (1-2))	12	-41738	0				
0x(1 2 (1x(2 2)))	10	-41722	0				
0x(1 2 (0x(1 2)))	10	-41310	0				
0x((2-1-2 (2-1-2))	12	-39150	0				
0x(1 2- (1x(2 2)))	12	-38888	0				
0x(1 (1x(2 2)))	8	-38640	0				
0x(1 (1x(2 2 2)))	10	-38622	0				
0x(1 2 2)	6	-37603	0				
0x(1 1 2 2)	8	-37572	0				
0x(((1x(2 2)) 2))	8	-37496	0				
0x(1 1 (0x(2 2)))	10	-37466	0				
0x(2 (1-2))	6	-37463	0				
0-2-1-2	6	-37463	0				
0x((1-2) 2 2)	8	-37428	0				
0x(1 2)	4	-34958	0				
0-0x(1 2)	6	-34954	0				
0x(1 1 2)	6	-34952	0				
0-1-2	4	-33495	0				

Table S 2. Ranking of the reaction mechanisms for **23TAG**.



Figure S 41. Top panel, concentration of the **23TAG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x((1-2))(1-2)) model**. Middle panel, simulation of the abundances of the species at (E) 100, (F) 200 μ M, (G) 500 μ M and (H) 1 mM KCl. Bottom panel, reconstructed CD spectra of the different species (I) with the 1a species response fixed to 0.


Figure S 42. Top panel, concentration of the **23TAG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x(1/2/2/(1-2)) model**. Middle panel, simulation of the abundances of the species at (E) 100, (F) 200 μ M, (G) 500 μ M and (H) 1 mM KCl.



Figure S 43. Top panel, concentration of the **23TAG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x((1-2)/(1x(2/2))) model**. Middle panel, simulation of the abundances of the species at (E) 100, (F) 200 μ M, (G) 500 μ M and (H) 1 mM KCl.



Figure S 44. Top panel, concentration of the **23TAG** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x(1/2/(1-2)) model**. Middle panel, simulation of the abundances of the species at (E) 100, (F) 200 μ M, (G) 500 μ M and (H) 1 mM KCl.

model AIC w(AIC) nP 0x((1-2)|(1x(2|2))) -37508 10 1 0x(1|2|(1x(2|2))) 10 -37385 0 0x(1|2|(1-2)|(1-2)) 12 -37163 0 0x((1-2)|(1-2)) -37128 0 8 0x(1|(1x(2|2))) 8 -37065 0 0x(1|(1x(2|2|2))) -37056 10 0 0x(1|2-|(1x(2|2))) -37054 0 12 0x(1|2|(1-2)) -36948 8 0 0x(1|2|(0x(1|2))) 10 -36869 0 0x(((1x(2|2))|2)) -29147 0 8 0-0x(1|2) -27552 0 6 0-1-2 -24383 0 4 0x(1|1|(0x(2|2))) -16554 0 10 0x((2-1-2|(2-1-2)) 12 -15608 0 0x((1-2)|2|2) -15569 8 0 0x(1|2|2) 6 -15554 0 0-2-1-2 6 -15243 0 0x(2|(1-2)) -12039 0 6 0x(1|2) 4 -11756 0 0x(1|1|2) 6 -11753 0 0x(1|1|2|2) -11748 0 8 0x(1|2|2|(1-2)) 42258.9 10 0

Table S 3. Ranking of the reaction mechanisms for **Pu24**.

Sequence Pu24

0x((1-2)|(1x(2|2)))



Figure S 45. Top panel, concentration of the **Pu24** complexes with 0, 1 or 2 potassium in green, blue and orange, respectively, as a function of the time after adding KCl to reach a final concentration of (A) 100, (B) 200 μ M, (C) 500 μ M and (D) 1 mM KCl. The fit displayed on the graphs was obtained using the **0x((1-2))(1x(2/2))) model**. Middle panel, simulation of the abundances of the species at (E) 100, (F) 200 μ M, (G) 500 μ M and (H) 1 mM KCl. Bottom panel, reconstructed CD spectra of the different species (I) with the 1a species response fixed to 0.



Figure S 46. Full scale MS and IMS spectrum of 2.5 μ M dT6 and 10 μ M 24TTG in 100 mM TMAA + 1 mM KCl, one hour after KCl addition.



Figure S 47. Arrival time distribution as a function of the time (from green (3min) to red (24h) through grey) for **24TTG** for the **4- charge state** for the 2-K stoichiometry (A, C, E and G) and for the 1-K stoichiometry (B, D, F, H). (A) and (B) at 100 μ M, (C) and (D) at 200 μ M, (E) and (F) at 500 μ M and (G) and (H) at 1 mM KCl.



Figure S 48. Arrival time distribution as a function of the time (from green (3min) to red (24h) through grey) for **24TTG** for the **5- charge state** for the 2-K stoichiometry (A, C, E and G) and for the 1-K stoichiometry (B, D, F, H). (A) and (B) at 100 μ M, (C) and (D) at 200 μ M, (E) and (F) at 500 μ M and (G) and (H) at 1 mM KCl.



Figure S 49. Arrival time distribution as a function of the time (from green (3min) to red (1h) through grey) for **23TAG** for the **4- charge** state for the 2-K stoichiometry (A, C, E and G) and for the 1-K stoichiometry (B, D, F, H). (A) and (B) at 100 μ M, (C) and (D) at 200 μ M, (E) and (F) at 500 μ M and (G) and (H) at 1 mM KCl.



Figure S 50. Kinetics at 20 °C monitored by CD of 10 μ M 22GT after adding KCl to reach a final concentration of (A) 200 μ M, (B) 500 μ M and (C) 1 mM. Inserts, the CD at the indicated wavelength as a function of the time.



Figure S 51. Kinetics at 20 °C monitored by CD of 10 μ M 22AG after adding KCl to reach a final concentration of (A) 200 μ M, (B) 500 μ M and (C) 1 mM. Inserts, the CD at the indicated wavelength as a function of the time.



Figure S 52. Kinetics at 20 °C monitored by CD of 10 μ M 21G after adding KCl to reach a final concentration of (A) 200 μ M, (B) 500 μ M and (C) 1 mM. Inserts, the CD at the indicated wavelength as a function of the time.



Figure S 53. Kinetics at 20 °C monitored by CD of 10 μ M 26TTA after adding KCl to reach a final concentration of (A) 200 μ M, (B) 500 μ M and (C) 1 mM. Inserts, the CD at the indicated wavelength as a function of the time.



Figure S 54. Kinetics at 20 °C monitored by CD of 10 μ M 24TTG after adding KCl to reach a final concentration of (A) 100 μ M, (B) 200 μ M, (C) 500 μ M and (D) 1 mM. Inserts, the CD at the indicated wavelength as a function of the time.



Figure S 55. Kinetics at 37 °C monitored by CD of 10 μ M 24TTG after adding KCl to reach a final concentration of (A) 1 mM and (B) 25 mM. The first recorded spectrum after the addition of KCl is highlighted in black. Inserts, the CD at the indicated wavelength as a function of the time.



Figure S 56. Kinetics at 20 °C monitored by CD of 10 μ M 23TAG after adding KCl to reach a final concentration of (A) 100 μ M, (B) 200 μ M, (C) 500 μ M and (D) 1 mM. Inserts, the CD at the indicated wavelength as a function of the time.



Figure S 57. Kinetics at 20 °C monitored by CD of 10 μ M Pu24 after adding KCl to reach a final concentration of (A) 200 μ M, (B) 500 μ M and (C) 1 mM. Inserts, the CD at the indicated wavelength as a function of the time.



Figure S 58. Kinetics at 37 °C monitored by CD of 10 μ M Pu24 after adding KCl to reach a final concentration of (A) 1 mM and (B) 25 mM. The first recorded spectrum after the addition of KCl is highlighted in black. The right panels are zooms on the 280-320 nm range. Inserts, the CD at the indicated wavelength as a function of the time. Additionally, a kinetic monitored at a single wavelength permitted to monitor the appearance and disappearance of the peak at 298 nm.

Additional references

- 1. D'Atri,V., Porrini,M., Rosu,F. and Gabelica,V. (2015) Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. J. Mass Spectrom., **50**, 711–726.
- May,J.C., Goodwin,C.R., Lareau,N.M., Leaptrot,K.L., Morris,C.B., Kurulugama,R.T., Mordehai,A., Klein,C., Barry,W., Darland,E., *et al.* (2014) Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Anal. Chem.*, **86**, 2107–2116.
- Wong,A. and Wu,G. (2003) 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., 125, 13895–13905.
- Sun, J., Kitova, E.N., Wang, W. and Klassen, J.S. (2006) Method for distinguishing specific from nonspecific protein-ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal. Chem.*, **78**, 3010–8.
- Marchand, A., Granzhan, A., Iida, K., Tsushima, Y., Ma, Y., Nagasawa, K., Teulade-Fichou, M.-P. and Gabelica, V. (2015) Ligand-induced conformational changes with cation ejection upon binding to human telomeric DNA G-quadruplexes. J. Am. Chem. Soc., 137, 750–756.
- Largy, E., Marchand, A., Amrane, S., Gabelica, V. and Mergny, J.-L. (2016) Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. J. Am. Chem. Soc., 138, 2780–2792.
- 7. Gabelica, V., Rosu, F. and De Pauw, E. (2009) A simple method to determine electrospray response factors of noncovalent complexes. *Anal. Chem.*, **81**, 6708–15.
- 8. Gray,R.D. and Chaires,J.B. (2011) Analysis of Multidimensional G-Quadruplex Melting Curves. In *Current Protocols in Nucleic Acid Chemistry*. John Wiley & Sons, Inc., Hoboken, NJ, USA, p. 17.4.1-17.4.16.

SUPPORTING INFORMATION ARTICLE 3

Selective and Cooperative Ligand Binding to Antiparallel Human Telomeric DNA G-quadruplexes

A. Marchand, D. Strzelecka & V. Gabelica

Chemistry – A European Journal **2016**, 22, 9551-9555

CHEMISTRY A European Journal

Supporting Information

Selective and Cooperative Ligand Binding to Antiparallel Human Telomeric DNA G-Quadruplexes

Adrien Marchand,^[a, b, c] Dominika Strzelecka,^[a, b, d] and Valerie Gabelica^{*[a, b, c]}

chem_201601937_sm_miscellaneous_information.pdf

Supplementary Data Table of Contents

		Page
Section 1. Experimenta	l section	
Table S1 List of	the sequences used in this study and their structure in K ⁺ solutions.	S2
Mater	ial and methods	
Section 2. Metal compl	exes synthesis	
Synthe	esis details	S4
Figure S1-S10 Charac	cterization of the ligands: HRMS spectra/X-ray structures	S5-S10
Table S2 Crysta	k data and structure refinement for Cu-ttpy and Cu-Clptp	S11
Section 3. Binding affin	ities and cooperativity: supplementary ESI-MS data	
Figure S11 Verific	ation that the responses factor are equals	S12
Figure S12 ESI-MS	S titration of 10 μ M 22GT, Pu24 or $[G_4T_4G_4]_2$ by Cu-ttpy in 100 mM	S13
TMAA	+ 1 mM KCl.	
Figure S13 ESI-MS	S titration of 10 μM 22GT, Pu24 or $[G_4 T_4 G_4]_2$ by $\mbox{Cu-Clptp}$ in 100 mM	S14
TMAA	+ 1 mM KCl.	
Figure S14 ESI-MS	S titration of 10 μM 22GT by $\textbf{Cu-tpy}$ in 100 mM TMAA + 1 mM KCl	S15
Table S3 $K_{\rm D}$ of li	igand binding to G-quadruplexes	S16
Figure S15 MS s	pectra of 23TAG with 1 eq Cu-ttpy, 360A, Phen-DC3, PDS	S17
(pyrido	ostatin) and TMPyP4	
Figure S16 MS sp	ectra of 10 μM A) ckit1, B) HIV-PRO1 and C) TBA in 100 mM TMAA	S18
+ 1 mN	M KCl with 10 μM Cu-ttpy .	
Figure S17 MS sp	ectra of 10 μ M DNA in 100 mM NH ₄ OAc or 100 mM TMAA + 1 mM	S19
KCl wit	th 10 μM Cu-Clptp .	
Figure S18 Compa	arison of the cooperativity of binding of different metal complexes	S20
on 220		
Figure S19 MS spo	ectra of 10 μ M (A) 22CTA, (B) 23AG and (C) 22AG in 100 mM TMAA	\$21
+ 1 mi	VI KCI WITH 10 µM CU-TTPY	622
Figure S20 CD of I	Pu24 and [G414G4] ₂ in presence of increasing amount of Cu-ttpy	522
Figure S21 Compe	etition experiment: MS spectra of 10 μ M 22GT in 100 mM TMAA +	\$23
ligariu.	Support of $AO = MA (G, T, G)$ in 100 mM NH OAC with $FO = MA (A)$	\$24
Figure 322 IVIS/IVI	$3 \text{ spectra of 40 } \mu\text{W} (341434)_2 \text{ in 100 mW NH4OAC with 60 } \mu\text{W} (A)$	324
Figure S23 MS/M	S spectra of 40 μ M (G ₄ T ₂ G ₄) ₂ in 100 mM NH ₄ OAc with 60 μ M (A)	\$25
Cu-ttn		525
Section 4. Supplemento	ary references	S26

Section 1. Experimental section

Table S1. List of the sequences used in this study and their structure in K^+ solutions. The structure published and the loops arrangement (from 5' to 3') are also referred *Lat*, *Diag* and *Prop* stand respectively for *Lateral*, *Diagonal* and *Propeller* loop type.

	Name	Sequence	Structure	Loops type
	22GT	(GGGTTA)₃GGGT	Antiparallel 2-quartet ⁵	Lat-Diag-Lat
	23TAG	TA(GGGTTA)₃GGG	Hybrid ⁶	Prop-Lat-Lat
	25TAG	TA(GGGTTA)₃GGGTT	Hybrid ⁶	Prop-Lat-Lat
e.	24TTG	TT(GGGTTA)₃GGGA	Hybrid ⁷	Prop-Lat-Lat
hТ	26TTA	TTA(GGGTTA)₃GGGTT	Hybrid ⁸	Lat-Lat-Prop
	23AG	A(GGGTTA)₃GGGT	Antiparallel 2-quartet ⁹	Lat-Diag-Lat
	22AG	A(GGGTTA)₃GGG	Hybrid / Mixture	Mixture
	22CTA	A(GGGCTA)₃GGG	Antiparallel 2-quartet ¹⁰	Lat-Lat-Lat
/c	СМТ	TTG GGA GGG TAG GGA GGG T	Parallel ¹¹	Prop-Prop-Prop
m	СМА	TAG GGA GGG TAG GGA GGG T	Parallel ¹¹	Prop-Prop-Prop
-	Pu24	TGA GGG TGG GGA GGG TGG GGA AGG	Parallel ¹²	Prop-Prop-Prop
	ckit1	AGG GAG GGC GCT GGG AGG AGG G	Parallel ¹³	Prop-Prop-Prop
	26CEB	AAG GGT GGG TGT AAG TGT GGG TGG GT	Parallel - 1 long loop ¹⁴	Prop-Prop-Prop
1's	[TG ₄ T] ₄	TGGGGT	Tetramolecular parallel ¹⁵	No loop
ir G	[G4T4G4]2	GGGG TTTT GGGG	Bimolecular antiparallel ^{16,17}	Diag/Diag
he	[G ₄ T ₃ G ₄] ₂	GGGG TTT GGGG	Bimolecular antiparallel ¹⁷	Lat/Lat
Ot	G4(T4G4)3	GGGG (TTTT GGGG)₃	Antiparallel 4-quartet ¹⁷	Lat-Diag-Lat
	HIV-PRO1	TGG CCT GGG CGG GAC TGG G	Antiparallel 2-quartet ¹⁸	Lat-Lat-Lat
	ТВА	GGT TGG TGT GGT TGG	Antiparallel 2-quartet ¹⁹	Lat-Lat-Lat
uc V	ds26	CAA TCG GAT CGA ATT CGA TCC GAT TG	Hairpin	-
bu	24nonG4	TGG GAT GCG ACA GAG AGG ACG GGA	Single stranded ²⁰	-

Oligodeoxynucleotides were purchased from Eurogentec (Seraing, Belgium) in RP-cartridge purified and lyophilized form. After solubilisation, the concentration was measured from the absorbance at 260 nm using the molar extinction coefficients from the IDT website using the Cavaluzzi-Borer correction.²¹ The buffer used to fix the ionic strength was trimethylammonium acetate (TMAA, Ultra for UPLC, Fluka). TMAA and potassium chloride (KCl, >99.999%, Sigma) were purchased from Sigma-Aldrich (Sant-Quentin Fallavier, France). Ammonium acetate (BioUltra for Molecular Biology, Fluka) was also used for comparison. Injected solutions contained 10 μ M DNA with either 100 mM TMAA + 1 mM KCl, or 100 mM NH₄OAc.

A LCT Premier (Waters, Manchester, UK) was used in negative ion mode. The ESI source voltage was set to 2200 V and the cone voltage to 200 V. Because ammonium is easily released from the complexes a lower voltage for the sample cone (70 V) was used for the experiments in ammonium

acetate. This value was adequate to keep intact the fragile dimeric $[G_4T_4G_4]_2$ G-quadruplex. The source temperature was 60 °C and the pressure was increased to 45 mbar and measured using a Center Two probe (Oerlikon Leybold Vacuum, Cologne, Germany). The flow rate for the syringe pump was 200 µL/h. Shown MS spectra are the sum of ~200 scans, smoothed and baseline-subtracted using the MassLynx software.

MS/MS experiments were performed on a Bruker Amazon SL (Bremen, Germany). The ions were trapped with a trap drive voltage of 92 V, maximizing the signal of the 6- charge state. We isolated the G-quadruplex before activation for 70 ms with 1.1 V for the free G-quadruplex and 1.3V for the G-quadruplex with one and two ligands.

The CD samples were prepared as for the MS experiments. Experiments were run at 20 °C for 3 accumulations. CD spectra were normalized to molar circular dichroic absorption ($\Delta\epsilon$) based on the G-quadruplex concentration following equation 3:

$\Delta \epsilon = \theta / (32980 \times C \times I)$

Where θ is the ellipticity in millidegrees, C is the G-quadruplex concentration (10 μ M) and I is the optical path length of the cell (0.2 cm).

Section 2. Metal complexes synthesis



Scheme 1. Representation of the synthesis of the complexes.

*For Cu-complexes:*²² The terpyridine derivative was solubilized in a vial in a minimum of dry dichloromethane. The metal salts previously solubilized in acetonitrile were added gently dropwise on the ligand solution to form two phases. The vial was put at 4 °C in the dark for one day. Crystals are formed at the interphase corresponding to the desired product. The purifications were performed washing three times with anhydrous acetonitrile, three times with dichloromethane and three times with diethyl ether. The salts were then dried under vacuum. Chemical yields were between 35 and 50%.

*For Zn-complexes:*²³ The same procedure as for the Cu-complexes was used except that the reaction was performed stirring under Ar for 24 hours and purified using diethyl ether. It was then dried under vacuum overnight.

<u>For Pt-complexes</u>:²³ The suspension of $Pt(COD)Cl_2$ in MeOH was heated at 50 °C under Ar. One equivalent of the terpyridine compound in dichloromethane was added and the mixture was stirred under Ar for 3 hours, protected from light. The solid was centrifuged and washed using EtOH and diethyl ether. It was then dried under vacuum overnight.

*For Pd-complexes:*²² The same procedure as for the Cu-complexes was used except that the reaction was performed stirring under Ar for 48 hours, protected from light, and purified using dimethylformamide, dichloromethane and diethyl ether. It was then dried under vacuum for two days.



Figure S1. HRMS of Cu-tpy Cl⁻ salt. A) full range spectrum, B) zoom in the dimer, C) selected experimental spectra and D) theoretical spectra for the indicated formula.



Figure S2. HRMS of Cu-tpy NO₃⁻ salt. A) full range spectrum, B) selected experimental spectra and C) theoretical spectra for the indicated formula.



Figure S3. Characterization of Cu-ttpy Cl⁻ salt. A) full range HRMS spectrum, B) zoom in the dimer, C) selected experimental spectra and D) theoretical spectra for the indicated formula. E) X-ray crystal structure of the metal complex. The Cu^{2+} core is pentacoordinated: 3 coordinations come from the ttpy and 2 from two Cl⁻ anions.



Figure S4. HRMS of Cu-ttpy NO₃⁻ salt. A) full range spectrum, B) zoom in the dimer, C) selected experimental spectra and D) theoretical spectra for the indicated formula.



Figure S5. HRMS of Cu-Clptp Cl⁻ salt. A) full range spectrum, B) zoom in the dimer, C) selected experimental spectra and D) theoretical spectra for the indicated formula.



Figure S6. Characterization of Cu-Clptp NO_3^- salt. A) full range HRMS spectrum, B) selected experimental spectra and C) theoretical spectra for the indicated formula. D) X-ray crystal structure of the metal complex. The Cu²⁺ core is pentacoordinated: 3 coordinations come from the Clptp and 2 from two NO_3^- anions. One of the NO_3^- can also be replaced by a H₂O molecule.



Figure S7. HRMS of Zn-ttpy Cl⁻ salt. A) full range spectrum, B) selected experimental spectra and C) theoretical spectra for the indicated formula.



Figure S8. HRMS of Pd-ttpy Cl⁻ salt. A) full range spectrum, B) selected experimental spectra and C) theoretical spectra for the indicated formula.



Figure S9. HRMS of Pt-ttpy Cl⁻ salt. A) full range spectrum, B) selected experimental spectra and C) theoretical spectra for the indicated formula.



Figure S10. HRMS of Pt-Clptp Cl⁻ salt. A) full range spectrum, B) selected experimental spectra and C) theoretical spectra for the indicated formula.

Table S2. Crv	ystal data and	structure	refinement f	for Cu-ttp	y and (Cu-Clr	otr	э.
	,	0		0. 00. ccp	,		~~~~	

Identification code	Cu-	ttpy	Cu-C	Clptp
Empirical formula	C ₆₆ H ₅₇ Cl ₆ Cu ₃ N ₉ (D ₃	C ₄₂ H ₃₂ Cl ₂ Cu ₂ N ₁₀ C	D ₁₄
Formula weight	1427.53		1098.76	
Temperature	103(2) K		103(2) K	
Wavelength	1.54187 Å		1.54187 Å	
Crystal system	Monoclinic		Monoclinic	
Space group	P21/n		P21/c	
Unit cell dimensions	a = 10.824(6) Å	a= 90°.	a = 14.497(5) Å	a= 90°.
	b = 23.346(13) Å	b= 101.313(11)°.	b = 14.961(5) Å	b= 104.458(12)°.
	c = 24.349(12) Å	g = 90°.	c = 20.208(7) Å	g = 90°.
Volume	6033(6) Å ³		4244(3) Å ³	
Ζ	4		4	
Density (calculated)	1.572 Mg/m ³		1.720 Mg/m ³	
Absorption coefficient	4.130 mm ⁻¹		3.129 mm ⁻¹	
F(000)	2916		2232	
Crystal size	0.07 x 0.04 x 0.01 m	וm ³	0.03 x 0.01 x 0.01 m	1m ³
Theta range for data collection	6.39 to 68.25°.		6.33 to 68.25°.	
Index ranges	-12<=h<=9, -27<=k<	<=28, -29<=l<=29	-17<=h<=12, -18<=k<	<=12, -19<=l<=24
Reflections collected	41936		23117	
Independent reflections	10912 [R(int) = 0.12	.34]	7668 [R(int) = 0.043	35]
Completeness to theta = 68.25°	98.90%		98.70%	
Absorption correction	Semi-empirical fro	om equivalents	Semi-empirical fro	om equivalents
Max. and min. transmission	0.9599 and 0.760	9	0.9694 and 0.912	0
Refinement method	Full-matrix least-	-squares on F ²	Full-matrix least-	squares on F ²
Data / restraints / parameters	10912 / 0 / 796		7668 / 0 / 647	
Goodness-of-fit on F2	1.04		1.019	
Final R indices [I>2sigma(I)]	R1 = 0.0661, w	R2 = 0.1551	R1 = 0.0315, wl	R2 = 0.0685
R indices (all data)	R1 = 0.1165, w	R2 = 0.1802	R1 = 0.0498, wl	R2 = 0.0720
Largest diff. peak and hole	1.605 and -0.767 e.	Å-3	0.332 and -0.328 e.	Å-3
Summary of Data CCDC	1475809		1475808	





Figure S11. Evolution of the ratio between the total DNA signal of the 5- charge state and the internal standard dT6 in function of the ligand concentration in percentage of deviation around the mean for (A and D) 22GT, (B and E) Pu24 and (C and F) $[G_4T_4G_4]_2$ with (A to C) Cu-ttpy and (D to F) Cu-Clptp. In black the experimental points, in red the mean and grey the standard deviation. There is no systematic deviation from the mean value upon titration, indicating that the response factors are equal for the different complexes.



Figure S12. Top: titration of 10 μ M 22GT, Pu24 or [G4T4G4]₂ by Cu-ttpy in 100 mM TMAA + 1 mM KCl. Zooms on the 5- charge state. Bottom: quantification and fitting.



Figure S13. Top: titration of 10 μ M 22GT, Pu24 or [G4T4G4]₂ by Cu-Clptp in 100 mM TMAA + 1 mM KCl. Zooms on the 5- charge state. Bottom: quantification and fitting.



Figure S14. Left, titration of 10 μ M 22GT by Cu-tpy in 100 mM TMAA + 1 mM KCl. Zooms on the 5-charge state. Right, quantification and fitting.

Table S3. Ko's obtained for the Cu-complexes (In red, the values coming from titrations, in blue from single concentration experiments). The sequences are ranked according to their K_{D1}/K_{D2} ratio for Cu-ttpy. The global K_Ds of the telomeric sequences (repeats of TTAGGG) are highlighted in blue. We calculated the equilibrium constants for each single point in the titration and calculated the standard deviation so that we can estimate the errors made on the single points K_D values: the error is less than 10%.

		Cu-t	tpy			Cu-	Clptp				Cu-tpy	
	10д К р1	וסס א ני	^{za} X* ^{za} X)60	K ^{D7} ,K ^{D2}	וס ס א ^{מז}	וספ א ניי	od(K ^{eז} *K ^{bs}	K _{p1} /Kp2	וסס א ^{נה}	וסט א ^{בז}	^{za} y* ^{za} y)60	K _{a1} /Ka3
22GT	-4.7±0.4	-6.5 ± 0.2	-11.2	73	-4.0±0.2	-5.9 ± 0.3	-9.9	69	-4.3 ± 0.1		1	I
25TAG	-5.5	-7.2	-12.7	49	-5.6	-7.1	-12.8	33	-4.6			
23TAG	-5.7	-6.7	-12.4	6	-5.8	-6.4	-12.2	4	-4.6		ı	ı
24TTG	-5.5	-6.4	-11.9	7	-4.9	-5.6	-10.4	ß	-4.3			
26ТТА	-5.1	-5.7	-10.8	5	-4.8	-5.8	-10.6	6	-5.0		-	-
23AG	-5.8	-5.9	-11.7	1.4	-5.4	-5.7	-11.0	2.1	-4.8			T
[G4T3G4] ₂	-4.6	-4.6	-9.3	1.0	-4.7	-4.5	-9.2	0.6	-4.0		ı	1
22AG	-6.1	-5.9	-12.0	0.7	-6.3	-5.9	-12.2	0.4	-4.5		ı	ı
[G4T4G4] ₂	-5.0±0.2	-4.8 ± 0.2	-9.8	0.7	-4.6 ± 0.1	-4.4 ± 0.3	-9.0	0.6	-4.2			1
24nonG4	-4.7	-4.3	0.6-	0.48	-4.6	-4.5	-9.0	0.78	-4.0			
CMA	-5.2	-4.8	-10.0	0.37	-5.0	-4.5	-9.5	0.33	-4.8		ı	1
Pu24	-5.5 ± 0.7	-5.0 ± 0.3	-10.4	0.35	-4.8±0.3	-4.1 ± 0.5	-8.9	0.21	-4.8		ı	ı
22CTA	-5.2	-4.7	-9.9	0.34	-5.3	-4.8	-10.1	0.27	-4.7			
CMT	-5.8	-5.3	-11.0	0.33	-5.5	-5.0	-10.5	0.36	-5.1	-4.8	-9.9	0.4
26CEB	-5.2	-4.3	9.6-	0.13	-5.1	-4.0	-9.2	0.08	-5.0	-4.1	ı	0.1
[TG4T]₄	-5.2	-4.1	-9.4	0.08	-5.0	-4.3	-9.3	0.20	-3.7		ı	1
G4(T4G4) ₃	-6.0	-4.5	-10.5	0.03	-6.1	-4.6	-10.7	0.04	-5.3		ı	1
ds26	-4.2	i.	·	1	-4.0				-4.2		ı	1
NIH		,	·						-4.2			
ckit1		,		1					1			1
TBA		1							1	•	-	

S16



Figure S15. MS spectra of 23TAG with 1 eq Cu-ttpy, 360A,^{1,2} Phen-DC3,³ PDS (pyridostatin)⁴ and TMPyP4, respectively, in 100 mM TMAA + 1 mM KCl. The amount of ligand bound was quantified: Cu-ttpy binds the hTel G4 with affinities as good as 360A and Phen-DC3.



Figure S16. MS spectra of 10 μ M A) ckit1, B) HIV-PRO1 and C) TBA in 100 mM TMAA + 1 mM KCl with 10 μ M Cu-ttpy. The ligand does not bind to TBA and only to the species that do not contain potassium for ckit1. On the other hand the fraction of DNA with 1-K⁺ is increased when bound to the ligand, which could indicate that G-quadruplex folding is induced.



Figure S17. MS spectra of 10 μ M DNA in 100 mM NH4OAc or 100 mM TMAA + 1 mM KCl with 10 μ M Cu-Clptp. The results are similar in both conditions: the binding of this ligands is cooperative for the human telomeric sequences.


Figure S18. Comparison of the cooperativity of binding of different metal complexes on 22GT. ESI-MS of 10 μ M 22GT in 100 mM TMAA + 1 mM KCl with 10 μ M of ligands.



Figure S19. MS spectra of 10 μ M (A) 22CTA, (B) 23AG and (C) 22AG in 100 mM TMAA + 1 mM KCl with 10 μ M Cu-ttpy ligand (L). When two ligands are bound the main potassium stoichiometry is 2-K⁺.



Figure S20. CD of (A) Pu24 and (B) [G4T4G4]₂ in presence of increasing amount of Cu-ttpy ligand.



Figure S21. Competition experiment: MS spectra of 10 μ M 22GT in 100 mM TMAA + 1 mM KCl with 10 μ M (A) Cu-ttpy, (B) Cu-Clptp and (C) 10 μ M of each ligand. When two ligands are bound the main potassium stoichiometry is 2K. There is a statistical distribution of stoichiometries: 2x Cuttpy, 2x CuClptp and 1+1. The red-colored peaks can only be attributed to the heterodimer bound to 22GT.



Figure S22. MS/MS spectra of 40 μ M (G₄T₄G₄)₂ in 100 mM NH₄OAc with 60 μ M (A) Cu-ttpy. Top, selection of [(G₄T₄G₄)₂+2/3NH₄⁺]⁶⁻ and activation with 1.1 V, middle, selection of [(G₄T₄G₄)₂+**1L**+2/3NH₄⁺]⁶⁻ and activation with 1.3 V and bottom, selection of [(G₄T₄G₄)₂+**2L**+2/3NH₄+]⁶⁻ and activation with 1.3 V.



Figure S23. MS/MS spectra of 40 μ M (G₄T₃G₄)₂ in 100 mM NH₄OAc with 60 μ M (A) Cu-ttpy. Top, selection of [(G₄T₃G₄)₂+2/3NH₄⁺]⁶⁻ and activation with 1.1 V, middle, selection of [(G₄T₃G₄)₂+**1L**+2/3NH₄⁺]⁶⁻ and activation with 1.3 V and bottom, selection of [(G₄T₃G₄)₂+**2L**+2/3NH₄+]⁶⁻ and activation with 1.3 V.

Section 4. Supplementary references

- 1 A. De Cian and J.-L. Mergny, *Nucleic Acids Res.*, 2007, **35**, 2483–93.
- 2 C. Granotier, G. Pennarun, L. Riou, F. Hoffschir, L. R. Gauthier, A. De Cian, D. Gomez, E. Mandine, J.-F. Riou, J.-L. Mergny, P. Mailliet, B. Dutrillaux and F. D. Boussin, *Nucleic Acids Res.*, 2005, **33**, 4182–90.
- 3 A. De Cian, E. Delemos, J.-L. Mergny, M.-P. Teulade-Fichou and D. Monchaud, J. Am. Chem. Soc., 2007, **129**, 1856–7.
- 4 D. Koirala, S. Dhakal, B. Ashbridge, Y. Sannohe, R. Rodriguez, H. Sugiyama, S. Balasubramanian and H. Mao, *Nat. Chem.*, 2011, **3**, 782–7.
- 5 K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu and A. T. Phan, *J. Am. Chem. Soc.*, 2009, **131**, 4301–9.
- 6 A. T. Phan, V. Kuryavyi, K. N. Luu and D. J. Patel, *Nucleic Acids Res.*, 2007, **35**, 6517–25.
- 7 K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2006, **128**, 9963–70.
- 8 J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927–4940.
- 2. Zhang, J. Dai, E. Veliath, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2010, **38**, 1009–1021.
- 10 K. W. Lim, P. Alberti, A. Guédin, L. Lacroix, J. F. Riou, N. J. Royle, J. L. Mergny and A. T. Phan, *Nucleic Acids Res.*, 2009, **37**, 6239–6248.
- 11 M. Trajkovski, E. Morel, F. Hamon, S. Bombard, M.-P. Teulade-Fichou and J. Plavec, *Chem. A Eur. J.*, 2015, **21**, 7798–7807.
- 12 A. T. Phan, V. Kuryavyi, H. Y. Gaw and D. J. Patel, *Nat. Chem. Biol.*, 2005, **1**, 167–173.
- A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle and D. J. Patel, J. Am. Chem. Soc., 2007, 129, 4386– 92.
- 14 S. Amrane, M. Adrian, B. Heddi, A. Serero, A. Nicolas, J. Mergny and A. T. Phan, *J. Am. Chem. Soc.*, 2012, **134**, 5807–5816.
- 15 G. Laughlan, A. Murchie, D. Norman, M. Moore, P. Moody, D. Lilley and B. Luisi, *Science (80-.).*, 1994, **265**, 520–524.
- 16 P. Schultze, F. W. Smith and J. Feigon, *Structure*, 1994, **2**, 221–233.
- 17 F. Balthasart, J. Plavec and V. Gabelica, J. Am. Soc. Mass Spectrom., 2013, 24, 1–8.
- 18 S. Amrane, A. Kerkour, A. Bedrat, B. Vialet, M.-L. Andreola and J.-L. Mergny, *J. Am. Chem. Soc.*, 2014, **136**, 5249–52.
- 19 R. F. Macaya, P. Schultze, F. W. Smith, J. a Roe and J. Feigon, *Proc. Natl. Acad. Sci.*, 1993, **90**, 3745–3749.
- 20 A. Marchand, A. Granzhan, K. Iida, Y. Tsushima, Y. Ma, K. Nagasawa, M.-P. Teulade-Fichou and V. Gabelica, *J. Am. Chem. Soc.*, 2015, **137**, 750–6.
- 21 M. J. Cavaluzzi and P. N. Borer, *Nucleic Acids Res.*, 2004, **32**, e13.
- 22 E. Largy, F. Hamon, F. Rosu, V. Gabelica, E. De Pauw, A. Guédin, J.-L. Mergny and M.-P. Teulade-Fichou, *Chemistry*, 2011, **17**, 13274–83.
- 23 H. Bertrand, D. Monchaud, A. De Cian, R. Guillot, J.-L. Mergny and M.-P. Teulade-Fichou, *Org. Biomol. Chem.*, 2007, **5**, 2555.

SUPPORTING INFORMATION ARTICLE 4

Ligand-Induced Conformational Changes with Cation Ejection upon Binding to Human Telomeric DNA G-Quadruplexes

A. Marchand, A. Granzhan, K. Iida, Y. Tsushima, Y. Ma, K. Nagasawa, M-P. Teulade-Fichou & V. Gabelica

Journal of the American Chemical Society **2015**, 137, 750-756.

Ligand-induced conformational changes with cation ejection upon binding to human telomeric DNA G-quadruplexes

Adrien Marchand^{1,2}; Anton Granzhan³; Keisuke Iida⁴, Yamato Tsushima⁴, Yue Ma⁴, Kazuo Nagasawa⁴; Marie-Paule Teulade-Fichou³; Valérie Gabelica^{1,2}

¹ Univ. Bordeaux, IECB, ARNA Laboratory, 33600 Pessac, France.

² Inserm, U869, ARNA Laboratory, 33000 Bordeaux, France.

³ Section Recherche, Institut Curie, CNRS UMR176, Centre Universitaire Paris XI, Bat. 110, 91405 Orsay, France.

⁴ Tokyo University of Agriculture & Technology. 2-24-16 Naka-cho. 84-8588 Koganei-shi, Tokyo, Japan.

Supporting Information - Table of contents:

	<u>Page</u>
Figure S1: Ligands structures and ligand molar extinction coefficients	S2
Figure S2: CD spectra and UV-melting of the four Htel sequences in 100 mM TMAA/1 mM KCl and in 100 mM KCl	S3
Figures S3-S7: Full-scale ESI mass spectra of ligand-DNA 5µM-5µM mixtures	S3-S7
Figure S8: Circular dichroism spectrum of 22nonG4 without ligand	S8
Mathematical treatment done on the raw MS spectra + Figure S9	S8-S10
Figure S10: Bar graphs showing the specific K+ distribution for the 4 human telomeric sequences without and with ligands	S11
Figure S11: Mass spectra of Pu24-myc without and with ligands	S12
Figure S12: Circular dichroism spectra of Pu24-myc without and with ligands	S13
Figure S13-S15: Circular dichroism spectra of the 4 human telomeric sequences without and with ligands	S14-S16
Figure S16-S17: Melting experiments monitored by CD for 24TTG and 26TTA	S17-S18
Figure S18: Titration of 22GT by KCl done by CD and MS	S19







Compound	Wavelength (nm)	Epsilon (M ⁻¹ cm ⁻¹)
360A	260	40 000
Phen-DC3	370	62 400
PDS	227	61 450
TrisQ	257	70 900
L2H2	257	46 600
TMPyP4	424	226 000

Figure S1. Chemical structures of the ligands and molar extinction coefficients. Salts were: 360A iodide salt, Phen-DC3 methylsulfonate salt, PDS trifluoroacetate salt, TrisQ bromide salt, L2H2 trifluoroacetate salt and TMPyP4 *p*-toluenesulfonate.



Figure S2. A-D, CD spectra at 20 °C of 26TTA, 23TAG, 24TTG and 22GT, respectively, in 100 mM TMAA + 1 mM KCl in red and in 20 mM K₂HPO₄/KH₂PO₄ + 70 mM KCl in purple. E-H, their respective melting experiments followed by the absorbance at 295 nm. Melting temperatures are indicated on the graphs and the difference between the two conditions is referred as well.



Figure S3. Mass spectra of 5 μM 24TTG in 100 mM TMAA + 1 mM KCl after 3 days incubation without (A) and with 5 μM ligand 360A (B), Phen-DC3 (C), PDS (D), TrisQ (E), L2H2 (F) and TMPyP4 (G).



Figure S4. Mass spectra of 5 μM 26TTA in 100 mM TMAA + 1 mM KCl after 3 days incubation without (A) and with 5 μM ligand 360A (B), Phen-DC3 (C), PDS (D), TrisQ (E), L2H2 (F) and TMPyP4 (G).



Figure S5. Mass spectra of 5 μM 23TAG in 100 mM TMAA + 1 mM KCl after 3 days incubation without (A) and with 5 μM ligand 360A (B), Phen-DC3 (C), PDS (D), TrisQ (E), L2H2 (F) and TMPyP4 (G).



Figure S6. Mass spectra of 5 μM 22GT in 100 mM TMAA + 1 mM KCl after 3 days incubation without (A) and with 5 μM ligand 360A (B), Phen-DC3 (C), PDS (D), TrisQ (E), L2H2 (F) and TMPyP4 (G).



Figure S7. Mass spectra of 5 μM ds26 hairpin duplex (self-complementary sequence d(CA₂TCG₂ATCGA₂T₂CGATC₂GAT₂G)) in 100 mM TMAA after 3 days incubation without (A) and with 5 μM ligand 360A (B), Phen-DC3 (C), PDS (D), TrisQ (E), L2H2 (F) and TMPyP4 (G).



Figure S8. Circular dichroism spectrum of 5 μM 22nonG4 (grey), 23nonG4 (green), 24nonG4 (blue) and 26nonG4 (orange) in 100 mM TMAA + 1 mM KCl. These sequences are not forming G-quadruplexes under these conditions.

Mathematical treatment on the non-specific adducts. The 22nonG4 sequence was used as a reference to obtain the non-specific K^+ distribution. We used the method described by Klassen's group (Sun, J., Kitova, E. N., Wang, W., & Klassen, J. S. (2006). Analytical Chemistry, 78(9), 3010–8.) in order to distinguish specific from non-specific adducts. The following procedure has been used:

- 1- The integrals of the K⁺ distribution for the non-G4 forming sequences were obtained
- 2- They were normalized according to the 0 K⁺ stoichiometry
- 3- The values have been used to reconstruct the specific K⁺ stoichiometry thanks to the following formula:

$$0K = i0K + i0K * x + i0K * y$$

$$1K = (i1K - i0K * x) + (i1K - i0K * x) * x + (i1K - i0K * x) * y$$

$$Equation 2$$

$$2K = (i2K - i0K + x) - (i1K - i0K + x) + (i2K - i0K + x) + (i$$

2K = (i2K - i0K * y - (i1K - i0K * x) * x) + (i2K - i0K * y - (i1K - i0K * x) * x) * x + (i2K - i0K * y - (i1K - i0K * x) * x) * yEquation 3

- $\begin{aligned} 3K &= (i3K (i1K i0K * x) * y (i2K i0K * y (i1K i0K * x) * x) * x) + (i3K (i1K i0K * x) * y (i2K i0K * y (i1K i0K * x) * x) * x) * x + (i3K (i1K i0K * x) * y) * y + (i2K i0K * y (i1K i0K * x) * x) * y \\ &= Equation 4 \end{aligned}$

Where 0K, 1K, 2K, 3K, 4K and 5K reflect the specific K⁺ distribution. x and y are the values of the normalized integrals of the 1 and 2 non-specific K⁺ (See Table S1, below). i0K, i1K, i2K, i3K, i4K and i5K are the initial integrals for the 0, 1, 2, 3, 4 and 5 K⁺ stoichiometries.

The specific K^+ intensity is therfore egal to the total area of the peak for this stoichiometry minus the non-specific adducts of the peaks with lower K^+ stoichiometries plus the non-specific adducts on the considered peak.



Figure S9. Schematic representation of the mathematical treatment done to obtain the specific K^{+} distributions.

Name	Sequence	I _o	I _{1К+} (<i>x</i>)	I _{2К+} (у)	I _{зк+} (<i>z</i>)
22nonG4	d(GGGATGCGACAGAGAGGACGGG)	1	0.4521	0.1578	0
23nonG4	d(TGGGATGCGACAGAGAGGACGGG)	1	0.5367	0.2087	0
24nonG4	d(TGGGATGCGACAGAGAGGACGGGA)	1	0.5961	0.3027	0.1459
26nonG4	d(TTGGGATGCGACAGAGAGGACGGGAT)	1	0.7804	0.4476	0.3241

Table S1. Values of the normalized areas for the different control sequences used.



Figure S10. Reconstructed K⁺ distributions from the 5- charge state of 24TTG (Panel 1), 26TTA (Panel 2), 23TAG (Panel 3) and 22GT (Panel 4) in 100 mM TMAA + 1 mM KCl after 3 days incubation without (A) and with 5 μM ligand TrisQ (B), L2H2 (C), 360A (D), Phen-DC3 (E) and PDS (F).



Figure S11. Mass spectra of 5 μ M c-myc in 100 mM TMAA + 1 mM KCl after 3 days incubation without (A) and with 5 μ M ligand 360A (B), Phen-DC3 (C), PDS (D), TrisQ (E), L2H2 (F) and TMPyP4 (G). Zooms on the 5- charge state showing the K⁺ distribution for each complexes.



Figure S12. Circular dichroism spectrum of 5 μM c-myc in 100 mM TMAA + 1 mM KCl (A) and in 20 mM KH₂PO₄/K₂HPO₄ + 70 mM KCl (B) without (Black) and with ligand 360A (Red), Phen-DC3 (Yellow), PDS (Pink), TrisQ (Blue), L2H2 (Orange) and TMPyP4 (Purple). All spectra are almost superimposed and indicates no change in c-myc conformation.



Figure S13. Circular dichroism spectra of 5 μ M 26TTA in 100 mM TMAA + 1 mM KCl (A) and in 20 mM KH₂PO₄/K₂HPO₄ + 70 mM KCl (B) after 3 days without ligand (black) and in the presence of Phen-DC3 (yellow), 360A (red), PDS (green), L2H2 (orange) and TrisQ (blue).



Figure S14. Circular dichroism spectra of 5 μ M 23TAG in 100 mM TMAA + 1 mM KCl (A) and in 20 mM KH₂PO₄/K₂HPO₄ + 70 mM KCl (B) after 3 days without ligand (black) and in the presence of Phen-DC3 (yellow), 360A (red), PDS (green), L2H2 (orange) and TrisQ (blue).



Figure S15. Circular dichroism spectra of 5 μ M 22GT in 100 mM TMAA + 1 mM KCl (A) and in 20 mM KH₂PO₄/K₂HPO₄ + 70 mM KCl (B) after 3 days without ligand (black) and in the presence of Phen-DC3 (yellow), 360A (red), PDS (green), L2H2 (orange) and TrisQ (blue).



Figure S16. Melting experiment obtained by circular dichroism of 5 μM 24TTG in 100 mM TMAA + 1 mM KCl without and with ligand 360A, TrisQ, Phen-DC3, L2H2 and PDS. Blue curves are the curves at 4 °C, the lowest temperature, and the red ones are at the highest temperature, 90 °C. T_{1/2} is the temperature where half of the maximum signal is lost.



Figure S17. Melting experiment obtained by circular dichroism of 5 μM 26TTA in 100 mM TMAA + 1 mM KCl without and with ligand 360A, TrisQ, Phen-DC3, L2H2 and PDS. Blue curves are the curves at 4 °C, the lowest temperature, and the red ones are at the highest temperature, 90 °C. T_{1/2} is the temperature where half of the maximum signal is lost.



Figure S18. A, CD spectra of 5 μ M 22GT in 100 mM TMAA with no KCl, 200 μ M, 1 mM and in 20 mM KH₂PO₄/K₂HPO₄ + 70 mM KCl (from light to dark, increasing KCl concentration). Mass spectra of 5 μ M 22GT in 100 mM TMAA with no KCl (B), 200 μ M KCl (C) and 1 mM KCl (D).

Etude de G-Quadruplexes par Spectrométrie de Masse: Chemins de Repliement et Modes de Liaison de Ligands

Résumé : Un G-quadruplex (G4) est une structure non-canonique d'acides nucléiques formée par des séquences riches en guanines. Certains G4s sont polymorphiques, une même séquence peut former des G4s de différentes topologies. Les G4s sont proposés comme régulateurs de processus biologiques car ils sont trouvés dans des régions génomiques clés telles que dans des promoteurs de gènes et au niveau des télomères. Stabiliser ces G4s par rapport à la forme duplexe est une stratégie proposée pour combattre le cancer. Pour ce faire, des ligands spécifiques et affins sont utilisés. Le design de ces ligands implique habituellement de larges plans aromatiques, optimisés pour se lier par des interactions π - π sur les G-quartets extérieurs. Cependant, si ce type d'interaction était le seul mode de liaison, tous les ligands auraient des affinités similaires pour tous les G4s.

Afin de caractériser les structures ciblées et de quelle manière les ligands vont interagir avec celles-ci, nous avons utilisé la spectrométrie de masse de type native (MS).

D'abord, nous avons développé une méthode de préparation d'échantillons en conditions KCI pour former les G4s dans des conditions biologiquement pertinentes. Ensuite, nous avons caractérisé les équilibres de liaison du K⁺ aux G4s et caractérisé leur mécanisme de repliement. Ce mécanisme implique la présence d'une impasse constituée de G4s antiparallèles à 1- et 2-K⁺ qui sont formés rapidement. Enfin, nos études de liaison de ligands ont montré que certains des ligands les plus affins pouvaient influencer la structure des G4s comme observé par le nombre d'ions potassium liés. Les ligands Phen-DC3, 360A et PDS sont capables de déplacer les équilibres vers la forme à 1-K⁺ antiparallèle. La structure antiparallèle à 2-K⁺ est favorisée par la liaison coopérative de deux ligands Cu-*ttpy*. Ces résultats démontrent l'importance de la caractérisation des stœchiométries de complexes ternaires (G4:ligand:K⁺), obtenue par la spectrométrie de masse native.

Mots clés : Spectrométrie de masse, G-quadruplexe, potassium, mécanisme de repliement, mode de liaison, ligand

Mass Spectrometry Study of G-Quadruplex Nucleic Acids: Folding Pathways and Ligand Binding Modes

Abstract: A G-quadruplex (G4) is a non-canonical nucleic acids structure formed by guanine-rich sequences. Some G4s are polymorphic, a given sequence can form G4s of different topologies. G4s are proposed to be biological regulators because they are found in key regions of the genome, for example, in gene promoters or at the telomeres. Stabilizing G4s formed in those regions as compared to the duplex form is a strategy to fight cancer. To do so, specific and affine ligands are used. Ligand design usually implies the optimization of large aromatic planes to π - π stack on external G-quartets. However, if this was the only binding mode, all ligands would bind with similar affinities to all G4s.

To characterize which structures should be targeted and how the ligands interact with these structures, we used native mass spectrometry (MS).

First, we developed a MS-compatible sample preparation method in KCI conditions in which G4s are folded with similar topologies as compared to those obtained in biologically relevant conditions. Then, we characterized the K⁺ binding equilibria and G4s folding pathways. This folding pathway involves the presence of a dead-end constituted by antiparallel G4s with either 1- or 2-K⁺ cations that are folded first. Finally, our ligand binding studies showed that some of the most affine ligands can influence G4's structures, as probed by the number of K⁺ ions bound. Ligands Phen-DC3, 360A and PDS are able to shift the equilibria towards the 1-K⁺ antiparallel G4s. The formation of antiparallel with 2-K⁺ complexes is induced by the cooperative binding of two Cu-*ttpy* ligands. Our results demonstrate the importance to characterize ternary complex stoichiometries (G4:ligand:K⁺) as obtained from native mass spectrometry.

Keywords: Mass spectrometry, G-quadruplex, potassium, folding pathway, binding mode, ligand

Unité de recherche Inserm U1212 / CNRS UMR5320 – ARNA : Régulations Naturelles et Artificielles Université de Bordeaux, 146 rue Léo Saignat, 33076 Bordeaux cedex