NMR structural studies of G-quadruplexes nucleic acids from KRAS and their interaction with ligands.

Julien Marquevielle

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTE

SPÉCIALITÉ BIOCHIMIE

Par Julien MARQUEVIELLE

Etudes structurales par RMN des acides nucléiques G-quadruplexes de KRAS et leur interaction avec des ligands.

Sous la direction de : Dr. Gilmar SALGADO

Soutenue le 18 décembre 2019

Membres du jury :

Pr. GIANCOLA, Concetta	Professeur à l'Université de Naples, Federico II	Présidente
Dr. FIALA, Radovan	Professeur associé, CEITEC, Brno	Rapporteur
Dr. MONCHAUD, David	Directeur de recherche, CNRS, ICMUB, Dijon	Rapporteur
Dr. MACKERETH, Cameron	Directeur de recherche, CNRS, IECB, Bordeaux	Examinateur
Dr. SALGADO, Gilmar	MCU à l'Université de Bordeaux	Directeur

Remerciements

A toutes les personnes qui m'ont soutenu et aidé,

Voici pour vous ces quelques mots que je souhaite vous dédier.

Avant tout, merci pour m'avoir permis de commencer cette aventure inouïe,

Reconnaissant d'avoir pu rejoindre l'équipe dirigée par Jean-Louis.

N'oubliant pas l'équipe administrative et leur aide pour quelques péripéties

Ainsi que Philippe pour m'avoir soutenu afin d'obtenir un sursis.

Longtemps au même bureau j'ai été installé,

En profitant de ce qu'Aurore me racontait,

Que ce soit du domaine privé ou bien pour me former.

Une fois descendu, tout cela était terminé.

Inoubliables sont ces moments avec Anne, Carmelo ou bien Samir,

Pour certains y repenser me fait toujours rire,

Espérant également, par moment, avoir pu les faire sourire.

En plus de ces personnes à mes côtés toute l'année,

Tous ces stagiaires que j'ai pu rencontrer,

Unis autour de ce magnifique projet,

Doivent aussi savoir que je tiens à les remercier.

Impossible sans eux d'avoir pu autant avancer,

Ainsi, Pierre, Mona, Louise, Coralie ou bien Olivier,

Ne vous inquiétez pas je ne vous ai pas oublié.

Très heureux d'avoir eu la chance de vous côtoyer,

Sachez que je vous souhaite toute la réussite que vous méritez.

Longtemps ils m'ont accompagné pendant mes études,

Et pourtant tout cela n'était qu'un prélude.

Survivants des plus grandes crises entre amis,

Avec vous j'ai partagé de sacrés moments de vie.

Mille mercis à vous Damien, Ismaël, Arthur, Jules, Fabien ou encore Jean-Sébastien,

Impossible sans vous ce projet n'aurait pu être mené à bien,

Surtout dans les périodes les plus sombres, guidé par votre soutien.

De tous ceux que je dois remercier je n'allais pas t'oublier,

Impossible de te dissocier de tout ce qui est arrivé.

Réel guide, tu as su très souvent m'aiguiller,

En permettant ainsi que je marche sur la route du succès.

Certains diront que tes méthodes ne sont pas adaptées,

Tu peux cependant avoir confiance en tes capacités à diriger.

Encore merci Gilmar pour m'avoir formé pendant ce voyage,

Une page se tourne vers un nouvel Age,

Reste à espérer pour nous de bons présages.

Les derniers concernés par ces mots sont pour moi les plus importants,

Avec notamment ceux qui ne sont plus là pour partager ce moment.

Fiers de moi, j'espère qu'ils le sont,

A toi mon papa mais surtout à toi mon grand tonton.

Ma chère maman, à toi, je dois plus que mille mercis,

Impossible d'oublier tout ce que pour moi tu as sacrifié dans ta vie.

Le cadeau le plus précieux de cette thèse c'est ma rencontre avec Coralie,

La femme qui partage ma vie et qui de moi a tout compris,

En espérant qu'elle veuille encore de moi après avoir lu ce manuscrit.

A tous ceux que je n’ai pas cités dans ces vers,

Veuillez m’excuser, je ne manie pas les mots comme Jacques Prévert.

Je vous remercie cependant sincèrement,

A Vimala, Ludovic et Thomas pour tous ces chaleureux moments.

A Clément et Willy avec qui j’ai partagé le même bateau,

On y est arrivé, il n’a pas coulé ce vieux rafiot !

Table of contents

List of figures

Abstract

KRAS gene codes for a highly mutated GTPase protein acting as a « switch » between an active and an inactive state, a mechanism found to be important in biological processes such as cell replication and proliferation. When misregulated, these processes are found to be at the origin many types of cancer. KRAS mutations are particularly implicated in lung (30%), colorectal (44%) and pancreatic (97%) cancers. Despite the fact that these mutations are well known, KRAS is still an undruggable target because all the actual strategies (RAS activator inhibitors, membrane association inhibitors, and so on) are not efficient enough as cancer therapies. That is why new strategies have emerged recently, such as directly targeting the KRAS promoter region and especially some specific structures called G-quadruplexes (G4). Although we do not understand well the phenomena, there are plenty of evidence in the literature that these structures can assemble both in vitro and in cellular conditions. It was shown that G4 within KRAS promoter region can bind transcription related proteins and disturb transcription process acting as a block when transcription machinery is reading the genetic sequence. Stabilization of these structures, using small chemical ligands for example, could become a new area of therapy. In my thesis work, I am focused on a 32 nucleotide sequence (KRAS32R) which can form G4 and also corresponds to the minimal interaction domain of transcription proteins such as MAZ or hnRNP1. This last protein is capable of binding to KRAS32R G-quadruplexes and possibly unfolding it, favoring the transcription of KRAS. This project is divided into two major parts. One part is to understand the folding of this KRAS32R G-quadruplex at atomic level. In another part I want to understand how these DNA structures can interact with small organic molecules that prevent the interaction with transcription factors that have been associated with the G-quadruplex motifs found in the promoter region.

Résumé en français

L'oncogène KRAS code pour une protéine GTPasique hautement mutée qui agit comme un « interrupteur » entre des états actifs et inactifs, un mécanisme important dans les processus comme la réplication ou la prolifération cellulaire. Quand ils sont dérégulés, ces processus sont à l'origine des cancers. Les mutations de KRAS sont particulièrement impliquées dans les cancers des poumons (30%), colorectaux (44%) et pancréatiques (97%). Malgré le fait que ces mutations soient bien connues, aucune molécule ne cible KRAS car toutes les stratégies actuelles ne sont pas assez efficaces pour les thérapies contre le cancer. C'est pourquoi de nouvelles stratégies ont émergé il y a quelques années visant directement la région promotrice de KRAS et plus précisément des structures appelées G-quadruplexes (G4). Même si le phénomène n'est pas encore parfaitement compris, de nombreux exemples dans la littérature montre que ces structures peuvent se former in vitro et dans les conditions cellulaires. Il a été montré que les G4 formés dans la région promotrice de KRAS peuvent lier des facteurs de transcription et perturber le processus en agissant comme un bloc lorsque l'enzyme vient lire la séquence. La stabilisation ou la destruction des G4, en utilisant de petits ligands chimiques par exemple, pourrait devenir une nouvelle voie de thérapie. Ce travail se concentre sur une séquence de 32 résidus (KRAS32R) qui peut former un G4 et correspond également au domaine minimal d'interaction de certains facteurs de transcription comme MAZ ou hnRNP A1. Cette dernière est capable de lier les G4 de KRAS32R et de les défaire favorisant ainsi la transcription de KRAS. Ce projet est divisé en deux parties majeures. L'une est de comprendre la formation du G4 de KRAS32R au niveau atomique. Pour les autres, nous voulons comprendre comment il interagit avec de petites molécules organiques qui pourraient empêcher l'interaction avec les facteurs qui sont associés aux motifs G4 de cette région promotrice.

Résumé long en français

L'oncogène KRAS code pour une protéine GTPasique hautement mutée qui agit comme un « interrupteur » entre des états actifs et inactifs, un mécanisme important dans les processus comme la réplication ou la prolifération cellulaire. Quand ils sont dérégulés, ces processus sont à l'origine des cancers. Les mutations de KRAS sont particulièrement impliquées dans les cancers des poumons (30%), colorectaux (44%) et pancréatiques (97%). Malgré le fait que ces mutations soient bien connues, aucune molécule ne cible KRAS car toutes les stratégies actuelles ne sont pas assez efficaces pour les thérapies contre le cancer. C'est pourquoi de nouvelles stratégies ont émergé il y a quelques années visant directement la région promotrice de KRAS et plus précisément des structures appelées G-quadruplexes (G4). Les G-quadruplexes sont de structures inhabituelles des acides nucléiques. Ils sont formés via un empilement d'au moins deux tétrades qui sont composées de quatre guanines liées dans un réseau de liaisons hydrogène de type Hoogsteen et stabilisées le plus souvent par un cation monovalent. Même si le phénomène n'est pas encore parfaitement compris, de nombreux exemples dans la littérature montre que ces structures peuvent se former in vitro et dans les conditions cellulaires. Il a été montré que les G4 formés dans la région promotrice de KRAS peuvent lier des facteurs de transcription et perturber le processus en agissant comme un bloc lorsque l'enzyme vient lire la séquence. La stabilisation des G4, en utilisant de petits ligands chimiques par exemple, pourrait devenir une nouvelle voie de thérapie. Ce travail se concentre sur une séquence de 32 résidus (KRAS32R) qui peut former un G4 et correspond également au domaine minimal d'interaction de certains facteurs de transcription comme MAZ ou hnRNP A1. Cette dernière est capable de lier les G4 de KRAS32R et de les défaire favorisant ainsi la transcription de KRAS. Cette étude a pour but de déterminer et valider des cibles G-quadruplexe au sein de l'oncogène KRAS. De plus, elle vise à identifier les groupes chimiques caractéristiques de ligands pouvant interagir avec ces G4. Cependant, afin de créer de bons ligands et d'identifier des groupes chimiques clé, il est nécessaire d'avoir des informations structurales à un niveau atomique concernant les différentes conformations de G-quadruplexe. L'identification d'éléments structuraux spécifiques de ces G-quadruplexes permettraient de créer des ligands spécifiques de G4s comme ceux formés au sein de l'oncogène KRAS. Ce projet est donc divisé en deux parties majeures avec dans un premier temps la détermination de la structure des G-quadruplexes de KRAS32R pour obtenir des

éléments structuraux et dans un second temps l'étude de l'interaction de ces G4s avec des protéines comme hnRNP A1 ainsi que plusieurs ligands provenant de différentes familles. Comme expliqué précédemment, obtenir des informations structurales sur la cible est très important dans le concept de Drug Design. En effet, cela permet de créer des molécules qui sont capables de venir cibler spécifiquement les sites de liaison identifiés lors de l'obtention de la structure. Ensuite lorsque les ligands sont testés, la structure permet l'identification des groupements essentiels à l'interaction. Concernant la structure de KRAS32R, plusieurs études ont tenté de la résoudre sans y parvenir. Afin de comprendre comment s'organisent les séquences riches en guanine de la région NHE de la région promotrice de l'oncogène KRAS, nous avons commencé par étudier ce qui a été montré comme étant la plus petite séquence pouvant former des G4s aussi appelée KRAS21R. KRAS21R représente donc le plus petit G4 de la région NHE qui est capable, tout comme KRAS32R, d'interagir avec des facteurs de transcription comme MAZ. En utilisant la séquence KRAS22RT qui possède une adénine supplémentaire en 3' et en incorporant une mutation de la guanine 16 en thymine, nous avons réussi à garder la conformation principale de KRAS21R tout en améliorant la qualité des spectres RMN (notamment 2D) pour la résolution de la structure. Grâce à la détermination de cette structure et à de nombreux spectres RMN de la séquence KRAS32R, nous avons tenté de comprendre comment se formaient les G-quadruplexes de cette séquence. Plus d'une centaine de mutants ont été testés avant de comprendre la formation de ces G4s pour commencer la détermination de structure. A l'issue de toutes ces analyses, il était clair que la séquence KRAS32R n'adoptait pas une seule conformation mais qu'il y avait un mélange de deux conformations majoritaires. Ces deux conformations ont été identifiées en tant que G9T et G25T correspondant aux résidus mutés menant aux conformations correspondantes. D'un côté, G9T forme une structure unique dans laquelle G9 n'est pas impliqué mais le dernier résidu G32 l'est. De plus, elle est stabilisée par la formation d'une triade. De l'autre côté, G25T est très différente avec G9 impliqué dans la formation du G-quadruplexe sans la présence d'une triade. De plus, la structure reste toujours polymorphique dû à un glissement dans le dernier tract de guanines entre G26 et G29. Il est également possible que G9 soit responsable en partie du polymorphisme qui permet l'échange entre ces deux états. Dans cette étude, nous avons finalement compris la formation des G-quadruplexes au sein de la séquence KRAS32R et nous avons déterminé ces structures G-quadruplexe qui sont connues pour être hautement impliquées dans la transcription et pouvant interagir avec plusieurs protéines

durant le processus. Ainsi nous avons pu apporter des informations cruciales concernant les structure de G-quadruplexes les plus probables au sein de la séquence KRAS32R de la région NHE. Ces structures tridimensionnelles pourraient représenter de nouvelles cibles essentielles pour le développement de nouvelles molécules spécifiques aux G-quadruplexes de la région promotrices de KRAS. En effet, elles pourraient interférer avec la régulation de la transcription en stabilisant ces G4s ou bien en entrant en compétitions avec des facteurs de transcription comme MAZ ou hnRNP A1. Grâce aux nouvelles informations concernant KRAS22RT et JRAS32R, nous avons ensuite testé un panel de ligands. En sélectionnant quelques candidats parmi les meilleurs testés et en étudiant leur interaction avec les G4s à un niveau atomique, nous pourrions comprendre comment les différents groupements chimiques participent à la stabilisation globale de la structure G4. En obtenant ce genre d'information nous voulons comprendre comment les ligands pourraient interagir de façon plus efficace avec les boucles, les sillons ou encore les tétrades. Comme les ligands pour G-quadruplexes sont souvent trop toxiques pour les cellules ou bien à l'inverse pas assez car ils ne peuvent pas passer les différentes barrières, c'est aussi un bon moyen d'identifier les groupements chimiques qui peuvent être remplacés pour surpasser ces limites. Comme ce travail a été réalisé en parallèle de la détermination de la structure de KRAS32R, la majorité des résultats des tests avec ligands a été obtenu avec KRAS22RT. Une fois les structures de KRAS32R G9T et G25T obtenues, nous avons également obtenu quelques résultats préliminaires. Dans le cas de KRAS32R, le but était de trouver des ligands capables de stabiliser les G4s mais également des composés qui seraient capables de déplacer l'équilibre entre les deux conformations G9T et G25T. Un autre objectif était de trouver des molécules capables de prévenir l'interaction avec UP1, une protéine contenant uniquement les deux domaines de liaisons à l'ARN de hnRNP A1. Parmi tous les ligands testés, aucun ne possédait toutes les propriétés que nous recherchions. Plusieurs d'entre eux ont cependant montré de bons résultats en terme de stabilisation, de protection contre la chaine complémentaire ou bien contre la protéine UP1 ou encore en terme de cytotoxicité et de capacité à rentrer dans la cellule. C8, un dérivé acridine, est probablement notre meilleur ligand notamment avec son effet stabilisant et sa capacité à protéger le G4 contre la chaîne complémentaire. Cependant, il n'est pas capable d'empêcher l'interaction de UP1 et n'est pas suffisamment spécifique des cellules cancéreuses en terme de toxicité. D'un autre côté, le ligand AG, un composé salphen, n'est pas notre meilleur ligand mais nous avons décidé de mener des études structurales plus poussées du complexe formé

avec KRAS22RT car il était le plus indiqué pour cela. D'autres composés comme le ligand PhenDC3, de la famille des phénantrolines, ont montré qu'il était possible d'éviter l'interaction avec UP1. Cependant, les mécanismes impliqués n'ont pas encore été identifiés. En conclusion, ce travail de thèse a permis d'apporter de nouveaux éléments sur l'organisation des G-quadruplexes de la région NHE de l'oncogène KRAS avec la résolution de trois structures par RMN dont deux structures en équilibre dans le cas de KRAS32R. La résolution de cette structure a longtemps été considérée comme un réel challenge. Ces trois structures ont le potentiel d'être utilisées dans une approche de Drug Design pour les G4s de KRAS. Une fois la cible identifiée et caractérisée, il est donc possible de se concentrer sur le développement de ligands avec une affinité accrue pour ladite cible. Ce travail représente une étape importante pour obtenir des informations qui permettront plus tard de trouver de nouvelles molécules pour répondre aux problèmes sociétaux majeurs que représentent les cancers pancréatiques, des poumons ou encore colorectaux. Ces cancers sont parmi les plus mortels et ce travail pourrait être une piste supplémentaire pour lutter contre leur croissance et le développement de tumeurs.

Abbreviations

AO: Acridine Orange

ASO: AntiSense Oligonucleotide

Bcl-2: B-cell lymphoma 2

CD: Circular Dichroism

c-KIT: cellular- Receptor tyrosine kinase

c-MYC: cellular- MYeloCytomatosis

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

DNA: DeoxyriboNucleic Acid

FRET: Förster Resonance Energy Transfer

G4: G-Quadruplex

GalNac: N-Acetylgalactosamine

GTP: Guanosine TriPhosphate

HIV: Human Immunodeficiency Virus

hnRNP A1: heterogeneous nuclear RiboNucleoProtein A1

hTERT: human TElomerase Reverse Transcriptase

HSQC: Heteronuclear Single Quantum Coherence spectroscopy

HMBC: Heteronuclear Multiple Bond Correlation spectroscopy

ITC: Isothermal Titration Calorimetry

KRAS: Kirsten RAt Sarcoma viral oncogene homolog

MAZ: MYC Associated Zinc Finger Protein

NDI: Naphthalene DiImide

NHDF: Normal Human Dermal Fibroblasts

NHE: Nuclear Hypersensitive Element

NMR: Nuclear Magnetic Resonance

NOESY: Nuclear Overhauser SpectroscopY

PAGE: Poly-Acrylamide Gel Electrophoresis

PEG: PolyEthylene Glycol

PARP1: Poly [ADP-Ribose] Polymerase 1

PDB: Protein Data Bank

Pot 1: Protection of telomeres protein 1

RNA: RiboNucleic Acid

RISC: RNA-Induced Silencing Complex

sgRNA: single Guide RNA

siRNA: Small Interfering RNA

TFO: Triplex Forming Oligonucleotide

TSS: Transcription Start Site

VEGF: Vascular Endothelial Growth Factor

UP1: Unfolding Protein 1

Introduction

I would like to start this chapter with a paramount statement that captured my imagination during the past three years or so: nucleic acids are considered the cornerstone of life. In this chapter I shall introduce the readers to the some of the concepts needed for the understanding of my results. One of the main reasons seems obvious and it is related to the fact that DNA holds the genetic information that each individual needs to survive, grow and reproduce. So it is without surprise that we expect a tight, sophisticated and intricate level of "management" of chromosomic DNA. During my early years in academic formation, especially during my master, I learnt that it was possible thanks to complex mechanisms and unusual structural motifs. For a long time, these unusual structures were largely unstudied and their role in all different biological processes seemed to remain rather unknown, especially at molecular level. In addition, thanks to their involvement in many metabolism processes, they became very attractive to be studied as possible targets for therapeutic purposes. Nowadays nucleic acids and their non-canonical structures represent one of the major fields in fundamental research. Hereafter, I will introduce you to this fascinating field of research that is the centrepiece of my dissertation.

I. Nucleic Acids

I.1. General information - History

Unlike public opinion, DNA double helix was not discovered by Watson et Crick in the 1950s. Indeed, the DNA discovery is one century older thanks to Johann Friedrich Miescher, a young Swiss chemist[1, 2]. As he was trying to isolate and purify proteins from leucocytes in order to characterize them, he discovered a substance obtained from the nuclei being insoluble in acidic conditions. The substance could be dissolved again in alkaline conditions, so without surprise he called it "nuclein" and though that it represented some storage mechanism for cells. Although this discovery represented a major step forward in biological sciences, Miescher could not properly publish his results and it remained hidden from the public for several years. Then the composition of "nuclein" was discovered by Albrecht Kossel[3, 4] who was working on the hypothesis that "nuclein" was participating in the formation of new biological tissue. A few years later he was able to identified the five nucleobases: Guanine, Adenine, Cytosine, Thymine only in DNA and Uracil only in RNA. He received Nobel Prize in Medicine in 1910 for his seminal work. Later in 1919, Phoebus Levene, a Russian biochemist,

proposed that DNA is composed of four different nucleotides which correspond to a phosphate group link to sugar with one of the four bases attached. This model is very closed to the one we know with ribose (RNA) or deoxyribose (DNA) linked at 5' –OH group to the phosphate group with a phosphoester bond and at 1'-OH to the bases A, G, C and T (DNA) or U (RNA) with an osidic bond[5] **(Figure 1)**.

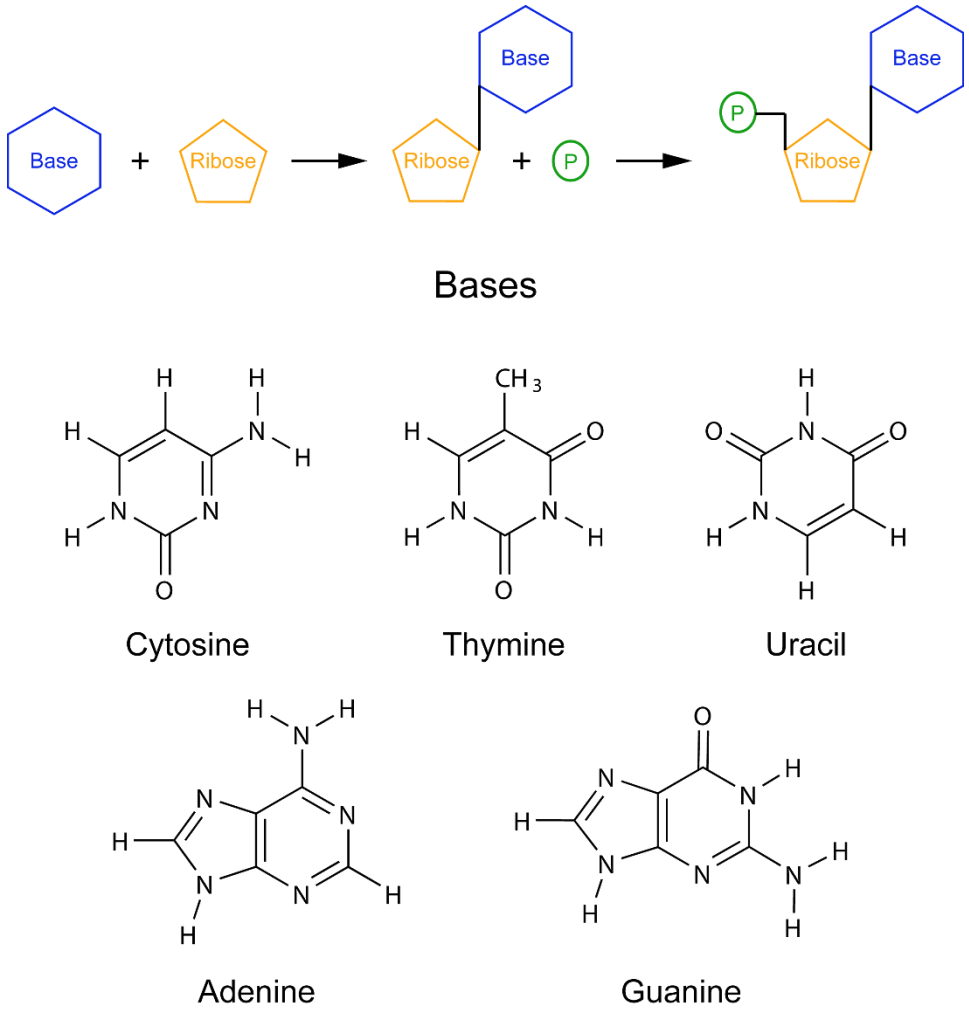


Figure 1. Chemical structure of a nucleotide with the different bases implicated in its composition.

In 1944, a huge discovery that shocked the nucleic acids world was published by Oswald Avery and colleagues [6] in a seminal paper which proved that DNA is the carrier of genetic information. Inspired by this new breakthrough, Erwin Chargaff discovered the "Chargaff rules" which states that in DNA, the amount of purine bases is nearly equal to the amount of pyrimidine bases with amount of A similar to T and amount of G similar to C but at this point he did not know that A is paired to T and G to C[7] **(Figure 2)**.

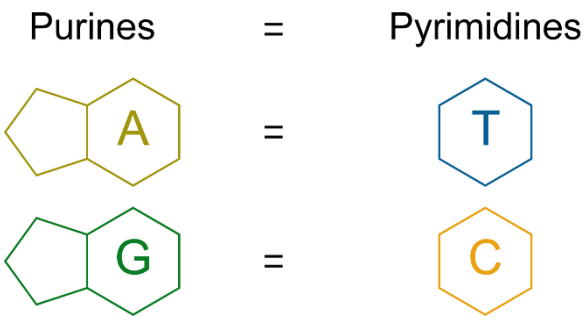


Figure 2. Simple view of Chargaff's rule

Finally, in 1953 Watson and Crick, and also thanks to researchers such as Rosalind Franklin, Maurice Wilkins or Linus Pauling, proposed their model of the DNA double helix [8] which leads to the actual knowledge about DNA **(Figure 3)**.

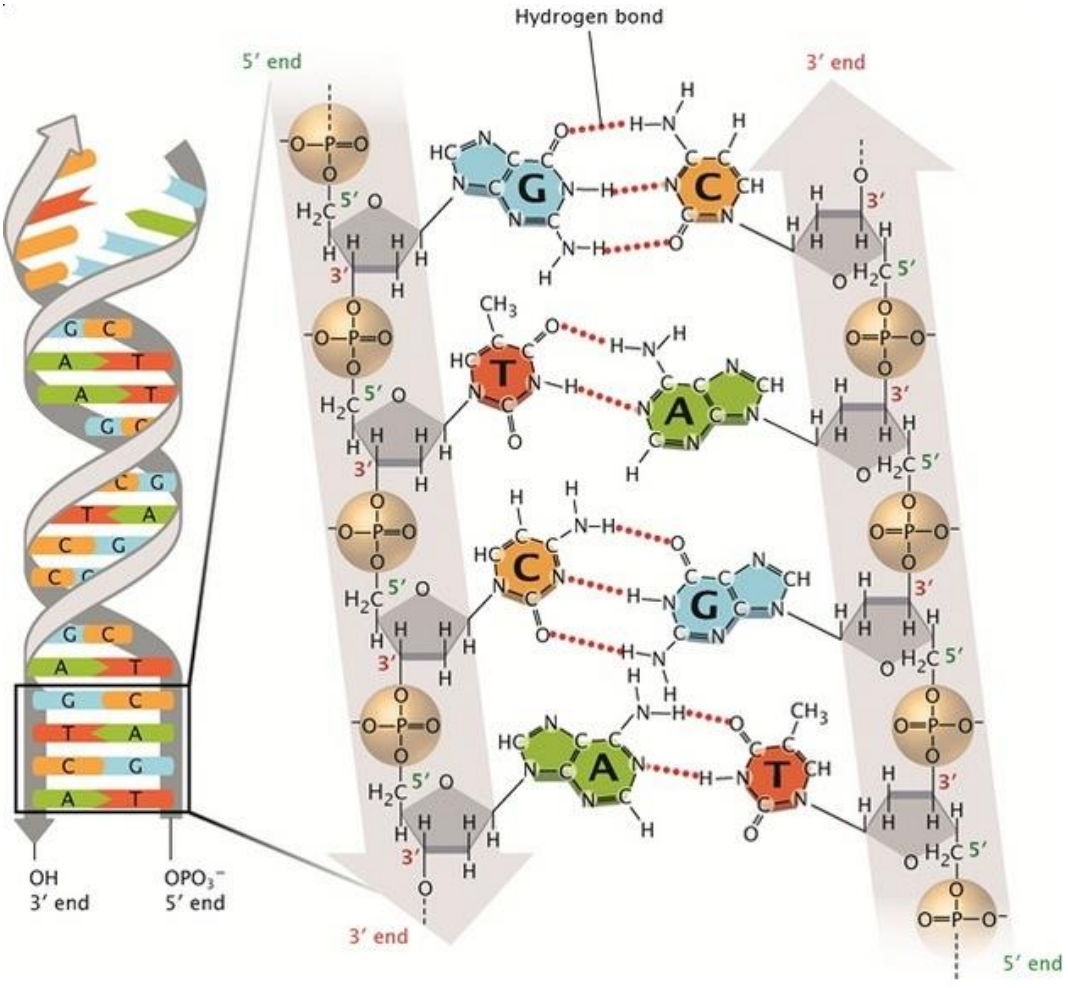


Figure 3. Structure of the double-helical structure of DNA discovered by Watson and Crick

From © 2013 Nature Education.

In DNA structure, the phosphate backbone is outward facing with bases inside the structure and allowing formation of hydrogen bonds. A is complementary to T and they are linked with two hydrogen bonds and C is complementary to G with three hydrogen bonds. In AT base pair, they are formed between N1(A)/O8(T) as donors and N3(T)/N10(A) as acceptors. In GC base pair the donors are O6(G)/N3(C)/O7(C) and the corresponding acceptors are N8(C)/N1(G)/N10(G). The DNA double helix is anti-parallel, which means that the 5' end of one strand is paired with the 3' end of its complementary strand. The model proposed by Watson and Crick was a right handed helix corresponding to canonical DNA form know as B-DNA which is the most common form in living cells. But there are also two other conformations which can be adopted by DNA double helix which are A-DNA also canonical corresponding to a smaller and wider form of the helix due to the fact that it is found in dehydrated environment and Z-DNA which is a non-canonical left handed helix know to be a transient form[9].

I.2. Nucleic acids based unusual structures

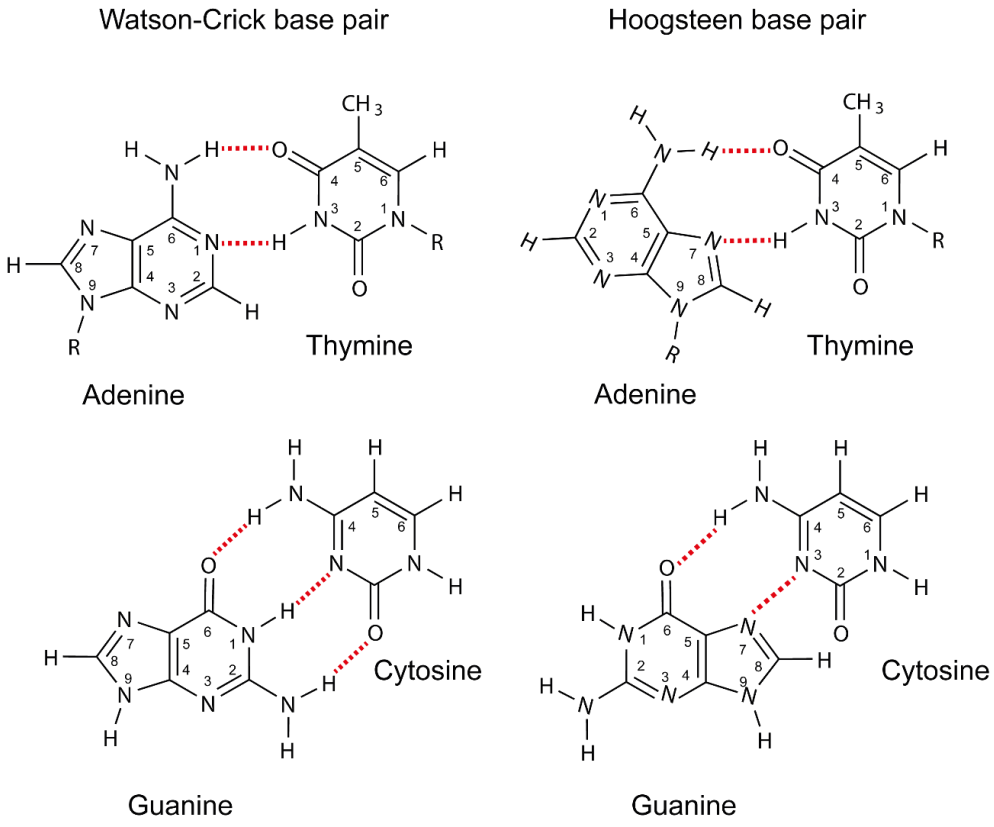


Figure 4. Differences between Watson-Crick base and Hoogsteen base pairing

Nucleic acids structures are commonly represented by a double helix such as the DNA model discovered by Watson and Crick but they can form several other structures. As double-helical structure, secondary structures are formed via hydrogen bonds and also for example stacking interaction with π-electron system of nucleobase aromatic rings with high contribution in terms of energy. Hydrogen bonds implicated in these structures can be Watson-Crick hydrogen bonds corresponding to the base pairing know in their helix structure but also arranged in some different configuration called Hoogsteen hydrogen bonds.

I.2.1. Hoogsteen hydrogen bonds

They have been discovered in 1959 by Karst Hoogsteen who wanted to study the structures of co-crystals containing 9-methyladenine and 1-methylthymine [10]. He noticed some new type of hydrogen bonds between A and T where the adenine was flipped and even if one of the two hydrogen bond was identical to the one describes by Watson and Crick, the other one implicated different atoms on another face of the adenine (between N3 and N7 whereas it should be N3 and N1). Several nucleic acids secondary structures (hairpin, i-motifs, G-quadruplexes…) formations are possible because of the existence of these hydrogen bonds **(Figure 4)**.[11]

I.2.2. Stem-loop structures

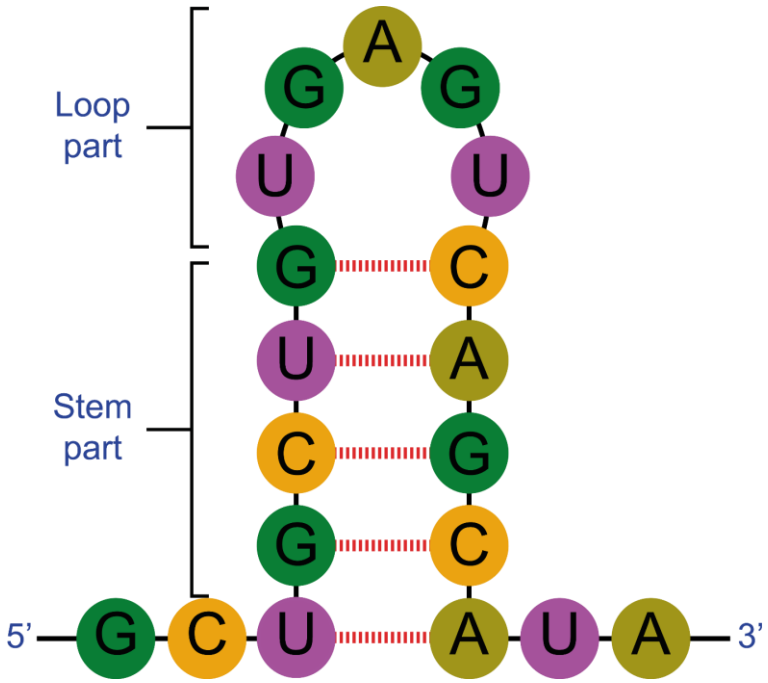


Figure 5. Schematic view of a stem loop structure with the stem and the loop parts (respectively without and with base pairings)

Both DNA and RNA are capable of forming stem-loop structures even if they most of the time found in RNA sequences. Stem-loop which are also called hairpin or hairpin loop is formed in single-strand nucleic acid with a base paired part corresponding to the stem and an unpaired region forming a loop. Stem part can be formed if a part of the strand is complementary to another part in the sequence when read in opposite directions[12] **(Figure 5)**. These structures can be found in ribozymes (RNA enzymes) and messenger RNA having a role in biological processes such as translation[13-15]. Stem-loop structures can be implicated in long range interactions and participate in the diversity of RNA secondary structure by forming pseudoknots and kissing complexes via hydrogen bonds forming between a stem or a loop part from a stem-loop unit to the stem or the loop part of another stem-loop unit. In the case of pseudoknots formation **(Figure 6)**, several conformations are possible but the most common form is called H-type fold where some base of a hairpin loop is going to establish intramolecular hydrogen bonds with other bases outside the stem.

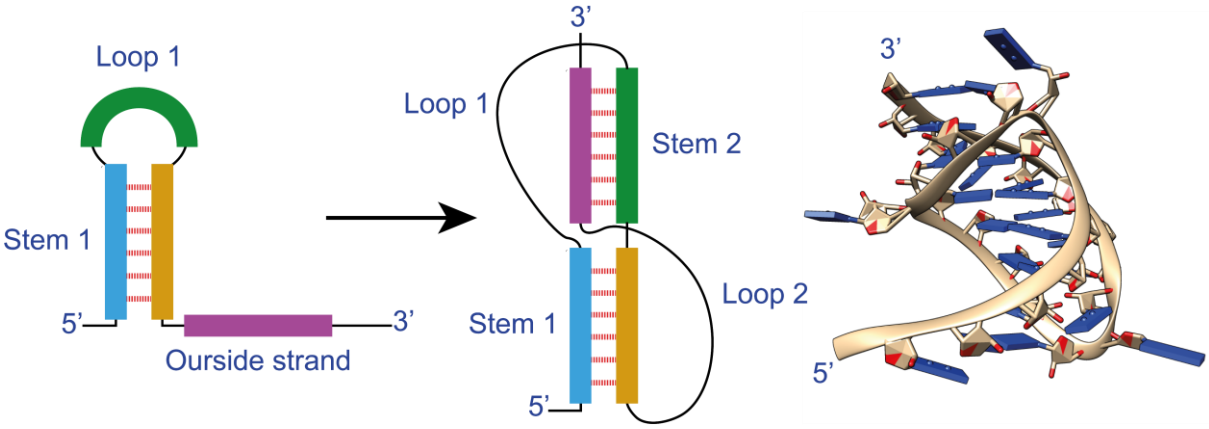


Figure 6. Schematic view of the formation of a H-type pseudoknot with the crystal structure of a viral RNA pseudoknot (code PDB: 1L2X)

This process lead to a pseudoknot with one more stem and one more loop formed[16]. Another kind of RNA-RNA dimer can be formed thanks to tertiary interactions which is called kissing complex **(Figure 7)**. It is the result of the base pairing between the terminal loops of two hairpins. As simple stem-loop unit, pseudoknots and kissing complexes can be implicated in the catalytic core of ribozymes in the translation process but one of their main interesting role is that they can induce a ribosomal frameshifting during translational process especially in virus[17-19]. This phenomenon is possible because of a higher energetic barrier to cross for the translating ribosome due to their structural geometry[20]. In the case of kissing complexes, one

of their most known implication is during genome dimerization in HIV-1[21, 22]. Indeed, for retroviruses, this phenomenon allows them to keep their RNA copies together by their 5' end and overcome immune system defenses [23].

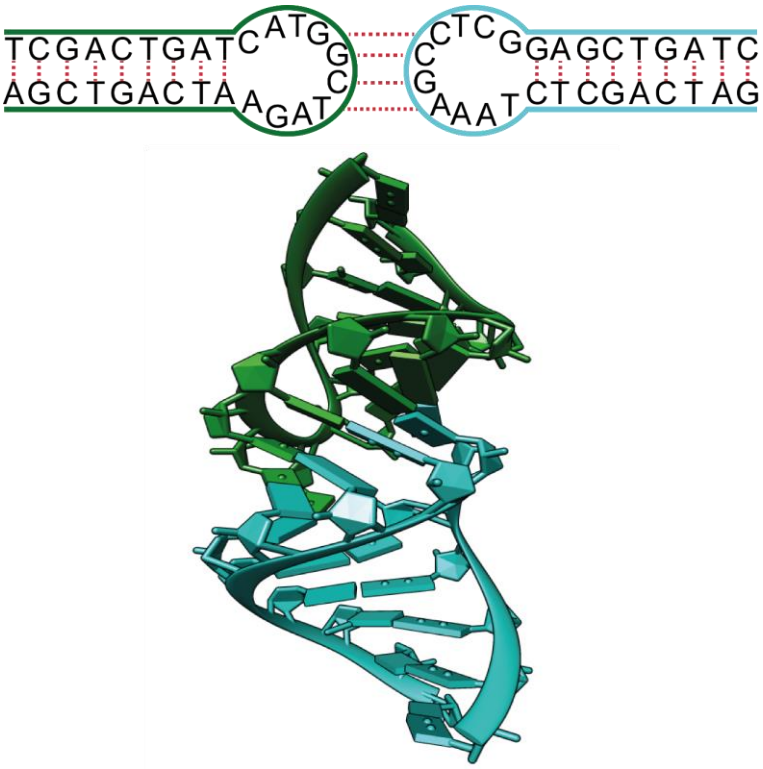


Figure 7. Schematic view of a kissing complex with the crystal structure of an TAR-R06 RNA kissing complex (code PDB:2JLT)

I.2.3. Triplex structure

DNA triplex is one of the most characteristic secondary structure adopted by nucleic acids in which Watson and Crick hydrogen bonds and Hoogsteen hydrogen bonds are involved. Triplex nucleic acids were first discovered in 1957 by Felsenfeld and Rich[24] who found that a complex can be formed between polyuridylic acid and polyadenylic strands if they are put together in a 2:1 ratio. Then in 1986, by designing DNA-binding molecule[25], Dervan discovered a DNA sequence capable of forming a stable specific triple helical DNA complex called triplex forming oligonucleotide (TFO). It binds to the major groove of the DNA duplex and then can form a DNA triplex if there are the needed requirements such as physiological conditions and especially Mg^{2+} salt concentration[26]. TFO should be complementary to one of the two DNA strand but unlike DNA duplex it is not mandatory to have antiparallel formation. DNA triplexes can be intermolecular or intramolecular depending of the TFO's nature **(Figure 8A)**. If TFO

comes from a second DNA molecule, then an intermolecular triplex can be formed. But the third strand can be provided by one of the strands of the same duplex if there is some mirror repeat sequence and the resulting triplex will be intermolecular and is known as "H-DNA"[27, 28]. Depending on the bases implicated in the triad formation (from the TFO and from the duplex), two types of DNA triplexes can be formed: parallel and anti-parallel[29]. Generally, parallel triplexes would form in case of pyrimidine rich triad such as TAT or CGC where the TFO bring pyrimidine base and anti-parallel would form in case of purine rich triad such as GGC or AAT where base from TFO Is purine **(Figure 8B)**.

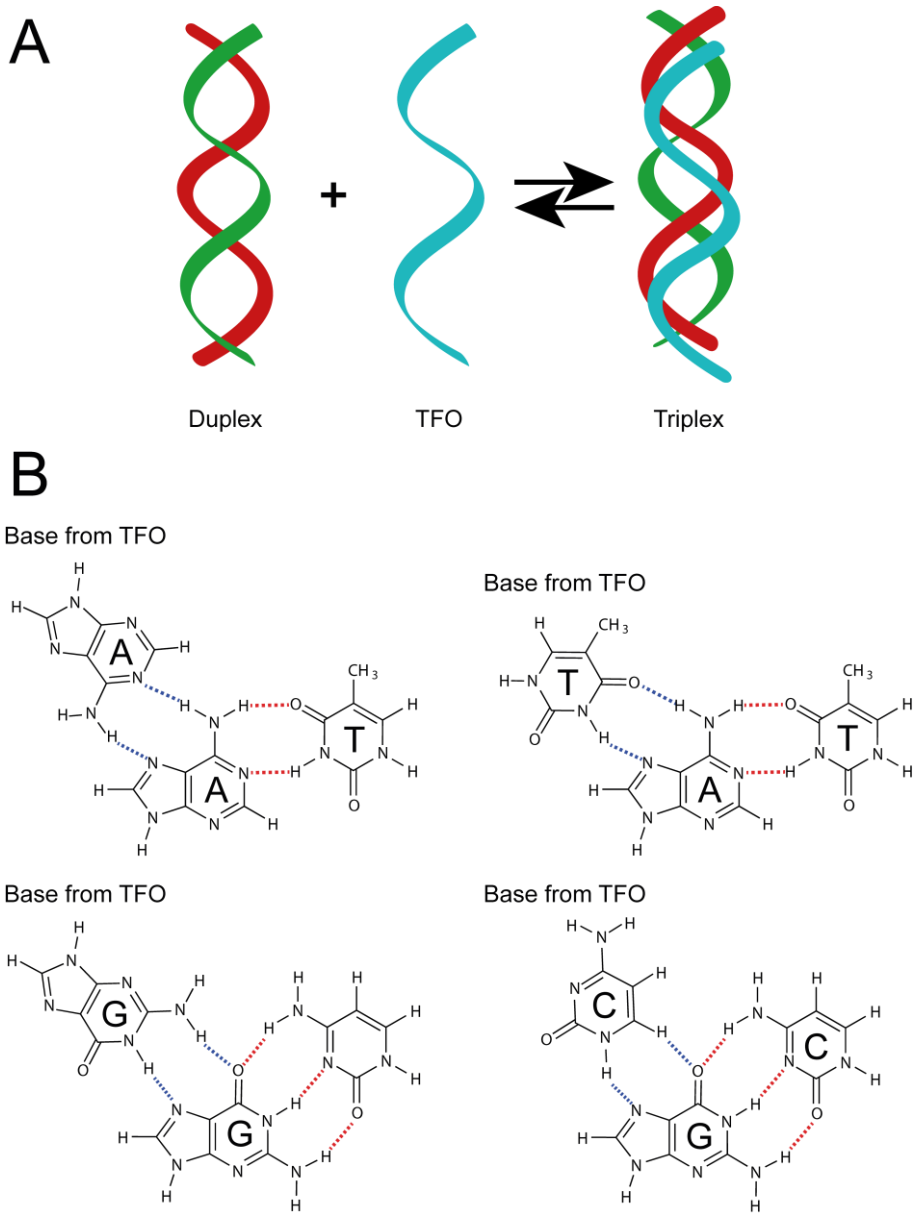


Figure 8. (A) Formation of a triplex from a duplex and an external TFO; and in (B) Representation of the different possible motifs depending on the base coming from TFO (purine motif if A or G; pyrimidine motif if T or C). Watson and Crick (red) and Hoogsteen (blue) base pairings are represented

Several studies about association and dissociation of DNA triplexes showed that the association can simply be described with a simple model Duplex + TFO → Triplex, and is highly depending on pH essentially due to protonation of cytosine which can lead to one-step formation or sequential formation[30]. For dissociation, a model was proposed with two different steps with; first a dissociation step of TFO from DNA duplex followed by the dissociation of the duplex : Triple → Duplex + TFO → 2 Single Stand + TFO. Indeed, studies revealed that DNA triplex is thermodynamically less stable than bound DNA duplex[31]. DNA triplexes are not only some kind of unusual DNA secondary structure because they also have some important implication in biological process especially because over 2 million TFO binding sites have been identified in mammalian genome[32, 33].

I.2.4. I-motif structure

The i-motif also called i-tetraplex is a four-stranded structure that can be formed within C-rich strand with intercalated C-C base pairs at slightly acidic pH. It has been first described in 1993 by Guéron and co-workers[34] but formation of C-C pairs base in cytosine-rich polymer was already demonstrated more than 50 years ago [35, 36]. However, i-motif is the only know nucleic acid structure which can be formed via a system of intercalated base pairs. This structure is formed with two parallel duplexes linked via hemi-protonated cytosine-cytosine base pairs that are intercalated in an antiparallel orientation[37, 38] **(Figure 9)**.

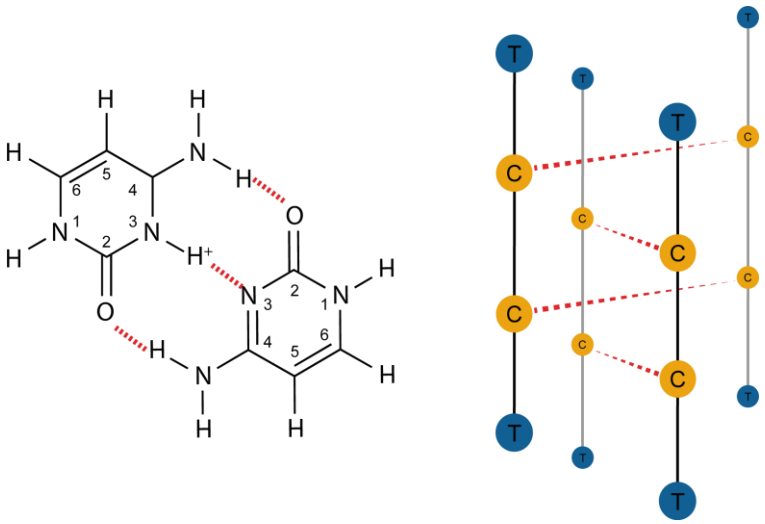


Figure 9. C-C base pairing with the protonated cytosine and schematic representation of the i-motif intercalation with TCCT as a simple model

Formation of i-motif depends on several factors such as the number of implicated intercalated cytosine [39], loop length [40], or environmental condition. Indeed, it has been shown that i-motif can even be formed at neutral pH conditions [41] and supports the idea that i-motif are dynamic structure over a wide range of pH from a folded structure to a more disorganized conformation especially at higher pH values. Moreover, i-motif can be formed in a tetramer with four strand containing at least a stretch of cytosine[42], in a dimer of a DNA strand containing two cytosine stretches [43] or even in a monomer by forming an intramolecular i-motif with a single strand of four cytosine stretches [44] **(Figure 10)**.

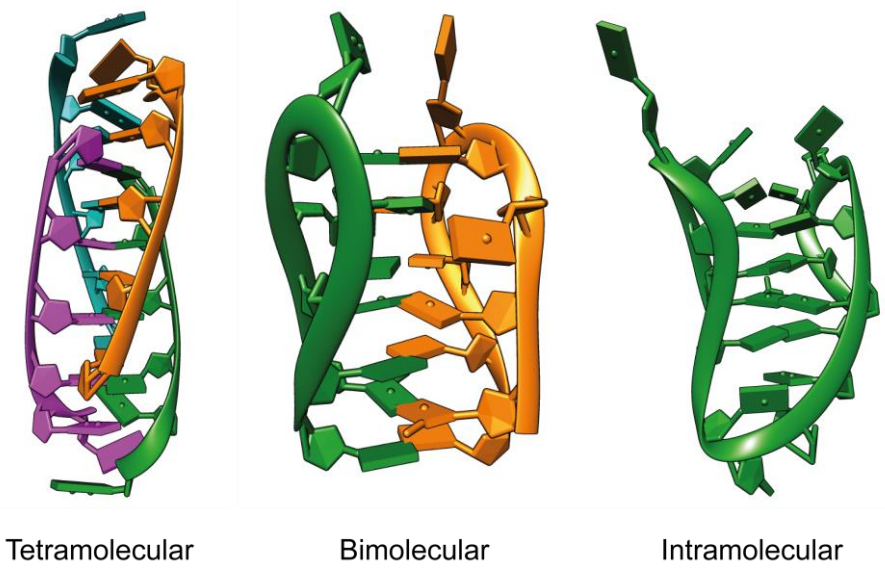


Figure 10. Examples of i-motifs with four, two and one strand. Tetramolecular i-motif is the NMR structure of 5' AACCCC 3' (PDB: 1YBL); bimolecular i-motif is the NMR structure of the dimer 5' CCTCACTCC 3' (PDB: 2AWV) and intramolecular i-motif corresponds to the NMR structure of 5' CCTTTCCTTTACCTTTCC 3'

The i-motif arrangement is quite stable and generally shows a high melting temperature in acidic conditions, nevertheless stability is highly dependent on the conditions especially pH conditions. Concerning kinetics, i-motif often show a very slow association and dissociation constants depending again on the conditions and the sequence[45]. Even if the role of i-motif remains unclear, there a lot of C-rich sequences such as promoter regions and human telomeric DNA which could be capable of forming i-motif[46]. Other C-rich regions can be found in minisatellite such as human insulin minisatellite or in the human centromeric α satellite[47, 48]. It has also been reported that some proteins such as the heterogeneous nuclear

ribonucleoprotein K can bind to the C-rich strand in c-myc NHE (Nuclease Hypertensive Element) promoter region[49].

I.2.5. G-quadruplex structure

I.2.5.1. From a nuisance to a molecule of interest

G-quadruplexes also known as G4 are nucleic acids secondary structures that can be formed within DNA or RNA guanine rich strands. Interestingly, their history is older than the double helix discovered by Watson and Crick in 1953. In 1910, Bang[50] discovered the that guanylic acid is capable of forming a gel at high concentrations suggesting that G-rich sequences could form auto associate and form high order structures. Around fifty years later, Gellert and colleagues[51] used X-ray diffraction and found that guanylic acids can assemble into tetrameric structures. After several studies[52, 53], the G-quartet (or G-tetrad) **(Figure 11)** was finally described as the basic structural motif of G-quadruplexes composed of four guanines associated in a planar square with Hoogsteen hydrogen bonds. Each guanine is associated to the two adjacent ones. Stacked G-quartets lead to the formation of a G-quadruplex. For years, even after being described, G-quadruplexes were seen as a nuisance first and then as an *in vitro* artefact until studies show that G-rich sequences at the ends of human telomeric DNA were capable of forming this kind of structure.

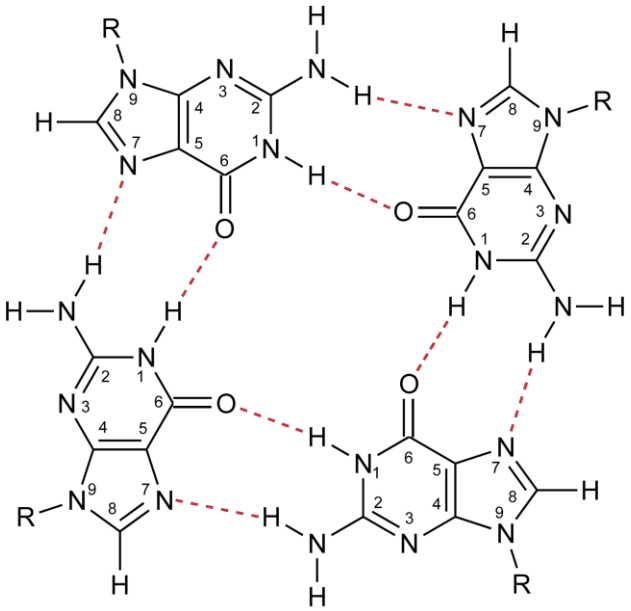


Figure 11. Representation of the G-quartet with the four guanines linked in a Hoogsteen hydrogen bonds network (red)

I.2.5.2. G-quadruplexes polymorphism

G-quadruplexes cannot be described as only a simple four-stranded structure. Indeed, G-quadruplexes are well known to be highly polymorphic structures. This polymorphism has several explanations such as the sequence[54], the number of strands implicated in the formation and their orientation[55], the loops size, orientation and connectivity[56-58], the *syn* or *anti* conformation of the guanines[59], the nature of the stabilizing cation[60-62] and some other factors depending on the environment.

Number of strands and orientation **(Figure 12)**

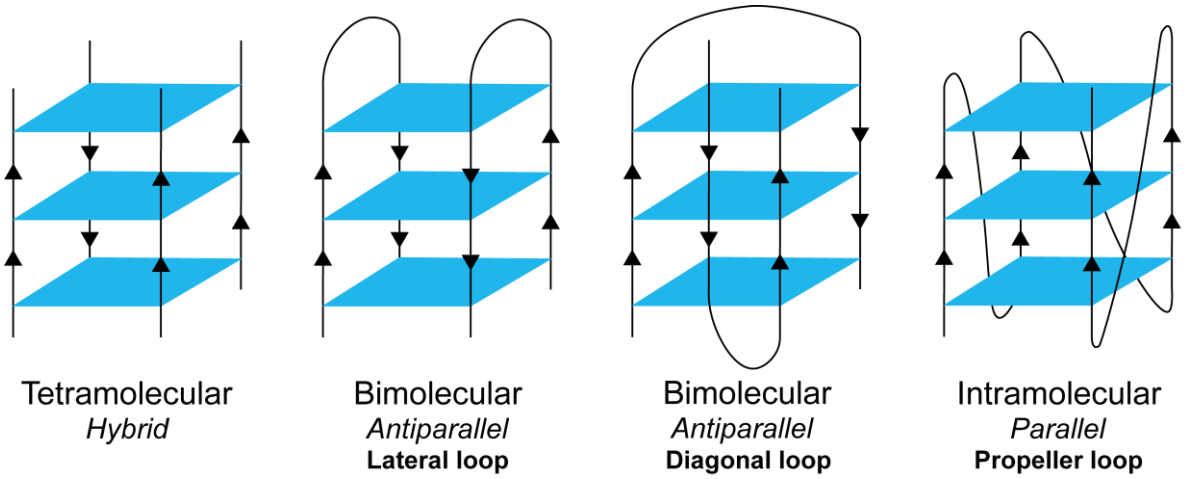


Figure 12. Examples of different G-quadruplexes depending on the number of implicated strands, their orientation and their connectivity (loops)

G-quadruplexes formed with only one strand are called unimolecular or intramolecular G-quadruplexes. They can be simply described with this kind of model $[G_nL_xG_nL_xGnL_xG_n]$ where G_n is the number of guanines in the tract that can be implicated in the G-tetrad and L_x is the number of residues which are implicated in the loops and can also be guanines if they are not part of a tetrad. Generally, G_n will contain at least three guanines because even if G-quadruplexes with only two stacks are possible, they are less stable[63]. G-quadruplexes can also be formed with several strands with bimolecular G-quadruplex (two strands) or tetrameric G-quadruplex (four strands). The strand sequence is not necessary the same and could be different but in most of the cases, sequences are identical. Among this three types of G-quadruplexes, the most polymorphic one is the bimolecular because the association of two strands increase topology variation. Recently an example of a triple stranded G-

quadruplex has been published[64] showing that polymorphism due to the number of strand in the G-quadruplex formation is not totally unsolved. In addition to the number of strands, polymorphism is also due to the orientation of these strands. Indeed, G-quadruplexes can be divided into three different groups[65]. When all strands are oriented in the same direction, G-quadruplex is called parallel[66]. If there are two strands in a direction and the two other in the opposite direction, then the G-quadruplex is anti-parallel. In the case of anti-parallel conformation, the two strands in the same direction can be on the same side in a "chair-like"[67] conformation or on opposite sides in a "basket-type"[68] conformation. Finally, the last case corresponds to only one strand in the opposite direction compared to the others strands and the G-quadruplex is called hydrid [69] also knowns as "3+1" mixed conformation.

Loops diversity and bulges **(Figure 12)**

Loops diversity also participates to increase G-quadruplex polymorphism. Loops can only be formed in bimolecular and intermolecular G-quadruplexes because there is no connection between strands in tetramolecular G-quadruplexes. Three different types of loops can be distinguished in G-quadruplexes[70] . First, the propeller type[71, 72] which corresponds to a loop linking adjacent parallel strands from a bottom G-tetrad to a top G-tetrad. This kind of loop can be formed in all type of G-quadruplexes. The two other types of loop that can exist only in anti-parallel and hybrid G-quadruplexes are known as lateral and diagonal loops. Lateral loops[73] connect two anti-parallel strands on the same side and can be separated in two types: "head-to-head" and "head-to-tail"[74, 75]. "Head" and "Tail" refer to the opposite faces of a guanine base. Diagonal loops[76] joins also two anti-parallel strands but on opposite sides. Another type of connectivity between guanines of tetrad which participate to G-quadruplexes polymorphism but not considered as loop is bulge[77, 78]. Bulges consist in, like loops, connecting guanines of G-tetrad but unlike loops which connect guanines of opposite sides, bulges connect two adjacent guanines of the G-core belonging to the same strand.

Guanine glycosidic conformation

In G-quadruplexes, guanine cap adopts two different conformations depending on the glycosidic angle χ: *anti* and *syn* [79] **(Figure 13)**. The glycosidic bond links the sugar and the base (A, T, C, G, U if RNA) between the C1' and N9 in the case of purine or N1 in the case of pyrimidines. The corresponding torsion angle χ may be really different depending on the

rotation of this bond. Torsion angle χ can be described as the angle between O4'–C1'–N9–C4 for purines and O4'–C1'–N1–C2 for pyrimidines. If χ is comprised between -120 and 180° then the corresponding conformation is *anti* whereas *syn* correspond to a torsion angle between 0 and 90°.

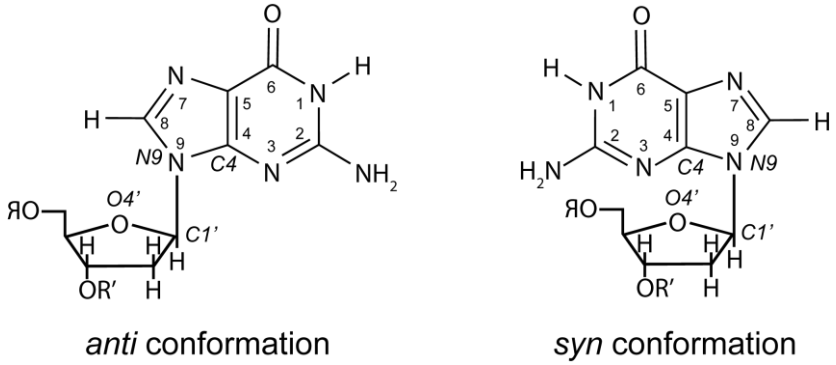


Figure 13. Representation of anti and syn conformations for guanosine (purines). For pyrimidines N9 and C4 are respectively N1 and C2

Parallel G-quadruplexes have all guanine in *anti* conformation but anti-parallel G-quadruplexes can both *anti* and *syn* conformation. There is no rule to predict if it would have formed one or the other conformation because glycosidic conformation will be unique for a given topology. The only exception concern RNA G-quadruplexes which generally have anti oriented guanines and harbor a parallel conformation due to the 2'-hydroxyl group[80, 81].

Nature of the cation

As described before, G-quadruplexes can be stabilized by monovalent cations and their role in G-quadruplexes stability and polymorphism have been deeply studied. Cations can contribute to neutralize the global negative charge of DNA but in the case of the G-quadruplexes their role become central[82, 83]. Indeed, during G-quadruplexes formation, the implication and orientation of guanines in the tetrad allow the creation of a central channel which is highly negatively charged. Due to the orientation of the O6, a strong negative electrostatic potential is created inside the channel which needs to be compensate by cations positive charge. Coordination by monovalent cations is possible due to the geometric arrangement of the guanine O6 pairs of electron. As the most biological relevant cations are K^+ and Na^+, they are the most known and studied cations in G-quadruplexes formation but there are several other cations, monovalent or divalent, that can stabilize G-quadruplexes,

sometimes in a better way. As a result, cations have been identified depending on their stabilizations potential: K^+, Na^+, Rb^+,NH_3^+, Ca^2+, Mg^{2+}, Pb^{2+}[84-87] for the most known of them, K^+ being the one with the higher stabilization effect[88]. Depending on the nature of the cations, position within the G-quadruplex channel can be different. For example, Na+, which is smaller compared to K+ can be in the same plane as guanines O6 in some case or it can adopt a configuration similar to K+. Potassium cation which is larger than sodium does not fit in a planar configuration and has to be equidistant between each tetrad plane be in a tetragonal bipyramidal electrostatic configuration. The nature of cations can be very important in G-quadruplexes polymorphism because in some cases, with the same sequence, different cations can lead to the formation of different conformations, such as in the case of human telomeric sequence 5' A(GGGTTA)3GGG 3'. In presence of Na^+ it forms an antiparallel G-quadruplex[89] but with K^+ it forms a parallel structure[90].

<u>High order G-quadruplexes structures</u>

Even of most of the studies have been focused on monomeric structures, G-quadruplexes can adopt a variety of multimeric forms from the smallest represented by a dimer to very large structures containing hundreds of G-quadruplexes monomers called G-wires[91, 92]. Regarding the potential high G-quadruplexes concentration availability in eukaryotes, formation of G-wires structures is particularly interesting. Different types of multimeric structures can be distinguished to monomeric G-quadruplexes[93, 94]: intramolecular or intermolecular; parallel, antiparallel or hybrid based structures depending on the conformation of G-quadruplex units. Moreover, each unit can be stacked in three different orientations: 5' to 3' "head-to-tail"[95], 5' to 5' "head-to-head"[96] or 3' to 3' "tail-to-tail"[97]. Formation of these higher structures is essentially possible thanks to the π-π stacking of the terminal faces of each G-quadruplex unit[60]. An example of this kind of structure has been described by Neidle and coworkers[98] who described consecutive parallel G-quadruplexes connected by a TTA loop in a "head-to-tail" configuration.

I.2.5.3. G4-motif localization

In vitro, G-quadruplexes are considered as highly stable structures with melting temperature much higher than that of duplex DNA[99, 100]. Due to their stability, their existence *in vivo* is still a critical point in G-quadruplexes studies. However, computational studies revealed more

than 375,000 G-quadruplexes forming motifs in the human genome[101, 102] suggesting that they may form *in vivo*. Moreover, these motifs are not randomly located within the genome but are concentrated in several regulation key regions[103-105]. G-quadruplexes motifs are essentially present at the end of the chromosome telomeres[106], in promoter regions[107], within UTRs (Untranslated Regions) of messenger RNA[108] and in several other regions such as mitotic and meiotic DSB (Double Strand Break) sites. All these localisations suggest that G-quadruplexes could have a biological role in several biological processes such as replication or transcription acting as a regulator. Several other studies have been performed in other organism especially in *Saccharomyces cerevisiae*[109, 110] where more than 1,400 G-quadruplexes forming motifs have been identified which remain quite high regarding the entire genome of *Saccharomyces cerevisiae*. Even if G-quadruplexes forming sequences have been identified in all over the genome, two localisations have been particularly studied: telomeres and gene promoter's regions.

G4 in telomeres

Telomeres are specialized DNA nucleoprotein complexes that cap the end of chromosomes[111-113]. Their primary role is to protect against gene erosion at cell divisions and attacks from nuclease[114]. For human (and vertebrate) telomeres consist of a tandem repeat of the hexanucleotide 5' $d(TTAGGG)_n$ 3' with a length of 5 to 10 kb terminating with a G-rich single strand capable of forming G-quadruplexes[115, 116]. In normal cells, each cell replication leads to the loss of 50 to 200 base in telomeres. This phenomenon is related with cancer, aging and genetic stability as the loss of bases directly corresponds to an information loss[117, 118]. In somatic cells, each cell replication cycle leads to the loss of 50 to 200 base in telomeres. After a critical shortening of the telomeres the cell undergoes apoptosis[119]. In most organisms, the telomerase, a reverse transcriptase enzyme, is capable of increase the length of telomere by using its RNA subunit as a template[120]. Telomerase is inactive is most somatic cells except when its functions are needed but it is upregulated in most cancer cells leading to immortal cancer cells[121]. It has been shown that the G-quadruplexes formation could influence the telomerase activity with different effects depending on the conformation: intramolecular antiparallel G-quadruplexes block telomerase whereas intermolecular parallel ones could favor telomerase activity[122, 123]. Moreover, in human, telomeres are associated with several related proteins such as Pot1 (protection-of-telomeres 1) which has been shown to avoid G-

quadruplexes formation by interacting and trapping the G-rich single strand[124, 125]. As a consequence, Pot1 can enhance telomerase activity. As the potential role of G-quadruplexes in telomerase activity is known for almost 40 years, several different of studies employing different methods have been performed in order to determine the structure adopted by telomeric G-quadruplexes **(Figure 14)**[126-128].

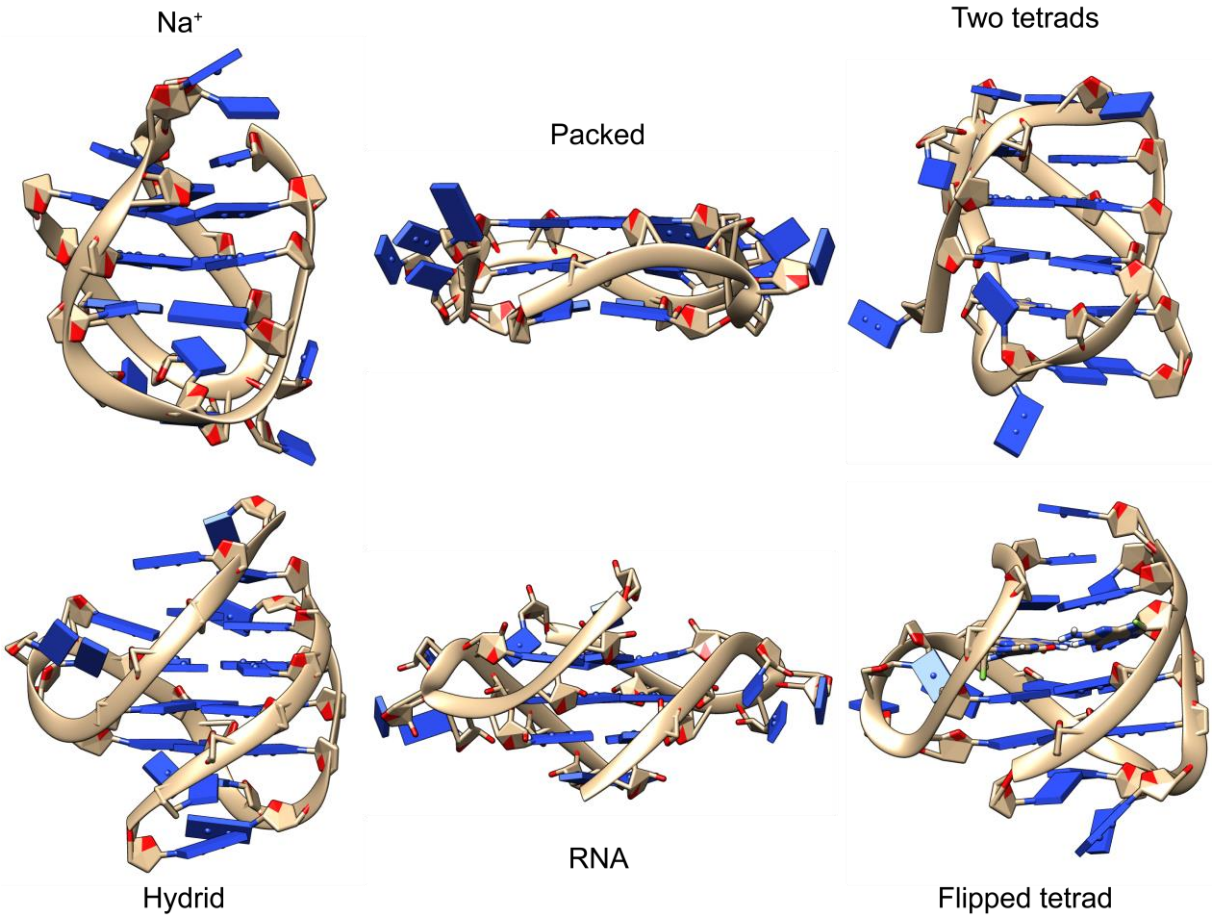


Figure 14. Several examples of telomeric G-quadruplex structures
PDB codes are (from left to right and from up to bottom): 143D, 1KF1, 2KKA, 2HY9, 3IBK and 5MBR.

Surprisingly, the variety of methods and conditions tested produced a variety of structures. The telomeric sequence can form different G-quadruplexes much depending on the salt composition. In Na^+ buffer seems to form parallel G-quadruplexes or hybrid G-quadruplexes even if the predominant form seems to be the hydrid one. The parallel conformation may have been induced by crystallization condition.

<u>G4 in promoter regions</u>

Thus, G-quadruplexes forming motifs are particularly present in promoter regions and suggest a potential function in transcription. Indeed, at least one G-quadruplexes motif have been found near (1000 nucleotides upstream) TSS (Transcription Starting Sequence) in more than 50% of the human genes[129]. This phenomenon is quite conserved across species because it can also be observed in yeast, plants or bacteria[130, 131]. During transcription process, there is only one strand used as a template and the other is the non-template strand. Depending on the situation, G-quadruplexes motifs can be located in one or in the other strand and two hypotheses can be proposed about the role of these structures. The presence of G-quadruplexes within the sequence can be compared to supercoiling[132] which is a well-known phenomenon that is thought to affect transcription[133]. Supercoiling can have both positive and negative effects on transcription that is why two hypotheses are possible **(Figure 15)**.

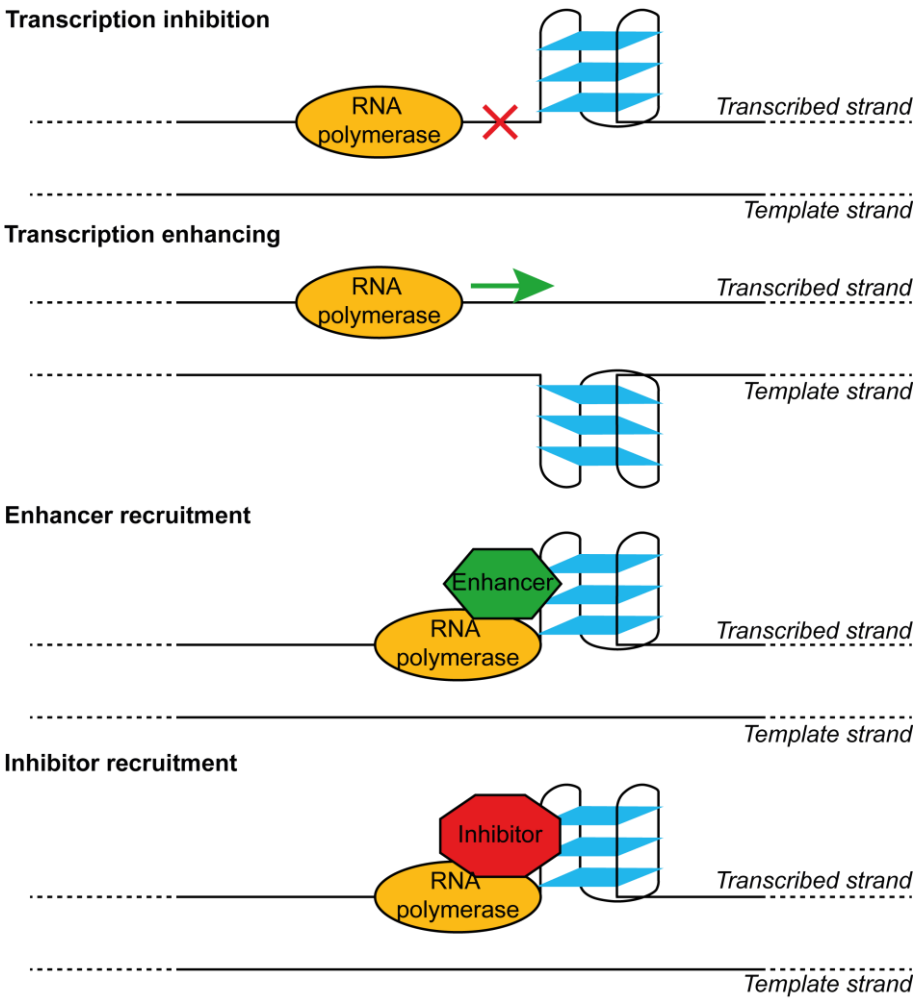


Figure 15. Hypothesizes about the putative role of G-quadruplexes during transcription in the promoter region

In the case where the G-quadruplex motif is located on the template strand, it could inhibit the transcription by acting as a roadblock and as a result disturb the transcription machinery. But, if the G-quadruplex motif is located on the other strand (non-template), it could enhance the transcription by maintaining the transcribed strand in a single-strand conformation. The main criticism of this model is that the G-quadruplex formation is too slow and the stability of these structures is too high to be considered as regulation elements. In addition to the two previous hypotheses, transcription may also be altered by protein binding to G-quadruplexes especially transcription-related proteins such as transcriptional enhancers or repressors[134]. These proteins could affect formation or unfolding of quadruplexes. If these proteins could act as chaperones[135, 136], then the time scale for G-quadruplexes formation could be adapted for a regulatory role. Several transcription-related proteins are known to bind to G-quadruplexes[137, 138] and make their role in transcription more relevant.

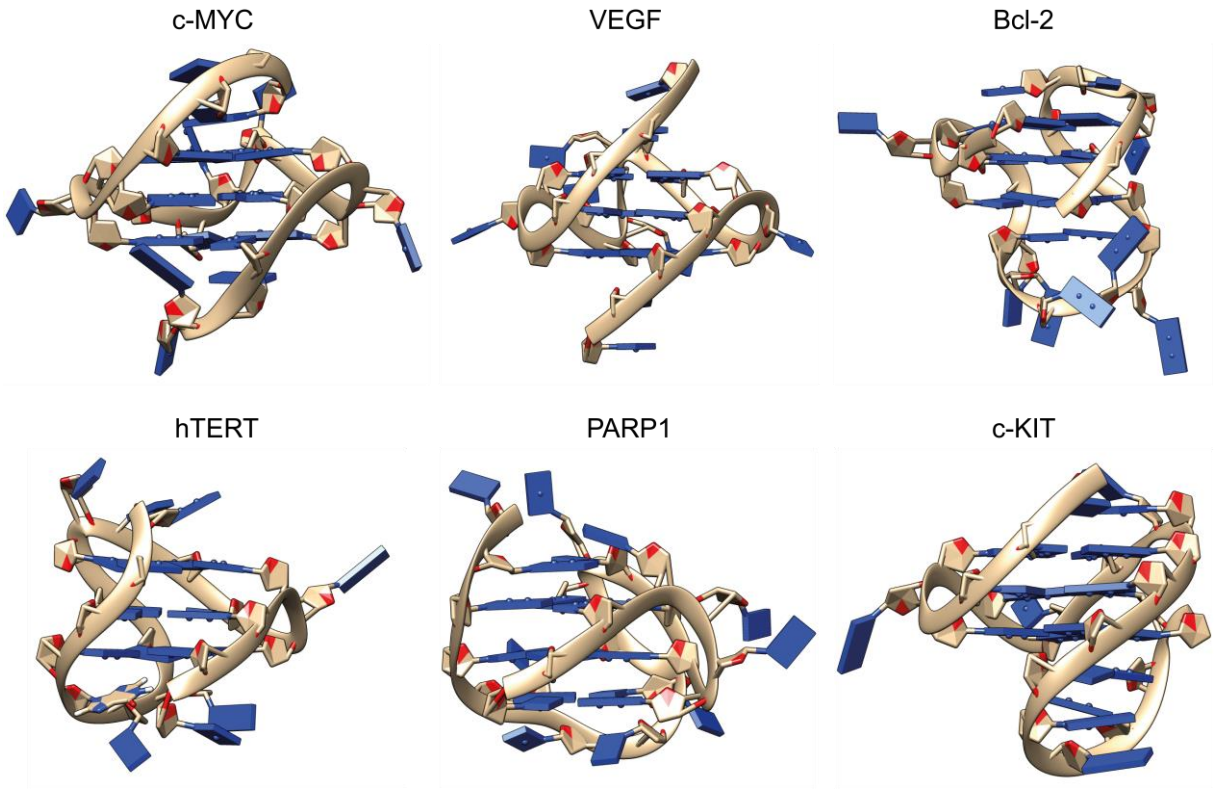


Figure 16. G-quadruplex structures from several oncogene promoter regions
PDB codes are (from left to right and from up to bottom): 1XAV, 2M27, 2F8U, 2KZD, 6AC7 and 2O3M.

One of the G-quadruplexes most exhaustively studied is the mammalian MYC (or c-MYC) oncogene which codes for a transcription factor associated with cell proliferation[139]. In more than 80% of human cancer cells, there is increased level of MYC expression[140, 141]. In MYC promoter, there is a sequence called Nuclease Hypersensitive Element III1 (NHE III1) which is essential for MYC transcription and capable of forming G-quadruplexes in vitro. Several structures adopted by sequences within the MYC NHE III1 and within several other oncogenes have already been studied **(Figure 16)**[142-144]. The role of the G-quadruplex has been studied by looking at the levels of expression of the gene with a wild type sequence containing the motif and compared to a mutated sequence without it and the results showed that the motif in NHE III1 represses transcription[145]. Several other oncogenes have been studied such as c-KIT[146], VEGF[147], BCL-2[148] or KRAS with similar results.

I.3. Use in therapeutics

In the last decade's we have seen some major developments in the field of therapeutics such as the development of monoclonal antibodies, immune modulators or replacement enzymes to target over 20,000 "druggable" targets which have been identified[149, 150] both by academia and industrial researchers all over the world. Unfortunately, a lot of those targets remain "undruggable". In order to overpass these limitations new approaches and strategies are necessary. It is without surprise that some actors in the field have turned to nucleic acids as possible targets, to the point of creating a new field by itself. This is due to the "explosion" of fundamental research performed in the field of nucleic acids in the past 30 years. Nucleic acids have many advantages for building artificial nanostructures in "drug Design" concept[151-153]: (1) their secondary structure could be determined; (2) they are easy and rapid to design; (3) their production is less costly involving chemical synthesis without variability of biologics; (4) they are stable and (5) they can be combined to increase drugs flexibility. A large variety of strategies have been developed based on nucleic acids and several applications in therapeutics and biotechnology have emerged with several patents being accepted both in Europe and in the USA.

I.3.1. Examples of nucleic acids applications

I.3.1.1. Antisense Oligonucleotide (ASO)

Antisense Oligonucleotides (ASOs) **(Figure 17)** are generally single-stranded nucleic acids (DNA, RNA or chemical analogs). They are chemically modified to have enhanced stability, affinity and cell penetration augmented and improved specificity to their target in order to use them as drugs. But, as they need to be recognized by cellular mechanisms, their modifications need to be carefully chosen and sometimes restricted which limited the improving of their properties[154]. ASO are usually used to bind pre-messenger RNA in order to modulate the splicing, cause a sequence-specific degradation or even to block polyadenylation to accelerate RNA decay [155, 156].

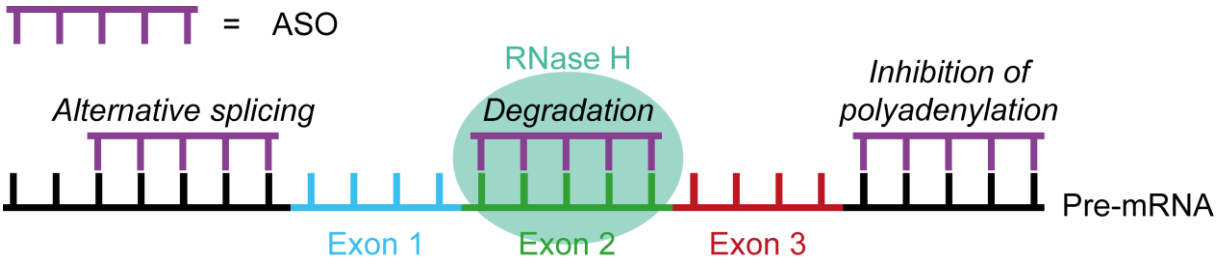


Figure 17. Principle of Antisense Oligonucleotides in therapeutics

Modified from *Nat Struct Mot Biol,* Judy Lieberman*, 2018*

Depending on the purpose of the ASOs, their chemical modifications will be different. For example, for a sequence specific degradation, ASO need to bind to the selected part of pre-messenger RNA and need to be recognized by RNase H which will cleave the targeted sequence. A good ASO candidate would be around 20 nucleotides long (corresponding to the classical length of ASOs) with phosphorothiate linkages between nucleosides to form the backbone. Several bases can be added to the flacking order to increase stability *in vivo* by protecting ASO from exonuclease and to improve binding to pre-mRNA. For slicing modulation or polyadenylation block, it is necessary to avoid degradation by RNase H that is why ASO would contains bases with other modifications. ASO can also be designed to target mRNA in order to inhibit or enhance translation. The main problem with this strategy is the ASO intracellular uptake remaining quite poor that is why only a few drugs antisense-based have been approved by FDA (USA): fomivirsen[157], an antiviral drug briefly marketed before effective anti-HIV drugs emerged and mipomersen[158, 159] which is used to treat familial hypercholesterolemia.

I.3.1.2. siRNAs (short interfering RNA)

Similar to ASO, siRNAs are operating by an antisense mechanism but they differ in their action and outcomes. RNA interference was first discovered in a nematode *Caenorhabditis elegans* where the delivery of a long, double-stranded RNA (dsRNA) silenced a gene expression by the degradation of the messenger RNA[160]. The mechanism was described as the degradation of this long double-stranded RNA into small (around 20 nucleotides) double-stranded RNA called siRNA which is capable of interacting with a multi-protein RNA-Induced Silencing Complex (RISC). Then, in the RISC, siRNA is unwound and the antisense strand binds to mRNA. AGO2, a component of RISC recognizes the complex and cleaves the mRNA around 10 nucleotides downstream from the 5'-end of the antisense strand[161] **(Figure 18)**.

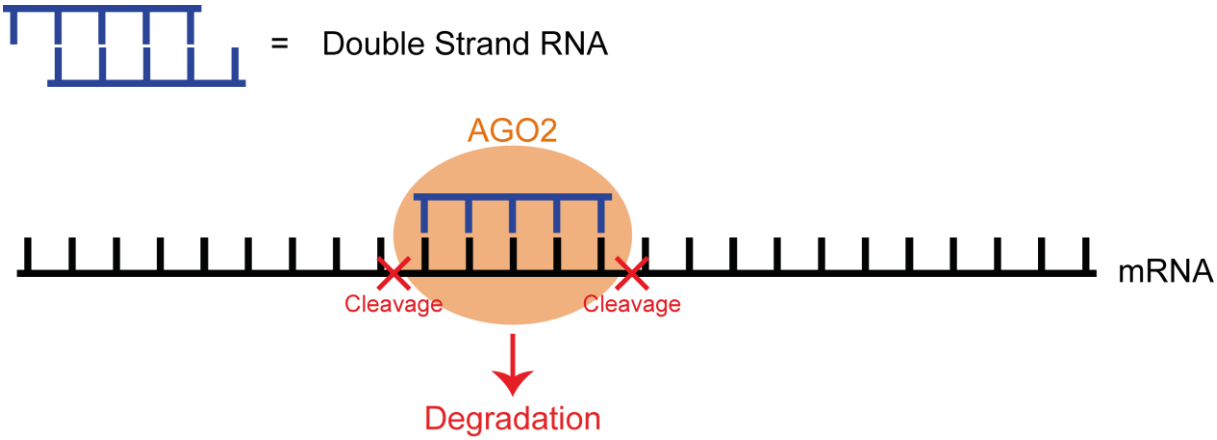


Figure 18. Strategy using siRNA leading to the degradation of messenger RNA (mRNA) Modified from *Nat Struct Mot Biol,* Judy Lieberman*, 2018*

This mechanism was thought not existing within mammalian cells but later studies showed that they just lack the degradation system of dsRNA and synthetic siRNA can be used to interact with RISC and degrade mRNA in mammalian cells[162, 163]. siRNAs were rapidly used in therapeutics with very encouraging results *in vitro* by inhibiting HIV replication[164] by depleting viral genes and *in vivo* by protecting mice from autoimmune hepatitis[165]. The principal problems found with siRNA strategies were similar to those found in ASO applications: they are rapidly degraded and the uptake into the respective targets is quite poor. In addition, unlike ASOs, chemical modifications are more limited and that is why in clinical trials, the development of siRNA strategies are focused on the delivery system part. Two major delivery systems have been studied: (1) using lipid nanoparticles to introduce siRNAs and (2) by conjugating siRNAs to trivalent N-acetylgalactosamine (GalNac)[166, 167]. These two systems target the liver which is the primary sites of several circulating proteins.

The next generation of siRNA delivery system is focusing on the second system as the lipid nanoparticles can be difficult and the system has side effects due to the immune system response.

I.3.1.3. CRISPR-Cas9 system

CRISPR is a defensive mechanism adopted by bacteria against viral infections[168, 169]. CRISPR is the abbreviation for Clustered Regularly Interspaced Short Palindromic Repeats. These clusters are separated by non-repeated DNA sequences corresponding to copies of foreign DNA encountered by the bacteria which are called CRISPR arrays and act as memory. When encountering a foreign genetic element again, bacteria produced RNA segments from CRISPR arrays to target directly pathogen and then use adaptive immune response thanks to nucleases coded by CIRSPR-associated (Cas) genes. CRISPR cluster has been first reported in 1987 by Yoshizumi Ishino[170], biotechnological application remained unexploited until 2013 when CRISPR-Cas9 was formerly applied for human gene editing[171]. CRISPR/Cas system has been classified into six main types (I-VI) and two main classes (1 and 2). Cas9 nuclease belongs to type II and class 2. By designing a guide RNA (sgRNA), it is possible to target genes and having an excised site at the correct location in the gene and then by triggering a cellular DNA repair mechanism having a precise sequence alteration.

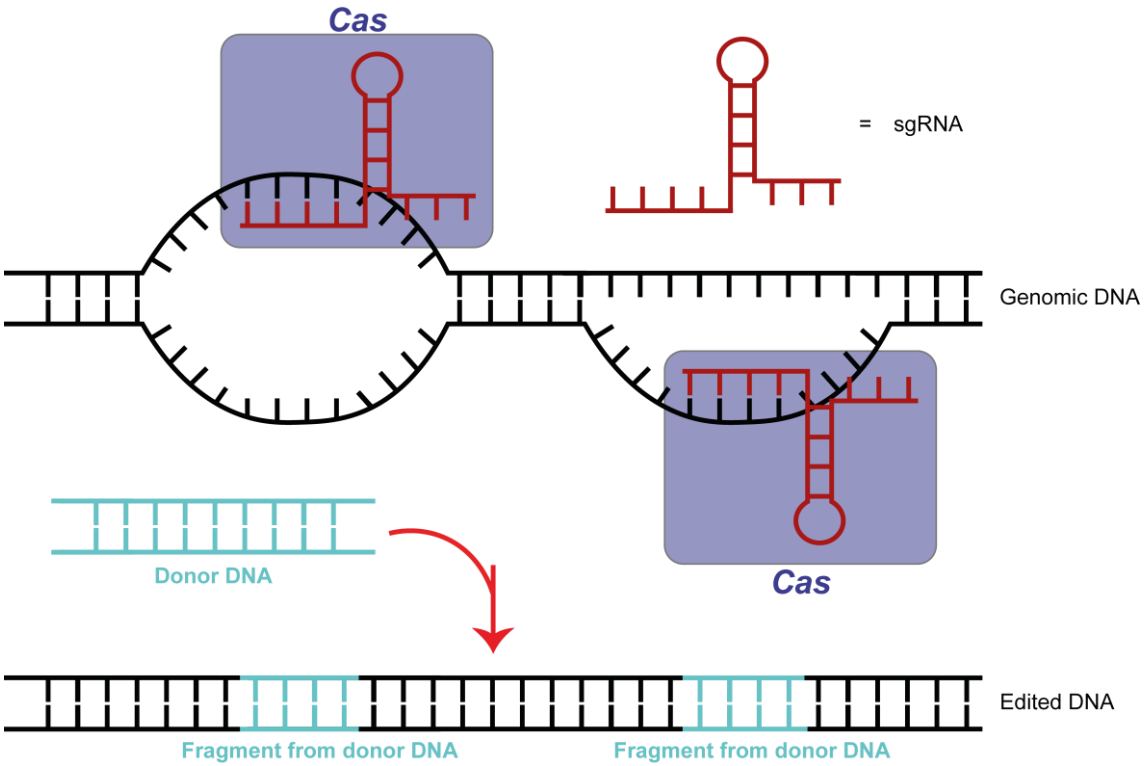


Figure 19. CRISPR-Cas9 technology principle leading to DNA editing

Modified from *Nat Struct Mot Biol,* Judy Lieberman*, 2018*

Once again, CRISPR-Cas9 technology **(Figure 19)** relies on the antisense pairing of sgRNAs but on DNA (chromosomal) instead of RNA for ASO and siRNA. CRISPR/Cas9 system lead to gene editing field which could allow to correct mutations in coding or regulatory regions of genes in therapeutics[172, 173]. However, as CRISPR-Cas9 modifies the genome, ethical and safety concerns remain a major problem depending in the effects of direct mutation in the genome. In order to use CRISPR-Cas9 efficiently, a sgRNA need to be designed to target the selected sequence and then add in the Cas9 plasmid construction. CRISPR system including sgRNA and Cas9 need to be delivered at a chosen target that is why delivery system have to be developed to allow gene editing only where it is needed. This technology remains very interesting especially in the case of cancers which are genetic diseases and progress for the last 5 years show that it can be rapidly developed.

I.3.1.4. Triplex DNA

Secondary nucleic acid structures also have applications in therapeutics. Triplexes DNA for example and more precisely intermolecular triplexes (with external TFO as described before) attracted attention because of their potential role in inhibiting gene expression[174]. Thus, they can be implicated in therapeutics concerning cancer or other genetic diseases. The idea is to design a TFO sequence specific to the target to allow the formation of triplex DNA leading to gene inactivation, stimulating DNA repair or homologous recombination. TFO represents a good candidate to bond to duplex DNA with high affinity and specificity as it can be perfectly complementary to the targeted strand and, similar to ASO, it can be chemically modified or link to a delivery cargo[175]. An advantage of using DNA rather than RNA is that there are fewer copies to target. Moreover, unlike ASO and siRNA, it is possible with TFO and triplexes formation to mutate or inactivate gene. Different approaches can be considered with triplex formation. The first application for DNA triplexes in therapeutics is the inhibition of transcription which could be very interesting in the case of oncogene **(Figure 20)**. This has been shown two decades ago by Cooney and colleagues[176] by using a TFO targeting c-MYC promoter region and since a lot of studies have been performed with several other targets[32]. Another way to use triplex DNA is for DNA repair. Indeed, thanks to its properties, it is possible to design a TFO that can target damage DNA and then the triplex formation will induce DNA repair to restore the sequence.

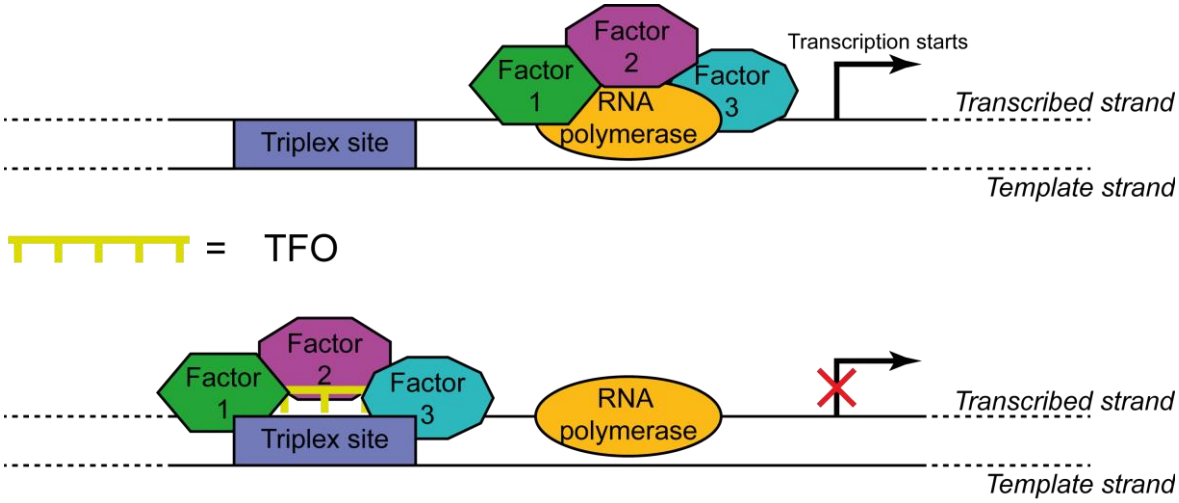


Figure 20. Schematic view of the putative role of triplex formation in transcription

This approach has been applied to c-MYC gene[177] in presence of gemcitabine, an antitumor antimetabolite, and the results were encouraging to inhibit growth of human breast cells. Finally, the last strategy is to simply use TFO as a probe to deliver drug to its target[178, 179]. The TFO has a high affinity for DNA and can be chemically modified to be linked to the drug and to increase stability, specificity or uptake. Limitations for this strategy remains quite the same compare to ASO or siRNA especially for cellular uptake but also some problem concerning triplexes formation. As cytosine N3 need to be protonated, triplex formation induced by TFO can be prevented by physiological pH which can in G-rich sequence promote G-quadruplexes formation[180, 181].

I.3.2. G-quadruplexes

Unlike strategies described before, G-quadruplexes are structures that are already present and formed in DNA sequences that is why there is no need to insert DNA or RNA material to promote their formation. However, some compounds, especially small chemical molecules (<1 kDa), will be used in most of the cases to promote G-quadruplexes formation and stabilize them. Considering that G-quadruplexes are implicated in several important biological processes, interfering with them could be worth exploring in order to find new possible therapeutic targets.

I.3.2.1. Overview of G-quadruplexes therapeutic strategies

Depending on the biological process and the G-quadruplexes (G4) implication, several strategies can be considered. Most of the time, the aim is to stabilize these structures to influence cells in different ways. In the case of transcription, especially for oncogenes, stabilizing G-quadruplexes would increase their implication in transcription regulation[182]. Indeed, there is a lot of proteins capable of interacting with G-quadruplexes that can unfold them and make G-quadruplexes more transient structures. Concerning their implication in telomeres, the aim is similar in order to reduce telomerase action[183, 184]. As G-quadruplexes in promoter, proteins can bind to them and affect their formation such as Pot1 which forces the single strand conformation. Another application to consider is the destabilization of G-quadruplexes which makes sense in the hypothesis of G-quadruplexes favoring transcription when formed on the non-template strand[185]. If the G4s are present and stable in the non-template strand, this favors the single strand conformation of the template strand and consequently increases the transcription. For these examples, destabilizing G4s would reduce the efficiency of transcription. Another way to use ligands to influence G-quadruplexes formation is to act directly on their interaction with proteins. The idea would be to avoid binding of proteins capable of unwinding G4 structures[186] and to increase interaction with proteins promoting G-quadruplexes formation, either by interacting with G-quadruplexes or the proteins. In the last case, the major issue remains in the fact that ligands could influence the activity of the protein which can lead to the defect of its functions or at the opposite an undesirable increase of their activity. All these strategies need ligands to be designed in ways adapted to G4 biophysical and structural properties, features that need to be designed case by case.

I.3.2.2. G-quadruplexes and G-quadruplex-Protein complex ligands

G-quadruplexes ligands are usually small chemical molecules that are supposed to bind to them with high affinity (binding constant which is generally lower than 10^{-6} mol.L^{-1}). Several ligands from different families have been designed over the years in order to interact and stabilize G-quadruplexes **(Figure 21)**[187, 188]. These ligands have specific features in common for G-quadruplex interaction. First, a polycyclic heteroaromatic core favouring π-π interaction with the planar aromatic faces of G-core and then, even it is not always the case, some charged hydrophilic groups can be added for electrostatic interaction with DNA negative charges and also for water solubility.

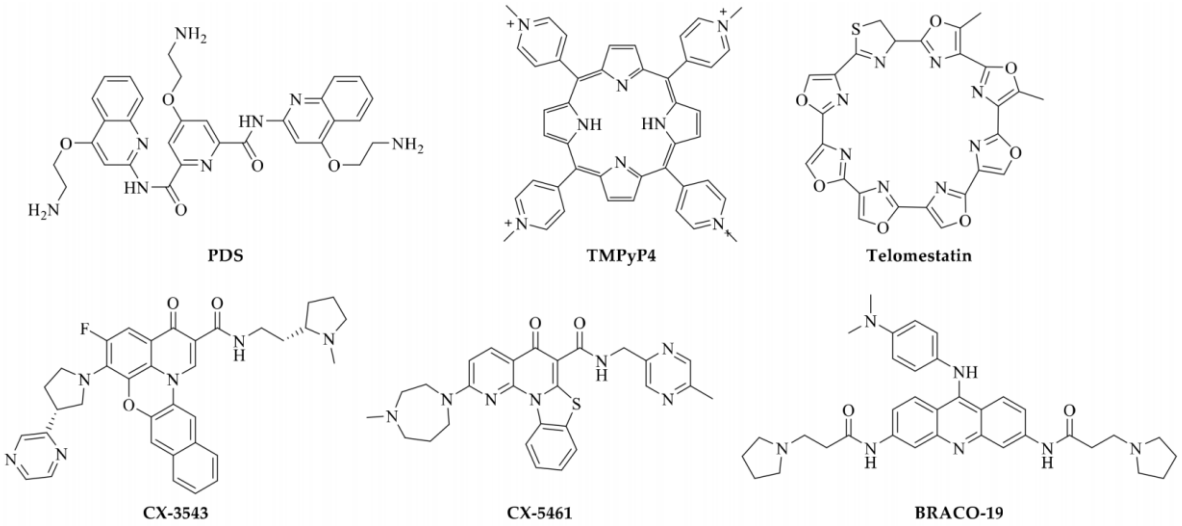


Figure 21. Examples of the most studied ligands for G-quadruplexes stabilization which belong to different chemical families (porphyrin, acridine, fluoroquinolone…)
From *Molecules, Zhi-Yin Sun et al., 2019*

These small molecules can bind to the planar quartets, the grooves, the loops or a combination of these possibilities. For example, adding positive charges will lead to improve grooves and loops interaction. As these chemical ligands are created to be used as drugs, they need to be stable under physiological conditions and their druggability need to be considered. Beside ligands designed for G-quadruplexes stabilization only, some other are used to block the interaction between proteins and G-quadruplexes **(Figure 22)** by interaction with the quadruplex such as PIPER or directly with the protein such isaindigotone (SYSU-ID-01)[189] derivative 37[190]. This ligand is capable of inhibiting interaction between NM23-H2 protein and G-quadruplexes. Protein NM23 (NME/nm23/NDPK) was discovered in 1988[191] and it is known for its implication in transcriptional regulation of gene expression. The original compound was known to interact with the protein but displays little binding affinity to G-quadruplexes[192] and the modification has been done to reduce the quadruplexes stabilization. A lot of studies have been performed with ligands and some of them show promising results that can be used in therapeutics.

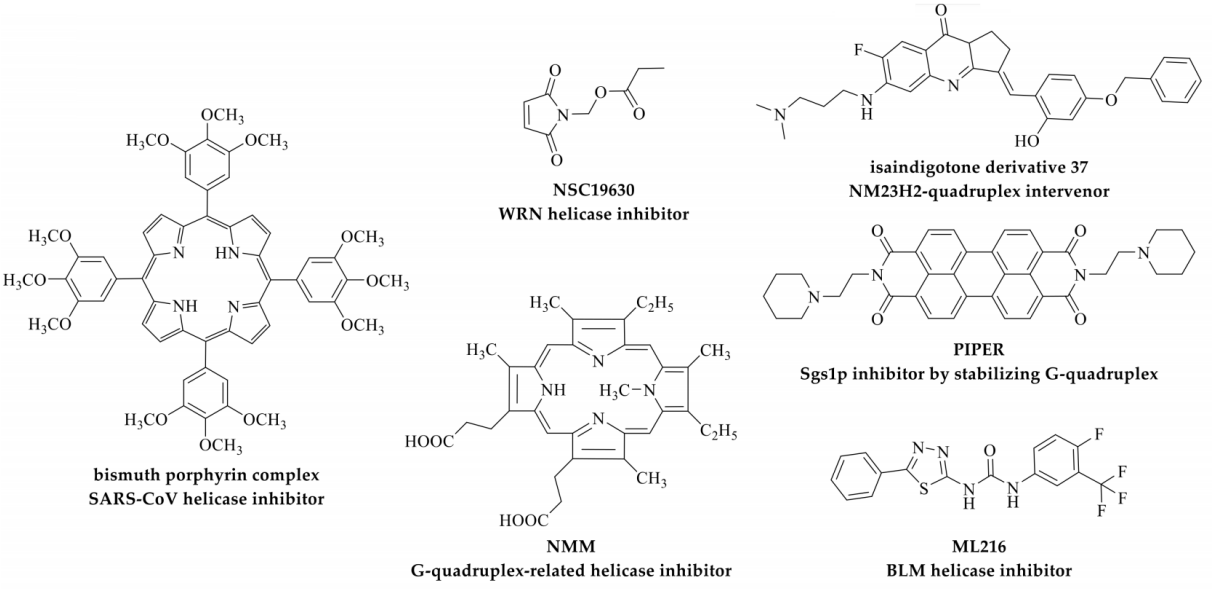


Figure 22. Examples of the most studied ligands designed to block interaction between G-quadruplexes and associated proteins

From *Molecules, Zhi-Yin Sun et al., 2019*

I.3.2.3. Examples of G-quadruplexes ligand applications

Braco-19 is a ligand optimized from the acridine derivative family. It is generally known as telomeric G-quadruplex binder that can interact by stacking and groove binding[193]. Braco-19 is mostly known to inhibit telomerase activity which lead to anti-tumor activity on human epidermoid carcinoma cells[194] or human uterus carcinoma cells[195]. The mechanism explaining this inhibition has been identified and Braco-19 is capable of uncapping 3' telomere ends[196]. Recently Braco-19 has been further studied as anti-HIV agent[197, 198]. Results showed that the ligand stabilized the viral RNA quadruplexes and during infection, was able to reduce virus genome copy in cells. The main limitation in using Braco-19 concerns its poor permeability across biological barriers that is why even if the ligand is soluble in aqueous environment and considered as a reference in G-quadruplexes research, pharmacological applications are quite restrained[199]. TMPyP4 is another well studied ligand for G-quadruplexes[200]. It has been designed and optimised from the porphyrin family which has a similar scaffold to the G-quartet and regroup strong quadruplexes ligands especially by π-π stacking. Several studies have shown that TMPyP4 have good affinity for G-quadruplexes and can inhibit telomerase and expression of several oncogenes such as c-MYC[201] or BCL-2. It has

also shown good anti-tumor activity in several tumor cells including retinoblastoma cell lines[202] or leukaemia cell lines[203]. The main criticism about this ligand concerns the fact that it can bind both quadruplexes and duplexes DNA and lacks specificity. Based on the knowledge acquired with G-quadruplexes ligand studies, ligands such as pyridostatin (PDS) have been designed[204]. The properties of this ligand are relying on a rich aromatic surface, a flat conformation and an ability to form hydrogen bonds by adding donors and acceptors groups. Even if this ligand is not so well studied compared to the ones described before, it shows interesting results by increasing telomere stability[205]. The last ligand is CX-3543 also called quarfloxin is one of the most important ligands for G-quadruplexes therapeutics. Indeed, this ligand is the first quadruplexes ligand to enter human clinical trials, even if it did not finish Phase II. It is a fluoroquinolone derivative which can bind to G4 DNA and disrupt their interaction with the nucleolin protein[206]. A derivative compound CX-5461 has been designed and show similar mechanism by inhibiting DNA replication and protein translation implicating Pol I[207]. This compound is now in advanced phase I clinical trials (Canadian trial, NCT02719977, started in May 2016).

II. KRAS oncogene

II.1. Generalities

As indicated before, G-quadruplexes motifs can be particularly enriched in several oncogene promoter regions including KRAS (Kirsten Rat Sarcoma 2 viral oncogene homolog). KRAS is a proto-oncogene that codes for a small GTPase protein. This protein belongs to RAS superfamily also called RAS-like GTPases which regroups a large number of triphosphate (GTP) binding proteins[208]. KRAS is one of front-line sensors that controls the activation of several signaling pathways by allowing the transmission of transducing signals from the cell surface to the nucleus. The most implicated cellular processes are cell differentiation, growth, chemotaxis and apoptosis. Understanding KRAS function, mechanism of action, regulation, partnership interaction, desensitization, and so on, is off upmost importance in order to tackle many forms of cancers

II.1.1. KRAS and RAS superfamily

In humans, the RAS superfamily consists of more than 100 members which can be divided in several subfamilies based in their structure, sequence and functions. The 5 main families are: RAS, RHO, RAN, RAB and ARF[209]. All proteins among the RAS superfamily share very similar molecular structures and a common ability to bind guanine nucleotides and hydrolyse them thanks to G domain. Three RAS genes codes for highly homologous Ras proteins: KRAS, HRAS and NRAS[210]. Expression of these three genes is highly conserved across species but their expression levels depends on the tissue[211]. For example, KRAS will be mainly expressed in pancreatic, colon or lung cells where NRAS and HRAS expression is quite low. In mice, the level of HRAS transcripts is highest in the brain muscle and skin but lowest in the liver. For KRAS, they are more abundant in gut, lung and thymus similar to humans and are in skin and skeletal muscle. Finally, NRAS transcripts are more present in testis and thymus. These three genes encode for four proteins: Hras, Nras, Kras4A and Kras4B[212]. The last two proteins are the result of alternative splicing of exon 4 but the dominant form is Kras4B commonly known as Kras[213]. KRAS oncogene is located on chromosome 12 at position 12.1[214]. Kras protein contains four different domains[215]: a N-terminal domain which is common to all Ras proteins and another domain which show low identity. These two regions are very important for the GTPase function because they form the needed G-domain. This domain includes the GTP-

binding pocket. The last important region is at C-terminal domain and it is implicated in posttranslational modifications and membrane anchoring. Contrary to Hras, Nras and Kras4A which traffic through the Golgi to the plasma membrane, Kras4B traffics through the cytosol[216]. Even if four Ras isoform proteins are highly homologous in their primary acid sequence (around 80%) the differences among them are more predominant in the hypervariable region (HVR) on the C-terminal domain[217].

II.1.2. KRAS GTPase mechanism

Newly synthesized Kras protein is a cytosolic and inactive protein. Localization and function of Kras, and more generally the Ras proteins, are regulated by several post-translational modifications in the C-terminal "CAAX" motif[218]. This tetrapeptide contains a Cysteine, two Aliphatic amino acids and a terminal X amino acid. The different modifications include a farnesylation of the cysteine by a farnesyltransferase, as well as a methylation by ICMT (Isoprenylcysteine Carboxyl MethylTransferase) and the proteolytic removal of the last three residues by RCE1 (Ras Converting Enzyme 1). Adding a methyl group prevents the plasma membrane repulsion by negating the negative charge. In addition, the plasma membrane localization of Kras requires a basic poly-lysine region which is located upstream of the C-terminus[219]. Unlike the Kras4A variant, Kras4B does not need palmitoylation by palmitoyl transferase to the targeting to the membrane. Once bound to the membrane, Kras cycles between an active state when bound to GTP and an inactive state when bound to GDP due to the hydrolysis of the GTP. Switches between these two different states are mediated by two different classes of proteins[220]: GEF (Guanosine nucleotide Exchange Factors) and GAP (GTPase-Activating Proteins). GEF proteins mediates the exchange between bound GDP with GTP and GAP proteins stimulates the hydrolytic ability of Kras to convert bound GTP to GDP. Among the GEF proteins, one of the most known is SOS (Son of Sevenless) which controls exchange between GDP with GTP within Kras. Under physiological conditions, Kras is mainly bound to GDP. Upon external stimuli, nucleotide binding of GEF is disabled and nucleotide is released. When bound to GTP, Kras GTPase activity is increased (around 100,000 times) because of changes in Kras interactions with GAP and the release of GTP is promoted by affecting interactions with GEF. These changes are mainly due to conformational changes when Kras binds GTP. One the most well-understood model for studying Kras activation is

activation implicating epidermal growth factor receptor (EGFR)[221, 222], part of the receptor tyrosine kinases (RTK) family **(Figure 23)**.

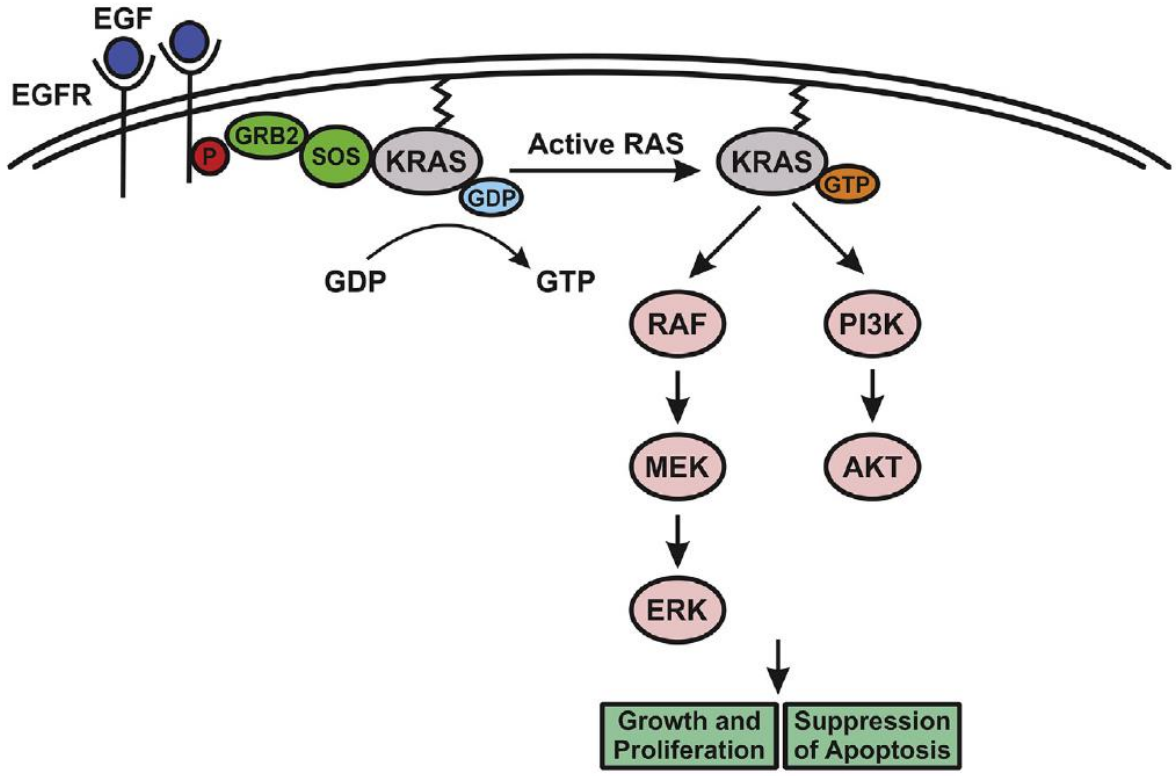


Figure 23. Kras signaling activation in EGF pathway

From Genes & Diseases, Kyle Knickelbein and Lin Zhang, 2015

The binding of EGF to its receptor will induce its phosphorylation and the fixation of GRB2 (Growth-Factor-Receptor-Bound protein 2) which will recruit SOS in the cytosol. Then Kras is bound to GTP in an active state. Kras mediates a lot of downstream pathways of kinases including RAF (Rapidly Accelerated Fribrosarcoma)/MEK/ERK and PI3K (PhosphoInositide 3-Kinase)/AKT cascades. The first cascade will propagate the growth signal and the second one will control apoptosis suppression[223]. Studies have shown that GTP-bound Kras could directly bind RAF protein recruiting another RAF protein to dimerize allowing activation[224, 225]. Then activated RAF induce phosphorylation cascade to activate MEK and ERK.

II.1.3. KRAS mutations and implications in cancer

RAS is one of the most frequently mutated oncogenes in human cancer but the mutations are not uniform[226]. Indeed, 85% of RAS mutations correspond to KRAS mutations[227]. Kras4B is

the most mutated isoform present in 90% of pancreatic cancers, 35-40% of colorectal cancers and 15-20% of lung cancers. As a comparison, mutations in HRAS and NRAS in colorectal cancers are around 5%. The mutated isoform is also present in other cancer types but with less frequency such as cervical or liver cancer. Perkins and colleagues showed that the most frequent mutations in KRAS gene occur at codons 12,13 and 61 and with less frequency at codons 63, 117, 119 and 146[228]. All mutations lead to a reduced GTPase activity maintaining Kras in an active state but depending on the implicated mutation, the mechanism could be different. For example, mutation of glycine 12 affects GAP binding affecting GTP hydrolysis and mutations of residue 13 sterically clash with arginine with the same consequences. Glutamine 61 is directly implicated in the catalytic site by positioning a water molecule to facilitate GTP hydrolysis and that is why G61 mutation directly affects the reaction. Beside the localization of these mutations, the mutation type is also important[229, 230]. As an example, mutation of glycine 12 by a valine has been shown to have a worse prognosis in lung cancers than the mutation by aspartic acid. In addition of the maintained active state by mutation, there is another important fact to consider in KRAS-driven cancer. Indeed, it has been shown for years that the association between wild-type and mutant Kras has in important role in malignancy[231]. Kras wild-type exhibits a restraining function of tumor growth in KRAS mutant cancers because it is capable to antagonize oncogenic Kras resulting in reduced activity. Unfortunately, this inhibitory effect is overcome during tumor progression because of KRAS gene allele loss and increasing copy number of the oncogenic form. It has been recently shown that this effect of Kras wild-type is due to its capacity of dimerization with the mutant form. Finally, the last type of mutations concerns the tumor microenvironment which increase the cancer malignancy[232]. Indeed, cancer cells that expressed mutated Kras are capable of inducing production of chemokines, cytokines and growth factors implicating in stroma reprogramming.

II.2. KRAS oncogene therapeutics

For therapies, the common strategy is to used chemotherapy that has important cytotoxic side effects. It generally uses pyrimidine analogs such as 5-fluorouracil (5-FU) which inhibits thymidylate synthase required for nucleotide synthesis or other cytotoxic drugs[233, 234]. Regarding the huge side effects of this method and its efficiency which is often limited, the need for more efficient therapies have been pursued by private to governmental institutions

with dozens of billions of dollars spend annually, and hundreds of billions in indirect social costs (World Health Organization). As the central role of KRAS oncogene has been broadly studied and better understood in the last past decades, there were also a tremendous attempt to target it through different ways. All studies covered the multiple aspects of KRAS activation in order to finally find a possible Achilles's heel in order to treat efficiently Kras implicated cancers **(Figure 24)**.

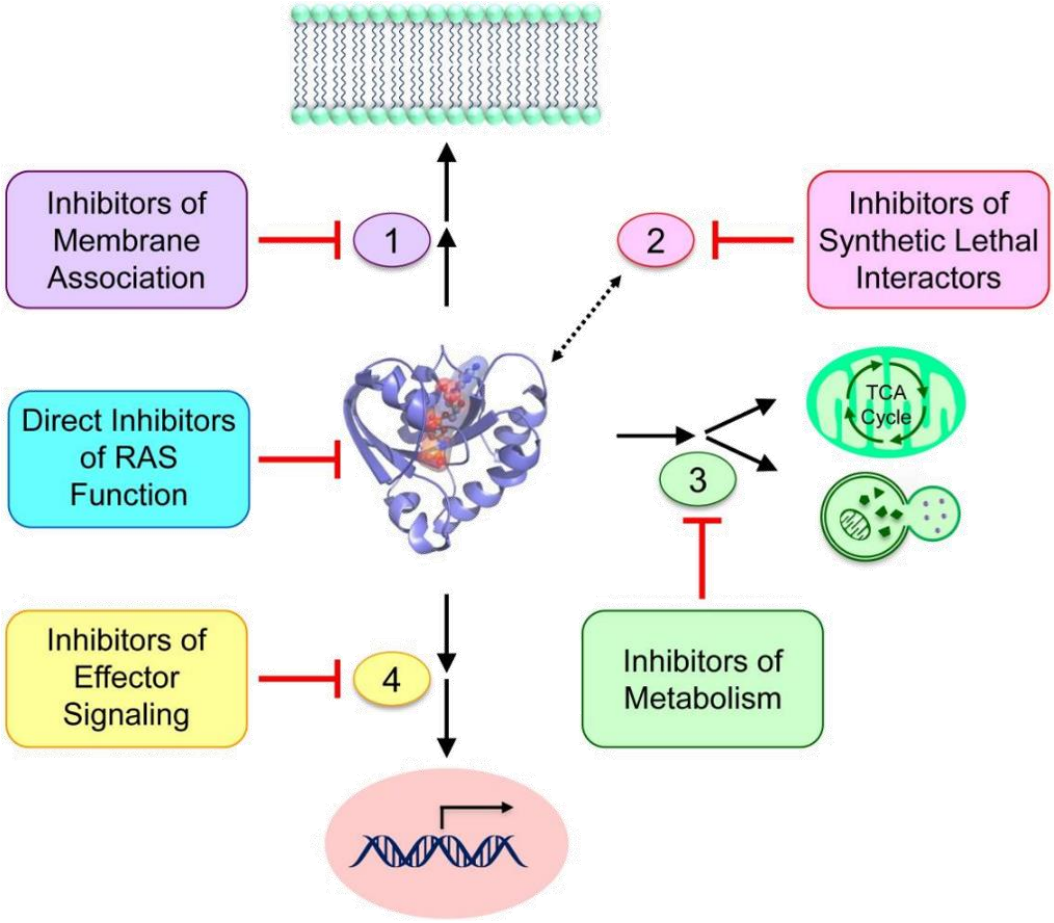


Figure 24. Overview of strategies in KRAS-driven cancers

From Trends in Cancer, Meagan B. Ryan et al., 2015

II.2.1. Directly targeting Kras

The first approach to be considered is to directly target Kras and its function. In principle, designing small molecules that directly compete with GTP for Kras interaction in order to decrease Kras activity. However, considering different factors difficult to bypass such as

shallow binding pockets, the high affinity of Kras for GTP which is in the 10^{-11} molar range, together with concentrations ranging in the hundreds of µM for GTP and GDP, it is almost impossible to find a molecule able to compete with GTP. Over time, as the new strategies to target Kras have been failing one after the other, Kras has been associated as an unreachable mountain summit, an "undruggable "target. Recently, a new strategy has emerged specifically targeting Kras G12C mutation which represents 10 to 20% of all G12 mutations[235]. In Kras structure, this mutation is close to the nucleotide binding pocket and the switch regions implicated in effector interactions. Several compounds have been designed to covalently bind with the cysteine. Studies show that these compounds bind to the GDP state of Kras, impair action of SOS, decrease the affinity for GTP and block Kras-RAF association. Among all the designed compounds, ARS-853 gives good results *in vitro* but not *in vivo*[236] and another last generation designed ARS compound ARS-1620 shows *in vivo* anticancer activity[237]. Both compounds are actually in preclinical trials. Several limitations can be found in this strategy. The first one is that it is limited to Kras G12C mutations but recent studies also reported Kras G12D inhibitors[238]. There are also strategies aiming to reduce KRAS expression by using siRNA as previously described. The major part of the successful compound is in preclinical trials except one called AZD-4785 which is studied in Phase I.

II.2.2. Membrane association

Kras requires membrane association for the oncogenic activity that is why several strategies aim to directly targeting post-translational modifications that modulate Kras membrane association. One of the first drugs discovered concerns farnesyl-transferase inhibitors (FTIs) that inhibit the prenylation of Kras required for membrane attachment[239]. FTIs triggered growth suppression in cancer cells in pre-clinical studies but phase II and phase III trials were disappointing because FTIs did not exhibit clinical efficiency as single agents[240]. It was explained by Kras4B prenylation through alternative mechanisms involving geranyl transferase I (GGT I)[241]. Geranylgeranylation of Kras is an alternative 20-carbon isoprenylation that can help Kras bioactivity when farnesylation is impaired. It has been proven by simultaneous genetic inactivation of farnesyl-transferase (FTase) and geranyl transferase (GTTase) which reduced KRAS-driven lung tumorigenesis in mice[242]. Dual inhibitors of FTase and GTTase have been developed (L778, 123) showing efficiency in KRAS-mediated cancers but also exhibiting high toxicity leading to the prohibition of further clinical

development[243]. Another strategy is to inhibit the interaction between Kras and cyclic GMP phosphodiesterase δ (PDEδ) required for the correct localization and signalling of farnesylated Kras[244]. PDEδ increase Kras signalling by enriching Kras at the plasma membrane. PDEδ binds and solubilizes farnesylated Kras enhancing diffusion in the cytoplasm. Recent inhibitors[245] such as deltarasin have been designed to interfere with binding of mammalian PDEδ to Kras and avoid Kras localization to endomembrane. They showed inhibition of oncogenic Kras signalling and supress *in vitro* and *in vivo* proliferation of human pancreatic cells. Then a new generation of compounds has been designed named daltazinon 1 showing high selectivity and less unspecific cytotoxicity but also less stability[246].

II.2.3. Acting on metabolism pathways

KRAS-mediated cancers are highly connected to several metabolic pathways[247, 248]. As examples, KRAS-mediated colorectal cancers are associated with increased expression of glutamine metabolic proteins and in human pancreatic cancers, KRAS mutation lead to the reprogramming of the glutamine metabolism[249, 250]. These cancer cells use a pathway using aspartate derived from glutamine metabolism. Aspartate is transported into the cytoplasm and converted into oxaloacetate by aspartate transaminase (GOT1). This pathway is essential for cancer cells growth by maintaining a favourable cellular redox state. KRAS mutation such as Kras G12D has also been recently reported to reprogram lipid homeostasis which supports tumorigenesis[251, 252]. All these implications related to metabolism create a new way to treat KRAS-driver cancers by studying metabolic inhibitors. Several compounds have been test in preclinical trials and one of them CB-839 which is a glutaminase inhibitor is a compound in Phase II trial.

II.2.4. Synthetic lethality

An emerging strategy for targeting oncogenic mutations, including KRAS mutations, which shows great promise in specifically targeting cancers cells is the exploitation of synthetic lethality[253, 254]. This mechanism also known as conditional genetics is based on the interaction of two genes that both contribute, even in a nonlinear way, to essential processes[255]. When only one gene is mutated the cell is viable but if mutations occurs in the two genes, it leads to cell death. This process is known as synthetic lethality because cells with both gene mutations are not viable and it is not possible to directly isolate these cells. In the case of cancer

therapeutics, it is very interesting because cells without the cancer-inducing genotype are unaffected by such targeting as only one gene exhibit mutations. This approach can obviously be envisaged for KRAS-mediated cancers **(Figure 25)**. Several tools have been used such as siRNA, shRNA (small hairpin RNA) and CRISPR library to screen synthetic lethal interactors with the KRAS oncogene[256, 257].

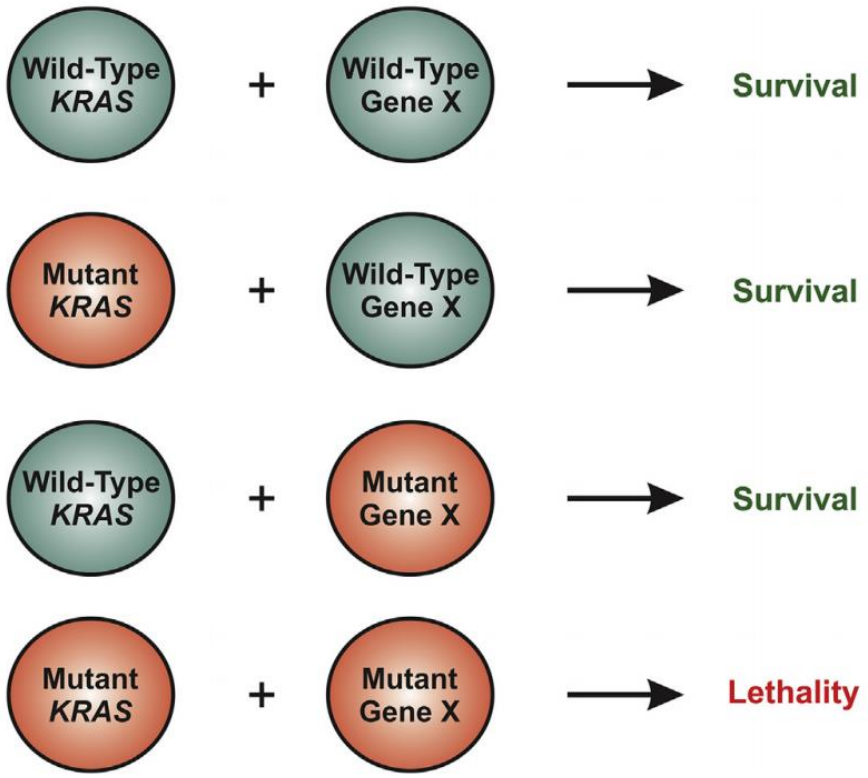


Figure 25. General principle of synthetic lethality in KRAS-driven cancers

From Genes & Diseases, Kyle Knickelbein and Lin Zhang, 2015

Several agents have been found to have synthetic lethality effects. For example, in a recent study, a shRNA which screened KRAS-driven pancreatic cells identified several synthetic lethal interactions with mutant KRAS including depletions of the mitotic protein polo-like kinase-1 (PLK1), anaphase-promoting complex subunits and other components of the proteasome[258]. There is increasing number of studies focused on synthetic lethality for several reasons. One of them concerns combinational therapy. Indeed, even if synthetic lethality agents could be used as single method, it seems that the best solution is to used them in combination with cytotoxic chemotherapy and radiotherapy[259]. In some cases, the effect of synthetic lethal interaction will be increased when combined with a genotoxic stress. It would be possible to

accentuate the efficiency of cytotoxic therapies at lower doses and as a consequence to reduce side effects and increase therapeutic indices.

II.2.5. Immunotherapy

Cancer immunotherapy is the strategy drawing the most attention in cancer treatment. Indeed, several immunotherapy strategies using antibodies have been performed several years ago. For examples, several antibodies have been used as anti-EGFR antibodies known as cetuximab and panitumumab. These anti-EGFR antibodies can inhibit ligand binding and its dimerization which is necessary for its auto-phosphorylation for its full activation[260, 261]. Cetuximab is a chimeric human-murine antibody of the IgG1 isotype and panitumumab is a humanized antibody of the IgG2 isotype[262, 263]. In addition to their role against the ligand binding, these antibodies are also capable of inducing EGFR internalization and its degradation resulting in a decrease level of EGFR at the cell surface. In clinical trials, these two antibodies have shown some therapeutic benefit as stand-alone agent and in combinational therapies. However primary and secondary resistance to this therapy has led to the end of the trials. Indeed, KRAS mutations lead to resistance to anti-EGFR antibodies even if the reason remains unsolved[264, 265]. However, this failed attempt did not stop efforts in immunotherapy field applied to KRAS mutations. Recently, antibodies against anti-programmed cell death protein 1 (PD-1)[266] named pembrolizumab have been designed and tested in KRAS mutant NSCLC (Non-Small Cell Lung Cancer). They have been used in combination with trametinib which is a MEK inhibitor but it is too early to conclude about some therapeutic benefit in clinical trials. For the next antibodies generation, studies may focus on FBP1 (Fructose-1,6-BisPhosphatase 1) which has an aberrant expression in natural killer (NK) cells and inhibits glycolysis and impairs viability. Indeed, it has been shown that KRAS-driven cancer lung cancer are closely associated with the dysfunctional state of NK cells[267].

II.3. G4 within KRAS promoter region

Despite all strategies used in KRAS-driven cancers, there is actually no good strategies passing to late stages of clinical trials. As previously described, in the last decades G-quadruplexes became a possible target at different levels to develop new cancer therapies. Bioinformatics and other genomic analysis tools showed that promoter regions of several oncogenes contains G-rich sequence which are capable of forming G-quadruplexes. The most studied oncogene is

c-MYC but KRAS promoter region has also been examined, although with less attention, probably due to its more complex DNA-sequence. The promoter region contains a nuclease polypurine-polypyrimidine hypersensitive element (NHPPE or simply NHE)[268, 269] which is essential for transcription and capable of forming secondary structures such as G-quadruplexes. The NHE sequence is comprised from -285 to -341 from the TSS in human and from -273 to -332 from the TSS in mouse. It has been shown that the deletion of this NHE region led to a significant downregulation of KRAS transcription. Moreover, KRAS promoter has been targeted by TFO in the triplex strategy for therapeutics[270, 271]. Unfortunately, the obtained results were not satisfying and the main hypothesis was that the triplex formation encountered a serious obstacle which could be the formation of G-quadruplexes. G-quadruplexes forming sequences have been investigated in the promoter region around 300 bases from the TSS[272]. Three main regions have been identified: a region near the TSS with a length of around 30 bases; a mid-region of around 50 bases and a far region of around 35 bases **(Figure26)**.

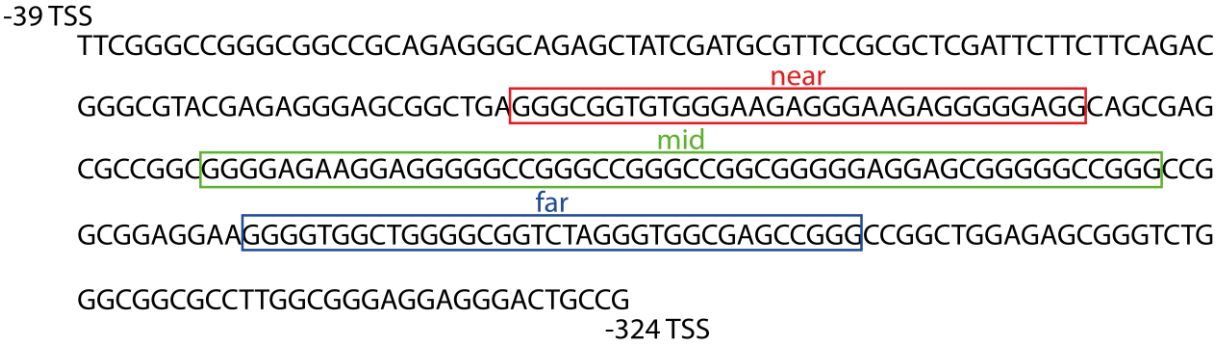


Figure 26. Identification of the three main G-quadruplex forming sequences within KRAS promoter region

Modified from BBA, Morgan et al., 2016

The near and mid regions form stable G-quadruplexes. Moreover, the near region is being part of the KRAS NHE region and the absence of G-quadruplex formation within this sequence (by inserting mutations G into T) leads to an increase in transcription efficiency[273]. That is why this sequence became the main target of studies for the formation of G-quadruplexes. It can form G-quadruplexes via five tracts of guanines including an another sequence called KRAS21R was also identified as G-quadruplex forming sequence via four tract of guanines[274]. It has been shown that this near sequence name KRAS32R can fold into a dynamic intra-molecular

G-quadruplex with two conformations in equilibrium. Xodo and colleagues[275] show that this KRAS32R G-quadruplex is capable of competing double strand DNA from NHE sequence for the binding of several nuclear proteins. They used nuclear extract of pancreatic adenocarcinoma (Panc-1) against KRAS32R G-quadruplex and dsNHE sequence and they identified several proteins which bind to quadruplex such as Poly [ADP-Ribose] Polymerase 1 (PARP-1), subunits of ATP-dependent DNA helicase 2 (Ku70 and Ku86) and heterogeneous nuclear RibonucleoProtein A1 (hnRNP A1). For the last protein, it has also been shown that it can, with its derivative, destabilize KRAS32R G-quadruplexes and facilitate hybridization to the complementary NHE strand. This point will be discussed later in the manuscript. Considering the failure of numerous strategies for cancer therapy and all evidences for G-quadruplexes roles, we thought that to understand and study if KRAS G4s were a possible new target for developing new cancer molecules, I had to understand its possible structure, dynamics and its polymorphism in the NHE region.

III. Project overall

This thesis project aims at participating in Drug Design for anticancer therapy applied to KRAS-driven cancers by targeting G-quadruplex forming sequences within the KRAS promoter region. The principal goal is to identify a compound capable of becoming a drug in anticancer therapy by screening and testing several ligands from different families. To this end, we need to evaluate and validate possible G4s targets in the NHE region of KRAS. My project fits in this category. This study aims at determining and validating a G-quadruplex target and identify the best chemical molecules to interact with and stabilize G4 from KRAS NHE region. However, in order to design good ligands and identify chemical moieties implicated in the interaction, it is necessary to have structural information at atomic level about the KRAS G-quadruplexes conformations. Identification of characteristic structural elements of these G-quadruplexes can allow researchers to design optimized ligands against a specific G-quadruplex such as KRAS G-quadruplexes. This project is divided in two major parts: the first one is devoted to the determination of KRAS32R G-quadruplexes structure to obtain atomic structural elements, and the second one is focused on the study of their interactions with proteins such as hnRNP A1 and several ligands from different families.

III.1. Determination of KRAS32R G-quadruplexes structure

As explained before, obtaining structural information on the target is very important in the concept of Drug Design. Indeed, it can help designing molecules that target specific binding sites identified in a resolved structure. Then, when ligands are tested against the target, the structure allows to determine the chemical moieties that are important in the interaction. For KRAS32R G-quadruplexes, structural determinations have been attempted but there is no structure of the wild type KRAS32R. However, Plavec and colleagues[276] determined the structure of a modified sequence from KRAS32R which is called KRAS32R-3n corresponding to the removal of the three first residue at 5' end and the addition of three residues at the 3' end[277]. It corresponds to a shifts within the NHE sequence. They showed that this sequence can form a stable dimeric G-quadruplex giving important structural information for ligand design. The main criticism concerning this work is that the shift of three residues could lead to major changes in the G-quadruplex formation compared to KRAS32R wild type. It has been shown by Xodo and colleagues[275] via DMS footprinting experiments that guanines 2 (G2) and

3 (G3) are implicated in KRAS32R G-quadruplexes conformations. We can assume that the G-quadruplex formation from 32R-3n could be different from the wild type sequence. Thus, the first objective of this work is to determine the structure of the major scaffold adopted by KRAS32R.

III.2. KRAS G-quadruplexes interaction

The second part of this work is focused on KRAS G-quadruplex interactions. The first goal is to perform interactions between our structure of KRAS32R quadruplex and the hnRNP A1 protein which has been shown to unfold KRAS32R G-quadruplexes it. We did use a truncated version of hnRNP A1 to study the interaction called UP1, a shorter version containing both RRM motifs but without the unfolded domain Gly-rich region. In parallel, I performed ligand binding assays to find good candidates that can efficiently stabilize KRAS32R G4. These ligands could be used to increase transcription disturbance or to compete with proteins such as hnRNP A1 to avoid G4 unfolding and reduce KRAS transcription. Once the best ligands identified, we tried to obtain structural information to identify the most important chemical moieties in interaction to design or modify ligands which can overcome common limits of G-quadruplexes ligands.

Part 1: Determination of KRAS32R G-quadruplexes structure

I. Introduction

To understand how are organized the guanine rich sequences in the NHE region in KRAS promoter we started by studying what is thought to be the smallest sequence that can form a stable G-quadruplex. Described previously by L. Xodo and coworkers, the 21R sequence represented the smallest G4 from the NHE region that interact with the transcription factor MAZ[277]. we continued the studies from previous work in the laboratory where the structure of KRAS 22RT, a mutated version of 21R, was determined at 20 °C **(Annexes Article 1)**. The first step was to determine the structure at 37 °C in order to have more biological relevance (structure deposited in the protein data bank with code 6T51). By analyzing this sequence in a G-quadruplex algorithm such as G4Hunter[278] which gives a score depending on the probability of forming G-quadruplex, it showed a similar score compared to KRAS21R (1.45 for KRAS22RT versus 1.41 for KRAS21R) known to form G-quadruplexes. Then, I started acquiring NMR and CD spectra of the complete sequence 32R, and I tried to understand how both the longer and smaller sequence could form G-quadruplexes. Several dozens of mutants were analyzed and I believed I could solve the structure of the longer sequence. By the end of the preliminary analysis, it was clear that KRAS32R WT did not adopt a single conformation but rather a mixture of two main conformers. Discovering how the 32R sequence would organize and how both conformers 3D structure looks like (at 37 °C) could be important to understand G4s structures in KRAS promotor region and their interplay with transcription factors. This become then the central core of my research project.

II. Determination of KRAS22RT G-quadruplex

II.1. KRAS22RT studies at physiological temperature (Article 1)

The article below represents the studies that we conducted in order to determine the structure of KRAS22RT at physiological temperature. What we found out was that the increasing of temperature from 20 to 37 °C did not affect much the overall folding of the structure. Nevertheless, even minor changes are important if we want to use the structure for rational drug design. In next section I will describe in details the findings.

Biomolecular NMR Assignments
https://doi.org/10.1007/s12104-017-9793-0

ARTICLE

^{1}H, ^{13}C, and ^{15}N chemical shift assignments of a G-quadruplex forming sequence within the *KRAS* proto-oncogene promoter region

Julien Marquevielle[1] · M. V. Vasantha Kumar[1] · Jean-Louis Mergny[1] · Gilmar F. Salgado[1]

Received: 22 July 2017 / Accepted: 28 November 2017

Abstract
Single stranded guanine rich DNA (or RNA) sequences adopt noncanonical secondary structures called G-quadruplexes (G4). Functionally, quadruplexes control gene transcription and regulate activities such as replication, gene recombination or alternative splicing. Hence they are potential targets for cancer, neuronal, and viral related diseases. KRAS is one of the most mutated oncogenes in the genome of cancer cells and contains a nuclease hypersensitive element (NHE) sequence capable of forming G-quadruplexes via its six runs of guanines. In our work, we are interested in the NMR structure of the major G4 scaffold formed in the KRAS NHE region with a mutated sequence of 22 residues. Here, we report ^{1}H, ^{13}C and ^{15}N chemical shift assignments the G4 formed within KRAS22RT sequence.

Keywords KRAS oncogene · Promoter region · G-quadruplex · Structure resonance assignments

Biological context

Single stranded guanine rich DNA (or RNA) sequences can adopt non canonical secondary structures called G-quadruplexes (G4). They are characterised by a stacking of at least two guanines tetrads. In each tetrad four guanine nucleotide bases are held together by Hoogsteen hydrogen bonding network. In Hoogsteen bonds, guanines are linked by the Hoogsteen face (compared to Watson–Crick face) involving an amino group at position 1 from one guanine and an oxo group at position 7 from another guanine (Sen and Gilbert 1988). Tetrads are further stabilized by staking (π–π interaction) and by monovalent cations such as potassium (K^{+}) or, to a lower extent, Na^{+} or NH_4^{+} (Largy et al. 2016; Wlodarczyk et al. 2005). G4 are highly polymorphic due to sequence composition, strands (number and orientation), loops (size and orientation) and ionic environment. Moreover, guanines are covalently linked to sugar by glycosidic bond giving two possible torsion angles: *syn* or *anti*. Recently, different approaches using antibodies or specific ligands allowed G4 visualisation inside cells (Henderson et al. 2013; Hansel-Hertsch et al. 2017). In addition, numerous studies have demonstrated that G4 play key roles in several biological processes such as gene transcription, replication, genomic instability or alternative splicing. Hence they are potential targets for cancer, neuronal, and viral related diseases (Bochman et al. 2012). Indeed, bioinformatics searches in human genome suggests that G-rich sequences are not randomly localized within genomes but are specifically enriched in particular regions such as telomeres, the first introns and promoters of genes (Stegle et al.2009; Bedrat et al. 2016). Proto-oncogenes are particularly enriched with G4 motifs (Huppert and Balasubramanian 2007; Eddy and Maizels 2006) such as BCL2, VEGF, KRAS, MYC and WNT1 which generally adopt intramolecular conformation. Presence of G4 in these regions raises important questions concerning their function, and could be used as target for inhibition of transcription. Indeed, depend of their localisation, G4 could slow transcription by disturbing polymerase either by blocking it directly or by recruiting transcription repressors (Qin and Hurley2008). Stabilization of these structures, using small chemical ligands for example, could become a new area of therapy, especially for KRAS which is still undruggable target. KRAS is one of the most mutated oncogenes in the genome. KRAS gene is located on chromosome 12 and is responsible for the coding of K-Ras, a 23.2 kDa G-protein that regulates the cell division (Brito et al. 2015). KRAS

✉ Gilmar F. Salgado
g.salgado@iecb.u-bordeaux.fr

[1] ARNA Laboratory, European Institute of Chemistry and Biology (IECB), Université de Bordeaux, Inserm U1212 – CNRS UMR 5320, 2, rue Robert Escarpit, 33607 Pessac, France

Published online: 30 November 2017

mutations terminate its coding and the signalling activity of K-ras proteins that potentially affects the normal cells growth rate. KRAS oncogene is particularly implicated in lungs or colorectal cancers (30% of these cases) (Bamford et al. 2004). It is thought that the KRAS oncogene promotes glycolysis through the activation of downstream signaling pathways to sustain the energy requirements for uncontrolled cellular proliferation, thus contributing to survival of cancer cells. The KRAS GTP/GDP binding site has a very high affinity (picomolar range) (McCormick 2015) that is why it is quite difficult to synthesize molecules to compete with GTP, at millimolar range inside cells, to reduce KRAS activity. It is no surprise that after many decades of unsuccessfully battling against the Ras proteins, new strategies emerged such as targeting promoter region with small ligands. KRAS promoter contains a nuclease hypersensitive element (NHE) sequence capable of forming G-quadruplexes via six short domains containing guanines (Lavrado et al. 2015). In this NHE sequence, a 32 residues sequence can form G4 via the first five domains. In this work, we are interested in the NMR structure of the major G4 scaffold formed in the KRAS NHE by this 32 residues sequence with a shorter mutated sequence of 22 residues with a guanine mutated with a thymine at position 16 which prevents the formation of a minor species and makes it simpler to study by NMR (Fig. 1a). In a previous work we have determined the 22 residues KRAS sequence NMR structure at 20 °C (Kerkour et al. 2017) (Fig. 1b) and we want to understand the behaviour of this structure at physiological temperature (37 °C).

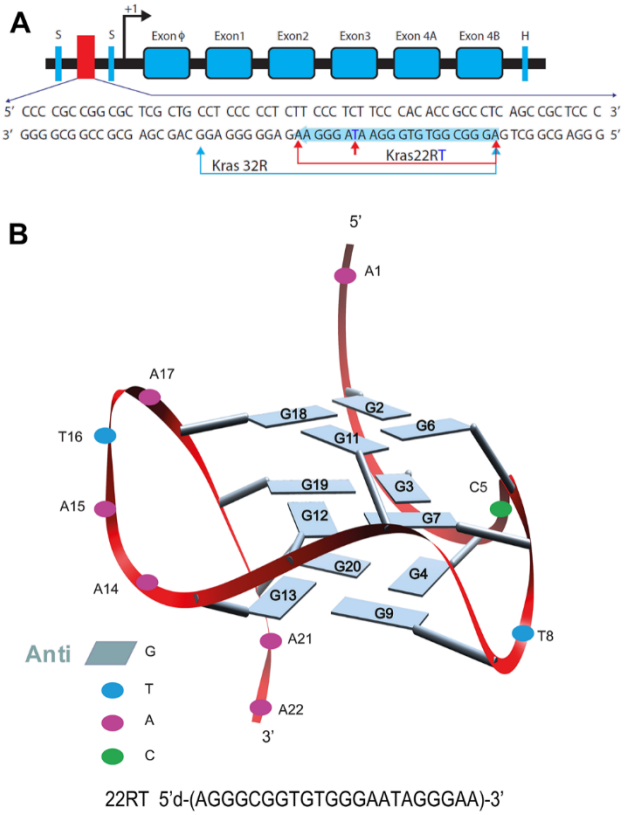


Fig. 1 **a**, Schematic view of KRAS oncogene and magnified on the NHE sequence with 32 residues sequence (blue arrow) and the shorter G4 sequence comprising 22 nucleotides (red arrows). **b**, Schematic representation of the KRAS22RT G-quadruplex

Materials and methods

DNA oligonucleotides synthesis and preparation

The DNA oligonucleotide used for NMR experiments contained 5% ^{15}N, ^{13}C site-specifically labelled nucleotides. It was synthesized in our laboratory (INSERM U1212, Bordeaux) on an automated Expedite 8909 DNA synthesizer at 1 µmol scale on 1000-Å primer support (Link Technologies SynBase CPG). All the standard phosphoramidites (dABz, dT, dGiBu, dCAc), reagents, and solvents used during the synthesis were purchased from Glen Research. The dGiBu-phosphoramidite (U-$^{13}C_{10}$, 98%; U-$^{15}N_5$, 98%; CP 95%) was purchased from Cambridge Isotope Laboratories. After the synthesis, the oligonucleotides were cleaved from the support and the nucleobases were deprotected with ammonium hydroxide at 55 °C for 16 h and then lyophilized. For HSQC ^{1}H–^{13}C and HSQC ^{1}H–^{15}N, 5% ^{15}N, ^{13}C uniformly labelled DNA oligonucleotides were synthesized in our laboratory by enzymatic synthesis using MuLV Reverse Transcriptase (Moloney Murine Leukemia Virus Reverse Transcriptase). Reaction is catalyzed by the enzyme during 20 h at 37 °C with ^{15}N, ^{13}C triphosphate nucleosides and a template sequence containing a ribosomal base just before the target G4 sequence in a specific buffer for optimal enzymatic conditions. Then, newly synthesized strand by cleaving the ribose base by adding an excess of potassium hydroxide (KOH) during 6 h, purified by HPLC (Dionex Ultimate 3000) with anionic exchange column DNAPac PA100. After lyophilization all samples were stored at − 20 °C. All the sequences used for NMR were prepared in potassium NMR buffer (10 mM K_2HPO_4/KH_2PO_4, 50 mM KCl, pH 6.5, 10% D_2O)(Tables 1, 2). After dissolving in buffer, the oligonucleotides were heated for 5 min at 95 °C and chilled on ice several times and then stored at 4 °C until further use.

NMR spectroscopy, resonance assignment and data deposition

NMR spectra were recorded on Bruker Avance instruments 700, 800 and 950 MHz instruments equipped with cryogenically cooled probes. Experiments were performed at 37 °C. For solution NMR, standard 3 mm NMR tubes

Table 1 ^{1}H NMR chemical shifts of KRAS22RT G-quadruplex at 310 K in 950 MHz spectrometer (90% NMR buffer 50 mM KCl; 10 mM KPi; pH 6.5 and 10% D_2O)

Residue	H1	H2	H8/H6	H5/H7	H1′	H3′	H4′	H5′/H5″	H2′/H2″
ADE^{1}		8.14	7.94		5.98	4.87	4.14	3.70/3.70	2.51/2.65
GUA^{2}	11.72		8.05		6.09	5.06	4.49	4.16/4.16	2.79/3.03
GUA^{3}	11.22		7.61		6.16	4.97	4.57	4.34/4.24	2.52/2.92
GUA^{4}	11.26		7.73		6.34	5.17	4.7	4.35/4.38	2.78/2.78
CYT^{5}			8.08	6.24	6.53	5.14	4.68	4.33/4.42	2.45/2.79
GUA^{6}	11.91		8.11		6.24	5.24	4.51	4.36/4.29	2.48/3.00
GUA^{7}	11.34		8.02		6.33	5.22	4.7	4.39/4.50	2.76/2.90
THY^{8}			8.02	1.83	5.83	4.7	4.45	4.38/4.14	2.22/2.28
GUA^{9}	11.25		7.72		6.51	5.05	4.71	4.45/4.35	2.53/3.11
THY^{10}			7.47	2.02	6.59	5.23	4.65	4.37/4.48	2.58/2.70
GUA^{11}	11.95		8.19		6.23	5.18	4.53	4.36/4.36	2.66/3.04
GUA^{12}	11.52		7.99		6.07	5.17	4.93	4.34/4.43	2.67/3.05
GUA^{13}	11.22		7.76		6.36	4.78	3.9	3.37/3.08	2.51/2.67
ADE^{14}		8.13	8.28		6.31	4.95	4.4	4.17/4.06	2.75/2.81
ADE^{15}		8	8.5		6.5	5.14	4.57	4.33/4.16	2.86/3.00
THY^{16}			7.57	1.95	6.11	5.14	4.82	4.14/4.31	1.73/2.20
ADE^{17}		7.81	8.13		6.28	5.11	4.39	4.12/4.00	2.80/2.86
GUA^{18}	11.48		8.2		6.14	5.1	4.54	4.26/4.26	2.75/3.00
GUA^{19}	11.36		7.95		6.12	5.12	4.59	4.59/4.34	2.79/3.00
GUA^{20}	10.88		7.65		6.07	4.99	4.46	4.33/4.17	2.52/2.78
ADE^{21}		7.92	7.85		6.01	4.87	4.33	4.16/4.16	2.39/2.53
ADE^{22}		7.89	7.99		6.09	4.89	4.58	4.11/4.06	2.36/2.48

Table 2 ^{15}N and ^{13}C NMR chemical shifts of KRAS22RT G-quadruplex at 310 K in 800 MHz spectrometer (90% NMR buffer 50 mM KCl; 10 mM KPi; pH 6.5 and 10% D_2O)

Residue	N1	C2	C8/C6	C5/C7	C1′	C3′	C4′	C5′	C2′
ADE^{1}		152.8	139.7		84.61	74.65	85.10	61.89	37.77
GUA^{2}	143.2		135.5		82.99	77.14	85.22	66.07	38.00
GUA^{3}	143.4		134.4		82.06	75.96	84.84	66.00	40.10
GUA^{4}	143.4		135.3		80.88	78.52		64.94	39.84
CYT^{5}			141.6	96.87	86.27	75.94	84.88	66.55	39.81
GUA^{6}	143.3		135.4		84.86	77.89	85.64	65.66	41.70
GUA^{7}	142.9		136.1		83.20	78.06		65.16	40.26
THY^{8}			137.0	11.97	83.76		83.43	64.71	37.22
GUA^{9}	143.3		135.5		83.70	77.35		66.58	40.29
THY^{10}			137.7	11.96	85.56	73.84	84.79	65.83	38.69
GUA^{11}	143.4		135.8		85.33	75.68	84.66	66.57	41.05
GUA^{12}	143.1		136.2		81.83	75.53	75.27	66.77	41.05
GUA^{13}	142.7		135.5		82.74	76.86	85.06	65.45	37.96
ADE^{14}		153.1	139.5		83.25	75.95	84.65	66.55	37.24
ADE^{15}		152.7	140.0		87.22	75.94	84.46	65.48	38.69
THY^{16}			136.7	11.97	85.08			65.78	37.73
ADE^{17}		152.4	139.4		82.76	76.66	85.33	65.60	37.93
GUA^{18}	143.4		135.9		82.75	76.43	84.87	66.10	38.93
GUA^{19}	143.1		135.3		82.55	76.65	84.67	70.05	38.91
GUA^{20}	143.3		134.8		82.27	76.63	84.43	66.07	37.27
ADE^{21}		152.5	135.2		82.52	76.42	83.95	64.91	38.39
ADE^{22}		152.5	139.1		82.03	76.20	84.62	64.66	39.58

were used. The samples were prepared in potassium NMR buffer (10 mM K_2HPO_4/KH_2PO_4, 50 mM KCl, pH 6.5, 10% D_2O). The oligonucleotide concentrations were between 1 and 3 mM depending on the experiment requirements. The concentrations of oligonucleotide were between 1 and 2 mM. Usually the signals from 10 to 12 ppm are attributed to imino protons of guanine residues involved in Hoogsteen hydrogen bond in tetrads arrangements. Unambiguously assignment of guanine imino protons (H1) and aromatic (H8) proton resonances were possible using 5–8% ^{15}N, ^{13}C enrichment of sequence-specific oligonucleotides using solid-phase chemistry. Individual 1D ^{15}N-HSQC spectra were for recorded at a time for each isotopically labelled guanine. 2D multidimensional spectra ^{1}H–^{13}C HMBC were used to correlate iminos with H8 protons using the $^{13}C5$ of the same guanine. In addition, NOESY, TOCSY and COSY experiments were used to identify the remaining resonances. Overall 83% of all protons (^{1}H); 96% of carbons (^{13}C); 35% of nitrogen's (^{15}N) were assigned. Unfortunately, the guanine amino groups were difficult to assign unambiguously at 37 °C due to fast exchange with water. All spectra were processed with TopSpin software version 3.5. Peak picking and frequency matching was performed within Sparky version 3.114. The ^{1}H, ^{13}C and ^{15}N resonance assignments have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu/) under the BMRB accession code 27,173.

Assignments and discussion

Among our twelve imino peaks, corresponding to H1 protons implicated in hydrogen bonds, some of them are very close or even superimposed such as guanines 3, 4, 9 and 13 which makes assignment more difficult. For this reason, we used HSQC ^{1}H–^{15}N experiment to identify each guanine. Two dimensional experiments with near 5% isotopically enriched guanines (^{15}N and ^{13}C) labelled samples allowed us to distinguish all guanines. Then we performed one dimensional ^{1}H–^{15}N HSQC experiments using 5% ^{15}N, ^{13}C site-specifically labelled KRAS 22RT on each guanine to identify individual imino peaks (Fig. 2). In the case of guanines 3 and 13 we also performed experiments at 20 °C where peaks are well separated (data not shown). Once the assignment of this region was successful, we identified NOE cross peaks between H1 protons of different guanines and then NOE's between H1 protons and H8 protons; finally it was possible to assign all guanines resonances (Fig. 3).

Other residues, especially cytosine 5 and thymine 16, gave distinct and strong signals which allowed us to unambiguously assign some of their resonances. Finally, to confirm our assignment and complete the remaining peaks assignments, we performed HSQC ^{1}H–^{13}C experiments with a 5% ^{15}N, ^{13}C all guanines labelled sample. Even

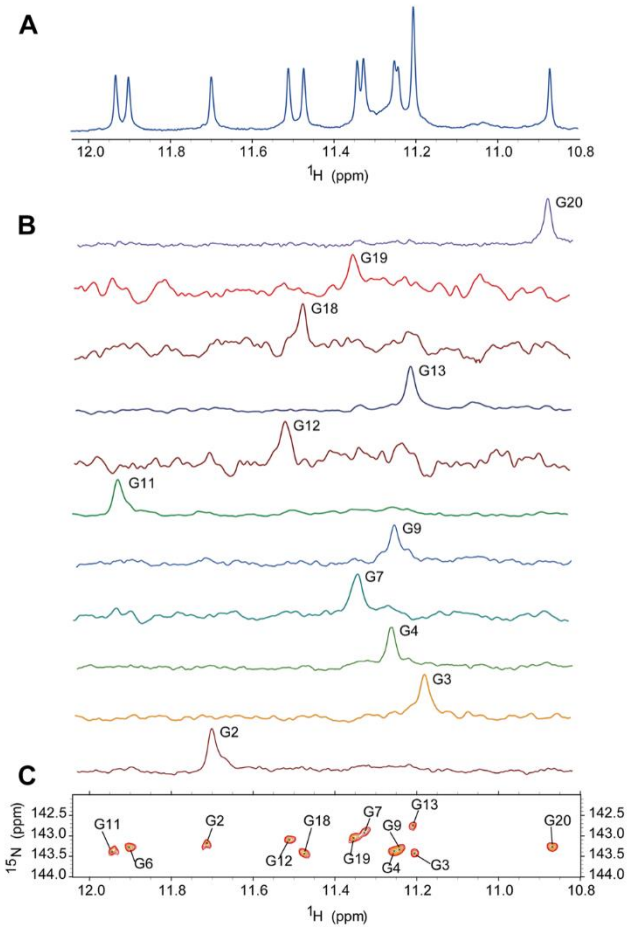


Fig. 2 **a**, One dimension imino protons region with some peaks not separated; **b**, examples of 1D HSQC ^{1}H–^{15}N spectrum used to assign guanines peaks; **c**, HSQC ^{1}H–^{15}N spectrum used to separate peaks in ^{15}N dimension; All experiments are done at 310 K, in NMR buffer using 700 MHz spectrometer

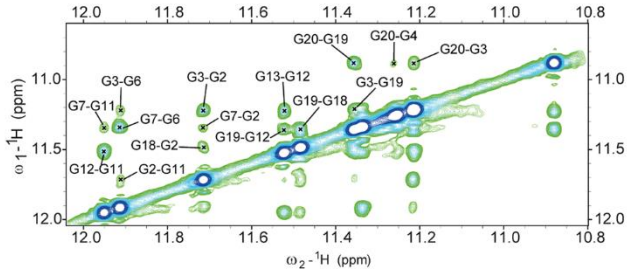


Fig. 3 2D ^{1}H–^{1}H NOESY spectrum obtained at 37 °C and acquired in a 950 MHz spectrometer. The spectra focus in the imino ^{1}H–^{1}H NOE correlations between the different guanines implicated in the formation of the G-quaruplex core

with these spectra, some resonances remained difficult to assign, especially protons of $–NH_2$ groups which exchange too fast with water at 37 °C and did not give a good NMR signal compared to the signal we observed at 20 °C in our

previous work (Kerkour et al. 2017). By careful Nosy spectra comparison, and thanks to the previous structure we determined for 22RT at 25 °C (PDB entry: 5I2V) and the spectra of circular dichroism we can tell that we have an all-parallel G4 with full *anti* guanines. In our ^{1}H–^{1}H NOESY spectra, the *anti*-conformation can be easily seen by looking at the intensities of correlations between H1′ (n) and H8/H6 (n + 1) protons of nucleotides consistent with *anti* glycosidic torsion angles (NMR spectroscopy of G-quadruplexes; Adrian et al. 2012) for all guanine residues. The assignments at 37 °C enable us to calculate the structure at a physiological relevant temperature, necessary to understand ligand binding in biological relevant conditions, also required for cellular drug toxicity testing.

Acknowledgements This work was possible thanks to financial supports from the IR-RMN THC Fr3050 CNRS facilities from Bordeaux (UMS3033/US001) and Gif-sur-Yvette. This work was also supported by iNEXT, Project Number 653706, funded by the Horizon 2020 from the European Union (Brno; Czech Republic). We are particularly grateful to R. Fiala from CEITEC for his help and his advices. Funds for conducting the research were also available from La Ligue contre le Cancer which we gratefully acknowledge. We also thank B. Vialet for synthesis of isotopically labeled oligonucleotides and our team for useful discussions.

References

Adrian M, Heddi B, Phan AT (2012) NMR spectroscopy of G-quadruplexes. Methods 57, 11–24

Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, Flanagan A et al (2004) "The cosmic (catalogue of somatic mutations in cancer) database and website". Br J Cancer 91(2):355–358

Bedrat A, Lacroix L, Mergny JL (2016) "Re-evaluation of g-quadruplex propensity with G4 hunter". Nucleic Acids Res 44(4):1746–1759

Bochman ML, Paeschke K, Zakian VA (2012) "DNA secondary structures: stability and function of G-quadruplex structures". Nat Rev Genet 13(11):770–780

Brito H, Martins AC, Lavrado J, Mendes E, Francisco AP, Santos SA, Ohnmacht SA et al (2015) "Targeting kras oncogene in colon cancer cells with 7-carboxylate indolo [3,2-B] quinoline trialkylamine derivatives". PLoS ONE 10(5):e0126891

Eddy J, Maizels N (2006) "Gene function correlates with potential for G4 DNA formation in the human genome". Nucleic Acids Res 34(14):3887–3896

Hansel-Hertsch R, Di Antonio M, Balasubramanian S (2017) "DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential". Nat Rev Mol Cell Biol 18(5):279 –284

Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, Brosh RM Jr, Sen D, Lansdorp (2013) "Detection of G-quadruplex DNA in mammalian cells". Nucleic Acids Res 42(2):860–869

Huppert JL, Balasubramanian S (2007) "G-quadruplexes in promoters throughout the human genome". Nucleic Acids Res 35(2):406–413

Kerkour A, Marquevielle J, Ivashchenko S, Yatsunyk LA, Mergny JL, Salgado GF (2017) "High-resolution three-dimensional nmr structure of the kras proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation". J Biol Chem 292(19):8082–8091

Largy E, Mergny JL, Gabelica V (2016) "Role of alkali metal ions in G-quadruplex nucleic acid structure and stability." Met Ions Life Sci 16:203–258

Lavrado J, Brito H, Borralho PM, Ohnmacht SA, Kim NS, Leitao C, Pisco S et al (2015) "Kras oncogene repression in colon cancer cell lines by G-quadruplex binding indolo [3,2-C] quinolines". Sci Rep 5:9696

McCormick F (2015) "Kras as a therapeutic target". Clin Cancer Res 21(8):1797–1801

Qin Y, Hurley LH (2008) "Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions." Biochimie 90(8):1149–1171

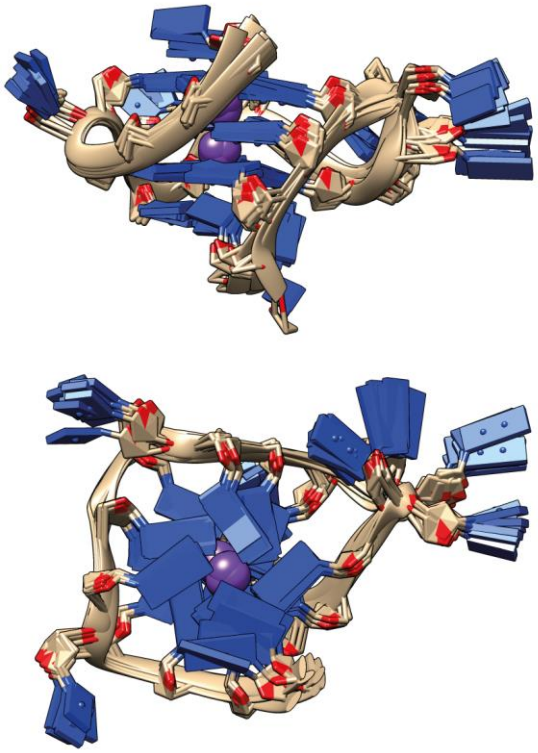
Sen D, Gilbert W (1988) "Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and Its implications for meiosis". Nature 334(6180):364–366

Stegle O, Payet L, Mergny JL, MacKay DJ, Leon JH (2009) "Predicting and understanding the stability of G-quadruplexes". Bioinformatics 25(12):i374–i382

Wlodarczyk A, Grzybowski P, Patkowski A, Dobek A (2005) Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. J Phys Chem B 109(8):3594–3605

II.2. Comparison between KRAS22RT G-quadruplexes at 20°C and 37°C

Based on the results obtained in our KRAS22RT studies at 37°C we determined the G-quadruplex structure of KRAS22RT with PDB deposition with code 6T51 and BMRB deposition with code 27173 **(Figure 27)**.



NMR Restrains	H_2O
distance restrains	489
intraresidue distance restrains	371
sequencial (i, i+1) distance restrains	77
long-range (i, i+2) distance restrains	20
short-range non sequencial distance restrains	21
dihedral restrains	12
H bonds restrains	24
Structural statistics	
structure calculation	
total calculated structures	750
NOE violations	
number (>0.3 Å)	0.04
RMSD of violations (Å)	0.11 ± 0.02
molecular dynamics	
simulation time (ns)	1.0
extracted structures (lowest energy)	10
RMSD	
All heavy atoms (Å)	0.6

Figure 27. The ten lowest structures deposited on PDB under code 6T51 with a table containing restrains used in structure calculation and statistics about the calculated structures

This 3D folding was determined after structure calculations with ARIA/CNS software based on distance retrains obtained by NMR. Usually 750 final structures were calculated and the ten lowest energy were selected. At the end, an NMR refinement protocol was performed with AMBER molecular dynamics software and the ten best structures based on energy deposited in the PDB. A more detailed protocol concerning structure calculation and refinement is described in the KRAS32R G-quadruplex structure determination part. As there is a difference of 17°C degrees between the previous determined structure and the new one, we compared these two structure features. Indeed, the additional energy brought to the system by this temperature difference did allow an increased motion, especially in the most external parts such as G-quadruplex ends or loops. This phenomenon could lead to the creation of new

binding sites or difficulties for ligands by changing the pockets accessibility or increasing the size of grooves and the azimuthal dispersion of loops. In the case of KRAS22RT G-quadruplex **(Figure 28)** the change in temperature did not trigger major changes in the structure but there are several minor changes that can be interesting for ligands studies.

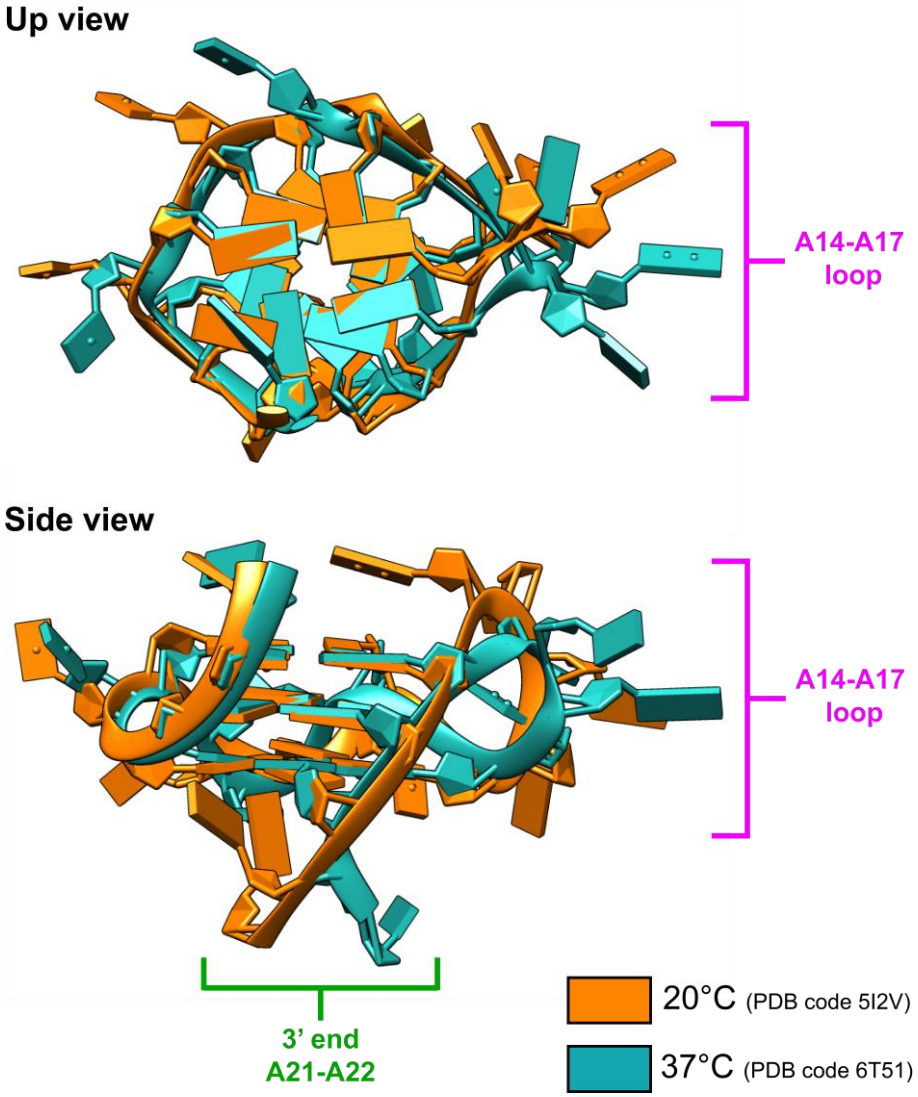


Figure 28. Comparison between KRAS22RT G-quadruplex at 20°C and 37°C with views from the top and from the side.

These minor changes essentially concern the loop between G13 form the bottom tetrad and G19 from the top tetrad. The bases in the two loops seem to be oriented at opposite directions leading in the 37°C temperature structure to a slightly narrower loop. The other difference that can be observed concerns the 3' end of KRAS22RT G-quadruplex. Indeed, 3'end extremities are oriented once again in opposite directions. At 20°C, A22 is oriented toward

the center of the G-quadruplex whereas A21 is oriented toward the outside. At 37°C, both of the residues are oriented toward the inside of the structure leading to a partial obstruction of a potential binding site. There is no other noticeable change between these two structures. Based on the results we obtained with KRAS22RT G-quadruplex structures, especially at 37°C, we then started to study structure KRAS32R G-quadruplex.

III. Determination of KRAS32R G-quadruplex

III.1. Structure of Two G-quadruplexes in equilibrium in the KRAS promoter (Article 2)

This article was submitted on November 6th (2019) where I am the first author. The main results indicated that two main conformations can be formed within KRAS32R. I determined the NMR solution structure of these two conformers called KRAS32R G9T and G25T. These results were obtained after analyzing over 10 NMR 2D spectra and more than 100 1D spectra. Several of them have been recorded at CEITEC (Central European Institute of Technology) where I went in order to perform NMR experiments and learn more about this technique. I would particularly like to thank Radovan FIALA for all these advices and the time he dedicated to me.

Structure of Two G-quadruplexes in equilibrium in the KRAS promoter.

Julien MARQUEVIELLE[1], Coralie ROBERT[1], Olivier LAGRABETTE[1], Mona WAHID[1], Anne BOURDONCLE[1], Luigi XODO[2], Jean-Louis MERGNY[1], Gilmar SALGADO[1*]

1. European Institute of Chemistry and Biology (IECB), ARNA laboratory, INSERM U1212 - CNRS UMR 5320, University of Bordeaux, France.

2. Department of Medicine, Laboratory of Biochemistry, 33100 Udine, Italy.

KEYWORDS: KRAS G-quadruplex; NMR structure.

ABSTRACT: KRAS is one of the most mutated oncogenes and still considered an undruggable target. An alternative strategy would consist in targeting its gene rather than the protein, via the formation of G-quadruplexes (G4) in its promoter. G4 are secondary structure implicated in biological processes, which can be formed among G-rich DNA (or RNA) sequences. Here we have studied the major scaffold of the commonly known KRAS32R or simply 32R, a 32 residue sequence within the KRAS NHE (Nuclease Hypersensitive Element) region. We have determined the structure of the two major conformers that 32R can adopt which are stable and in slow (> ms) equilibrium with each other. By using different biophysical methods, we found that G9, G25, 28 and G32are particularly implicated in the exchange between these two conformations. We also showed that a triad at the 3' end further stabilizes one of the G4 conformations, while the second conformer remains more flexible and less stable.

Cancer is one of the most predominant causes of death around the world representing a major issue for human health, according to the international Agency for Research on Cancer (IARC). Among all different types of cancer, some remain very deadly without therapeutic solutions such as lung, pancreatic and colorectal cancers, where *KRAS* is particularly implicated[1-4]. *KRAS* is one of the most mutated oncogenes located on chromosome 12 and is responsible for the coding of a G-protein of 23.2 kD [5-7]. It plays an important role in cell proliferation, differentiation and survival[8]. Recent studies have also found that *KRAS* reprograms the metabolism and controls the redox homeostasis in pancreatic cancer[8,9] The KRAS protein promotes oncogenic events switching from an inactive form bound to GDP to a constitutively active form bound to GTP[10,11]. However, even if its implication is well documented in a broad range of cancers, *KRAS* is known for being almost undruggable still remaining one of the least successfully therapeutically targets, with very limited efficient treatments and huge side effects[12,13]. A broad range of therapies against *KRAS*-mediated cancers have been developed in the past decades. Most of these therapies have been designed to target either the KRAS protein directly or the enzymes involved in its association to the plasma membrane or downstream effector pathways[9,14-17]. Moreover, recent discoveries on the effect of KRAS on cancer metabolism gave new therapeutic opportunities[18]. As these strategies have yet to be efficient enough for clinical applications [8], new therapeutic approaches focus on upstream elements, by targeting mRNA or directly the gene and its promoter regions[19]. The promoter sequence of several oncogenes, including KRAS, is highly rich in guanine thus allowing the formation of non-canonical structures called G-quadruplexes (G4)[20,21]. G4 consists in the stacking of at least two G-tetrads which are planar structures involving four guanines linked in a Hoogsteen hydrogen bonds network[22,23]. Due to their base composition, number of strands or loop length, these structures are highly often polymorphic adopting structures much different from those observed in canonical double helical DNA[24-27]. As they are present in several key regulatory regions of genes, G-quadruplexes structures have been tested as possible therapeutically targets[28-30], including KRAS[12,14,31-33]. It is particularly attractive to think that small ligands can bind specifically to G4 structures without having any influence on canonical DNA. Moreover, in certain cases the ligands can prevent the binding of proteins that modulate *KRAS* transcription[34-36], reducing the activity of this oncogene near the beginning of the network pathway in KRAS metabolism. Here we examined the major structure of the commonly known KRAS32R or 32R, a 32 nucleotide sequence within the KRAS NHE (Nuclease Hypersensitive Element) region[37]. Knowing the high resolution 3D structure of such unusual DNA element, we may be able to understand the folding and its role in *KRAS* regulation. In addition, the 3D structure is necessary for rational drug design and validation of new targets in *KRAS*-related cancers. The 32R sequence also corresponds to the minimal interaction domain of several transcription factors and related proteins such as MAZ, PARP-1 and hnRNP A1[38]. It has been shown that a small (21R) and a longer 32-mer (32R) region can form stable G-quadruplexes *in vitro* which were revealed using Taq polymerase extension assays[39] and shown to interact with these proteins[31,32,40]. Several studies have been conducted to answer questions about KRAS G4 including one of our previous studies where we proposed some structural insights by resolving structure of a G4 forming in a sequence similar to 21R[41,42] already used to screen for G4 ligands[42,43]. Concerning the longer sequence 32R, structual insights have been proposed using a truncated sequence[43] and Cogoi *et al*.,[40] showed that it is capable of forming two parallel G-quadruplex conformations. In this

work, we determined the high-resolution 3D structure of both 32R conformers that co-exist in equilibrium using solution NMR methods.

RESULTS and DISCUSSION

KRAS 32R WT forms two major distinct conformers with different topologies

In this work, we have used solution NMR spectroscopy to determine the high-resolution structure of the 32R sequence containing six tracts of guanines in KRAS NHE region (**Figure 1A)**. The 1D ^{1}H NMR spectrum of 32R is depicted in **Figure 1B.** The spectrum showed seven well-resolved single-peaks and four double-imino peaks. In addition, other non-resolved broad regions are present such as the one near 11 ppm. Taken together, the number of imino peaks and the peak broadening suggest an important structural polymorphism. We used CD spectroscopy to inspect the polarity of the chain. The CD spectrum shows a clear maximum positive and a negative around 260 nm and 240 nm respectively[44,45] (**Figure 1C**), indicative of a parallel structure. Next, we used 1D ^{15}N-labeling filtered NMR experiments with copies of 32R sequences, each bearing an individual labelled guanine residue. Each guanine was enriched with ~5% of 15N- isotope[46], enough for individual identification of individual guanines. We can observe some interesting results from the individual spectra plotted in **Figure 1D**. Some nucleobases produced clear individual peaks such as G25-G28, while the majority of the guanine bases such as G3, G7 or G13 clearly depicted two or more peaks. All samples presented in the figure were re-synthesized and each respective spectra repeated twice. Taken together, the NMR results suggested either the formation of an asymmetric dimer, the existence of a large G-tetrad core composed by more than 3 tetrads, or a mixture of conformers. Native polyacrylamide electrophoresis for the native sequence 32R and single substitutions G➔T mutations shown in **Figure 2A**, allowed us to exclude dimer formation, except for 9-25* that is a double substituent. The correspondent CD spectra for each mutant are displayed in **Figure 2B**. We also observed that the remaining guanines did not have ^{15}N-filtered imino peaks (**Figure S1**). Cogoi et al 2006, 2011 using DMS-footprinting reported a parallel G-quadruplex structure composed by three G-tetrads and three loops and excluded a dimer possibility. Although some mutants (G2T, G11T, G31T, G32T), showed a gel-band corresponding to a hypothetical dimer, these sequences presented a 1D ^{1}H NMR spectra far from what was observed for the 32R native sequence **Figure 2C**. These mutated DNA sequences have a very complicated 1D NMR spectrum with broad peaks supporting oligomerization. The results imply important structural rearrangements upon G➔T changes emphasizing the importance of those respective guanines. The exception is G2T that does not display any imino peaks as observed for the other G➔T mutants. The results indicated that high order and/or unstable species not observed in the NMR time scale, are present since the CD signal and the native gel indicates the presence of parallel G4 structure with exception for the CD spectrum of G32T. Other base mutations (X➔T) were tested (**Figure S2)** in order to understand the role of each residue in the structure of 32R. The respective ^{1}H NMR spectrum of the imino regions showed different results depending on the mutated residue and we could distinguish four different categories of mutations (**Figure S2**). (1) Mutations of residues already known to be important for the folding of the G-quadruplex core, confirmed by ^{15}N-filtered experiments. These mutations lead without surprise to a dramatic effect on the G-quadruplex imino region. For example, the G2 or G11 mutations seem to disrupt completely the imino folding pattern that otherwise should be responsible for the existence of a single conformation in the sample, but when analyzed by CD an all-parallel conformer seems to exist. The results point that the CD observes a broad mixture of parallel conformers, which reduce significantly the signal in the NMR spectra giving the impression that the G4 is almost inexistent. The respective native gel sample displays a band compatible with a dimer or some other high order species. (2) Mutation of some of the guanine not implicated in the tetrad core such as G18 and G23 resulted in much cleaner spectra with sharper peaks. We presume that these mutations lead to the elimination of minor confirmation sub-states. (3) The last three residues (A30 to G32) mutations unexpectedly had a great impact in the imino peak-pattern, although not as severe as G2 or G11 mutations. A G-quadruplex imino region was well visible but without sharp peaks, indicating that a large number of different conformations co-existed. (4) Finally, the last type of G➔T mutations we observed were those corresponding to residues identified to be part of the tetrad core, and showed well-resolved spectra such as G9 and G25. Among those, we can distinguish two patterns of imino peaks. One group composed by G9, G18-G23 and G29, and a second group by guanines G25, G26, G27 and G28. Hereafter and for simplification, we refer to those two groups of mutants as G9T and G25T respectively. The spectrum of G9T showed thirteen well-resolved imino peaks. For G25T it was more difficult to correctly quantify the number of peaks due to the superposition around 11.3 ppm. Both mutants had a parallel signature in CD (**Figure 2b**) and the native gel showed a unique band with the same migration as the wild type. The imino peaks profile for G25T is more complex and different compared to G9T. We can easily identify an overlapped region around 11.2 ppm similar to the one we can observe for 32R WT (**Figure 2C**). The formation of these two principal groups of imino-peaks pattern indicates that two major distinct conformers may coexist among other minor populations, forming together the species observed in the 32R WT sample. To test the possibility that KRAS32R WT adopts alternative stable-conformers besides the G9T and G25T structures, we performed double mutations and inspect those using NMR and CD spectroscopy (**Figure S3**). We observed that the double mutant G9TG25T could not form a proper G4, although some multiple peaks with much smaller intensities are visible in the imino region. As observed before, replacing G2 for a thymine severely affects both G4 conformers. We hypothesize that G2 may be important for the folding initiation and stabilization as observed before with WT sequence. Overall G9T is unique because no other guanine can take its place and give rise to a single well folded conformer. For example, G9T and G32T or G31T double mutants present multiple imino peaks with characteristic of multiple conformations. On the other hand, the G25TG26T double mutant still produces a clear imino pattern with twelve sharp peaks similar to what we found with G25T (**Figure S6**). The results indicate that there is a sliding in the last track passing from G25-G27 to G27-G29, which produces a less-stable conformer with an almost identical imino pattern.

Both KRAS 32R G9T and G25T adopt parallel conformations with propeller loops topologies. G9T contains a *syn* fold-back 3' end that increases its stability.

In order to get insight in the topologies adopted by G9T and G25T, we performed additional CD and NMR experiments. CD

and 1D ^{1}H NMR was used to perform melting experiments and obtain a T_m (**Figure S4A**). The CD melting results shows that 32R WT, G9T and G25T melt with temperatures of 59, 57.5 and 56 °C respectively for the forward (heating) experiment, and 50, 52, and 48 °C respectively, for the backward (cooling) experiments. The NMR melting (**Figure S4B**) confirms that G25T is the least stable conformer. These and the 1D 1H spectra results support our hypothesis that the G9T conformer is more stable than the G25T conformer. We proceeded to identify each guanine implicated in G9T and G25T conformers using site-specific low-enrichment ^{15}N-labeling (**Figure S5**). The results revealed that each imino proton in G9T mutant was implicated in a single conformation and, in the case of the G25T mutant, some of the imino peaks were involved in multiple conformations, such as G4, G7 and G11. Interestingly, the 1D 1H NMR spectra of G9T presented thirteen peaks instead of the twelve expected for a G-quadruplex with 3 tetrads of guanines. Together with through-bond HCCNH-TOCSY and the through-space {^{1}H-^{1}H} NOESY spectra (250 ms as mixing time), we were able to identify all imino protons and stablish the correlations with the respective aromatic protons (H8) for G9T and G25T mutants, (**Figure 3 and Figure S6** respectively). From there, all the H6/8-H1' NOE sequential connectivities could be mapped for all 32 bases (**Figures S7 and S8** respectively). Additional experiments with mutants by 1D ^{1}H NMR, CD and PAGE (**Figures S9, S10** for G9T and **S11, S12** for G25T) together with 2D classical NMR experiments such as {^{31}P-^{1}H} HSQC, {^{13}C-^{1}H} HSQC, {^{1}H-^{1}H} TOCSY and COSY were used to interpret and support the assignments. We identified that the three tetrad-layered guanine core was composed by G2-G6-G11-G25, G3-G7-G12-G26, G4-G32-G13-G27 for G9T and by G2-G6-G11-G26, G3-G7-G12-G27, G4-G9-G13-G28 for G25T. According to the H1′-H8 correlations observed in the ^{1}H-^{1}H NOESY spectrum, all guanine bases, except G32, were in the *anti* glycosidic position[47]. In G9T, G32 folds back within the third strand and completes the tetrad formed at the 3' tail. In addition to the tetrad core with a folded-back guanine, G9T showed NOEs connectivities between G28, A30 and G31 compatible with the formation of a triad. Some of the NOEs included correlations between A30 (H8) and G28 (NH2), between A30 (H8) and G31 (H8), and between G31 (H8) and G28 (NH2), among other diverse sugar correlations. Both mutants have three tetrads of guanines as central core with parallel strands connected differently in terms of loops. G9T has a short one-base propeller loop (C5), one three-base propeller loop (T8-T10) and a long propeller loop composed by eleven-nucleotides (A14-G24). Finally, G28 through G31 form a diagonal loop that encase a triad formed by G28-A30-G31. Together with the diagonal loop, the triad caps and stabilizes[51,52] the 3' tail of the G9T conformer. In the case of G25T conformer, it contains two short one-nucleotide propeller loops (C5) and (T10), a bulge (T8), and a long propeller loop composed by eleven-nucleotides (A14-T25). A simple folding topology of both models is depicted in **Figure 4**. The conformation exchange between both structures is achieved in G9T by G25 base sliding-out from the tetrad core and simultaneously G26 takes its place; G27 slides to the central tetrad and finally G28 sliding-in forming part of the tetrad core in G25T. This sliding effect in the same direction is also present in the case of a double mutant G25T-G26T where the imino peak of G29, which is not observed in G25T, is well visible in G25T-G26T (**Figure S13**). One of the consequences was that G32 could not fold-back and be part of the guanine tetrad core of G25T conformer. A second important consequence is the loss of the triad (G28, A30 and G31) at the 3'-end in the G25T mutant. Overall G25T has less stable structural elements, in agreement with CD and NMR melting experiments.

NMR solution structure of KRAS 32R G9T and G25T

Unless otherwise specified, the NMR solution structures of both G9T and G25T mutants were calculated with spectra collected at 37°C. The necessary restraints were extracted from {^{1}H-^{1}H} NOESY experiments with a mixing time of 250 ms. A total of 873 and 564 NOE-derived restraints were obtained for G9T and G25T, respectively (**Tables S1 and S2**). **Figure 5** depicts the ten lowest-energy structures superimposed obtained after MD simulations for structure refinement purposes in presence of water and K^{+} counter-ions (cf. Material and Methods). G9T had all principal structural motifs (loops and tetrad-core) converged during the MD simulation, with a final heavy-atom RMSD of 1.5 Å. On the other hand, due to the increased peak broadening and less amount of NOEs obtained, G25T have a looser global conformation with a heavy-atom RMSD of 1.9 Å. That was particularly evident in the long propeller loop and in the 3'-end. The results are supported by the difference in the number of NOEs collected from both mutants, and most importantly for the nucleotides in the long propeller loop. Consequently, the long propeller loop in G9T presents more π-π interactions between bases such as A24, G23 and A22 and less dispersion of the nucleotides (**Figure 5)**. In addition, the bases from A1, T10 and A24 nucleotides in G9T also participate in π-π interactions, further stabilizing the top of the first tetrad (G2-G6-G11-G25). At the 3'-end, G32 folds back and participates to the third tetrad (G4-G32-G13-G27). In the case of G25T, only A1 and T8 seem to provide the same sort of stabilization over the tetrad core. The triad observed in G9T runs diagonally and connects two opposite tracts composed by G6-G7-G32 and G25-G26-G27 respectively. The triad structure is stabilized by π-π interactions with the closest tetrad of guanines and by a network of hydrogen bonds (HB) comprising one formed between G28 with A30 and two between G28 and G31. In the case of G28-A3, one of the hydrogens (H21) of the amine group acts as donor and N6 from A30 is the acceptor. Two other HBs can also be observed between G28 and G31, involved the same type of atoms, and between G28 and G31 using H1 and O6, respectively. In addition, we observed that G28 and G31 have their respective H1 protons outwards in exchange with the solvent, and that may explain why we could not observe the corresponding imino peaks in ^{15}N-filtered 1D NMR experiments at 37 °C (**Figure S1**), nor when spectra were acquired at 4°C. The formation of the triad was also supported from observed inter-NOEs corresponding to H8 protons from both A30 and G31, which come closer and share several NOEs with respective sugar protons nearby. The triad structure has the inter-proton NOEs color mapped depicted in **Figure 6**. NMR spectra collected at 47 °C still had the respective H8-H8 inter-base NOEs between A30 and G31(data not shown). In addition, when we inspected the 32RWT {^{1}H-^{1}H} NOESY spectra we could identify to the exception of G31H8-G28H21/22, all the same respective cross-peaks correlations, indicating that the triad is well preserved in the WT sample, including G28 with A30 and between G28 and G31 (**Figure S14)**.

Conclusions

G-quadruplexes have emerged in the last decade as possible new therapeutically targets for developing new anticancer strategies[48-50]. Validating the G4's as possible targets requires a deeper understanding of their organization, structure and mode of action. In this study, we have determined the high-resolution structure, at physiological temperature and in K^+ solution, of two G-quadruplexes existing in a 32-mer G-rich sequence in the NHE region of the KRAS promotor sequence. We found that two G4s coexist in equilibrium giving rise to two different G4 structures. Both structures are parallel-stranded and have nearly identical stability. Both conformers contain long loops that have little influence in the global stability of each respective structure, but one of the structures presents a folded-back guanine and a 3-end triad that increases the thermal stability. Both structures support the DMS footprinting pattern exhibited by the 32-mer G-rich sequence of KRAS[39]. Our new results also demonstrate that simple DNA sequences rich in guanines can sample multiple conformations giving rise to stable ensembles of G-quadruplexes. They may look similar under certain biochemical methods, but high resolution structures can unveil important differences worth exploring in the development of small molecules for therapeutic purposes.

Material and methods

DNA Samples Preparation

Uniformly {^{13}C, ^{15}N}-labeled KRAS32R WT DNA was purchased from Silantes GmbH. and G9T and G25T samples were produced in our lab using M-MuLV enzyme from New England Biolabs and dNTPs from Eurisotop. Non labelled samples were purchased from IDT (Integrated DNA Technologies; USA). They were synthesized on the 250 nmole scale and purified by reverse HPLC. 5% ^{15}N, ^{13}C site-specifically labelled samples were synthesized in our laboratory (INSERM U1212, Bordeaux) on an automated Expedite 8909 DNA synthesizer at 1 µmol scale on 1000-Å primer support (Link Technologies SynBase CPG). All the standard phosphoramidites (dABz, dT, dGiBu, dCAc), reagents, and solvents used during the synthesis were purchased from Glen Research. The dGiBu-phosphoramidite (U-^{13}C10, 98%; U-^{15}N5, 98%; CP 95%) was purchased from Cambridge Isotope Laboratories. After the synthesis, the oligonucleotides were cleaved from the support and the nucleobases were deprotected with ammonium hydroxide at 55 °C for 16 hours and then lyophilized. Oligonucleotides solutions were prepared in 1X buffer (10 mM K_2HPO_4/KH_2PO_4; 50 mM KCl; pH 6,5). In order to form G-quadruplexes, samples are heated at 95°C during 5 minutes then quickly cooled in ice. This process is repeated 4 times and sample are conserved at 4°C. All NMR samples presented in the manuscript were re-synthesized and each spectra repeated more than once.

Circular Dichroism

All Circular Dichroism experiments were realized with a Jasco J-1500 CD spectrapolarimeter using Spectra Manager software. Quartz cuvets were used and contained 500 µL of 5 µM oligonucleotide samples in 1X buffer. The CD spectra were measured in the region between 220 and 320 nm with a scan speed of 50 nm/min and a response time of 1 s. Three scans were collected and averaged. All experiments were done at 37°C degrees.

CD melting studies were performed on 5 µM DNA sample alone. CD melting was performed in the single wavelength mode. The data were collected at several wavelengths. Temperature was changed from 4 to 90 °C with 1 °C interval, and 0.4 °C/min rate. These parameters lead to an overall acquisition time of 1h per 10 °C temperature change.

Native Gel Electrophoresis

Native gel experiments were performed using 8% acrylamide/bisacrylamide (19:1 ration) gel containing 10 mM KCl and 1X TBE buffer with a running buffer containing also 10 mM KCl and 1X TBE buffer in native conditions by cooling down the system at 4°C. A low molecular weight DNA ladder purchased from New England Biolabs had been used (range from 25 bp to 766 bp) in each gel. Each sample was diluted in 20 µL of 1X buffer with 6 µL sucrose to reach a final concentration at 25 µM. Before loading samples, a migration was performed in order to eliminate impurities. Then samples were loaded and migrated for 3 hours. Stains-All coloration was used to reveal oligonucleotides bands.

NMR Spectroscopy

NMR spectra were recorded on Bruker Advance 700, 850 and 950 MHz spectrometer equipped with cryogenically cooled probes. Experiments were performed at 37 °C. For solution NMR, standard 3 mm NMR tubes were used. The samples were prepared in 1X buffer with 10% D_2O. The concentration of DNA samples was between 0.35 mM and 2 mM. ^{15}N and ^{13}C HSQC experiments were performed using full labelled samples or partially with sing 5% ^{15}N, ^{13}C site-specific low-enrichment labelling for imino protons for assignments.

To unambiguous assign the intranucleotide exchangeable imino to the nonexchangeable purine H8 and pyrimidine H6 protons through-bond connectivities, we used HCCNH-TOCSY experiment in uniformly {^{13}C, ^{15}N}-labeled DNA[51]. Melting experiments were performed using 500 µM of oligonucleotides. The tube was put in spectrometer to a start temperature of 4°C (277K). After 5 minutes enough time for temperature stabilization, a 1H spectrum was recorded. The same process was repeated until a maximum temperature of 60°C (333K).

Structure calculation

32R G9T and G25T G-quadruplex structures were calculated using a ARIA2/CNS based on NMR-derived distance and dihedral constraints. Distance constraints were manually extracted from volume obtained by integrating peaks in Sparky and CCPNMR software and classified as weak (4.0–6.5 Å), medium (2.5–4.5 Å), and strong (1.8–3 Å). Dihedral angles were constrained based on intraresidue H1′-H8 NOE peak intensities and the canonical B-DNA backbone conformation for the stem-loop. After calculation two different structures which had lower violations were chosen for AMBER refinement. More detailed protocols are provided in Supplementary Information.

Data Deposition

The NMR chemical shifts of 32R G9T and G25T have been deposited in the Biological Magnetic Resonance Bank (accession code 34431 and 34441) and their coordinates have been deposited in the Protein Data Bank (**accession code 6SUU for G9T and 6T2G for G25T**).

ASSOCIATED CONTENT

Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Gilmar F. Salgado Gilmar.salgado@u-bordeaux.fr
Orcid: Gilmar F. Salgado: 0000-0002-0296-5979

Orcid: Julien Marquevielle 0000-0001-8155-5195

Orcid: Jean-Louis Mergny 0000-0003-3043-8401

Present Addresses

J.L.M.: Institut Curie, Univ. Paris Saclay, F-91405 Orsay, France.

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work on the 800 MHz NMR spectrometer (IECB) was possible thanks to support from TGIR-RMN-THC Fr3050 CNRS, CNRS UMS3033, Inserm US001, Univ. Bordeaux. The project also received support from iNEXT, grant number 653706, funded by the Horizon 2020 programme of the European Commission. We appreciate the numerous discussion with Radovan Fiala and the access to the National NMR Centre at CEITEC MU, Brno, Czech Republic. Finally, we appreciate the support from *La ligue contre le Cancer, comité de Gironde*.

ABBREVIATIONS

KRAS: V-ki-ras2 Kirsten Rat Sarcoma Viral Oncogene Homolog
NHE: Nuclease Hypersensitive Element
GDP: Guanosine Di-Phosphate
GTP: Guanosine Tri-Phosphate
G4: G-quadruplex
NMR: Nuclear Magnetic Resonance
CD: Circular Dichroism
DMS: DiMethylSulphate
MD: Molecular Dynamics
NOESY: Nuclear Overhauser Effect Spectroscopy
COSY: COrrelated Spectroscopy
TOCSY: TOtal Correlation Spectroscopy
HSQC: Heteronuclear Single Quantum Coherence spectroscopy
PAGE: PolyAcrylamide Gel Electrophoresis
TBE: Tris Borate EDTA

REFERENCES

(1) Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries., *CA Cancer J Clin* **2018**, 68(6):394-424.

(2) Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer statistics, 2018., *CA Cancer J Clin* **2018**, *68*, 7.

(3) Riely, G. J.; Kris, M. G.; Rosenbaum, D.; Marks, J.; Li, A.; Chitale, D. A.; Nafa, K.; Riedel, E. R.; Hsu, M.; Pao, W.; Miller, V. A.; Ladanyi, M., Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma., *Clin Cancer Res* **2008**, *14*, 5731.

(4) Karnoub, A. E.; Weinberg, R. A., Ras oncogenes: split personalities., *Nat Rev Mol Cell Biol* **2008**, *9*, 517.

(5) Riely, G. J.; Ladanyi, M., KRAS mutations: an old oncogene becomes a new predictive biomarker., *J Mol Diagn* **2008**, *10*, 493.

(6) Tan, C.; Du, X., KRAS mutation testing in metastatic colorectal cancer., *World J Gastroenterol* **2012**, *18*, 5171.

(7) Krasinskas, A. M.; Chiosea, S. I.; Pal, T.; Dacic, S., KRAS mutational analysis and immunohistochemical studies can help distinguish pancreatic metastases from primary lung adenocarcinomas., *Mod Pathol* **2014**, *27*, 262.

(8) Cox, A. D.; Fesik, S. W.; Kimmelman, A. C.; Luo, J.; Der, C. J., Drugging the undruggable RAS: Mission possible?, *Nat Rev Drug Discov* **2014**, *13*, 828.

(9) Ying, H.; Kimmelman, A. C.; Lyssiotis, C. A.; Hua, S.; Chu, G. C.; Fletcher-Sananikone, E.; Locasale, J. W.; Son, J.; Zhang, H.; Coloff, J. L.; Yan, H.; Wang, W.; Chen, S.; Viale, A.; Zheng, H.; Paik, J. H.; Lim, C.; Guimaraes, A. R.; Martin, E. S.; Chang, J.; Hezel, A. F.; Perry, S. R.; Hu, J.; Gan, B.; Xiao, Y.; Asara, J. M.; Weissleder, R.; Wang, Y. A.; Chin, L.; Cantley, L. C.; DePinho, R. A., Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism., *Cell* **2012**, *149*, 656.

(10) Yang, M. H.; Nickerson, S.; Kim, E. T.; Liot, C.; Laurent, G.; Spang, R.; Philips, M. R.; Shan, Y.; Shaw, D. E.; Bar-Sagi, D.; Haigis, M. C.; Haigis, K. M., Regulation of RAS oncogenicity by acetylation., *Proc Natl Acad Sci U S A* **2012**, *109*, 10843.

(11) Maurer, T.; Garrenton, L. S.; Oh, A.; Pitts, K.; Anderson, D. J.; Skelton, N. J.; Fauber, B. P.; Pan, B.; Malek, S.; Stokoe, D.; Ludlam, M. J.; Bowman, K. K.; Wu, J.; Giannetti, A. M.; Starovasnik, M. A.; Mellman, I.; Jackson, P. K.; Rudolph, J.; Wang, W.; Fang, G., Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity., *Proc Natl Acad Sci U S A* **2012**, *109*, 5299.

(12) Bournet, B.; Buscail, C.; Muscari, F.; Cordelier, P.; Buscail, L., Targeting KRAS for diagnosis, prognosis, and treatment of pancreatic cancer: Hopes and realities., *Eur J Cancer* **2016**, *54*, 75.

(13) Ryan, M. B.; Corcoran, R. B., Therapeutic strategies to target RAS-mutant cancers., *Nat Rev Clin Oncol* **2018**, Nov;15(11):709-720.

(14) Xie, C.; Li, Y.; Li, L. L.; Fan, X. X.; Wang, Y. W.; Wei, C. L.; Liu, L.; Leung, E. L.; Yao, X. J., Identification of a New Potent Inhibitor Targeting KRAS in Non-small Cell Lung Cancer Cells., *Front Pharmacol* **2017**, *8*, 823.

(15) Sun, Q.; Burke, J. P.; Phan, J.; Burns, M. C.; Olejniczak, E. T.; Waterson, A. G.; Lee, T.; Rossanese, O. W.; Fesik, S. W., Discovery of small molecules that bind to

K-Ras and inhibit Sos-mediated activation., *Angew Chem Int Ed Engl* **2012**, *51*, 6140.

(16) Leung, E. L. H.; Luo, L. X.; Liu, Z. Q.; Wong, V. K. W.; Lu, L. L.; Xie, Y.; Zhang, N.; Qu, Y. Q.; Fan, X. X.; Li, Y.; Huang, M.; Xiao, D. K.; Huang, J.; Zhou, Y. L.; He, J. X.; Ding, J.; Yao, X. J.; Ward, D. C.; Liu, L., Inhibition of KRAS-dependent lung cancer cell growth by deltarasin: blockage of autophagy increases its cytotoxicity., *Cell Death Dis* **2018**, *9*, 216.

(17) Cox, A. D.; Der, C. J.; Philips, M. R., Targeting RAS Membrane Association: Back to the Future for Anti-RAS Drug Discovery?, *Clin Cancer Res* **2015**, *21*, 1819.

(18) Pupo, E.; Avanzato, D.; Middonti, E.; Bussolino, F.; Lanzetti, L., KRAS-Driven Metabolic Rewiring Reveals Novel Actionable Targets in Cancer., *Front Oncol* **2019**, *9*, 848.

(19) Kim, W.; Lee, S.; Kim, H. S.; Song, M.; Cha, Y. H.; Kim, Y. H.; Shin, J.; Lee, E. S.; Joo, Y.; Song, J. J.; Choi, E. J.; Choi, J. W.; Lee, J.; Kang, M.; Yook, J. I.; Lee, M. G.; Kim, Y. S.; Paik, S.; Kim, H. H., Targeting mutant KRAS with CRISPR-Cas9 controls tumor growth., *Genome Res* **2018**, 28: 374-382.

(20) Bedrat, A.; Lacroix, L.; Mergny, J. L., Re-evaluation of G-quadruplex propensity with G4Hunter., *Nucleic acids research* **2016**, Feb 29;44(4):1746-59.

(21) Wong, H. M.; Stegle, O.; Rodgers, S.; Huppert, J. L., A Toolbox for Predicting G-Quadruplex Formation and Stability, *J Nucleic Acids* **2010**, *Article ID 564946, 6 pages, 2010*.

(22) Sannohe, Y.; Sugiyama, H., Overview of formation of G-quadruplex structures., *Curr Protoc Nucleic Acid Chem* **2010**, *Chapter 17*, Unit 17 2 1.

(23) Stegle, O.; Payet, L.; Mergny, J. L.; MacKay, D. J.; Leon, J. H., Predicting and understanding the stability of G-quadruplexes., *Bioinformatics* **2009**, *25*, i374.

(24) Tippana, R.; Xiao, W.; Myong, S., G-quadruplex conformation and dynamics are determined by loop length and sequence., *Nucleic Acids Res* **2014**, *42*, 8106.

(25) Cang, X.; Sponer, J.; Cheatham, T. E., 3rd, Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes., *Nucleic Acids Res* **2011**, *39*, 4499.

(26) Le, H. T.; Dean, W. L.; Buscaglia, R.; Chaires, J. B.; Trent, J. O., An investigation of G-quadruplex structural polymorphism in the human telomere using a combined approach of hydrodynamic bead modeling and molecular dynamics simulation., *J Phys Chem B* **2014**, *118*, 5390.

(27) Dai, J.; Carver, M.; Yang, D., Polymorphism of human telomeric quadruplex structures., *Biochimie* **2008**, *90*, 1172.

(28) Bochman, M. L.; Paeschke, K.; Zakian, V. A., DNA secondary structures: stability and function of G-quadruplex structures., *Nat Rev Genet* **2012**, *13*, 770.

(29) Lipps, H. J.; Rhodes, D., G-quadruplex structures: in vivo evidence and function., *Trends Cell Biol* **2009**, *19*, 414.

(30) Gomez-Marquez, J., DNA G-quadruplex: structure, function and human disease., *FEBS J* **2010**, *277*, 3451.

(31) Amato, J.; Madanayake, T. W.; Iaccarino, N.; Novellino, E.; Randazzo, A.; Hurley, L. H.; Pagano, B., HMGB1 binds to the KRAS promoter G-quadruplex: a new player in oncogene transcriptional regulation?, *Chem Commun (Camb)* **2018**, *54*, 9442.

(32) Cogoi, S.; Rapozzi, V.; Cauci, S.; Xodo, L. E., Critical role of hnRNP A1 in activating KRAS transcription in pancreatic cancer cells: A molecular mechanism involving G4 DNA., *Biochim Biophys Acta* **2017**, *1861*, 1389.

(33) Carvalho, J.; Pereira, E.; Marquevielle, J.; Campello, M. P. C.; Mergny, J. L.; Paulo, A.; Salgado, G. F.; Queiroz, J. A.; Cruz, C., Fluorescent light-up acridine orange derivatives bind and stabilize KRAS-22RT G-quadruplex., *Biochimie* **2017**, Jan;144:144-152.

(34) Neidle, S., The structures of quadruplex nucleic acids and their drug complexes., *Curr Opin Struct Biol* **2009**, *19*, 239.

(35) Cogoi, S.; Paramasivam, M.; Filichev, V.; Geci, I.; Pedersen, E. B.; Xodo, L. E., Identification of a new G-quadruplex motif in the KRAS promoter and design of pyrene-modified G4-decoys with antiproliferative activity in pancreatic cancer cells., *J Med Chem* **2009**, *52*, 564.

(36) Pattanayak, R.; Basak, P.; Sen, S.; Bhattacharyya, M., Interaction of KRAS G-quadruplex with natural polyphenols: A spectroscopic analysis with molecular modeling., *Int J Biol Macromol* **2016**, *89*, 228.

(37) Cogoi, S.; Ferino, A.; Miglietta, G.; Pedersen, E. B.; Xodo, L. E., The regulatory G4 motif of the Kirsten ras (KRAS) gene is sensitive to guanine oxidation: implications on transcription., *Nucleic Acids Res* **2018**, *46*, 661.

(38) Bossone, S. A.; Asselin, C.; Patel, A. J.; Marcu, K. B., MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination, *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 7452.

(39) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination., *Acta Crystallogr D Biol Crystallogr* **1998**, *54*, 905.

(40) Cogoi, S.; Paramasivam, M.; Spolaore, B.; Xodo, L. E., Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins., *Nucleic Acids Res* **2008**, *36*, 3765.

(41) Kerkour, A.; Marquevielle, J.; Ivashchenko, S.; Yatsunyk, L. A.; Mergny, J. L.; Salgado, G. F., High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation., *J Biol Chem* **2017**, *292*, 8082.

(42) Marquevielle, J.; Kumar, M. V. V.; Mergny, J. L.; Salgado, G. F., 1H, 13C, and 15N chemical shift assignments of a G-quadruplex forming sequence within the KRAS proto-oncogene promoter region., *Biomolecular NMR assignments* **2018**, *12*, 123.

(43) Podbevsek, P.; Plavec, J., KRAS promoter oligonucleotide with decoy activity dimerizes into a unique topology consisting of two G-quadruplex units., *Nucleic Acids Res* **2016**, *44*, 917.

(44) Brunger, A. T., Version 1.2 of the Crystallography and NMR system., *Nat Protoc* **2007**, *2*, 2728.

(45) Randazzo, A.; Spada, G. P.; da Silva, M. W., Circular dichroism of quadruplex structures., *Top Curr Chem* **2013**, *330*, 67.

(46) Vorlickova, M.; Kejnovska, I.; Bednarova, K.; Renciuk, D.; Kypr, J., Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes., *Chirality* **2012**, *24*, 691.

(47) Webba da Silva, M., NMR methods for studying quadruplex nucleic acids., *Methods* **2007**, *43*, 264.

(48) Sengupta, P.; Chattopadhyay, S.; Chatterjee, S., G-Quadruplex surveillance in BCL-2 gene: a promising therapeutic intervention in cancer treatment., *Drug Discov Today* **2017**, *22*, 1165.

(49) Francisco, A. P.; Paulo, A., Oncogene Expression Modulation in Cancer Cell Lines by DNA G-Quadruplex-Interactive Small Molecules., *Curr Med Chem* **2017**, *24*, 4873.

(50) Asamitsu, S.; Obata, S.; Yu, Z.; Bando, T.; Sugiyama, H., Recent Progress of Targeted G-Quadruplex-Preferred Ligands Toward Cancer Therapy., *Molecules* **2019**, *24(3), 429*.

(51) Sklenar, V.; Dieckmann, T.; Butcher, S. E.; Feigon, J., Through-bond correlation of imino and aromatic resonances in 13C-, 15N-labeled RNA via heteronuclear TOCSY., *J Biomol NMR* **1996**, *7*, 83.

Manuscript Main Figures

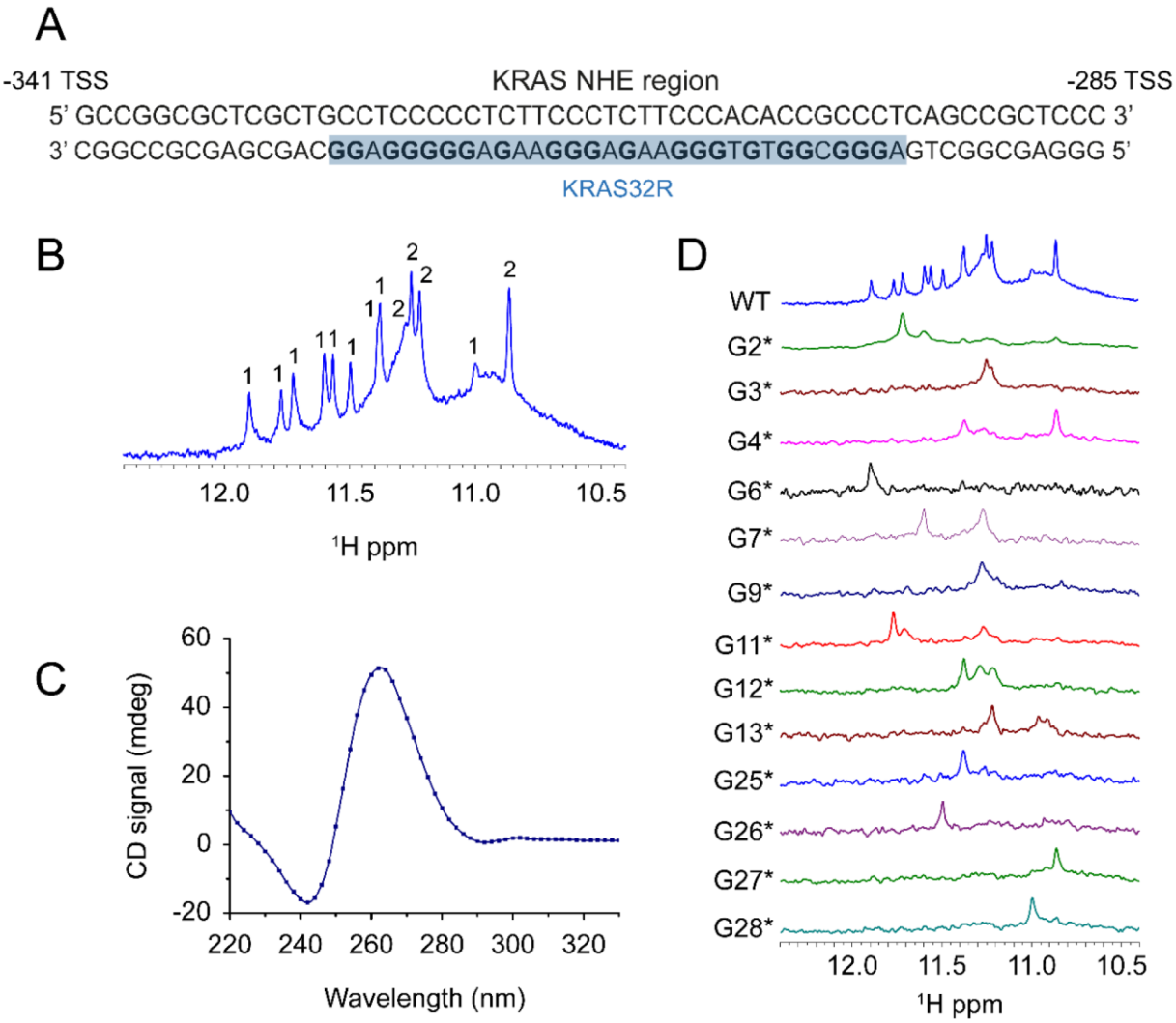


Figure 1. KRAS32R characterization with in **A**) its localisation within KRAS NHE sequence in the promoter region; **B**) its 1D proton NMR imino region at 37°C in 700MHz spectrometer with several countable peaks; **C**) its CD spectrum at 37°C from 220 to 320 nm with positive peak around 260 nm and negative peak around 240 nm characteristic of parallel G4 and **D**) determination of the guanines implicated in KRAS32R tetrads using 15N-filtered NMR spectra of samples containing ≈5% of 15N-enriched isotope. All experiments were performed in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

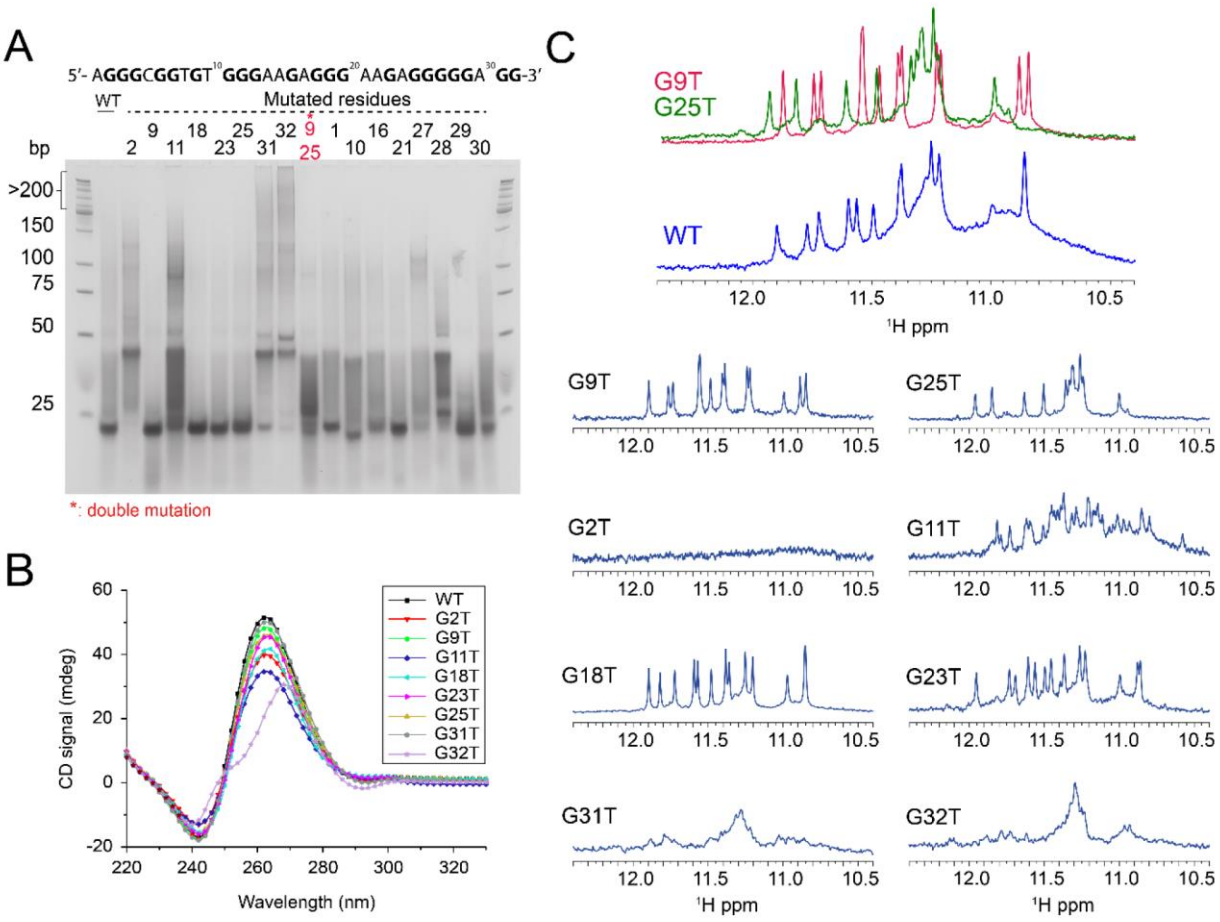


Figure 2. **A**) Native gel experiment of KRAS32R and several simple mutants and one double mutant showing that some mutants harbour the same conformation as wild type whereas several mutants such as G2T or G32T severely impact G4 formation; **B**) KRAS32R mutants' CD spectra at 37°C show that most of the mutations do not affect G4 conformation showing all parallel conformation with characteristic peaks around 260 nm (positive) and 240 nm (negative) except G32T showing signal similar to hybrid G4; **C**) KRAS32R mutants imino region at 37°C showing different effects of mutation in agreement with results obtained in native gel demonstrating the existence of two major conformations and the role of the 3'end residues. Imino region comparison between KRAS32R wild type (blue), G9T (red) and G25T (green) showing that profiles from G9T and G25T can be found in wild type spectra as an evidence of the model. All experiments were performed in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

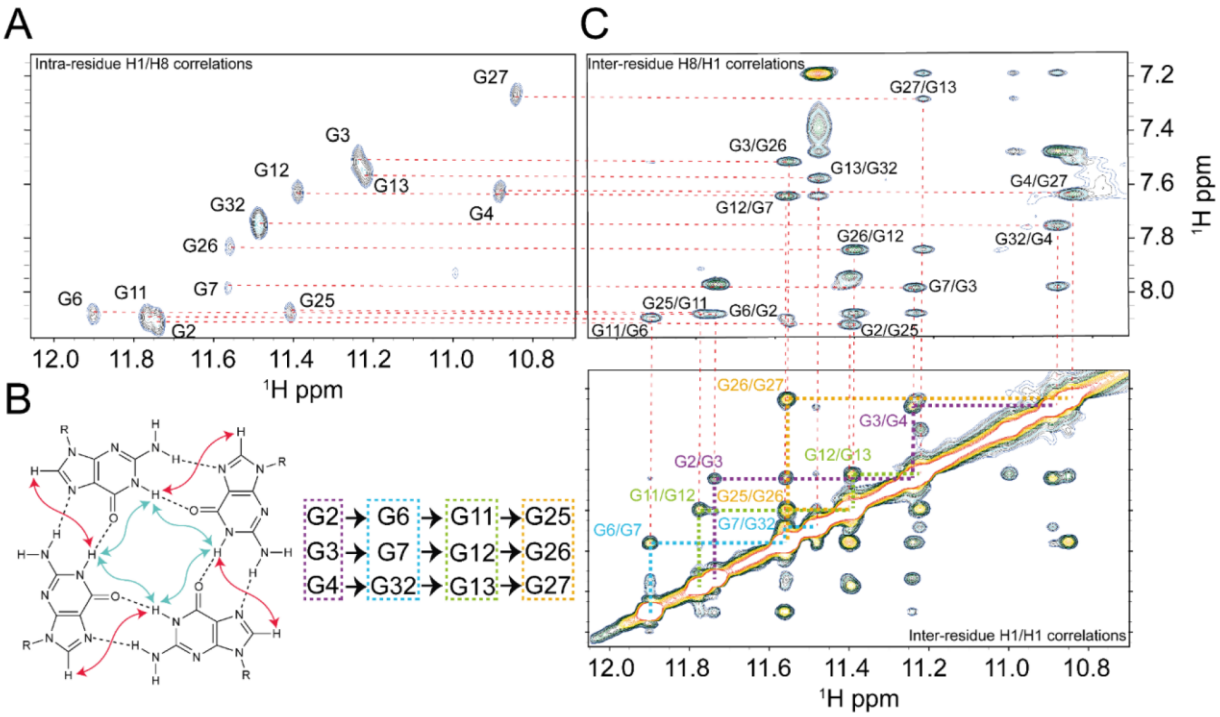


Figure 3. **A**) NMR HCCNH-TOCSY experiment of KRAS32R G9T showing intra-residue H1/H8 correlations in order to identify H8 of guanines implicated in G9T G-quadruplex and NOESY inter-residue H8/H1 correlations in **B**) and in **C**) NOESY inter-residue H1/H1 correlations were used to determine tetrads pattern in **D**). All experiments were performed at 37°C in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

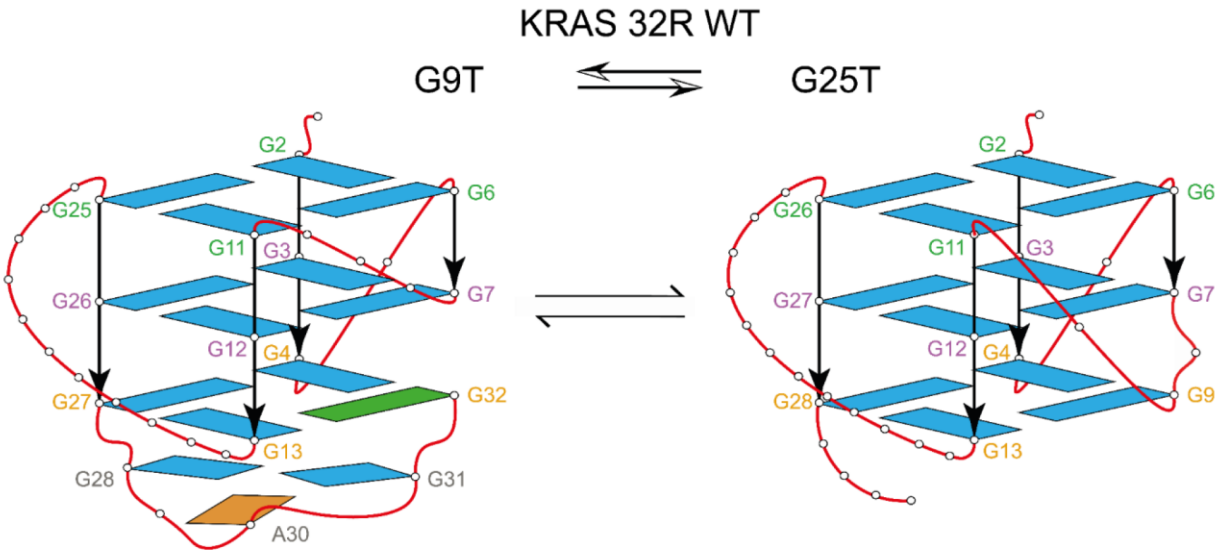


Figure 4. KRAS32R wild type contains two conformers in equilibrium, presenting different structural features depicted here by two families of mutants. Both mutants are formed by a three-layered tetrad core of guanines. G9T (*left*) has a fold-back guanine in *syn* glyosidic conformation (G32, green) and it is capped in the 3'-end with a triad (G28, A30 and G31). Conformer G25T (*right*) has all guanines in *anti*-position and contains three regular propeller loops.

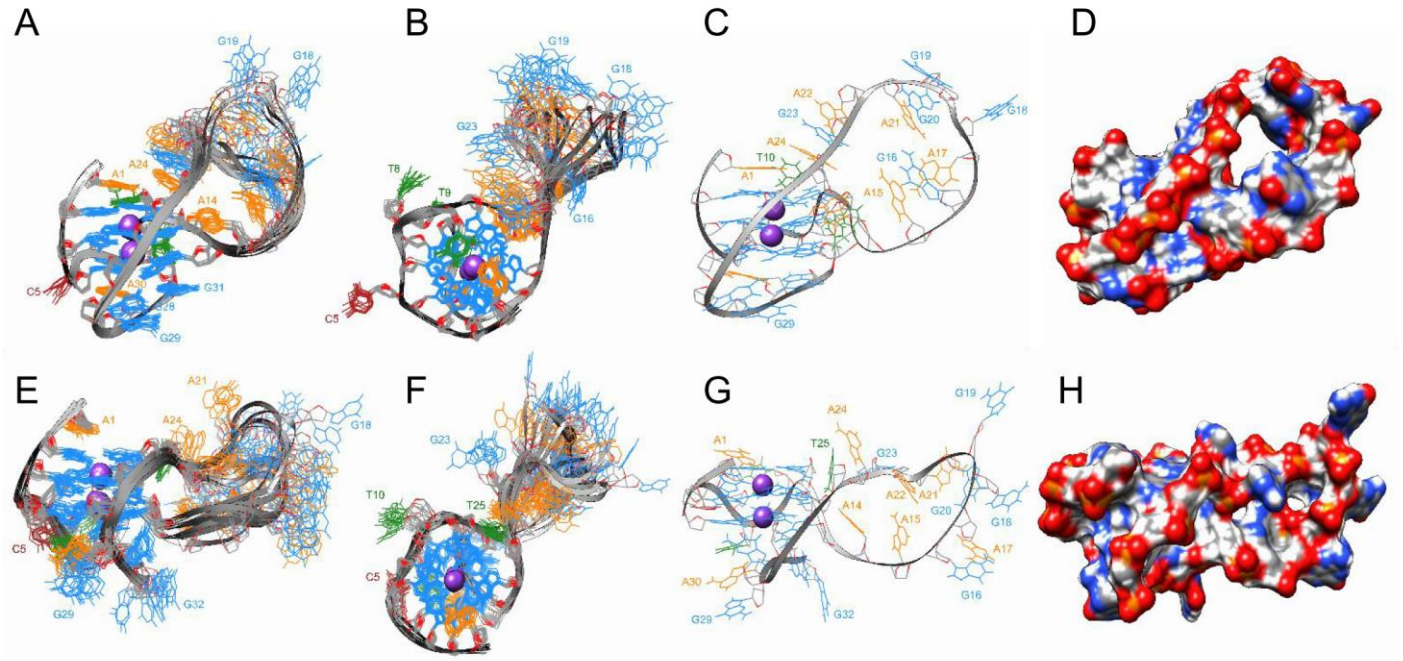


Figure 5. Side view of the superposition of the ten lowest-energy structures at 37°C based on all heavy-atoms for 32R G9T and G25T in A and E, respectively, and the corresponding up view in B and F (with guanines in blue, adenines in orange, thymines in green and the only cytosine in brown). C and G show just only one structure of G9T and G25T, respectively, and the geometric arrangement of the loop in each conformer. D and H show same structure as C and G with surfaces.

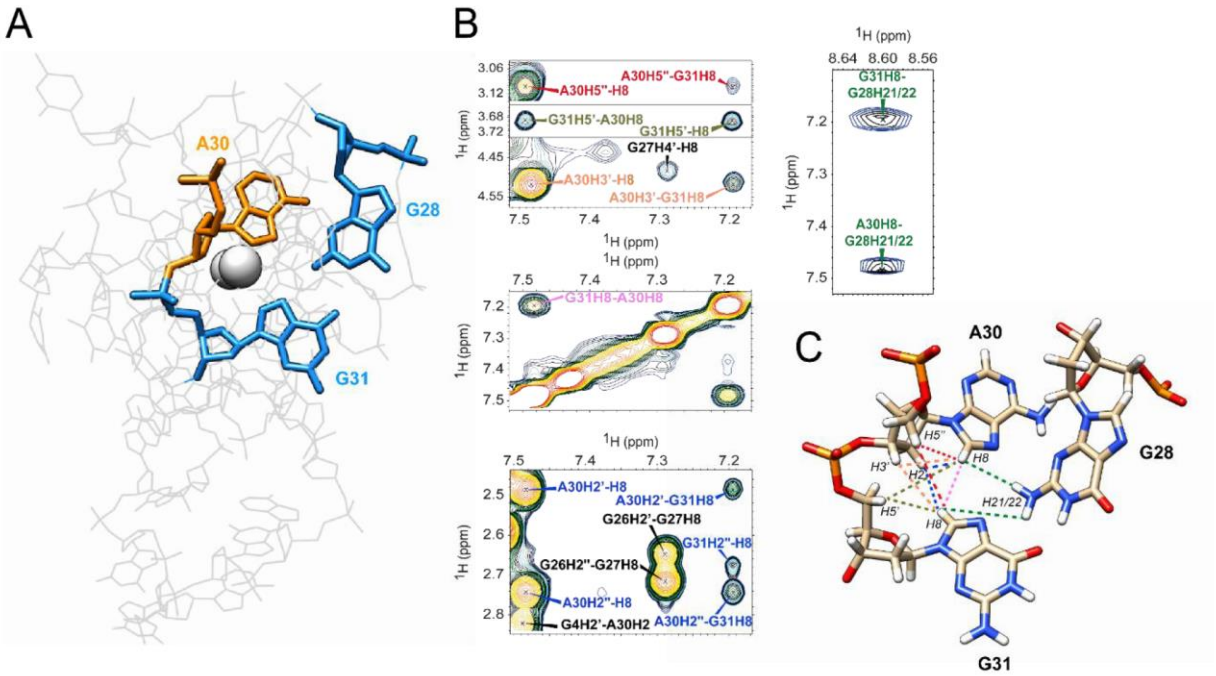


Figure 6. Representation of the triad within the KRAS32R G9T calculated structure in **A** with guanines 28 and 31 in blue and adenine 30 in orange. Several peaks proving the existence of the triad at 37°C found in KRAS32R G9T 1H-1H NOESY spectrum are show in **B** involving aromatic and sugar protons from the three bases implicated in the triad and the nearby residues. The corresponding restraints are shown in the calculated triad in **C** with the corresponding colours.

Supplemental Information

Structure of Multiple G-quadruplexes in equilibrium in Kras promotor

Julien MARQUEVIELLE[1], Coralie ROBERT[1], Olivier LAGRABETTE[1], Mona WAHID[1], Anne BOURDONCLE[1], Luigi XODO[2], Jean-Louis MERGNY[1], Gilmar SALGADO[1]

1. European Institute of Chemistry and Biology (IECB), INSERM U1212 - CNRS UMR 5320, University of Bordeaux, France.

2. Department of Medical and Biological Sciences, School of Medicine, 33100 Udine, Italy.

NMR structure calculation

NOE distance Restraints

Distance between the protons of KRAS32R G9T and G25T were obtained from NOESY experiments at 37°C with a different mixing time (250, 500 ms). In the final structure calculations, only data from the 250 ms mixing time was used. Peaks assignment was performed with Sparky and peaks volumes were extracted from the spectra with manual correction depending on the reliability of the integration. NOE distance restraints were then calculated using CCPN Analysis software using chemical shifts and peaks list from Sparky with Thymine H6-H7 distance as a reference.

Hydrogen Bonds

Hoogsteen hydrogen bonds between guanines were restrained using H21-N7 distance (2.0 ± 0.3) and N1-O6 distance (2.0 ± 0.3). Hydrogen bond between guanine and adenine in KRAS32R G9T triad used H21-N3 distance (2.0 ± 0.3)

Planarity restraints

Planarity restraints were used for G2→G6→G11→G25, G3→G7→G12→G26, G4→G32→G13→G27 tetrads and for G28→A30→G31 triad in the case of KRAS32R G9T conformer.

Planarity restraints were used for G2→G6→G11→G26, G3→G7→G12→G27, G4→G9→G13→G28 tetrads in the case of KRAS32R G25T conformer.

ARIA2/CNS structure calculation

All restraints files were included in ARIA2.3/CNS1.2 for structure calculation. Eight iterations of calculations were performed with 100 to 750 structures per iteration with mixed Cartesian and torsion angle dynamics during the simulated annealing runs. The protocol contains several stages. First, an initial high-temperature torsion angle simulated annealing of 10,000 steps at 10,000 K followed a torsion angle dynamic cooling stage of 10,000 steps from 10,000 to 2000 K. Then, a Cartesian dynamics cooling stage of 5,000 steps from 2000 to 1000 K, and finally a Cartesian dynamics cooling stage of 4,000 steps from 1000 to 50 K with 3 fs per step. For distances and hydrogen bonds, 10 kcal $mol^{-1}Å^{-2}$ was applied during the initial stage of dynamics and was increased up to 50 kcal $mol^{-1}Å^{-2}$ for the remaining steps of the dynamics. For dihedral restraints, different energy constants were applied depending in the calculation phase: 5 kcal $mol^{-1}Å^{-2}$ for the high temperature steps and then 25 and 200 kcal $mol^{-1}Å^{-2}$ respectively for the first and second cooling step. An energy constant of 25 kcal $mol^{-1}Å^{-2}$ was applied for planarity restraints. In the final run, the best twenty structures were extracted and analysed to choose structures for AMBER refinement.

Structure refinement in Explicit Solvent (AMBER12)

Before neutralizing the system by using K^+ cations, two K^+ cations were included in the G-tetrad core of the two best structures obtained from ARIA calculation. Then, the system was solvated with water

molecules (TIP3P) in a truncated octahedral box. Distance restrains and planarity restrains were imposed during molecular dynamics refinement. The system was first minimized with harmonic potential position restraints at 25 kcal $mol^{-1}Å^{-2}$ over 1000 steps of steepest descent minimization. Then it was heated from 100 K to 310 K during 20 ps with constant volume and the same potential position restraints as before. Then a several equilibrations were performed using gradually reduced positional restraints: 22, 20, 17, 15 and 12 kcal $mol^{-1}Å^{-2}$. Finally, the system was equilibrated without positional restraints for 5 ns at 310K and the ten lowest-energy structures were chosen.

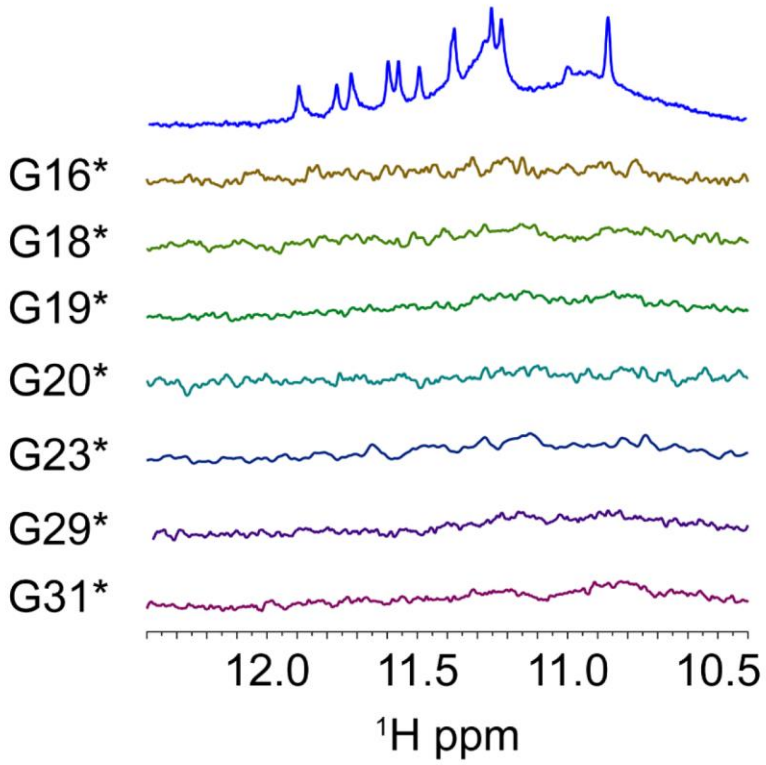


Figure S1. 15N-filtered NMR spectra at 37°C of samples containing 5% of 15N-enriched isotope of residues not implicated in KRAS32R G-quadruplex. All experiments were performed at 37°C in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

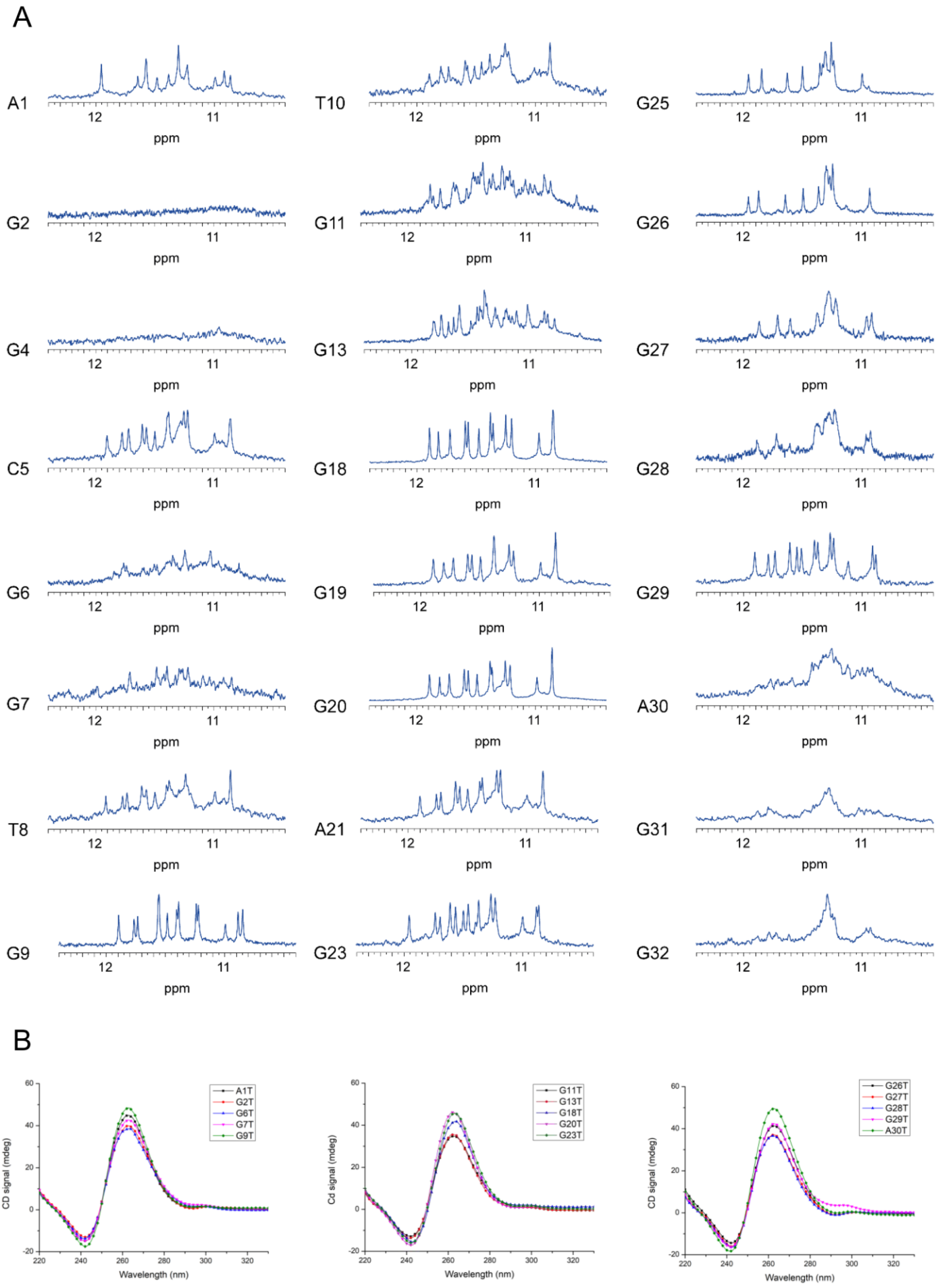
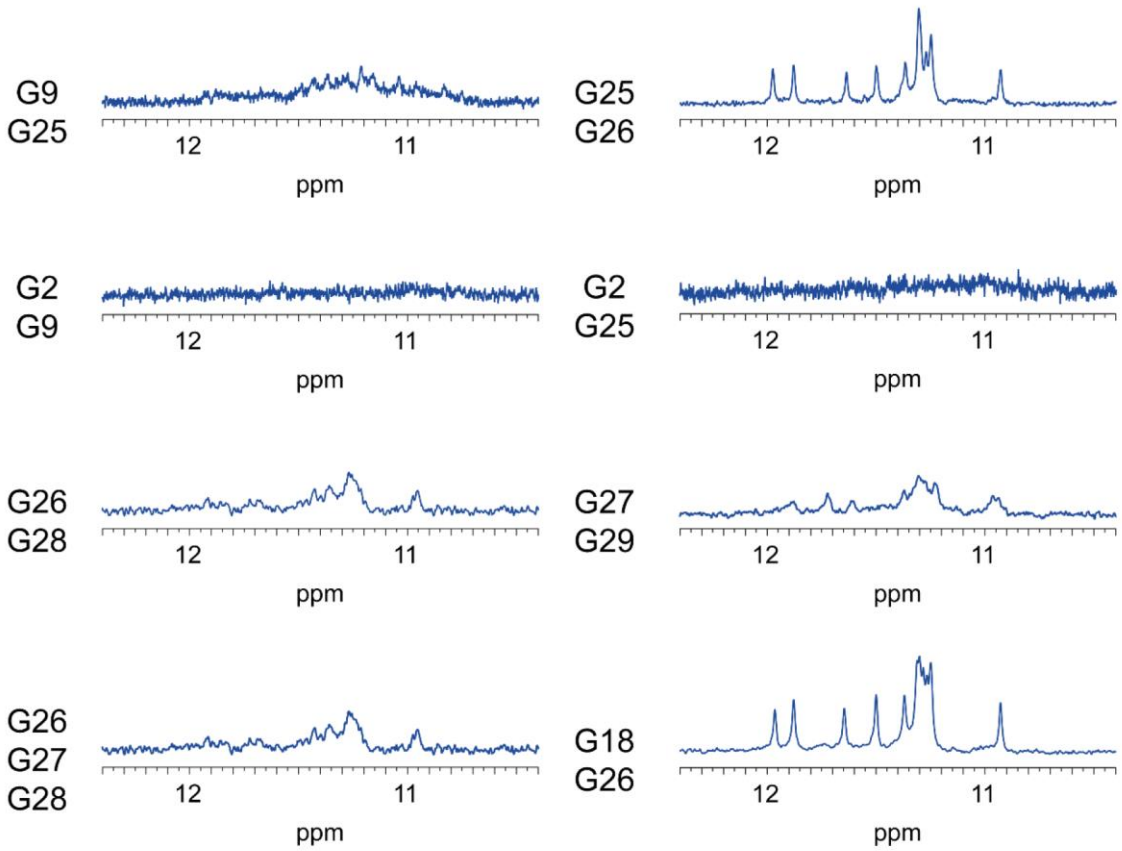


Figure S2. A) All NMR spectra imino regions of KRAS32R simple mutants at 37°C used in this study to better understand the KRAS32R G4 formation and the implication of the different residues with in **B**) the corresponding CD spectra at 37°C. All experiments were performed in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

A



B

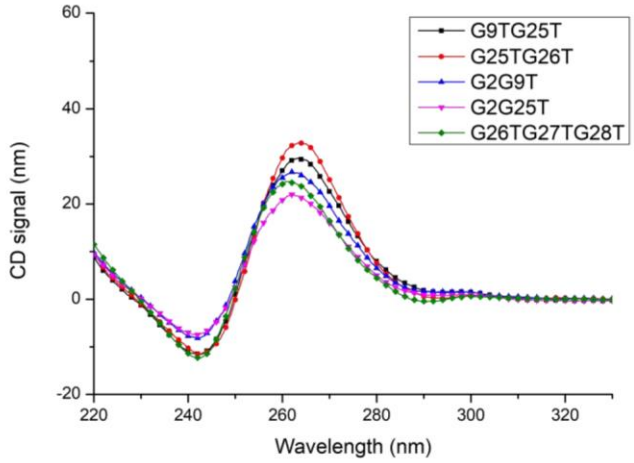


Figure S3. A) ^{1}H spectra imino regions of double mutants performed within KRAS32R WT at 37°C to identify exchanges among bases implicated in tetrads with in **B**) the corresponding CD spectra. These results confirm the models for G9T and G25T and give more information on the role of G2 and the guanines in the last tract. All experiments were performed at 37°C in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

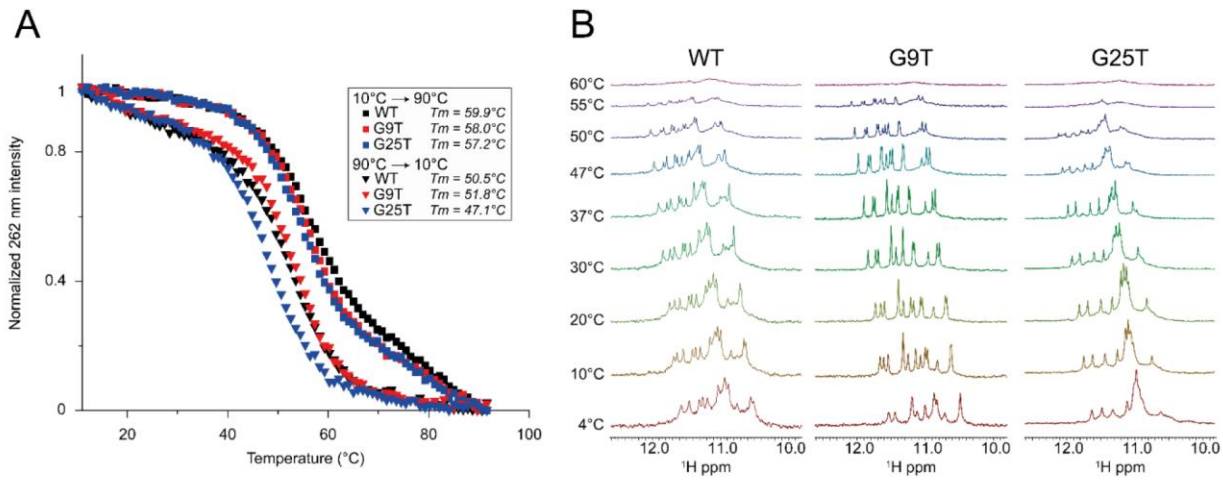


Figure S4. In **A**, CD spectra of wild type (black), G9T (red) and G25T (blue) from 10°C to 90°C (square) show very close melting temperatures but from 90°C to 10°C stabilization is the same compared to NMR observed in **B**) NMR melting experiment of wild type, G9T and G25T conformers in A from 4°C to 60°C showing that G9T conformer seems to be more stable than wild type which is itself more stable than G25T especially when looking at the 5°C spectrum. All experiments were performed in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

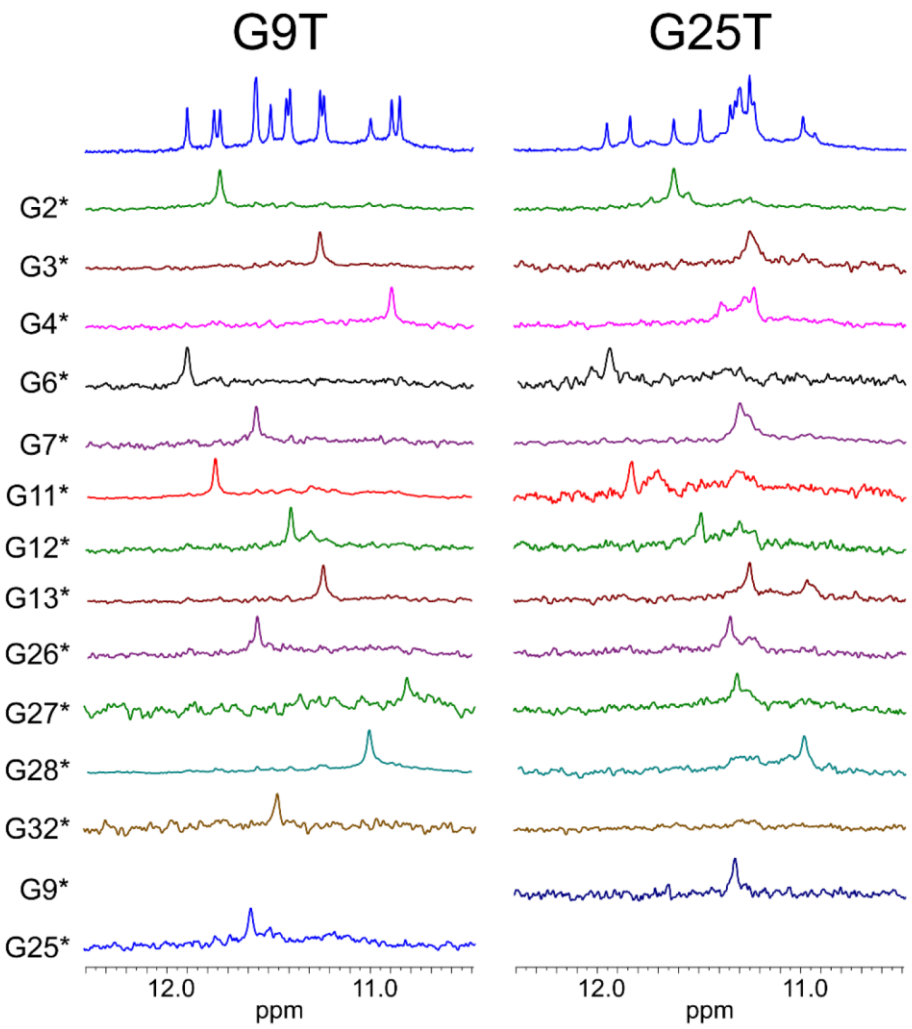


Figure S5. Determination of the guanines implicated in KRAS32R G9T and G25T tetrads using 15N-filtered NMR spectra at 37°C of samples containing ≈5% of 15N-enriched isotope. G9T imino protons give only one peak meaning that each of them is implicated in only one conformation whereas in G25T, several imino protons such as G4 or G7 are implicated in more than one conformation. All experiments were performed at 37°C in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

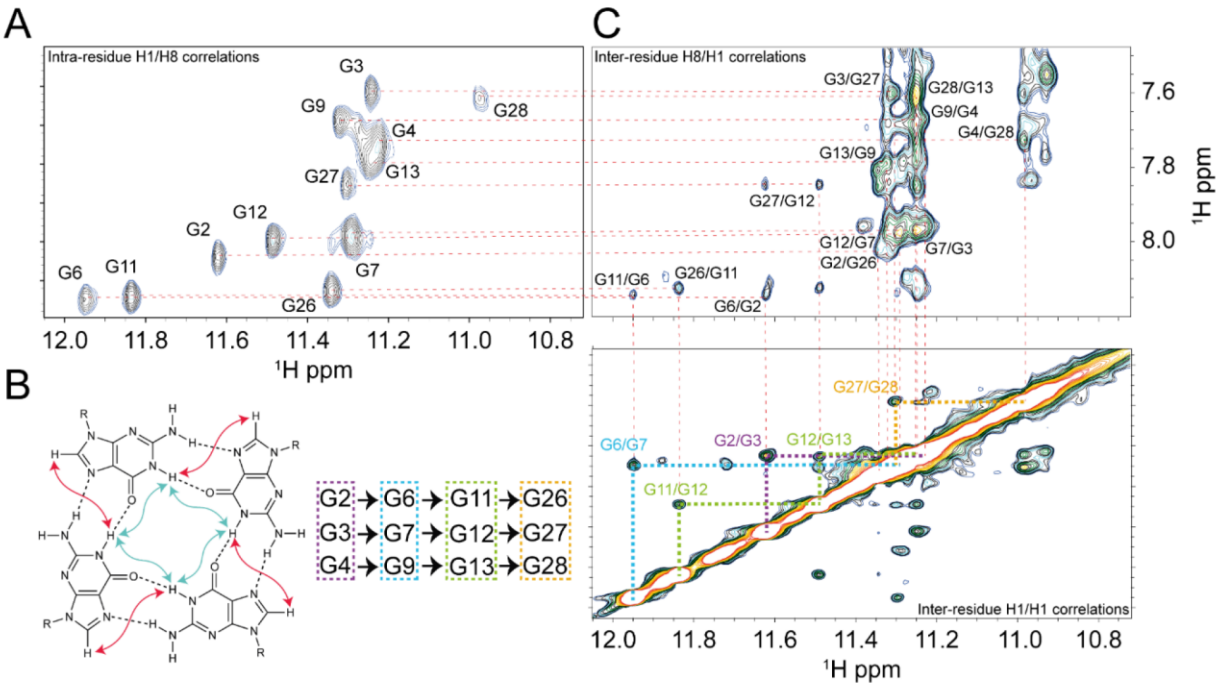


Figure S6. **A**) NMR HCCNH-TOCSY experiment of KRAS32R G25T showing intra-residue H1/H8 correlations in order to identify H8 of guanines implicated in G9T G-quadruplex and NOESY inter-residue H8/H1 correlations in **B**) and in **C**) NOESY inter-residue H1/H1 correlations were used to determine tetrads pattern in **D**).

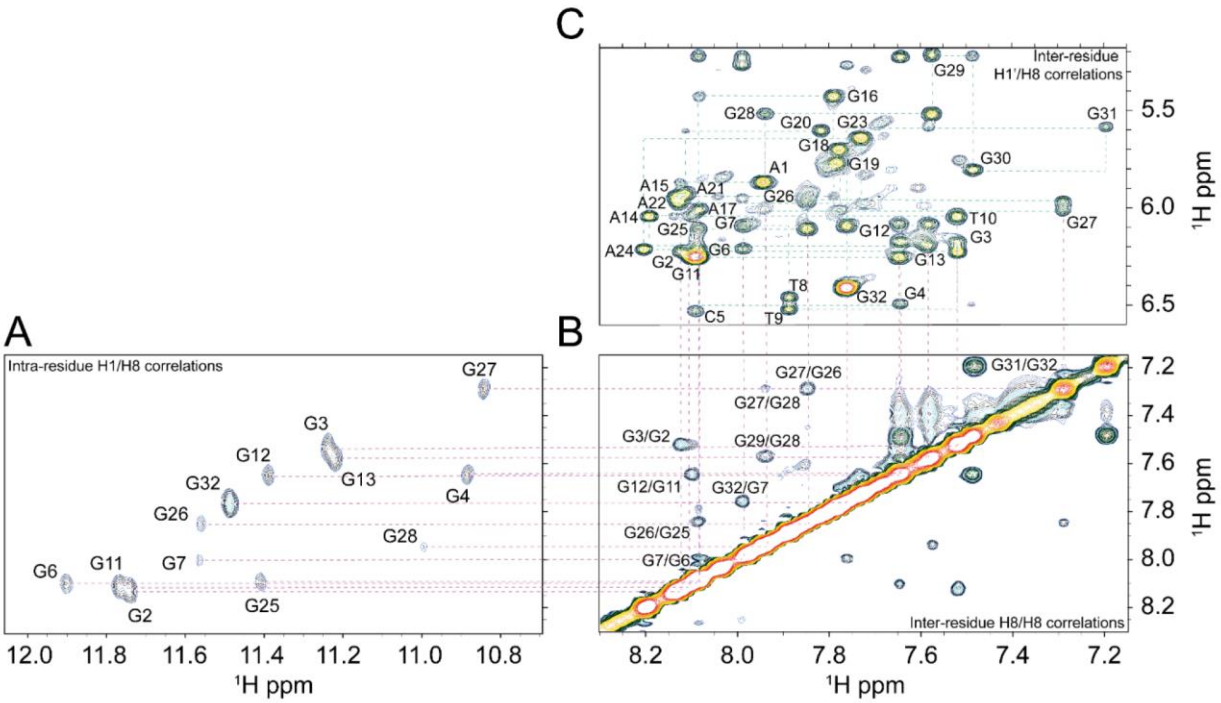


Figure S7. **A**) NMR HCCNH-TOCSY experiment of KRAS32R G9T showing intra-residue H1/H8 correlations used to identify H8 from guanines implicated in the G4. Then in **B**) Inter-residue H8/H8 correlations were identified between these guanines. All these results together help to determine KRAS32R G9T "walk" identifying inter-residue correlations H1'/H8 in **C**) between H1' from a residue and H8 from the next residue. Some of these correlations are missing, especially for residues implicated in triad formation due to its particular geometry and cross-peaks are also missing in bulges. All experiments were performed at 37°C in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

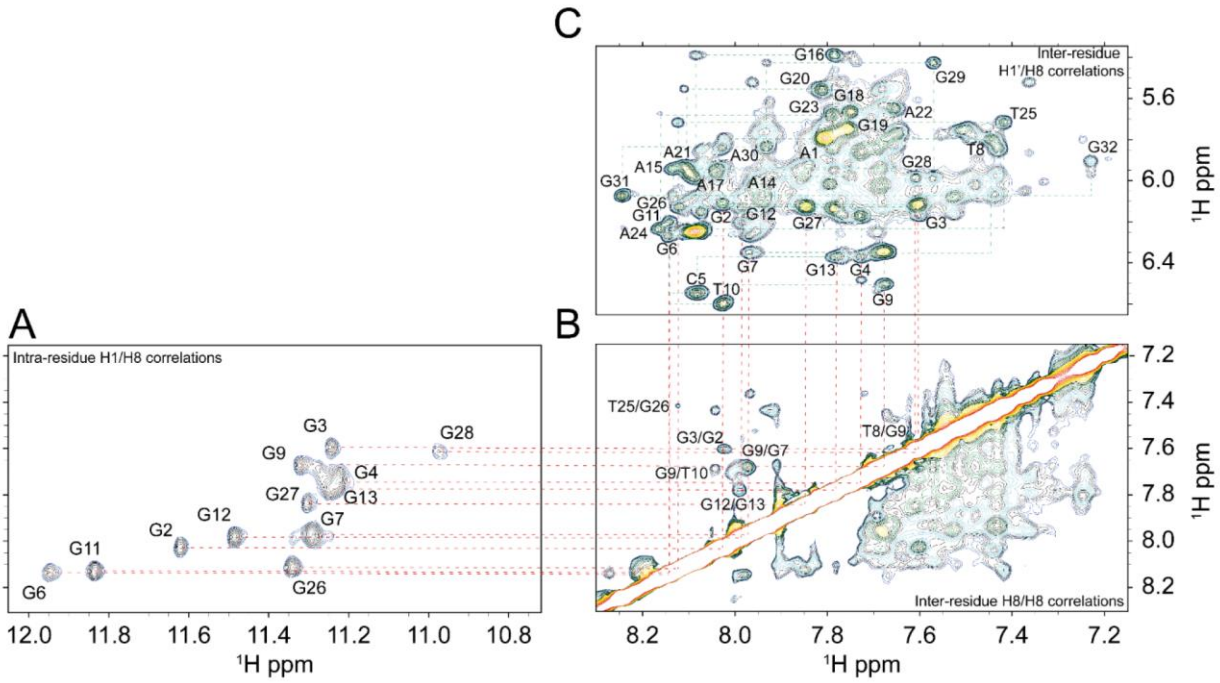


Figure S8. **A**) NMR HCCNH-TOCSY experiment of KRAS32R G25T showing intra-residue H1/H8 correlations used to identify H8 from guanines implicated in the G4. Then in **B**) Inter-residue H8/H8 correlations were identified between these guanines. All these results together help to determine KRAS32R G25T "walk" identifying inter-residue correlations H1'/H8 in **C**) between H1' from a residue and H8 from the next residue. Cross-peaks in bulges are mostly missing. All experiments were performed at 37°C in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

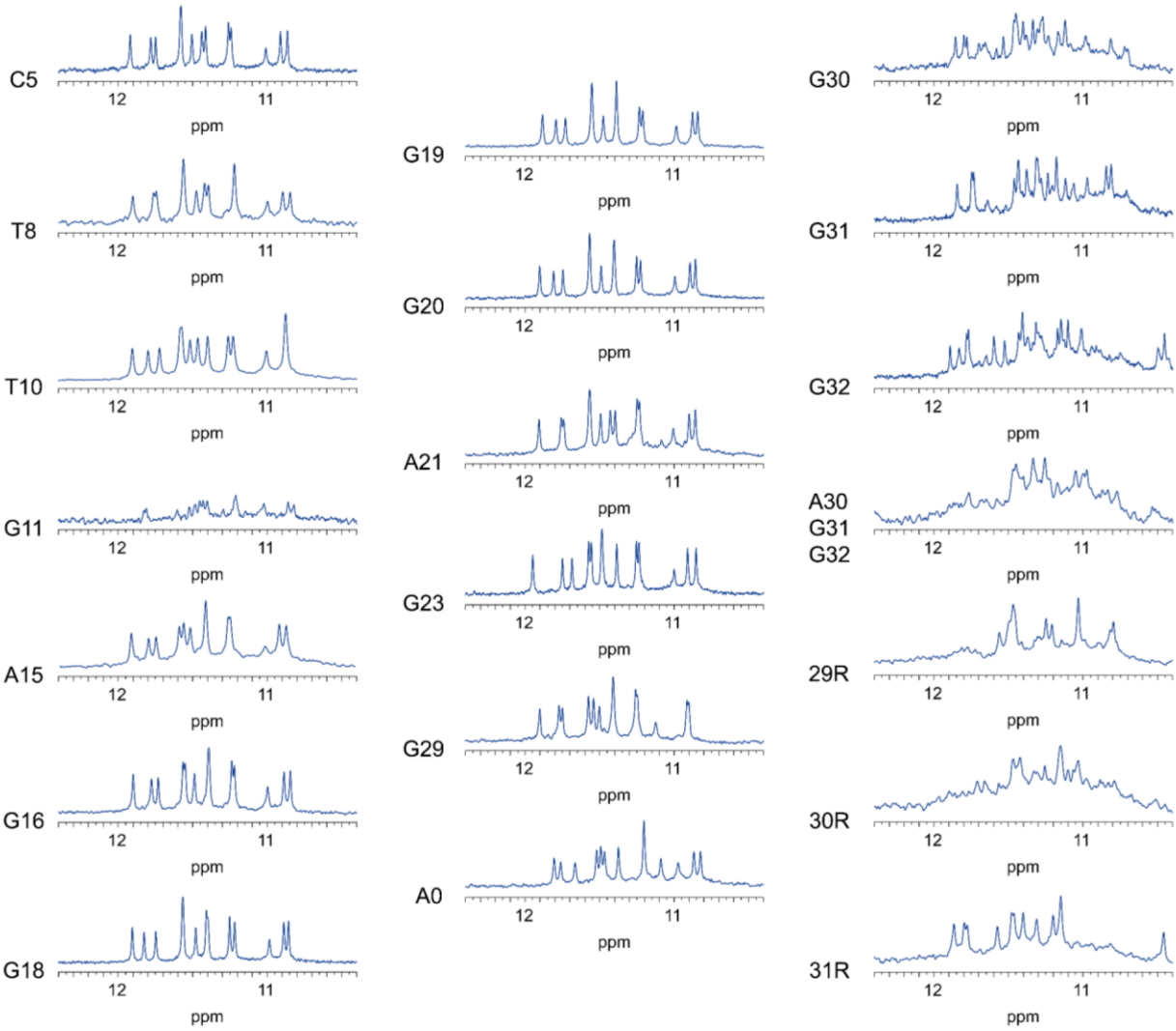


Figure S9. All NMR spectra imino regions of KRAS32R G9T mutants at 37°C used in this study. A0 corresponds to the addition of an adenine before A1 in the sequence. 29R, 30R and 31R correspond to the corresponding length of sequence. These results confirm the implication of G28, A30 and G31 in the structure and the participation of G32 in a tetrad. No residue from the loop has been identified to participate in any structural element. All experiments were performed in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

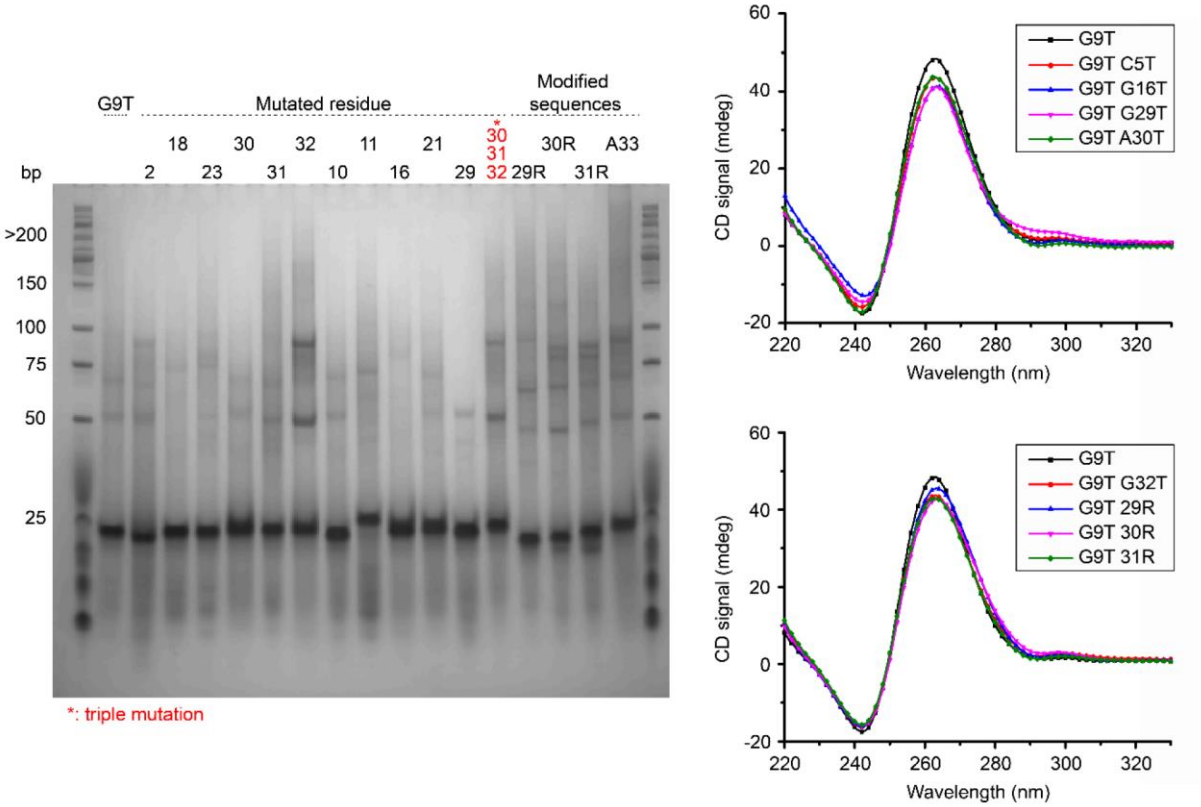


Figure S10. **A**) Native gel experiment of KRAS32R G9T and several mutants with also modified sequences. Most of the mutations do not affect migration but several of them especially mutations of the last four residues lead to higher weigh complexes. **B**) KRAS32R G9T mutants' CD spectra at 37°C show that most of the mutations do not affect G4 conformation showing all parallel conformation with characteristic peaks around 260 nm (positive) and 240 nm (negative). All experiments were performed in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

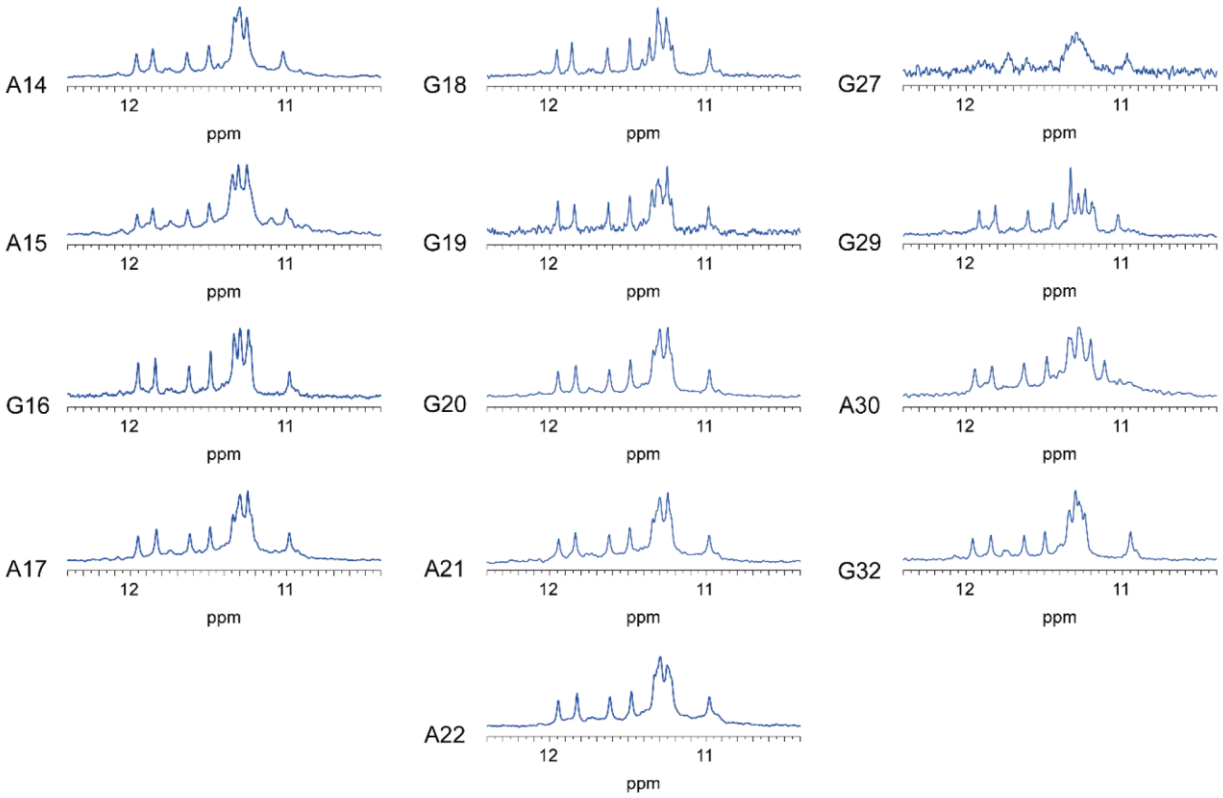


Figure S11. All NMR spectra imino regions of KRAS32R G25T mutants at 37°C used in this study. No residue from the loop neither the last four residues have been identified to participate in any structural element. All experiments were performed in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

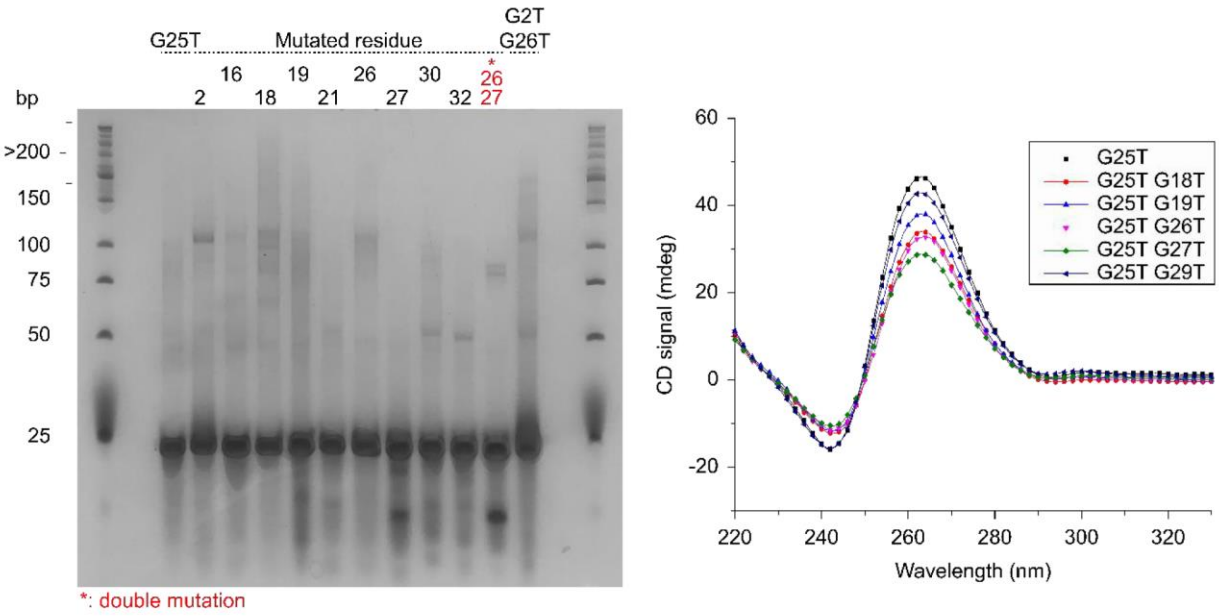


Figure S12. **A**) Native gel experiment of KRAS32R G25T and several. Mutations do not affect migration except mutations implicating G27 which is known to be part of G4 tetrad. **B**) KRAS32R G25T mutants' CD spectra at 37°C show that most of the mutations do not affect G4 conformation showing all parallel conformation with characteristic peaks around 260 nm (positive) and 240 nm (negative). All experiments were performed at 37°C in buffer 1X (50 mM KCl; 10 mM KPi; pH 6.66).

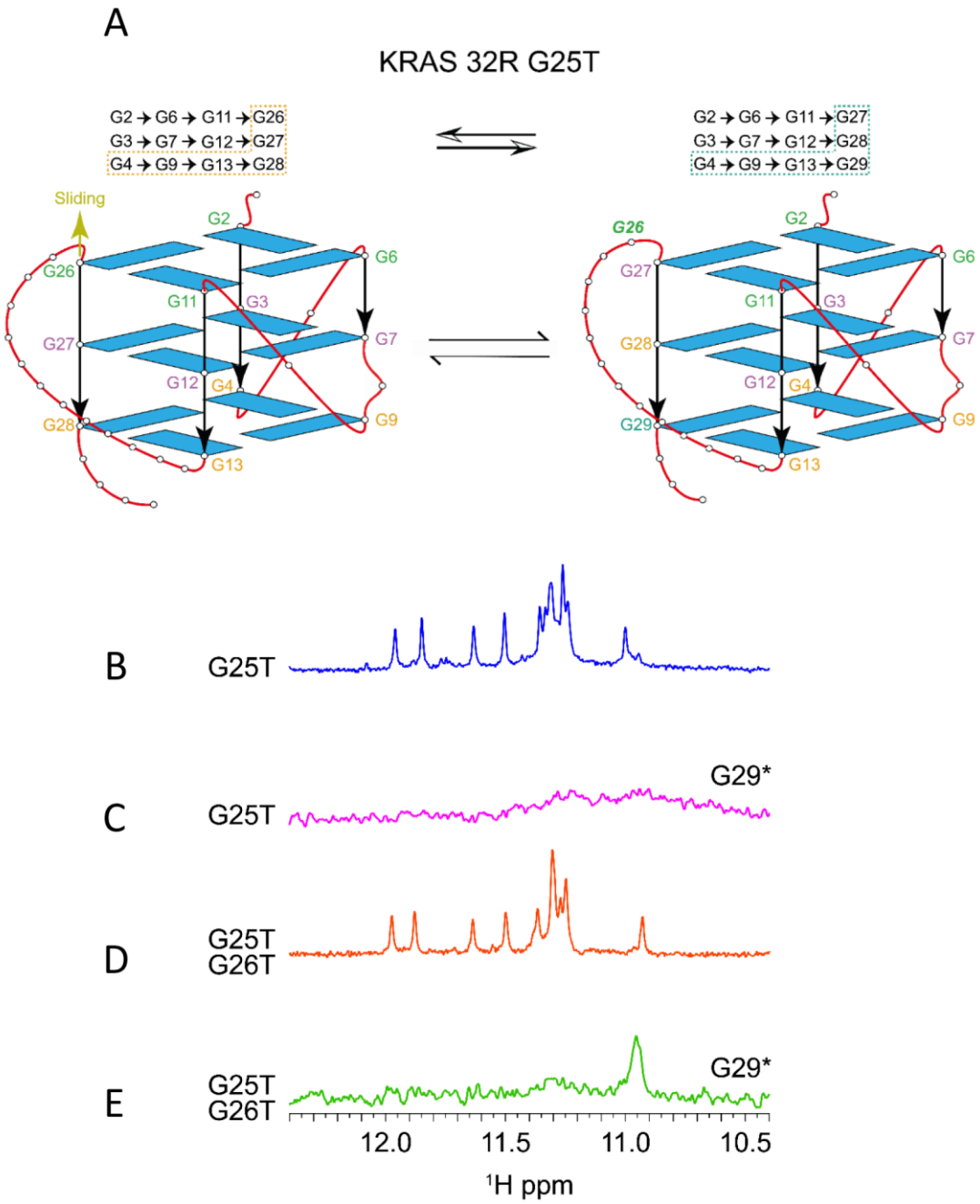


Figure S13 (**A**) Proposed model for G25T sliding into the tetrad when both G25 and G26 are simultaneously mutated into T. 1D NMR spectra comparing KRAS32R G25T simple mutant (**B**) with G25TG26T (**D**) double mutant imino region at 37°C. Results show that G29 is implicated in G-quadruplex formed within double mutant (**E**) but not in KRAS32R G25T (**C**).

NMR Restrains	H_2O
distance restrains	873
intraresidue distance restrains	503
sequencial (i, i+1) distance restrains	199
long-range (i, i+2) distance restrains	116
short-range non sequencial distance restrains	55
dihedral restrains	15
H bonds restrains	28
Structural statistics	
structure calculation	
total calculated structures	750
NOE violations	
number (>0.3 Å)	0.13
RMSD of violations (Å)	0.26 ± 0.03
molecular dynamics	
simulation time (ns)	5.0
extracted structures (lowest energy)	10
RMSD	
All heavy atoms (Å)	1.5

Table S1. Table containing NMR restrains for KRAS32R G9T G-quadruplex structure calculation with structural statistics after structure calculation and refinement by molecular dynamics

NMR Restrains	H_2O
distance restrains	564
intraresidue distance restrains	381
sequencial (i, i+1) distance restrains	109
long-range (i, i+2) distance restrains	37
short-range non sequencial distance restrains	37
dihedral restrains	12
H bonds restrains	24
Structural statistics	
structure calculation	
total calculated structures	750
NOE violations	
number (>0.3 Å)	0.09
RMSD of violations (Å)	0.33 ± 0.03
molecular dynamics	
simulation time (ns)	5.0
extracted structures (lowest energy)	10
RMSD	
All heavy atoms (Å)	1.9

Table S2. Table containing NMR restrains for KRAS32R G25T G-quadruplex structure calculation with structural statistics after structure calculation and refinement by molecular dynamics.

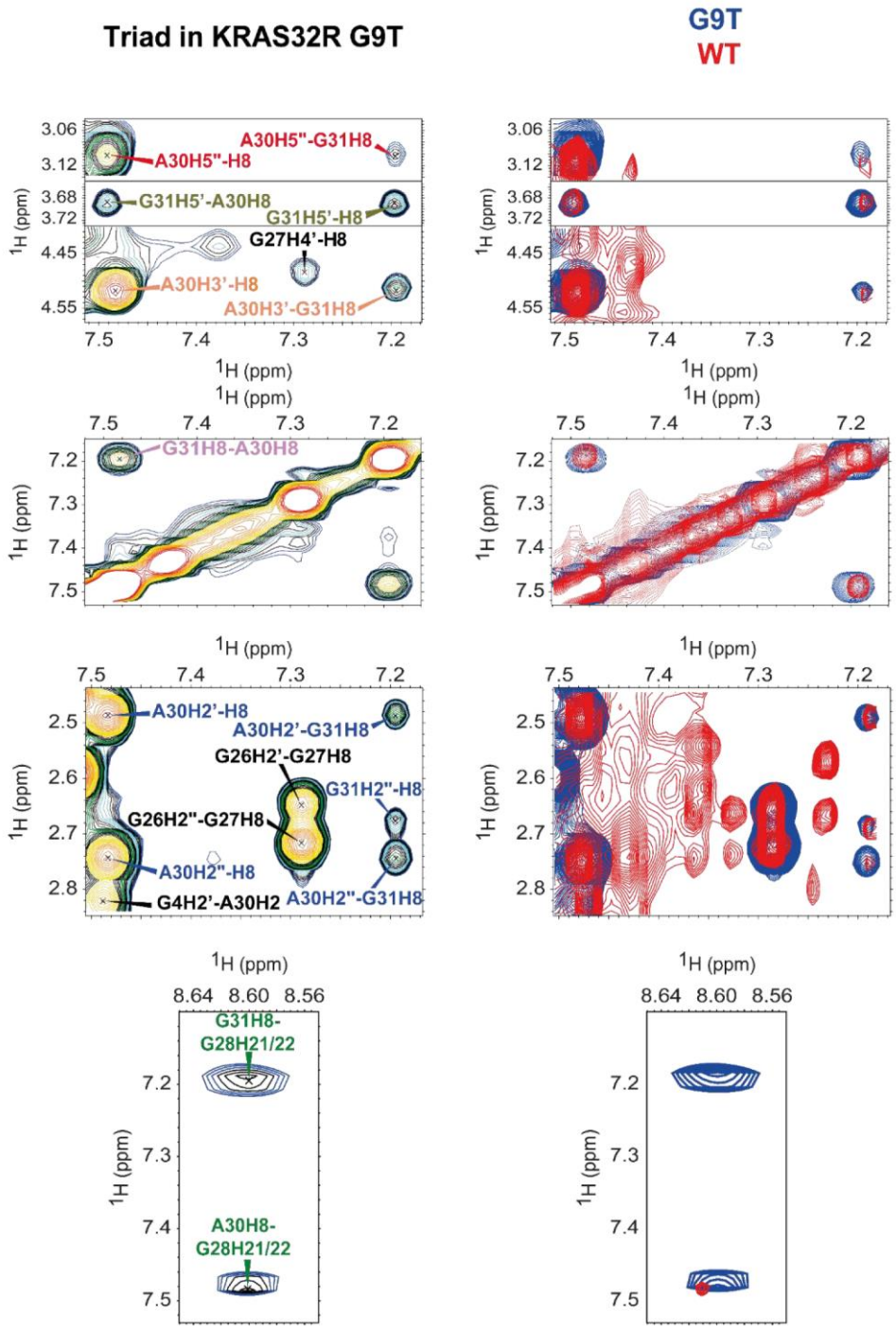


Figure S14. Cross-peaks indicating the presence of a triad in Kras32RWT sample. Peaks proving the existence of the triad at 37°C in KRAS32R G9T ^{1}H-^{1}H NOESY spectrum are also found in KRAS32R WTT ^{1}H-^{1}H NOESY spectrum.

III.2. Additionnal NMR dynamics relaxation experiments

III.2.1. Introduction

I used NMR solution state relaxation to understand the dynamics of exchange between the two conformers reported as G9T and G25T. In NMR, relaxation corresponds to the changes occuring to the signal with time. Generally, time causes deterioration of the signal which becomes weaker and broader. The deterioration of the signal can be splitted into two separate process T1 and T2. T1 (or longitudinal relaxation) is responsible for the loss of signal intensity, whereas T2 (or transverse relaxation) is associated with the brodening of the signal. The analysis of these two parameters allows to obtain site specific information about motion in G-quadruplexes, especially in the case of base exchange or formation of intermediate structures. The longitudinal and transverse relaxation rate constants (R1 and R2) and ^{15}N-^{1}H steady-state nuclear Overhauser effect (NOE) of a given imino ^{15}N are influenced mainly by dipolar interaction with its attached proton and by ^{15}N chemical shift anisotropy. The strength of the dipolar interaction is determined by the motion of the N-H bond vector.

III.2.2. Experimental parameters

Imino 15N T1, T2 relaxation and NOE/noNOE measure-ments data were acquired on an 850MHz spectrometer at 310 K with a fully labelled ^{13}C and ^{15}N KRAS32R (isotopic purity is >98 %) wild type sample produced by enzymatic synthesis. Full labelled KRAS32R WT was purchased from Silantes. Delays for relaxation were chosen based on the estimation of the relaxation time from the first increment (0.85 seconds for T1 and 170 ms for T2) so the intensity would drop by 10% in each of the 10 data sets to uniformly sample the relaxation curve. For T1, delays are: 10, 90, 190, 300, 430, 590, 780, 1020, 1370 and 1960 ms. For T2, delays are: 17, 34, 51, 85, 102, 136, 170, 220, 288 and 407 ms.

III.2.3. NMR dynamics results

In these experiments, we looked at the the imino region with signals corresponding to the guanines implicated in the KRAS32R wild type G-quadruplex structure. As we prevoously performed NMR experimentd using 5% 15N, 13C site-specific low-enrichment labelling, we had imino protons assignments. Even if we had several peaks for several guanines we were capable of assigning peaks corresponding to the major conformation within wild type that we

determined later as G9T. We analysed evolution of T1 and T2 for each residue as well as the value of their NOE. Results from our NMR relaxation experiments are presented in **Figure 29.**

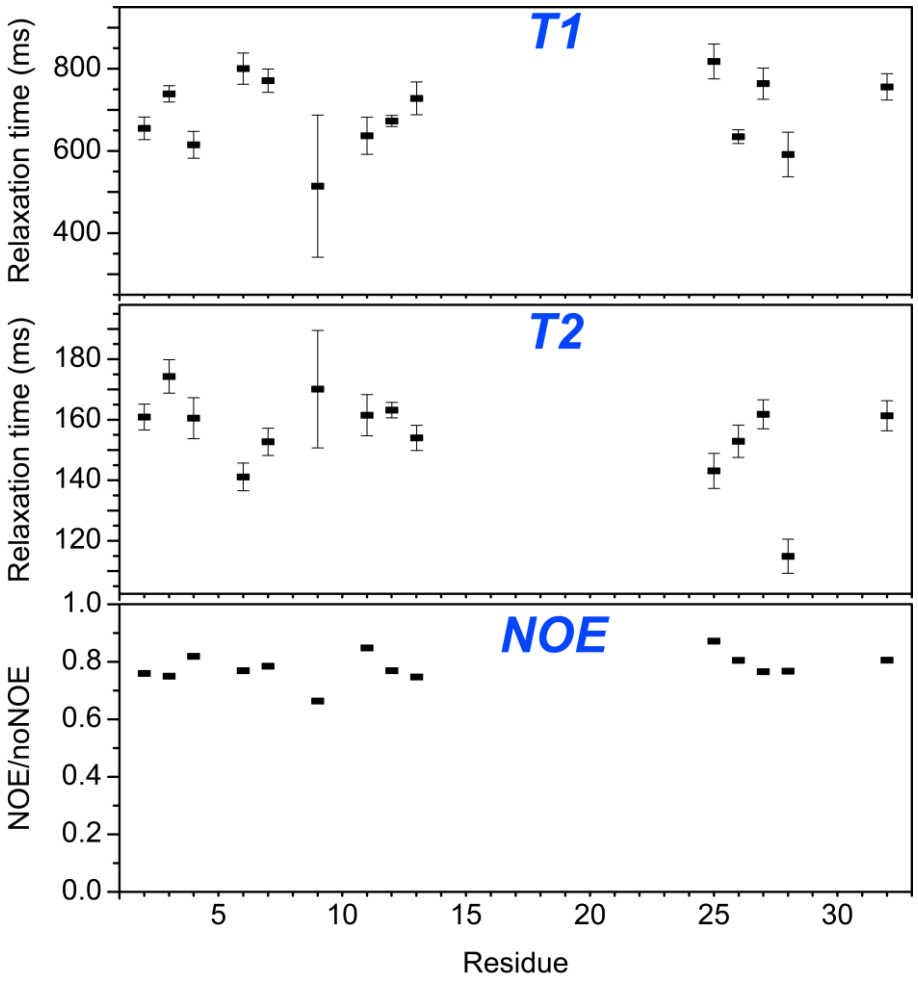


Figure 29. KRAS32R wild type relaxation measurements at 37°C in 1X buffer. Longitudinal (T1), transverse (T2) relaxations and imino ^{15}N NOE effect were plotted for each residue of KRAS32R G-quadruplex.

When we looked at T1 and T2 relaxations, it was pretty clear that G9 is the most involved residue in KRAS32R G-quadruplex dynamics. This result is also confirmed in NOE experiments because it has the more important variation in its NOE. Another residue which seems to be quite dynamic is G28. Indeed, even if it is not so clear by looking at T1 relaxation, its T2 relaxation is far less important compared to the other residue. However, it seems similar to the others in NOE experiments. The other residue we could identify implicated in KRAS32R dynamics is G25 especially in NOE experiments but results are not so clear compared the the previosu residues. As a conclusion, these NMR relaxation experiments supported the model we proposed with our two main conformation G9T and G25T. Indeed, the important motion

of G9 is in favour of the idea that it could be implicated in only one conformation and its exclusion of the structure coud lead to the other conformation. Relaxation results for G28 could be explained by its role in the triad formation in G9T conformation and not in G25T. Concerning G25, even it is pretty unclear, we could assume that its motion is due to the fact that it is implicated in G9T conformation and also in G25T with the slide betwwen the last guanines we observed in this conformation. However, we could observe that the results we had were not so clear as the results we expected. Indeed, we are clearly looking at a wrong time-scale here. The rates of T1, T2 and heteronuclear NOE can reveal the motions on the nanosecond time-scale or faster while the exchange between the conformers is likely to occur at time-scales slower by at least three orders of magnitude. As a conclusion, we need to think about other methods allowing us to observe dynamics in the ms range such as CPMG and XESY **(Figure 30)**.

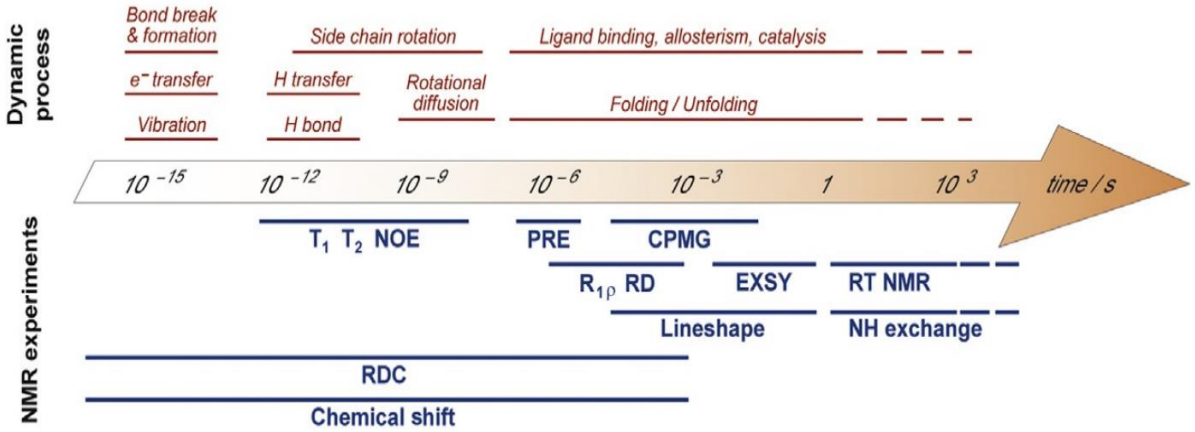


Figure 30. Different NMR methods that can be used for dynamics studies with the corresponding time scales and structural changes that can be observed in each experiment. From 10^{-12} to 10^{3} seconds time scale, NMR provide a panel of tools to study several mechanisms.

From Gabriel Ortega, Miquel Pons, Oscar Millet, Advances in Protein Chemistry and Structural Biology, 2013

Unfortunately, I han no more time to perform these experiments and bring new elements for the understanding of KRAS32R G-quadruplex dynamics.

IV. Conclusion

In this part, we wanted to determine the major scaffold adopted by the G-quadruplex formed within KRAS32R sequence. As this G-quadruplex seemed to be highly polymorphic, the structure determination, especially by NMR, represented a huge challenge. Despite the fact that KRAS22RT and KRAS32R G-quadruplexes shared common characteristics especially in the implication of the first and third tracts of guanines, the overall structure remained quite different. Indeed, in KRAS22RT, the last tract implicated in the G-quadruplex formation is G18-G19-G20 but in KRAS32R, these residues are not implicated in none of the conformations leading to the existence of a loop with a quite important size. Our study about KRAS32R G-quadruplexes showed that this sequence can adopt two major conformations that we identified as G9T and G25T corresponding to the mutated residue leading to the corresponding conformations. On one hand, G9T is a unique conformation stabilized by the formation of a triad by the 3'end residues in which the residue G9 is not implicated. On the other hand, G25T is very different as G9 is implicated in the G-quadruplex formation and there is no triad formation. Moreover, the structure is still polymorphic due to a slide in the last tract of five guanines between G26 to G29 residues leading to different implications of the guanines such as G26-G27-G28 or G27-G28-G29 in the G-quadruplex. It is also possible that G9 is also responsible for a part of the polymorphism by exchanging between two states. In the wild type structure, G9T conformation represents around 70% of the structure and G25T around 25% looking at the intensities in KRAS32R wild type two dimensional ^{1}H-^{1}H NOESY spectrum. Unfortunately, it seems that these two conformations are not the only ones formed by KRAS32R. Indeed, when we looked at the results with selectively labelled guanine samples in KRAS32R wild type we found that several peaks show more than the two expected peaks in the case of two conformations. **(Figure 31)**. In this study, we finally understand the G-quadruplex formation within KRAS32R sequence and we resolved the structure of this G-quadruplex which is known to be highly implicated in transcription and interact with several proteins during transcription. We brought some crucial information concerning the possible and most plausible structures of the 32R sequence in the NHE region of KRAS. The 3D structures could represent important new targets for the development of new drugs targeting

the organization of the G-quadruplex in the NHE region and interfere with regulation by proteins such as MAZ or hnRNPA1 transcription factors.

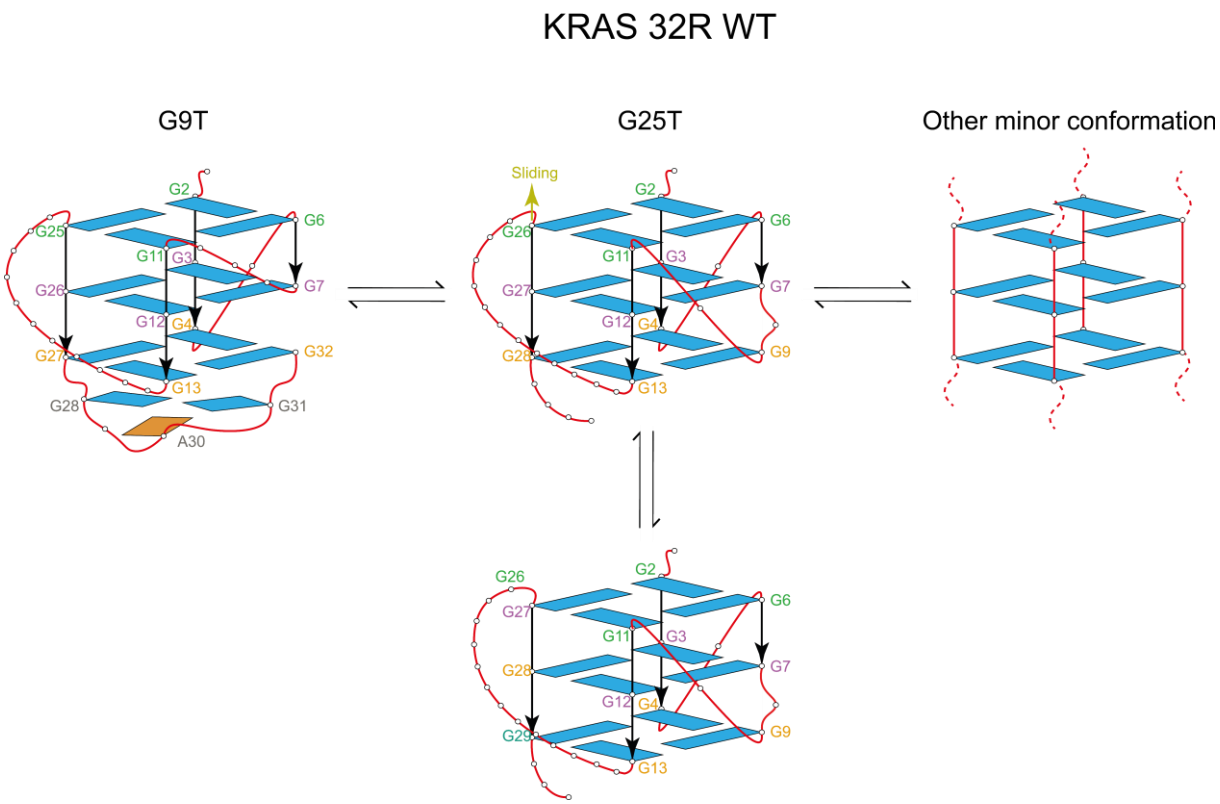


Figure 31. Schematic view of our model for the formation of KRAS32R G-quadruplex with G9T, G25T and the sliding phenomenon in equilibrium with other minor conformation

Part 2: KRAS G-quadruplexes interactions

I. Introduction

G-quadruplexes represent alternative targets in cancer therapies due to their localizations and their roles in biological processes such as transcription. They are mainly targeted by small chemical compounds that specifically bind to G-quadruplexes to stabilize them or inhibit their interaction with proteins. Many studies have been conducted to screen ligands from different chemical families. Thanks to the structural information we obtained by determining the 3D structures of KRAS22RT and KRAS32R G-quadruplexes, we can screen a panel of ligands. By selecting few good candidates and study their interaction at the atomic level, we might be able to identify how chemical groups contribute to the overall stabilization of the G4 structure. Our goal is to obtain such structural information and find how ligands could interact more strongly with certain loops, grooves or tetrad sites. As G-quadruplexes ligands often have cytotoxicity or cellular uptake issues, it is also a good way to identify and replace chemical groups that are not needed for interaction but implicated in these problems. In parallel to KRAS32R G-quadruplex structure determination, we also studied ligands against KRAS22RT structure to obtain structural features and performed a first screen against KRAS G-quadruplexes **(Figure 32)**.

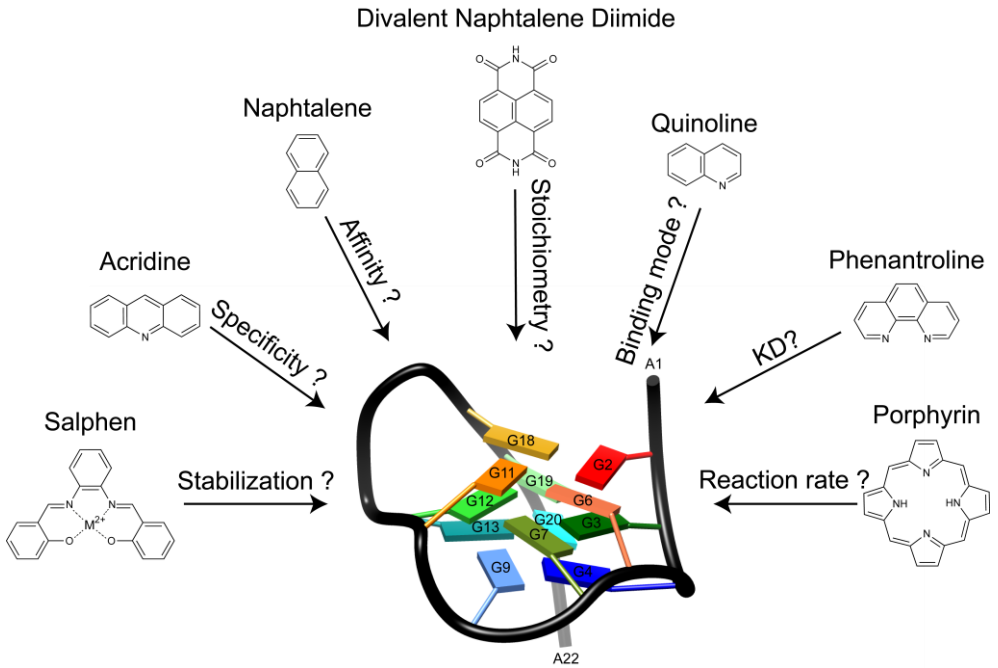


Figure 32. Concept of Drug Design against KRAS22RT G-quadruplex by studying the interaction with several ligands from different families and determining parameters such as, pharmacophoric-important chemical moieties, stabilization, overall binding mode or the K_D.

Recently, when the structures become available, we started using these data for studies against KRAS32R G-quadruplex conformers G9T and G25T as well as the wild type. The main goal of this project was to find ligands to stabilize G-quadruplexes but also to shift the equilibrium between conformers of KRAS32R G-quadruplexes. We also wanted to identify ligands able to inhibit interactions with proteins such as hnRNP A1, shown to unfold KRAS G-quadruplexes[279-281]. hnRNP A1 belongs to the hnRNP family that are ribonucleoproteins. This family is composed of 20 majors hnRNPs including hnRNP A1 in the family named A. This protein shares common structural elements found in hnRNP family with two RNA recognition motifs (RRM) that are the most common RNA-binding domains (RBDs) found in hnRNP family. It also has another common RBD that is a RGG box consisting of repeats of Arginine-Glycine-Glycine. In addition to these RBD domains, hnRNP A1 also has an auxiliary domain that is a glycine-rich domain whose role is to mediate protein-protein interaction or subcellular localization. hnRNP A1 is involved in RNA transport (thanks to its capacity to bind mRNA by interacting with the nascent transcript), is implicated in alternative mRNA splicing and has also a role in the transcription process (by associating with multiple promoter sequences and also by interacting with G-quadruplexes and unwinding them to facilitate transcription). Unwinding Protein 1 also called UP1 is an acid proteolytic product of hnRNP A1 of 196 amino acids which contains only the two RRM domains **(Figure 33)**.

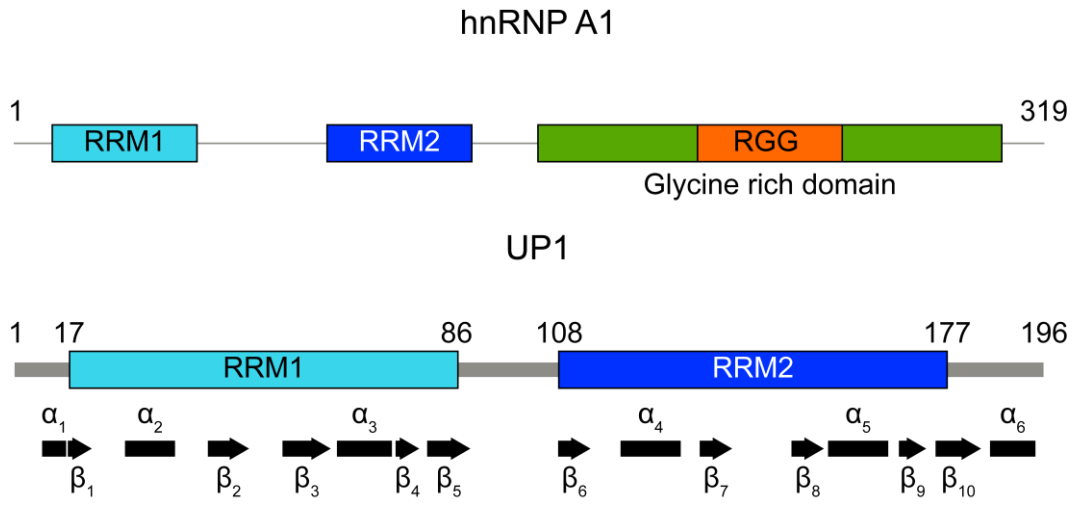


Figure 33. Schematic view of hnRNP A1 and UP1 proteins with two RRM domains and the glycine rich domain containing RGG box in hnRNP A1 whereas UP1 only contains the two RRM domains

It has been shown that this protein is also capable of interacting with KRAS G-quadruplexes and unwinding them. As a consequence, we wanted to confirm that the G-quadruplexes formed by our sequences (22RT, 32R wild type and conformers) could interact with UP1 and assess if these structures can be unfolded. The structure of UP1 has already been deposited in the PDB[282] (PDB code 1L3K) making the study of the interaction straight forward. The last part of this chapter will be focused on the preliminary experiment we performed using the best ligands we identified by FRET melting and 1D NMR, in order to assess whether they can compete and prevent binding of UP1 to different KRAS G4s. As we determined 22RT and 32R G4s structures at physiological temperature, all experiments with ligands have been performed at 37°C as well.

II. Material and methods

II.1. Oligonucleotides samples preparation

Non labelled samples were purchased from Eurogentec (Belgium) and IDT (Integrated DNA Technologies; USA). They were synthesised with 250 nmoles or 1 µmole scale and purified by reverse HPLC. Oligonucleotides solutions were prepared in 1X buffer (10 mM K_2HPO4/KH_2PO4; 50 mM KCl; pH 6,5). In order to fold G-quadruplexes, samples were heated at 95°C during 5 minutes then quickly cooled in ice. This process is repeated 3 times and sample are conserved at 4°C.

II.2. Labelled and Non-labelled UP1 Production

Recombinant protein UP1 tagged to with GST were expressed in Escherichia coli BL21 using plasmids pGEX-Up1. After transformation, the bacteria were grown for the night in Lysogeny Broth (LB) medium (10 g/L Yeast Extract, 10 g/L Tryptone, 10 g/L NaCl) at 37°C with 100 µg/ml ampicillin. Then bacteria were transferred in Terrific Broth (TB) medium (24 g/L Yeast Extract, 20 g/L Tryptone, 4 mL/L Glycerol, 100 mM Phosphate Buffer (KH_2PO_4/K_2HPO_4)) for non-labelled production and in M9T medium (300µM $CaCl_2$, 1mM $MgSO_4$, 6g Na_2HPO_4, 3g KH_2PO_4, 0,5g NaCl, 1mg Vitamin B1, 1g NH_4Cl ^{15}N, 2g Glucose ^{13}C) for ^{15}N, ^{13}C labelled production with the same concentration of ampicillin and grown to an A600 of at least 2.0 prior to induction with IPTG (1 mM final concentration). Cells were allowed to grow for the night before harvesting. The cells were centrifuged at 5000 RPM., and 4°C. After centrifugation the

supernatant was removed. The pellet was re-suspended in a solution of PBS with PMSF 100 mM and DTT 1 M. The bacteria were lysed by sonication (40%: 45sec ON, 45sec OFF for 4 min and 30 seconds). The lysate was then centrifuged for 10 min at 4°C at 10 000 r.p.m. Glutathione Sepharose 4B (GE Healthcare) (50% slurry in PBS) was added to the supernatant and incubated for 30 min at 4°C on a shaker. The mix was centrifuged for 5 min at 500 g and the pellet was washed 5 times in PBS and eluted with elution buffer containing 20 mM NaCl, 20 mM reduced glutathione, 200 mM Tris–HCl, pH 9.5 for A1 elution and pH 7.5 for Up1 elution. Alternatively, to re-move the GST tag, the mix was centrifuged for 5 min at 500g, washed with PreScission Cleavage buffer (GE Healthcare) and centrifuged 5 min at 500g. The pellet was incubated for 4 h at 4°C with PreScission protease to cleave the GST tag from the purified proteins. After PreScission cleavage, the UP1 moieties were detached from GST which remained bound to the Glutathione Sephadex beads. The reaction mixtures were centrifuged for 5 min at 500g, 4°C, and the untagged proteins collected from the supernatant. Finally, the purification of tagged and un-tagged UP1 protein was checked by SDS–PAGE.

II.3. Ligands origin

All the ligands used in this work had different origins with some of them commercially available, but the majority was issued from ongoing collaborations. Their structures are all presented in **Figure A1 (Annexes)** with their corresponding chemical families. Braco19, PhenDC3 and Quarfloxin were purchased from commercial companies such as Sigma Aldrich. Salphen ligands[283] were provided by the laboratory of Fabrice Thomas (Prof.) from Grenoble (France). Phenantroline ligands (except PhenDC3), IP5 and NaphTrip were provided by Carla Cruz group in Beira Interior (Portugal) and JL205[274] by Alexandra Paulo group in Lisboa (Portugal). Naphthalene DiImide (NDI) ligands[284-286] were provided by Mauro Freccero from Pavia (Italy). 20A was synthesised by chemists from our unit INSERM U1212.

II.4. FRET melting experiments

FRET melting experiments have been used to screen ligands by studying their capacity to stabilize KRAS G4s. In order to perform the experiment, we need to use labelled sequences with two different fluorophores that are compatible for emission and quenching phenomenon. Quenching is happening when the emission spectrum of one of the fluorophores is overlapping the absorption spectrum of the second fluorophore leading to

extinction of the second one. The condition for quenching is that the two fluorophores have to be close in space. Based on this concept, we can use our G4 forming sequences with fluorescein (FAM) as a 5' label and tetramethylrhodamine (TAMRA) as a 3' label observing FAM fluorescence. When the G-quadruplex is formed we cannot observe any FAM fluorescence and the by increasing temperature, we will recover fluorescence because of G4 denaturation which leads to an increased distance between fluorophores. All FRET melting experiments were done with Statagene Mx3005P device. Buffer contains 10 mM LiCaCo, 10 mM KCl et 90 mM LiCl à pH 7,2 for KRAS22RT experiments and we used 1X buffer for 32R experiments. Each well contains 0.2 µM of 22RT sequence or duplex sequence. Ligands are tested at different two concentrations 0.4 and 1µM corresponding to 2 and 5 ligand equivalence conditions. Negative control contains water instead of ligand and we used PhenDC3 ligand as a positive control knowing its stabilizer effects on G-quadruplexes. Each well is duplicated in each plate and experiments were repeated three times per ligand. In the case of testing ligands that are capable of absorbing the same wavelengths as our fluorophores, we would test their stabilization effect by Circular Dichroism melting experiments.

II.5. Circular Dichroism

All Circular Dichroism experiments were realized with Jasco J-1500 CD spectrometer using Spectra Manager software. Quartz cuve were used and contains 500 µL of 5 µM oligonucleotide samples in 1X buffer. The CD spectra were measured in the region between 220 and 320 nm with a scan speed of 50 nm/min and a response time of 1 s. Three scans were collected and averaged. All experiments were done at 37°C degrees.

II.5.1. Titrations

Ligands were diluted from their stocks to 1 mM with water and kept on ice during the titration. Some ligands where previously diluted in 10-20% D6 DMSO before diluted in water. Before starting the titration experiments a spectrum without ligand is recorded to check the G4 signal. Ligands were added starting at 0.5 equivalent and incubated for 10-15 min after which CD spectra were collected. Then addition was done at 1, 1.5, 2, 2.5, 3, 4 and 5 equivalents (total volume change for the G4 sample was ≤ 10 %).

II.5.2. Melting experiments

CD melting studies were performed with 5 µM DNA sample alone or in the presence of different equivalent of ligand. CD melting was performed using either a full wavelength or a single wavelength mode. In the former case, the data were collected in the wavelength range 350-220 nm with 0.5 sec averaging (DIT) time, 2 nm bandwidth, 100 nm/min scan speed, two accumulations and 0.2 nm step. Temperature was changed from 4 to 90 °C with 1 °C interval, and 0.4 °C/min rate. These parameters lead to the overall acquisition time of 1h per 10 °C temperature change. All data collected in this manner were examined for the presence of possible intermediates during the melting process. However, the overall shape of the CD signature remained unchanged for every sample examined. Thus after completion of the full wavelength scan for each sample additional melting data were collected at specified wavelengths of 264 nm (characteristic for a parallel G4) and of 330 nm (used as a reference to factor out instrument fluctuations), the averaging time 32 s, and the bandwidth was 2 nm. Temperature was changed from 4 ° C to 95 °C and back at 1 °C/min rate. This set of experiments allowed us to test the reversibility of the melting process and to obtained thermodynamic parameters.

II.6. Native gel experiments

Native gel experiments were performed using 8% acrylamide/bisacrylamide (19:1 ration) gel containing 10 mM KCl and 1X TBE buffer with a running buffer containing also 10 mM KCl and 1X TBE buffer in native conditions by cooling down the system at 4°C. A DNA ladder containing polythymine samples with different lengths (9 to 90 nucleotides) have been used. Each sample contains 25 µM of DNA and different equivalents of UP1 and ligands. Before loading samples, a migration was performed in order to eliminate impurities. Then samples were loaded and migrated for 3H hours. Stains all coloration was used to revealed oligonucleotides bands.

II.7. Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) experiments were performed using a Microcal VP-ITC instrument (GE Healthcare) at 37°C. All samples were thoroughly degassed while being stirred prior to use. Titrations were conducted with the different DNA sequences (KRAS32R WT, G9T, G25T and double mutant G9TG25T. For the ITC titrations, the sample cell was filled to capacity with a dilute solution of UP1 at 10 µM and titrated with DNA at 100 µM. A typical titration involved the injection of 16 (6−12 µL) aliquots of titrant with titrant injections made at 400 s

intervals. The integrated heat data were corrected for the heat of dilution and blank effects and the corrected data fit for a binding model by nonlinear regression. The binding isotherms were sigmoidal and well fit with the standard one-site binding model incorporated into the Microcal Origin VP-ITC software.

II.8. NMR experiments

NMR spectra were recorded on in house Bruker Avance NEO 700 and 800 MHz and 400 MHz (Bruker Avance III). In addition, some spectra were collected at CEITEC (Central European Institute of Technology, Masaryk University, Brno) of in 850 and 950 MHz spectrometers equipped with cryogenically cooled probes (850 and 950 MHz). Experiments were usually performed at 37 °C for routine experiments, otherwise temperatures from 4 °C up to 60°C were used for different experiments such as NMR melting profiles. For solution NMR, standard 3 mm NMR tubes were used. The samples were prepared in 1X buffer with 10% D_2O. Most of the ^{1}H 1D spectra were recorded using an echo pulse sequence, which selectively removes resonance due to water without affecting other resonances including those which are in fast exchange with water. Resonance assignments were made using 5% ^{15}N, ^{13}C site-specific low-enrichment labelling for imino protons. The resonances of the KRAS22RT G4-ligand complex were assigned through the peak assignments of the free G-quadruplex by using exchange cross-peaks (in NOESY) between the free and ligand-bound species obtained at 2 equivalents of ligand. Interproton distances were measured by using NOESY experiments.

II.8.1. NMR ligand titrations

Concentration of oligonucleotide samples used were 350 µM. Ligands were added at different equivalent increments depending on the ligand studied. Then ^{1}H 1D spectra were recorded after mixing oligonucleotide with ligand.

II.8.2. NMR ligand competition experiments

Samples contained 350 or 500 µM of KRAS22RT oligonucleotide (depend on the experiment). For the competition assay without ligand, 500 µM of 22RT complementary strand were added and sample put immediately in the spectrometer to measure spectra at different period times after ligand addition and follow kinetics of duplex formation and G4 disappearing. For completion with ligands, before adding complementary strand, we included 2 equivalent of

ligand. Then, we selected an unchanged peak, usually an isolated aromatic peak (before and after addition of complementary strand) to scale all different spectra and integrate the duplex region. We converted the integration data into concentration using the initial concentrations of 22RT and its complementary strand. Then we plot duplex concentration in function of time. To fit the data, we use a simple model for hybridization of two DNA independent chains in an exponential model equation (1):

$$y = a * (1 - \exp(-k * x))$$

where a is the maximal concentration of duplex formed and k the speed constant of the duplex formation.

We also determined t1/2 duplex using a second order equation (2):

$$\mathrm{t} = \frac{1}{2 * \mathrm{k} * \mathrm{C}_0}$$

II.8.3. 2D ^{1}H-^{15}N HSQC NMR UP1 titrations

Kinetics for the interaction between UP1 and KRAS32R conformers G9T and G25T have bene followed by using 2D 1H-15N HSQC NMR with 15N, 13C isotopically enriched samples of UP1 allowing us to identify each residue with its backbone –NH connection. Assignment was done by using the deposited data from PDB structure 1L3K. Then we added increasing amounts of G4 samples (0.25, 0.5, 0.75, 1 equivalent of 32R G9T or G25T) and we followed peaks shifting. We calculate shift for several peaks using the equation (3):

$$\delta = \sqrt{(\delta H^2 + ((\frac{\delta N}{5})^2)}$$

with δH the shift in ^{1}H dimension and δN the shift in ^{15}N dimension. Shifts were then plotted in function of the corresponding residue in Origin 8.6

II.9. Cell Viability Assay and Cellular Imaging

These experiments have been done in collaboration with Dr. Carla Cruz and her PhD student Josué Carvalho at the University of Beira Interior in Portugal and with David Santamaria and his colleague Benjamin Drogat at European Institute of Chemistry and Biology.

II.9.1. Cell Viability Assay

HeLa, NHDF non-malignant cells (normal human dermal fibroblasts) and KRAS-mediated cancer cells were seeded in 48-well plates (5×104 cells/well) and exposed to various concentrations of compounds. After 72 h, 50 µL of 1 mg/mL methyl thiazolyl tetrazolium (MTT) solution was added to each well, and the cells were further incubated for 4 h in the dark at 37 °C. At the end of the incubation period, dark purple formazan crystals were formed. These crystals were solubilized in dimethyl sulfoxide (DMSO) (250 µL per well for 10 min), and the absorbance was measured at 570 nm using Biorad xMark™ microplate reader. The experiment was repeated three times. The IC50 values were derived from the curves of the cell viability against the compound concentration. The extent of cell death was expressed as the percentage of cell viability in comparison with control cells.

II.9.2. Cellular Imaging

The HeLa cells were grown on µ-Slide 8-well plates from ibidi at a density of 3 × 104 cells/well for 24 h at 37 °C in a 5% CO2/air incubator. Cells were then incubated with 10 µM of AG and 1 µM of TriPropil for 3 h. Following incubation, cells were washed three times with PBS and stained with 1 µM of Hoechst 33342 (nuclear stain) for 15 min. Confocal images were recorded using a Zeiss AxioObserver LSM 710 confocal laser scanning microscope and ZEN software (Carl Zeiss Microimaging, Gottingen, Germany), using 4,6-diamidino-2-phenylindole (DAPI), green fluorescent protein (GFP) and Rhodamine/Texas Red channels at 405-, 488- and 514-nm laser excitation, respectively. An appropriate emission band was selected for the three fluorescent channels by using a Plan-Apochromat 63×, 1.4 N.A. oil immersion objective (Carl Zeiss) to ensure high-resolution images. Images were processed with LSM Image Browser (Zeiss) and Adobe Photoshop.

III. KRAS22RT G-quadruplex ligands study

Besides the study that we conducted with 22RT sequence, other studies have been performed by different laboratories using our deposited structure[287], including our collaborators with Carla Cruz lab **(Annexes Article 2)**. In this study, a novel panel of acridine orange (AO) derivatives was tested against KRAS22RT G-quadruplexes. These ligands have an AO moiety with an aliphatic side chain with a different number of ethylene groups and an iodobenzene moiety. Overall, all ligands from the AO series, i.e., C3, C5 and C8 named according to the number of carbons in the arm between the acridine and the iodobenzene group, show good

results in term of thermal stabilization: 18, 29, and 40°C for C3, C5 and C8 respectively, obtained using CD melting experiments. However, their specificity for KRAS G-quadruplex when tested against DNA duplex was very modest. In cytotoxicity assays, these ligands show high toxicity against HeLa cancer cells but also against non-malignant cells which has to be improved for the next generation of AO ligands. A positive point observed by confocal microscopy, is their good capacity to cross the cellular membrane and localize in the cell nucleus. This series of ligand and especially C8 seems to be good candidates for KRAS22RT G-quadruplexes as several other ligands we also tested.

III.1. FRET melting screening

The first step in ligand studies against G-quadruplexes is to screen a large number of ligands from different families with FRET melting assay **(Figure 34)**. For a matter of simplification in the analysis we decided that the temperature cut-off will be the following: < ≈5 °C, low stabilization properties, not worth investigating; 5 - 12 °C, average stabilization effect, worth pursuing studies; > 13°C good stabilization properties and worth testing by 1D NMR experiments).

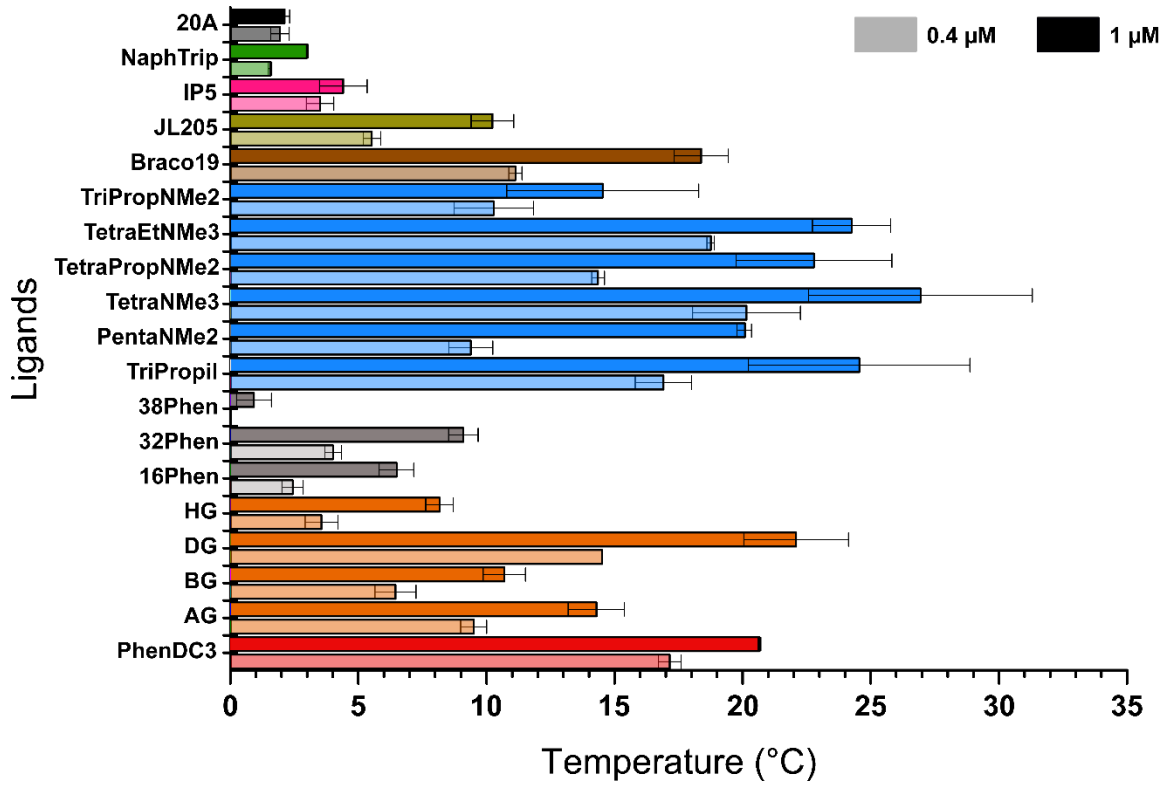


Figure 34. FRET melting experiments with a panel of ligands against KRAS22RT G-quadruplex. All ligands have been tested at 2 and 5 equivalent of ligand. ΔT_M have been obtained by subtracting T_M of negative control in T_M of each well.

From the figure 34 we can observe that the lowest stabilization effects have been observed with 38Phen, 20A, NaphTrip and IP5 (0, 2, 1.5 and 3.5°C respectively, at 2 molar equivalents of ligand). As a consequence, these ligands cannot be considered as good stabilizers for 22RT as well as JL205 which gave a stabilization temperature of 5.5°C. Braco19 can be considered as a good stabilizer (20°C) for 22RT. Among the salphen, family we tested several modifications such as aliphatic chain length, the nature of the coordinated metal, and the presence of additional chemical substituents. The ΔT_M for AG, BG, DG and HG are respectively 9.5, 6.5, 14.5 and 3.5°C at 2 ligand equivalents. Regarding the results, the most effective modification if we consider AG as the reference compound, is the addition of a dimethyl imidazole group **(Figure A1)**. With this modification, an extra planar conjugated system is available to interact via π-π stacking and can explain this increased stabilization. We also tested three different phenantroline ligands from the same family as PhenDC3 which is known as one of the best ligand for G4[288, 289]. All of them show relatively low stabilization temperature (9°C for 32Phen which is the best one) compared to PhenDC3 which means that this kind of modifications are not beneficial for 22RT G4 stabilization. These low stabilizations could be explained due to the size of 32Phen and 38Phen ligands with their long aliphatic chains. In the case of 16Phen, the absence of positive charges is probably the main reason for the low ΔT_M observed. We finally used ligands from NDI family. They all gave good stabilization results as they all have a ΔT_M superior to 15°C, the best one is TriPropil with a stabilization of 25°C which is better than PhenDC3. Among this family we chose TriPropil as the reference ligand on which modifications have been done such as addition of tertiary amines or side chains. Regarding the results, these modifications decreased the stabilization and did not increase specificity toward duplex DNA **(Figure A2)**. All the other ligands did not show significant DNA duplex stabilization except DG. After this first screening, the most promising ligand for 22RT G-quadruplex are PhenDC3, AG and DG from salphen family, TriPropil from NDI family and Braco19. Considering the DNA duplex stabilization obtained for DG we decided to use AG as principal ligand from this family to pursue further NMR studies.

III.2. Ligands effect on KRAS22RT G-quadruplex by CD titrations

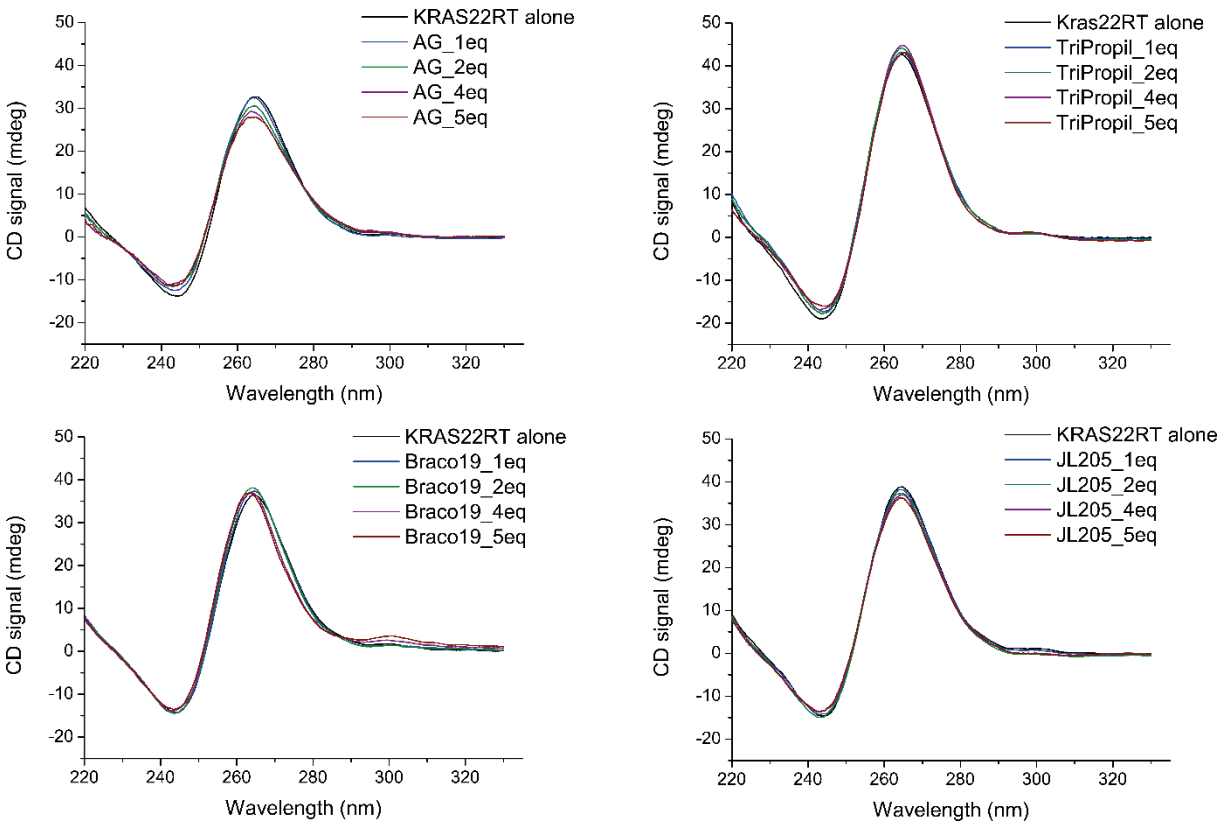


Figure 35. CD titrations of the best ligands identified in FRET melting experiments against 22RT G-quadruplex. We added increasing amount of ligand up to 5 equivalents

In parallel to FRET melting experiments, we also perform Circular Dichroism titrations to observe the effects on the conformation of 22RT G4. We looked at the modifications on the two characteristic peaks at 260 and 243 nm of parallel conformation obtained for 22RT. **Figure 35** is showing spectra of the best ligands identified in FRET melting experiments. The remaining spectra are presented in **Figure A3**. All CD titrations with all the ligands are showing similar results. Indeed, it is clear from observing the spectra that all the different ligands tested did not change 22RT G4 conformation. However, we could notice some slightly change with a decrease of the signal at 260 nm and also at 243 nm in spectra which could be a hint for an interaction but it is only a qualitative observation. The only exception is PentaNMe2 ligand because in the spectrum, we can observe a more pronounced decrease in intensity and a slight shift at 260 nm and 243 nm than the other ligands, together with an increasing at 290 nm, a characteristic signal of antiparallel conformation. This result means that this ligand seems to be able to change part of 22RT conformation.

III.3. NMR titrations

We also performed NMR titrations with ligands against KRAS22RT G-quadruplex. These experiments could provide useful information. First, as the assignment of KRAS22RT imino region is known, we could follow the peaks shifting by increasing the amount of ligand. Then, NMR titrations could be analysed to determine the exchange between the G-quadruplex free state and ligand bound state **(Figure 36)**.

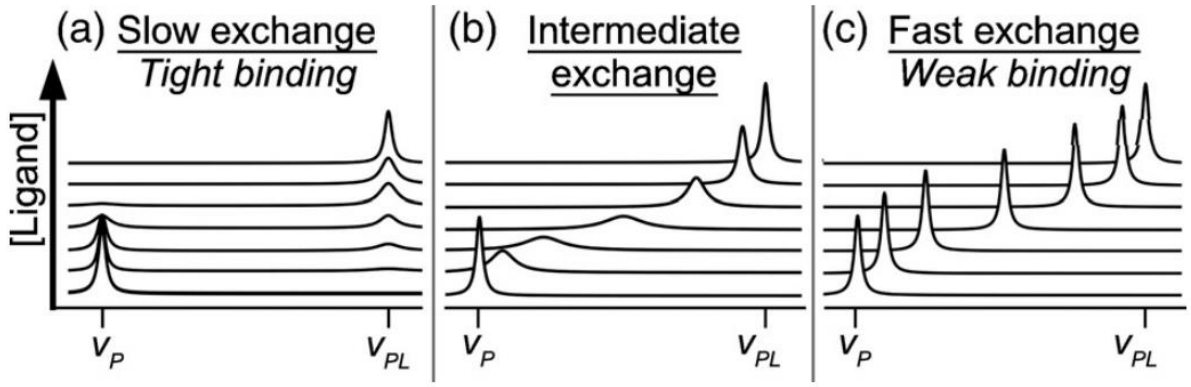


Figure 36. Different profiles we can observe in NMR titrations with ligands to determine the type of exchange we encounter

From BBA, I.R. Kleckner and M.P. Foster, 2011

In the case of slow exchange indicating a tight binding, transition is quite difficult to observe in the NMR signature with only initial and final states. Signals are disappearing and reappear only when saturation state is almost reached. In the case of fast exchange indicating a weak binding, we can easily follow the peaks when other factors such as aggregation do not broaden the peaks of interest. Intermediate exchange is characterised by a mix between the two others with peaks that almost disappear in a non-resolved region and that reappear faster than slow exchange. Another crucial information that can be obtained from the NMR titrations is the binding affinities for each ligand. Unfortunately, many of the tested ligands do not allow to correct follow individual peaks due to the exchange nature of the ligand complex, but also due to severe ligand-induced aggregation when interacting with DNA in a nonspecific manner. Finally, the last information we can extract from NMR titrations is the possibility of performing structural studies on the complex in 2D dimensional NMR depending on the quality of the spectrum during titrations. Among all the tested ligands, we selected the ones with the best ability to stabilize G4 and also some other criteria such as solubility, their chemical family, cytotoxicity and imino peak dispersion. Without a clear imino peak dispersion we would not

be able to proceed to 2D NMR without huge ambiguities. So with that in mind, the best 1D NMR titrations that I selected are presented in **Figure 37** and the remaining titrations are presented in **Figure A4**.

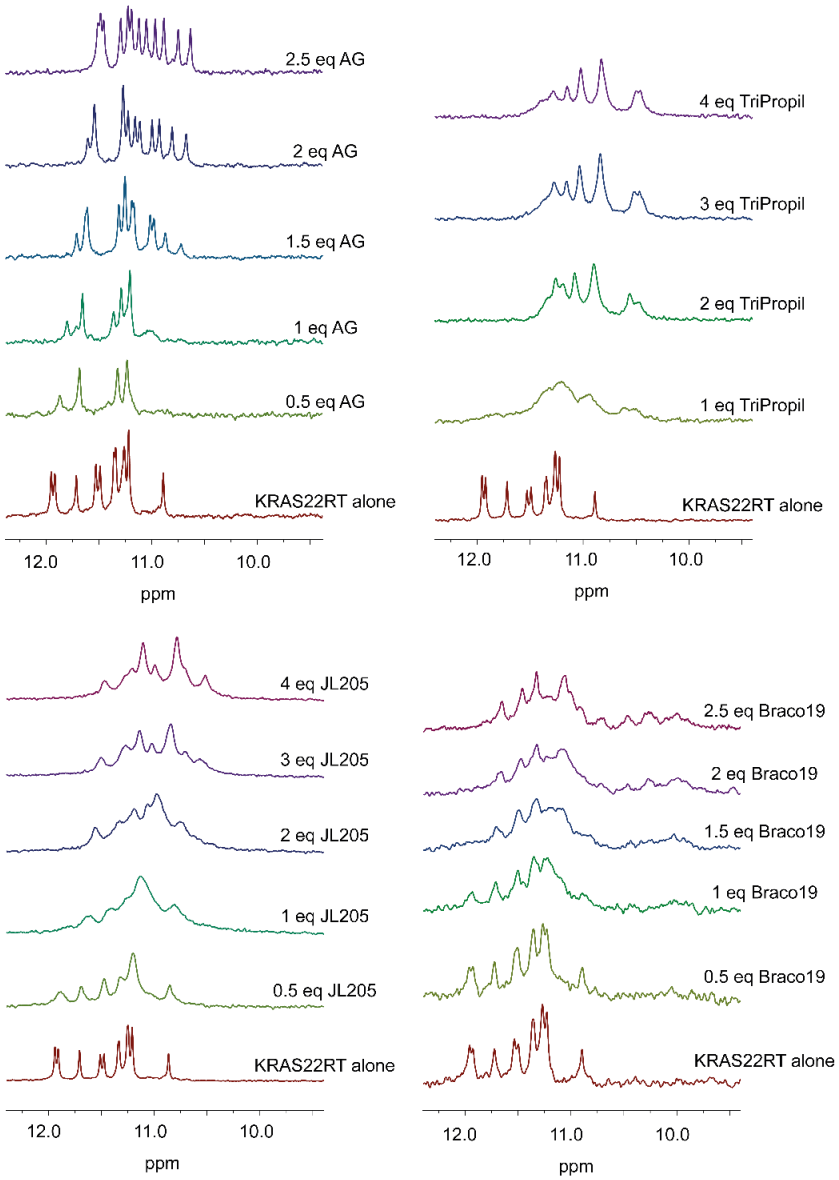


Figure 37. NMR titrations performed with the best stabilizers with increasing amount of ligands, depending of the saturation in each complex

From combining good stabilization with best 1D NMR peak dispersion, AG had the best results. All 12 imino peaks were well resolved indicating a single conformation in complex with 2 ligands molecules, probably with two independent binding sites. Looking at the profile of the transition, it seems that AG induces an intermediate exchange of 22RT. For the case of both TriPropil and Braco19 the imino profile did not had the 12 well resolved peaks as expected,

only a few peaks could be distinguished (around 6). Looking at the different transitions, we also have intermediate exchanges. For TriPropil, JL205 and Braco19 saturation is respectively reached at 2, 3 and 2 equivalents of ligand suggesting the corresponding number of binding sites. Among the other tested ligands, PhenDC3 which has one of the highest ΔTm values, did not produce a profile that could be properly analysed as the imino region became quite broad with no resolved peak. To the exception of AG that was a limiting factor for pursuing 2D NMR studies with different ligands. We could observe similar results with PentaNMe2. For the case of NaphTrip and IP5 we can observe sharp peaks up to 4 molar equivalents, and then an aggregation phenomenon where most probably the oligo desolvates and precipitates. Nevertheless, those ligands have not good ΔTm values and will be used in further studies. In the case of 20A and 16Phen, we could assume that ligands could partially unfold the G-quadruplex looking at the imino region but it is also possible that it is still in a transitional state with an unfinished titration, so the results need to be completed. As a results of these titration experiments, together with melting studies, we found that the best ligand for NMR structural studies in complex with 22RT was AG.

III.4. NMR competition experiments

In G-quadruplexes studies, it is not unusual to use competitors such as duplex DNA to probe the ligand efficiency to prevent hybridization of the G4 sequence with the complementary sequence. That is why we thought about performing NMR competition experiments in order to see the G4 protection from the complementary chain for each ligand and at same time get information about the first and the last tetrad to be unfolded. The complementary sequence represents the best competitor as it is capable of unfolding G-quadruplex by forming the corresponding double stranded DNA in cells. NMR would allow us to follow the disappearance of G-quadruplex along with the appearance of the corresponding duplex (new peaks in the region of 12.5-14 ppm). We wanted to know if the ligands we studied were also able to maintain the G4 integrity despite the presence of the complementary chain. By integrating the peaks corresponding to the duplex region at different periods of time we were able to determine the association constant of 22RT with its complementary strand to form a duplex. We also determined duplex formation half-life time ($t_{1/2}$ duplex) by using second order reaction equation. NMR competition results without ligand used as control are presented in **Figure 38**.

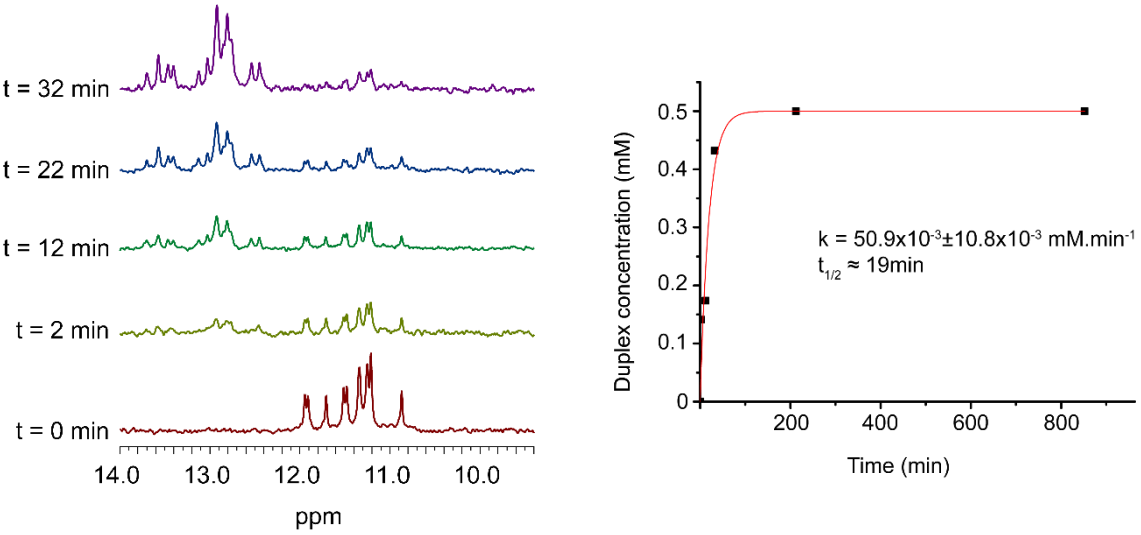


Figure 38. NMR competitions results (without ligand) with the disappearance of imino peaks corresponding to G4 between 10.5 and 12 ppm, and the appearance of peaks corresponding to the duplex formation between 12.5 and 14 ppm.

In the case of competition assay without ligand, the duplex formation is a very fast event since after 32 minutes the G4 is almost totally unfolded and hybridized. As a result of the data fitting using equation (1), we obtain a $t_{1/2}$ of 19 minutes without ligand, which can be used as reference to assess the efficiency of a ligand to protect G-quadruplexes against complementary strand. **Figure 39** shows the results we obtained with AG and TriPropil ligands. Even if TriPropil is a better stabilizer than AG in terms of ΔT_m, in NMR competition experiments, AG had the best protector properties and the longest with a $t_{1/2}$ of 170 min (k = $5.87x10^{-3}$±$8.12x10^{-4}$ $mM.min^{-1}$) whereas TriPropil had a half-life time of 46 min (k = $31.0x10^{-3}$±3.27x10-3 $mM.min^{-1}$). Compared to the control experiment, TriPropil only slightly increases the protection of the G-quadruplex meaning that it can well stabilize 22RT G-quadruplex but it is not able to avoid duplex formation in presence of complementary strand. AG is not the best stabilizer but increase the protection by almost a factor 10. We also tested several other ligands presented in **Figure A5** JL205 and NaphTrip. For NaphTrip, the obtained half-life is even worst compared to TriPropil which means that this ligand is no good in stabilization or protection neither. JL205 gave an average result with a $t_{1/2}$ of 90 min increasing the protection of the G-quadruplex by a factor 4.5.

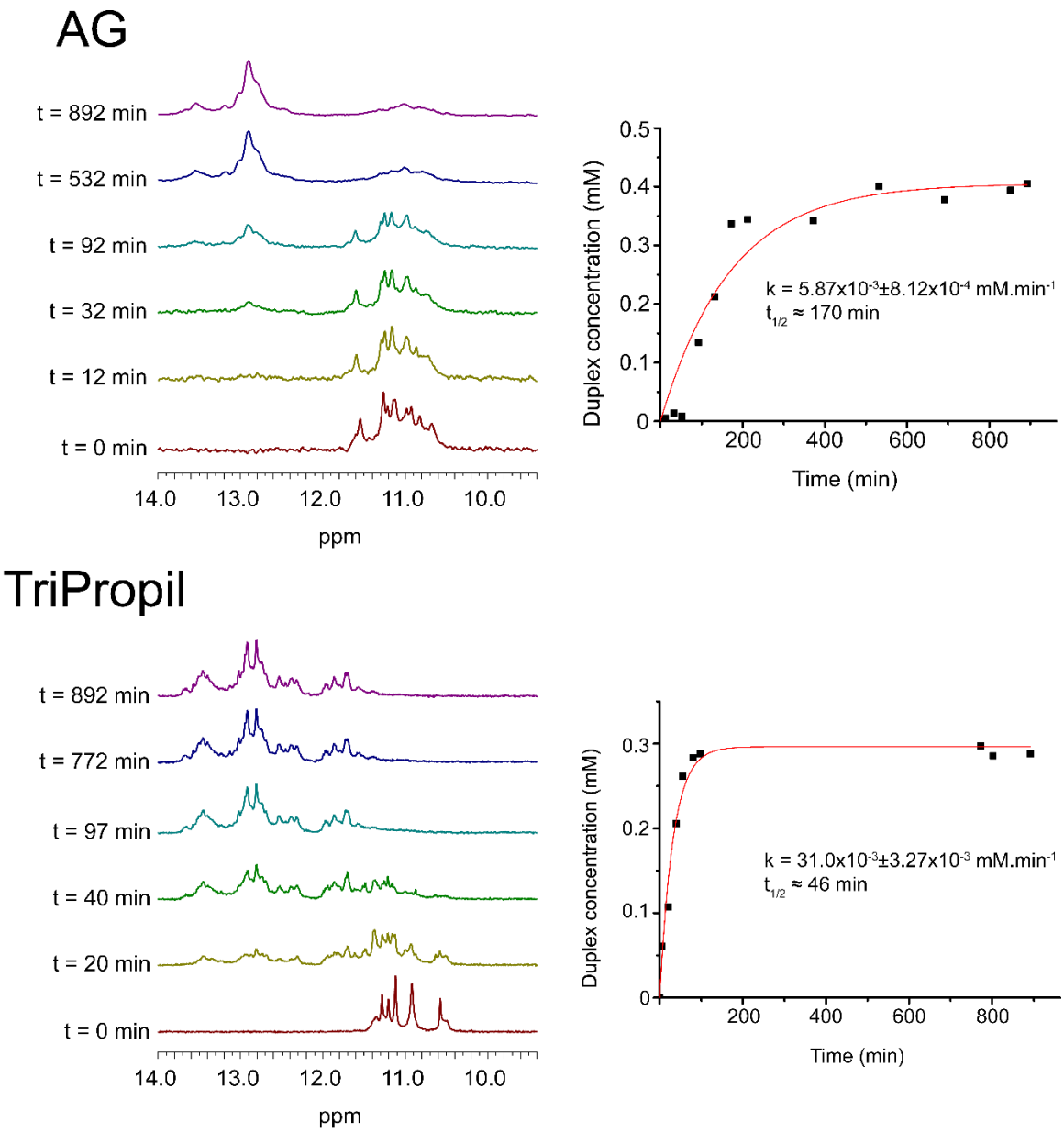


Figure 39. NMR competitions results for AG and TriPropil with NMR spectra and the corresponding fit

Nevertheless, this result needs to be carefully taken since it very difficult to correctly integrate the duplex signal. In the acridine orange paper **(Article 2 in Annexes)** we identified the C8 ligand as a very good stabilizer with promising cytotoxicity results and an interesting cellular colocalization. The NMR competition assays with 22RT are depicted in **Figure 40**.

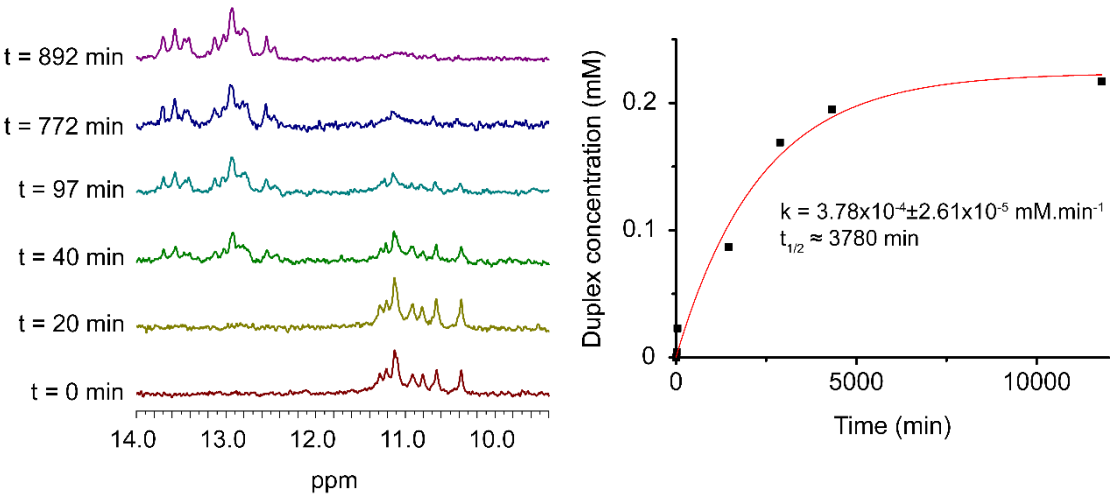


Figure 40. NMR competitions results for C8 an acridine orange derivative ligand tested in a previous study

It gave the best results in terms of NMR competition with a $t_{1/2}$ of 3780 minutes, which is more than 20 times superior to all the other ligands. Its capability to protect 22RT G4 from complementary hybridization looks very interesting as it gives hints about a possible binding pocket with several interacting points. All competition results in this study have been resume in **Figure 41**.

Ligand	k (10^{-3} mM.min^{-1})	$t_{1/2}$ (min)
None	50.9 ± 10.8	19
NaphTrip	33.8 ± 1.68	30
TriPropil	31.0 ± 3.27	46
JL205	11.1 ± 0.77	90
AG	5.87 ± 0.81	170
C8	0.38 ± 0.03	3780

Figure 41. NMR competitions results for all ligands tested in this study with also results without ligands. Compounds have been classified depending on their efficiency in protecting KRAS22RT G-quadruplex from complementary strand

Although we expected the hybridization profile (corresponding to duplex signal) to be similar with the different ligands, it was obvious that it was influenced and target by each ligand because the peaks profiles were quite different.

III.5. Cytotoxicity assays and cellular imaging

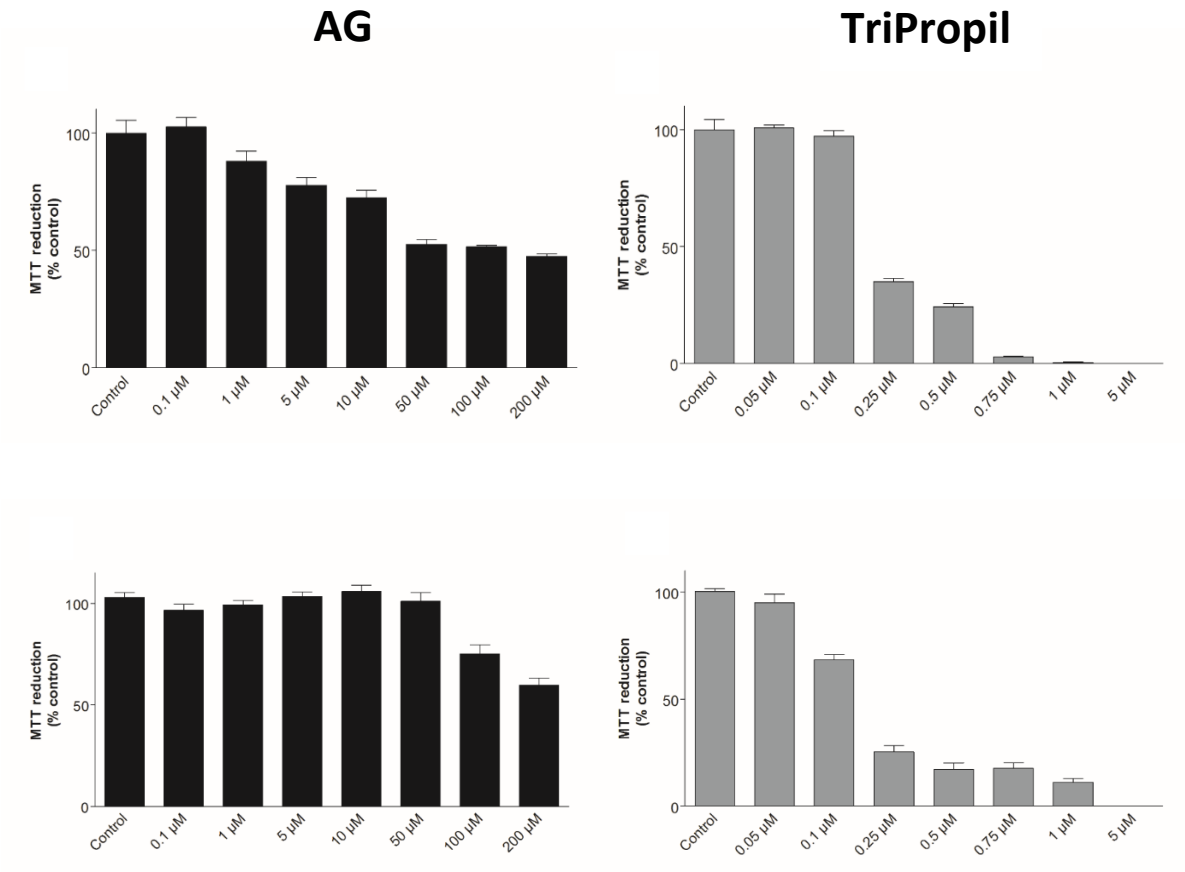
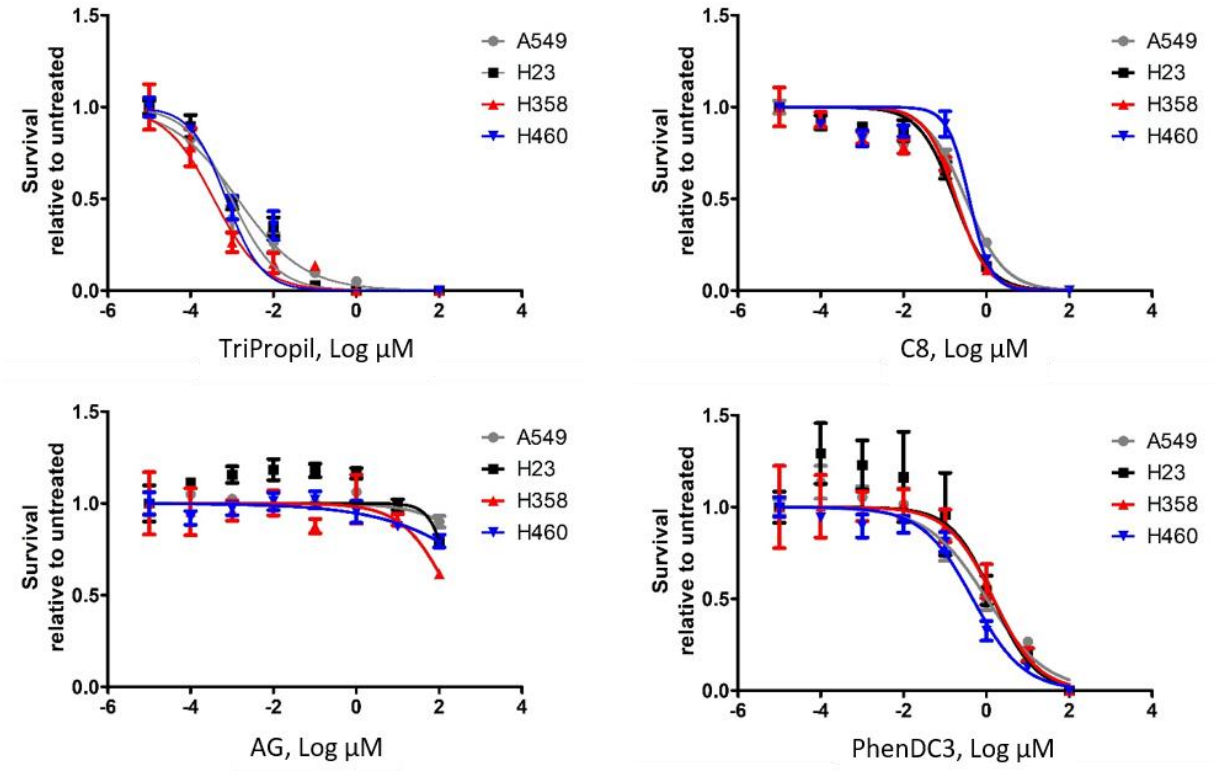


Figure 42. Relative cell viability of HeLa cancer cells (first line) and NDHF cells (second line) measured by the MTT assay after 72 h incubation with ligands AG and TriPropil at different concentrations. Bars represent the mean of cell viability relative to control cells (no treatment)

For these experiments, we first decided to test two ligands: AG and TriPropil in HeLa cells and NHDF cells. AG was chosen because of the results obtained in NMR titrations and competitions and TriPropil was chosen because despite the poor results we obtained in NMR, it is still a very good stabilizer for 22RT G-quadruplex. Results obtained for cell viability in HeLa and NHDF cells are presented in **Figure 42.** These experiments have been performed with our collaborators in Portugal. Results indicated that TriPropil caused a remarkable decrease in cell viability of HeLa with IC50 of 0.24 µM. At higher concentrations of TriPropil (1-5 µM) less than 5% of HeLa cells were viable but this compound also affected the normal cells even at low concentration, suggesting that TriPropil is not selective for HeLa cancer cells. The compound AG is less cytotoxic to HeLa and normal cells. At the highest concentration of AG (200 µM) about half of the cancer cells survived (47 %). In the case of normal cells, at the same

concentration, the cell viability is 59 % suggesting a modest selectivity for cancer cells. As a conclusion, none of these ligands is specific to cancer cells.



Cell line	TriPropil	C8	AG	PhenDC3
A549	1.352 nM	0.313 µM	>100 µM	1.020 µM
H23	0.962 nM	0.168 µM	>100 µM	1.458 µM
H358	0.359 nM	0.186 µM	>100 µM	1.478 µM
H460	0.718 nM	0.383 µM	>100 µM	0.467 µM

Figure 43. Relative cell viability of the different cancer cell line measured by the MTT assay after 72 h incubation with ligands at different concentrations. Results with different cells lines are plotted in graph and in a table with summarized values

Indeed, we studied G-quadruplex formed within KRAS promoter region that is why we decided to test again our ligands but in KRAS-driven cancers cell lines A549, H23, H358 and H460 respectively with mutations in Kras G12S, G12C, G12C and Q61H. The last one has been chosen in order to study another mutated residue which is not G12. In this assay, we tested as before AG and TriPropil but also PhenDC3 which is a good stabilizer and C8 the best ligand we tested **(Figure 43)**. These experiments have been done with our collaborators at IECB. AG and TriPropil showed similar results that the ones obtained in HeLa cancer cells with TriPropil which is too toxic with IC_{50} under nanomolar range and AG which is not enough toxic with IC_{50}

superior to 100 µM in every cell lines. Concerning the other ligands PhenDC3 and C8, they show interesting results with IC_{50} around 1 µM for PhenDC3 and C8 with an IC_{50} in the nanomoles range which corresponds to the desired value for an IC_{50}. Unfortunately, C8 also showed similar toxicity for NHDF cells in the previous studies. However, another critical point to assess whether a molecule is a good candidate considered as a potential drug is the cellular uptake and localization. Among all the G-quadruplexes ligands that have been tested cellular uptake remain a major issue and need to be improved. That is why we tested AG and TriPropil to see their localization by cellular imaging in HeLa cancer cells with confocal microscopy **(Figure 44)**.

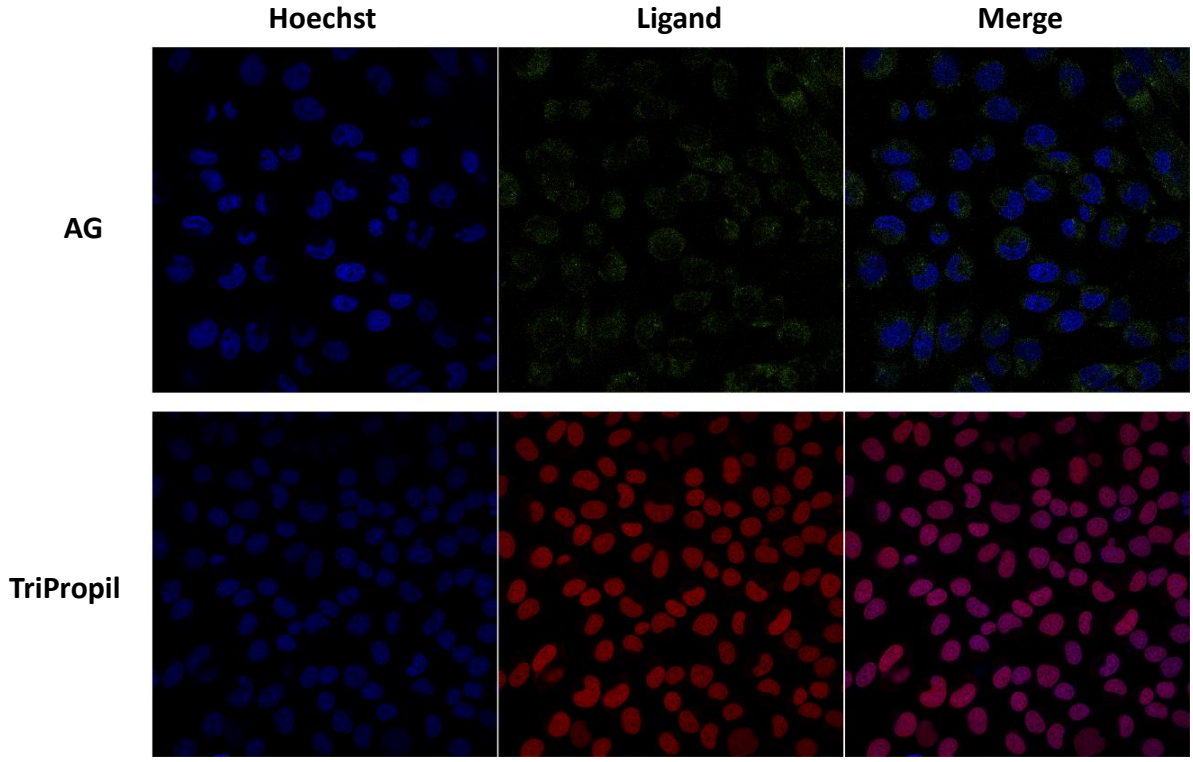


Figure 44. Confocal microscopy images of the subcellular localization of compounds AG and TriPropil in HeLa cells after 3 h incubation at 10 and 1 µM concentration, respectively. Cell nucleus was counterstained with Hoechst 33342 (blue)

We were able to study localization of these two compounds thanks to their fluorescence properties. AG emits green and TriPropil red fluorescence, upon excitation at 488 and 514 nm, respectively. Even if AG is a poor fluorophore compared to TriPropil, the emission is sufficient to clearly identify their cellular localization by merging with nuclear staining fluorescent compounds. As a results, we could identify TriPropil localization in the nucleus, whereas AG is

more abundant in the cytoplasm. These results indicate that AG does not have the ability to pass as easily through the nuclear pore nor diffuse over the nucleus double membrane. Looking at the structures of the two ligands, we can hypothesize that the cause can be related to both the aliphatic chain which makes AG longer and the Ni^{2+} cation. Taken together, the biophysical preliminary studies with different ligands allows to select a few ligands worth solving the structure in complex with 22RT. Among them, we identify AG, TriPropil and C8 as the most interesting ligands. We started the 2D NMR experiments with AG since it had the best peak dispersion in the imino region. To unambiguously identify the position of each imino we proceeded with ^{1}H-^{15}N 1D HMQC experiments in different individual samples containing an isotopically labelled guanine at each position.

III.6. Structural studies of KRAS22RT G4 in complex with AG ligand

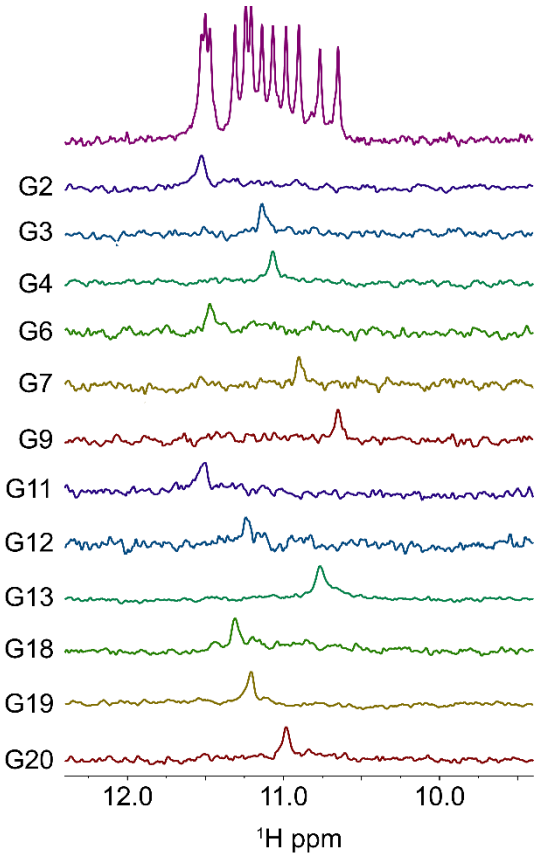


Figure 45. Imino peaks assignment of KRAS22RT in presence of AG using ^{1}H-^{15}N 1D HMQC -filtered NMR spectra of samples containing ≈5% of ^{15}N-enriched isotope

In order to directly find interactions between both molecules, we first needed to assign proton peaks from each molecule. I was able to identify all 12 imino peaks in presence of AG from the

1D ^{1}H-^{15}N HMQC **(Figure 45)**. One of the first observations from the 1D titration experiments with AG, was the fact that some peaks disappear and reaper later on. So, the 1D HMQC experiment was crucial to identify slow exchange peaks that underwent large chemical shift displacements such as G9 residue which had a chemical shift near 11.2 ppm and moved to 10.7 ppm, experiencing a difference of 0.5 ppm. Then, I proceeded to record 2D NMR spectra such as ^{1}H-^{13}C HSQC, ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY in order to characterize and assign all possible peaks and compare with free 22RT. Once assignments were completed from an exhaustive observation of different ^{1}H-^{1}H NOESY spectra acquired at different temperatures, we were able to assign (as much as possible) cross-peaks between 22RT and AG. In order to be able of doing that it was also necessary to assigned the individual protons of isolated AG samples. Fortunately, imidazole protons such as the methyl group, imid-H2, -H3 and -H4, together with aromatic protons (aro1 and aro2) and the aliphatic chain (-CH2 protons) were simple to identify from 2D NMR spectra. The last proton is located near to the cationic ion Ni^{2+} and called H_{JV} **(Figure 46)**.

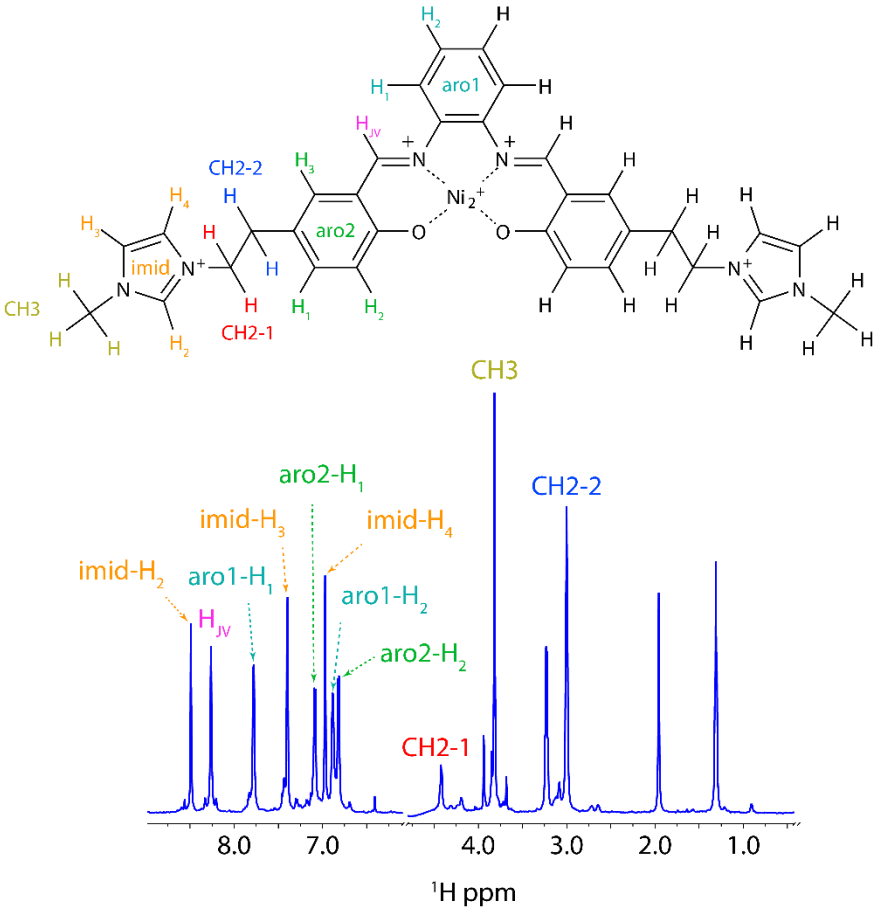


Figure 46. Structure of AG ligand with all protons and the corresponding peaks assigned in 1D ^{1}H NMR spectrum

Upon complexation formation with 22RT all peaks experience broadening and collapsing (**Figure A6**). Some of the cross peaks between 22RT and AG are depicted in **Figure 47**. The most important interacting peaks that we could unambiguously assign were imid-H2 and imid-H4 with H1' protons from G3, C5, G6, G7, G11, G12, A14, G18 and G19. Some cross peaks of imid-H2 and imid-H4 were also identified interacting with H3' of C5, G11, G18. Since AG is a symmetric molecule, we only named half of the protons because they are isochronous with the second half. So it is without surprise that we have several cross peaks from AG with so many different bases.

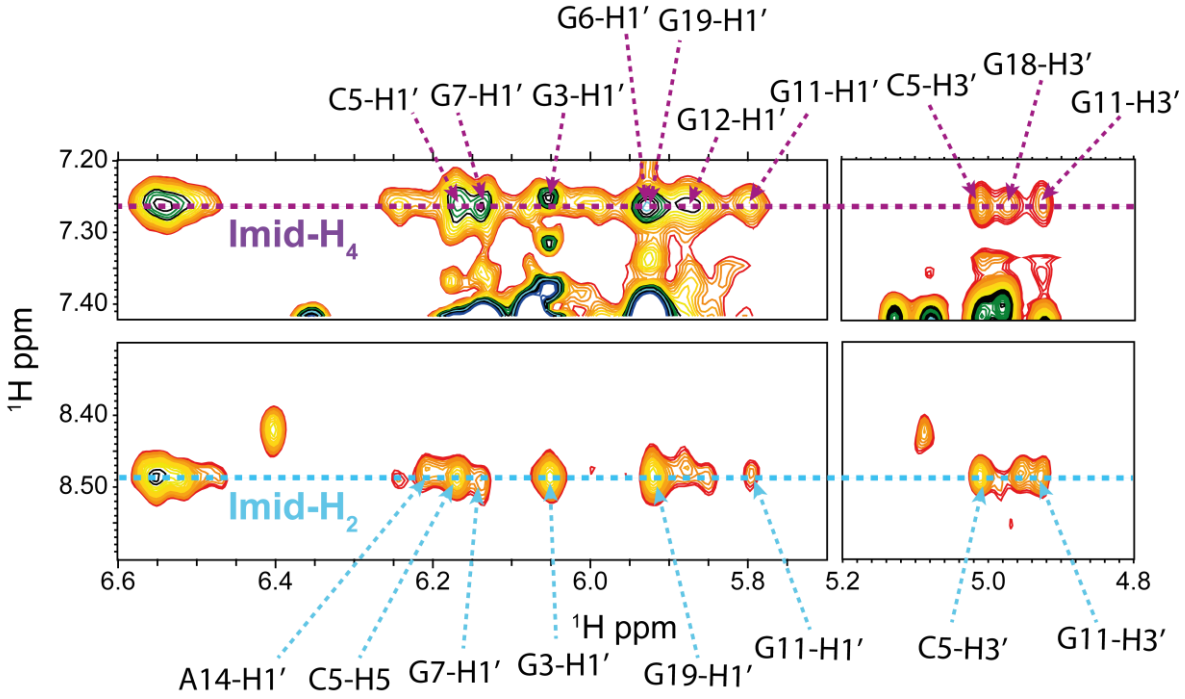


Figure 47. Interactions between AG ligand and KRAS22RT G-quadruplex observed in 1H-1H NOESY spectrum implicating protons from ligand imidazole group and several residues from KRAS22RT

Considering the interaction pattern and the number of ligand used until saturation, AG is likely interacting in two different binding sites that I was not able to solve. Further studies need to be performed before we could decipher the exact location of both binding sites. After validating the list of cross-peaks with AG a molecular dynamics run over several hundreds of ns would probably be necessary to understand correctly the binding mode of the different chemical moieties.

IV. KRAS32R G-quadruplex ligands study

By studying ligands against 22RT structure, we identified good candidates that could be used for ligand development. But before that we need to test them against longer sequences from KRAS that have also a propensity to fold into G4 structures. So we studied the same ligands against 32R, which is also a parallel G4 but contains a large propeller loop with 11/12 bases, depending on the conformer. We were expecting that the interaction behaviour could also be different. We tested the previous ligands against 32R wild type, G9T and G25T conformers. By looking at the results with the three different sequences, we wanted to know if we could find the same trend in the WT as in both G9 and G25 conformers or simply if any of the ligands could shift the equilibria of WT towards a single conformation. Similarly to what we did before, we tested a panel of ligands against all three sequences in order to obtain ΔT_M values from FRET and finally 1D NMR titrations.

IV.1. Stabilization assays against KRAS32R G-quadruplexes

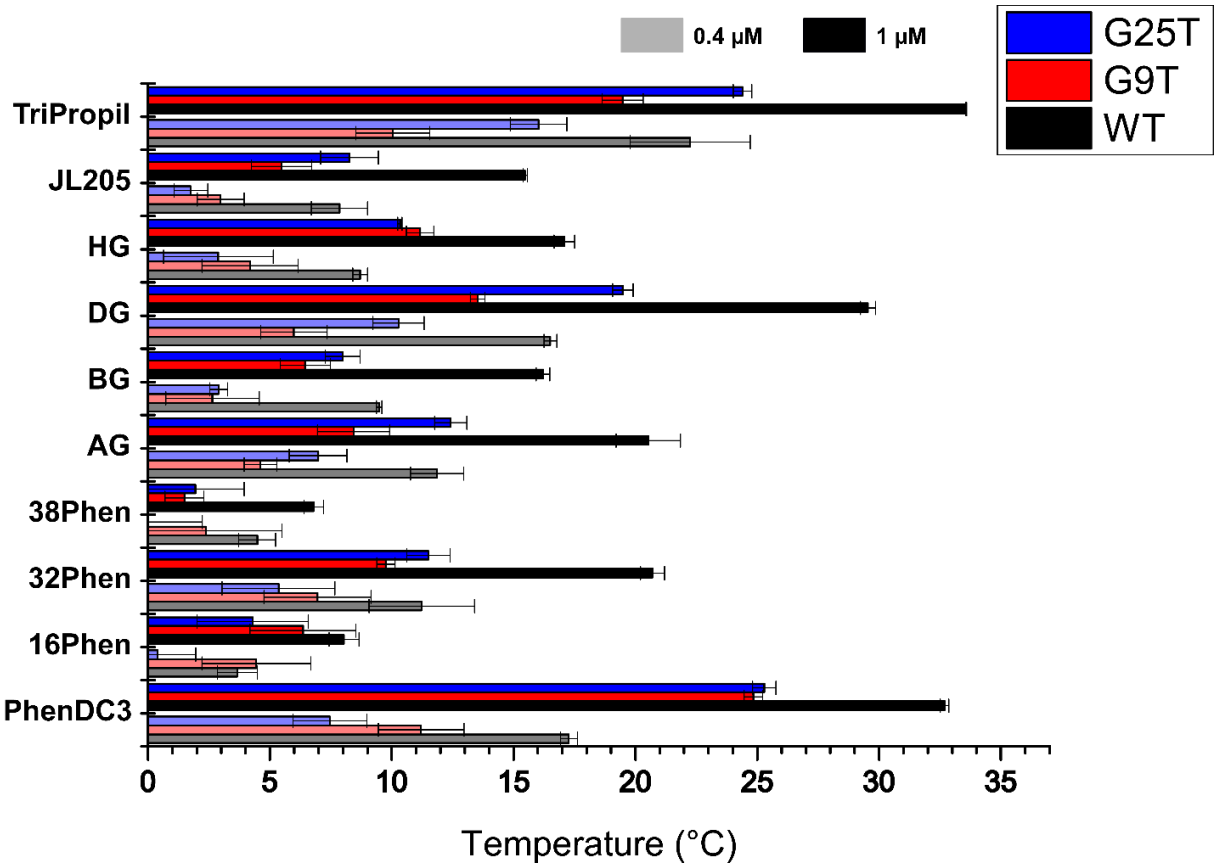


Figure 48. FRET melting experiments with a panel of ligands against 32R G-quadruplexes using wild type, G9T and G25T. All ligands have been tested at 2 and 5 molar equivalents of each ligand. ΔT_M have been obtained by subtracting the T_M of a negative control from the T_M of each well

Similarly to what has been done with 22RT, we performed the same type of FRET melting experiments to assess the stabilization of different compounds against KRAS32R. We also performed Circular Dichroism melting experiments to test C8 ligand as it cannot be tested by FRET because it absorbs at the same wavelength as the fluorophores used in KRAS sequences. The FRET melting results are presented in **Figure 48**, showing the stabilization temperatures with the three sequence, wild type (WT), G9T and G25T at 2 and 5 equivalents of ligands. Once again, PhenDC3 was used as positive control and gave very good results with a stabilization around 25°C for both G9T and G25T, and around 33°C for WT at 5 equivalents. Other ligands from the phenantroline family produced poor results (under 10°C for all three sequences at 5 equivalents with 16Phen and 38Phen). In the case of 32Phen the results indicate a stabilization of G9T and G25T by 10°C and more than 20 for the wild type. Another ligand which gave similar results to PhenDC3 is TriPropil with stabilization of more than 20°C for G9T and G25T and almost 35°C for WT. Concerning the salphen family, DG showed the best results with a minimum stabilization by 14°C for G9T and a maximum stabilization at almost 30°C for WT. Then AG and DG gave similar results (between 10 and 15°C for G9T and G25T and between 15°C and 20 and between 15°C and 20°C for WT at 5 equivalents). Still from the salphen family, BG is the ligand that gave the poorest results. Finally, the last ligand we tested so far in FRET melting, JL205 reported poor results for G9T and G25T and gave correct results for WT (around 15°C at 5 equivalents). After analysing the results in **figure 48**, a major observation that can be made is that all ligands gave a better stabilization with the wild type sequence. We can hypothesize that we were measuring the stabilization of the multiples conformations that could form aggregates in presence of ligands making some sort of macromolecular complex difficult to interpret. Results obtained with 32R G-quadruplexes are similar to 22RT for PhenDC3, TriPropil and AG ligands. As we obtained similar results with these ligands, we also wanted to assess the stabilization of C8 ligand which was the ligand with the best results in terms of ΔT_M with 22RT. As we could not use FRET melting we performed CD melting experiments **(Figure 49)**. Results from this experiment show that C8 is capable of stabilizing 32R G-quadruplexes but not as efficiently as 22RT. Even if the stabilization is lower than the one observed with 22RT (around 40°C at 3 equivalents) it could stabilize 32R WT, G9T and G25T with a respectively temperature of 14, 10 and 19°C at 5 ligand equivalents. It seems that C8 could stabilize with a better efficiency WT and G25T compared to G9T and this seems to be inversely proportional to T_M values of each sequence without ligand. This phenomenon

was also observed in FRET melting experiments but the difference is that in CD melting the stabilization for WT at 5 equivalents seemed to be an average between G9T and G25T values.

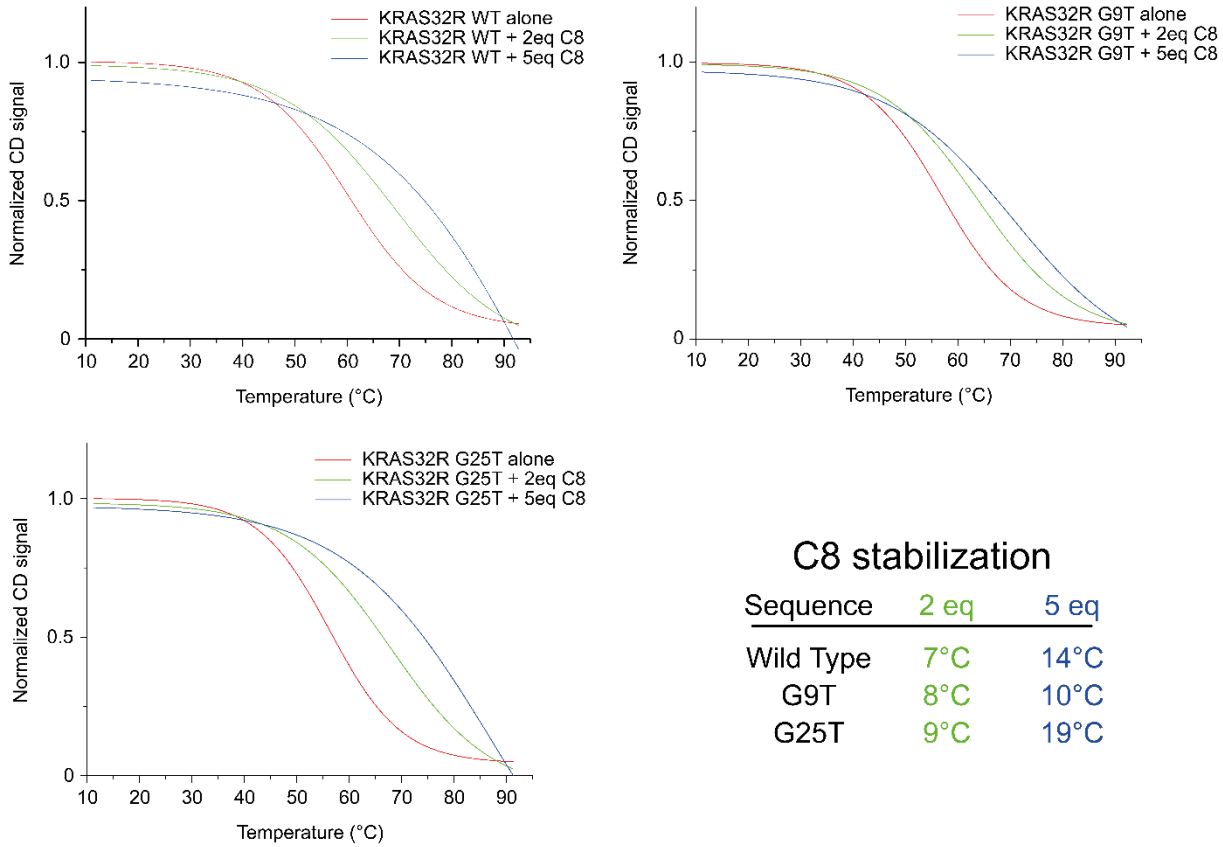


Figure 49. CD melting experiments with C8 ligand against 32R G-quadruplexes using wild type, G9T and G25T sequences. The table on the right summarizes the variation of T_M in presence of ligand.

Considering the differences between FRET and CD melting experiments, we could assume that our hypothesis concerning aggregation in FRET melting wells may be a valid point. Despite the fact that C8 had lower stabilization effects on 32R G-quadruplexes, it remains a good stabilizer. In addition, these results also reinforce the hypothesis that C8 ligand interacts in a tight binding pocket in 22RT made by a short propeller loop composed by 4-nucleotide (A14, A15, T16, A17), located near G11 which imino had an important cross-peak with C8 ligand (Figure 5, article 2, Annex). The 32R sequences do have a longer propeller loop that most probably does not make such tight binding pocket.

IV.2. NMR titrations

As previously reported, NMR titrations provide useful information by monitoring the chemical shifts of isolated peaks, inspecting possible exchange events and determine if samples have enough quality to be involved in further NMR structural studies. In order to have robust results about the interaction between 32R G-quadruplexes and ligands, we continued to study WT, G9T and G25T in the same conditions. Based in the FRET results and cellular cytotoxicity we decided to use PhenDC3 and C8 ligand as a starting point for the new set of NMR titrations resumed in **Figure 50**.

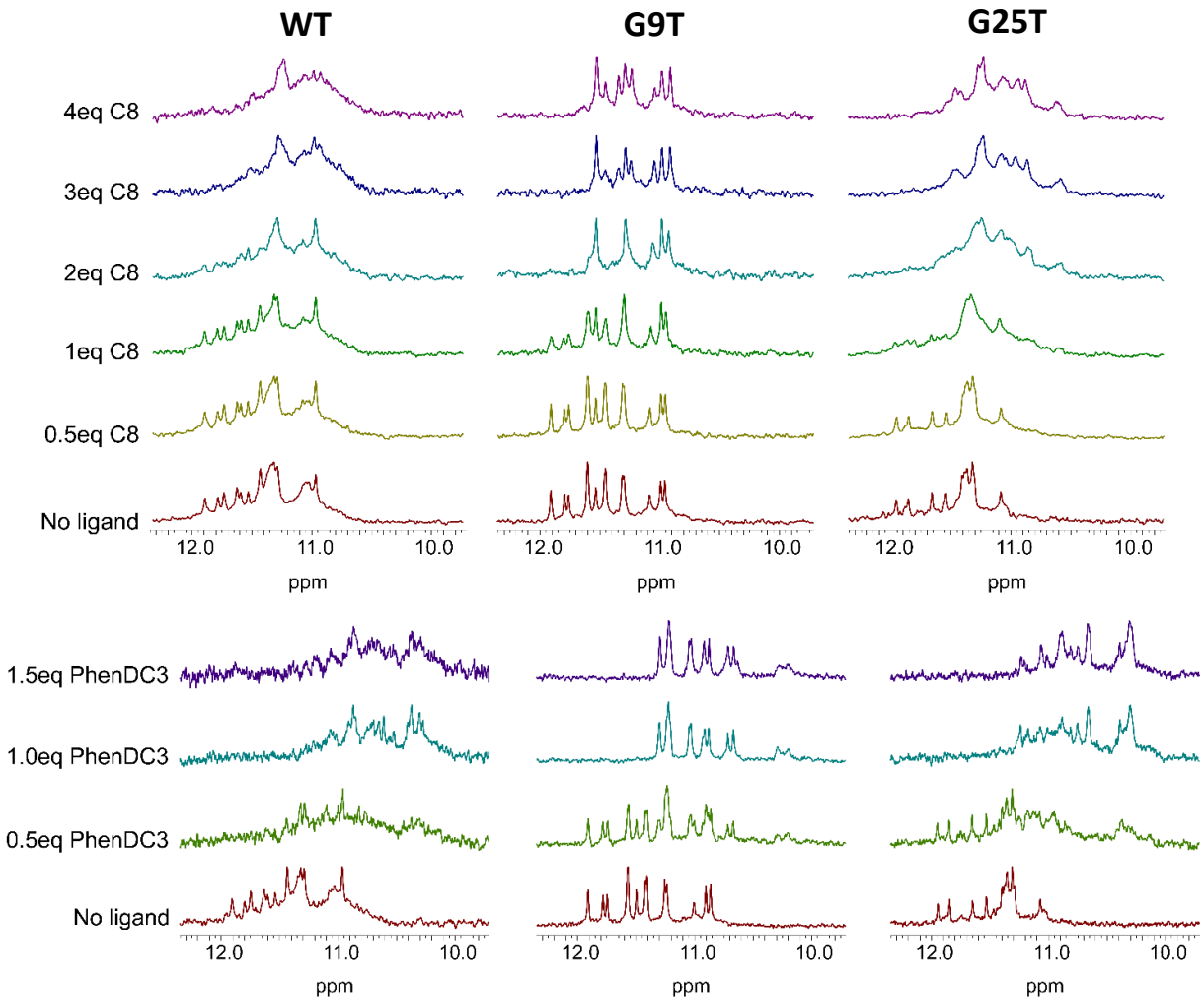


Figure 50. NMR titrations performed with the best stabilizers with increasing amount of ligands up to the point we started observing aggregate formation in the NMR tube.

From this figure it can be concluded that the interaction with 32R WT sequence induced the formation of multiple indistinguishable peaks. The presence of the two ligands does not induce the appearance of a single conformer and lead to small aggregates. In the case of G9T and C8, the titration depicts a nice profile with well-resolved peaks in which we could clearly

identify the isolated peaks even at 4 molar equivalents of ligands. At ratios off 1:1 (C8:G9T), it seems clear that intensities of the three imino peaks (G2, G6, G11) belonging to the 5' side tetrad together with G25, are severely diminished compared to the remaining residues (G25 is not unambiguously detected in 1D spectra). The behaviour is characteristic of an intermediate exchange processes (µs-ms) and could be indicative of a preferred binding site region. It is difficult to extract more information from 1D spectra, but it shows good quality to proceed with 2D NMR characterization. At the same amount of ligand, G25T showed a better imino profile, with a completely set of unresolved imino peaks, hinting that multiple conformations are in exchange in the NMR time scale. G25T shows severe peak coalescence, which is lower limit of intermediate exchange with multiple unspecific binding events. Unfortunately, PhenDC3 ligand interaction were not completed up to 3 or more equivalents of ligand. We can observe that for all 3 sequences the co-existence between the free and bound states of the G-quadruplexes that are shifted upfield. In G9T profile, a new set of well resolved peaks is appearing around 10-11 ppm. In G25T profile, even if not so well resolved compared to G9T, we also could distinguish a second set of peaks in the same region whereas WT presents multiple small and undistinguishable peaks. At 1 molar equivalent of PhenDC3, the free state of the G-quadruplex cannot be seen anymore indicating the saturation and as a consequence one preferred binding site. Regarding the evolution of the imino profile, the exchange is at intermediate regime, indicating a medium binding mode in the order of µM. Considering future NMR structural studies, WT will be almost impossible to follow, G9T could be studied with both ligands but for G25T, only PhenDC3 seems to be a valid candidate.

V. KRAS G-quadruplexes interaction with UP1

In the ligand interaction studies, we were able to identify good candidates that can be used as G4s stabilizers for further structure preliminary studies. Next we wanted to know if the same ligands were capable of interfering with transcription in cancer cells. But before that we need to observe if the same ligands can be used to inhibit the interaction with proteins such as hnRNP A1 *in vitro*. If ligands could have both stabilization and capping properties, i.e., could prevent transcription factors from binding and unwinding the G4 structures. Before directly using the ligands against KRAS addicted cancer cells, we performed several experiments in order to characterized the interaction between all KRAS G-quadruplexes we studied so far (22RT, 32R, G9T and G25T). Indeed, the role of hnRNP A1 and UP1 has been shown for 32R

WT and 21R[279, 281] but we did not have any data concerning the isolated conformers G9T and G25T. In this part, I describe the tests concerning binding and functional properties in presence of UP1, and a possible unfolding mechanism, using PAGE, ITC and NMR.

V.1. Interaction studies between KRAS G-quadruplexes and UP1 by native gel experiments

I performed native gel experiments to assess the interaction between KRAS G-quadruplexes and UP1. This assay would directly allow us to see if there is any interaction with a shift of the band corresponding to G4 in case of interaction. Moreover, by using a ladder including a panel of different sizes we could also have an idea of the stoichiometry of the complex formed. I performed the native gel experiment with 22RT, 32RWT, G9T and G25T. I also included a sequence of 32 residues corresponding the wild type sequence but with several mutations (G into T) in all guanines tract to avoid G-quadruplexes formation (32R mut) **(Figure 51)**.

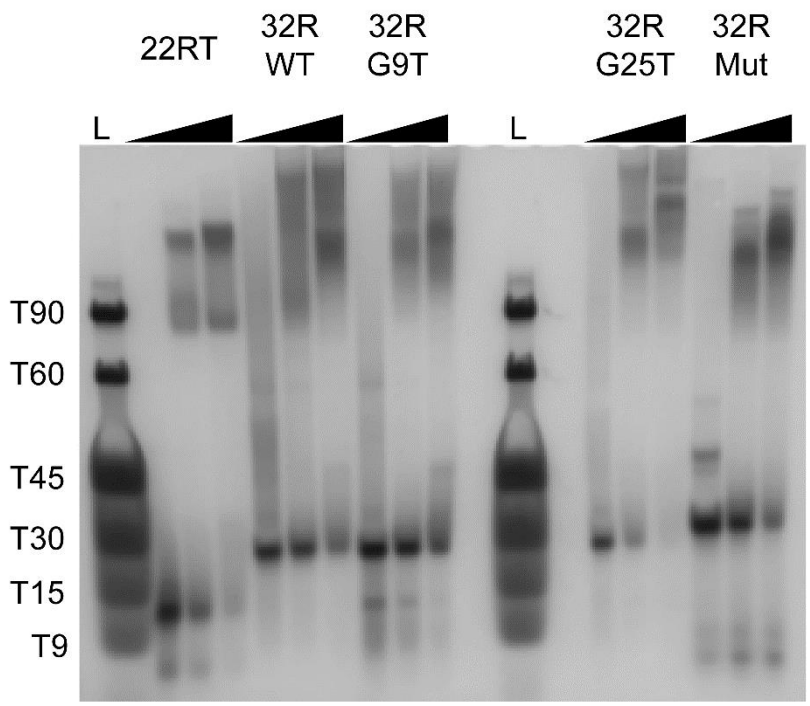
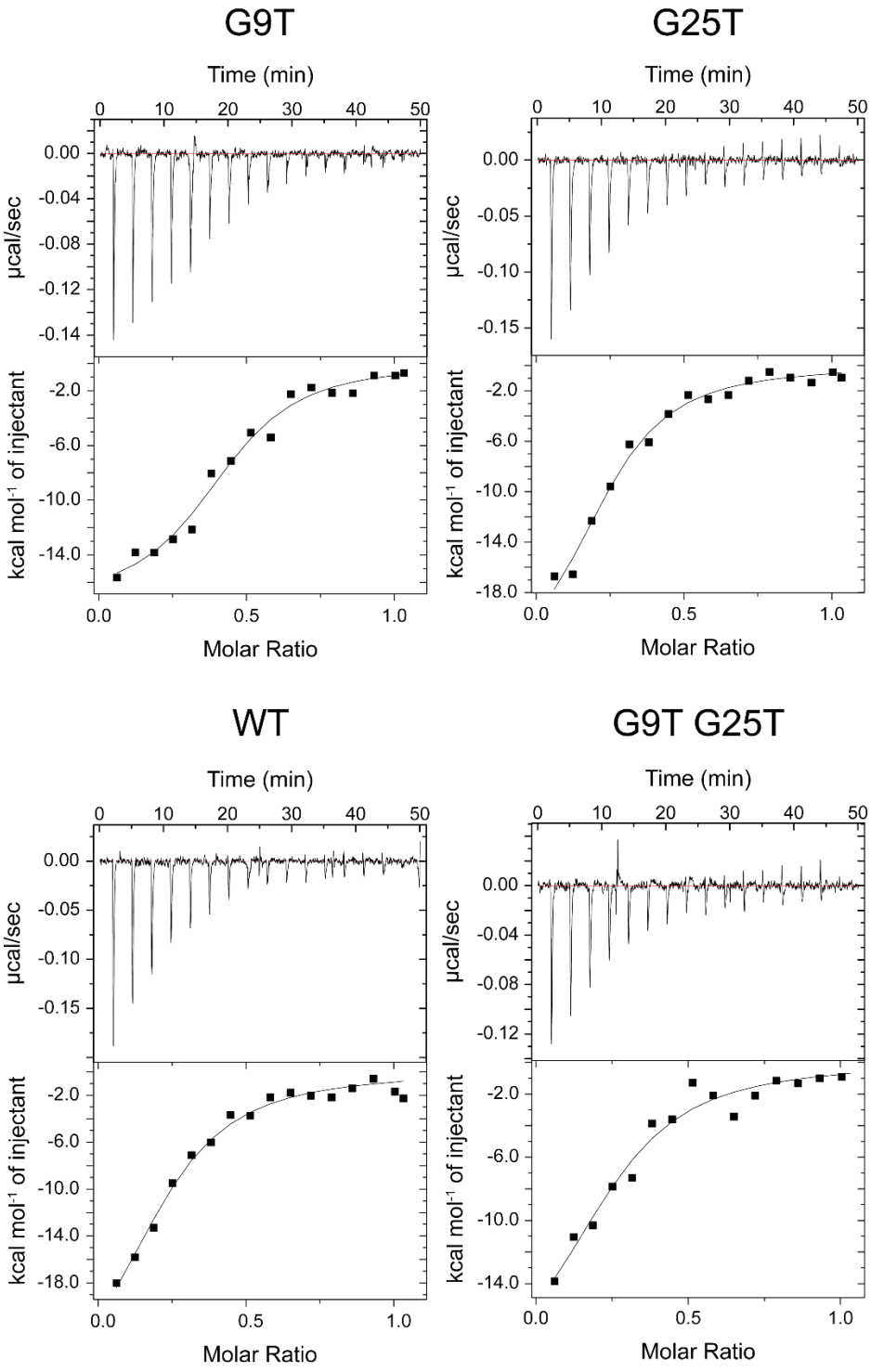


Figure 51. Native gel experiments to assess KRAS G-quadruplexes and UP1 by using different sequences: KRAS 22RT, 32RWT, 32RG9T, 32RG25T and 32RMut in presence of 1 and 2 equivalents of ligand. A ladder has been added containing several polythymine sequences T9, T15, T30, T45, T60 and T90 respectively corresponding to 2.7, 4.5, 9.0, 13.6, 18.2 and 27.3 kDa.

As G-quadruplexes are more compact structures compared to linear polythymine samples, the corresponding band are below the expected size (under T30 for 32R samples and under ≈ T15 for 22RT). In the case of 22RT, in the presence of 1 equivalent of UP1, two bands are visible ≈T90 (27 kDa) and above, while the band of the G-quadruplex becomes less intense. This means that the addition of UP1 induced the formation of two different complexes with the binding of one or two molecules of protein. By adding 1 more equivalent of UP1, the two bands corresponding to the different complexes sizes becomes more intense and the free G4 band almost disappeared. For the 32R G4 sequences, we observe similar results with additional complex bands with lower mobility in the gel than those observed for 22RT. We could also notice that UP1 is not specific for G4 sequences. Indeed, with the addition of UP1, the 32R mutated sequence is also forming a complex with the protein with the highest band appearing above the T90 mark. We can conclude that all G4s sequences have a pattern that is compatible with the formation of 1:1 and 1:2 complexes as observed before and the negative control seems to form predominantly a 1:1 complex. The nonspecific interaction of UP1 makes sense as it corresponds to two RRM domains which are known to bind to nucleic acids and not only to G-quadruplexes. Moreover, no sufficient differences can be seen between interaction with 22RT and 32R G4s despite their structural differences. Concerning the unfolding role of UP1, this experiment did not allow us to conclude about the capacity of UP1 to unfold G4s.

V.2. Interaction parameters obtained by ITC

After the confirmation that 32R WT and conformers are capable of interacting with UP1, we wanted to obtain interaction properties such as stoichiometry, K_D and other thermodynamics parameters in order to better characterize and understand this phenomenon. Due to low yields in expressing UP1, titrations were conducted with UP1 in the sample cell and DNA in the injection syringe as a reverse titration. In addition to WT, G9T and G25T, we also tested the double mutant G9TG25T known not to form a stable G4 but rather multiple different conformers in slow exchange. So, G9TG25T cannot be used as a pure negative control.



Sequence	n (sites)	K_D (µM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)
WT	**0.23 ± 0.02**	**1.1 ± 0.25**	**-8.5 ± 1.9**	**-29 ± 4.0**	**-20.5 ± 5.9**
G9T	**0.42 ± 0.02**	**0.49 ± 0.12**	**-8.9 ± 2.2**	**-17 ± 1.0**	**-8.1 ± 3.2**
G25T	**0.24 ± 0.02**	**0.79 ± 0.18**	**-8.6 ± 2.0**	**-25 ± 2.6**	**-16.4 ± 4.6**
G9T G25T	**0.25 ± 0.04**	**1.1 ± 0.43**	**-8.4 ± 3.2**	**-21 ± 4.5**	**-12.6 ± 7.7**

Figure 52. Thermodynamics parameters obtained from Isothermal Titration Calorimetry bonding studies with UP1 protein as target and WT, G9T, G25T and G9T G25T sequence as ligands at 37°C with the corresponding curves

The results **(Figure 52)** showed that UP1 has a moderate binding affinity for the KRAS quadruplexes in the order of μM, i.e., 1.1, 0.5, 0.79 and 1.1 μM for WT, G9T, G25T and G9TG25T respectively. The differences are not significant in view of the experimental errors. Interestingly, the double mutant G9TG25T has the same affinity for UP1 as the WT. Although there are in the literature examples with both 1:1 and 2:1 ratio (protein: DNA), in our case G9T displays a 2:1 stoichiometry. In the case of WT, G25T and G9TG25T the results deviate from this model and we relate that to the apparent extra degree of flexibility that those sequences seem to have from NMR spectra. The ΔG results have mild differences between all four sequences and the more structural-flexible sequences have a better enthalpic component. The entropic component seems to be far more important for the better folded model, G9T. With ITC experiments, we also expected to have information concerning the unfold of KRAS G4s by UP1. However, the ITC experimental data did not allow us draw a definitive conclusion. Regarding the results, we cannot clearly see a difference between interaction with the different conformers maybe because the way the titrations have been performed (reverse way) cannot reveal the diversity of the interactions with different conformers. I think this work needs to be repeated with the oligonucleotide in the cell compartment and the protein in the syringe.

V.3. Interaction studies by NMR

Next, we were interested in following the capacity of using previously identified ligands to disrupt and prevent UP1 binding to the KRAS G4s. It is more difficult and expensive in terms of preparing ^{15}N isotopically enriched oligonucleotides so we decided to proceed and express UP1 in E. coli. Isotopically labelling the protein also permits to follow the folding state, i.e., we can observe the formation of the complex. We were only able to do part of the experiments involving each G4 and UP1 to characterize their binding properties. After forming the complex at 1:1 molar equivalents, the sample started to precipitate and the ligand experiments were not conducted.

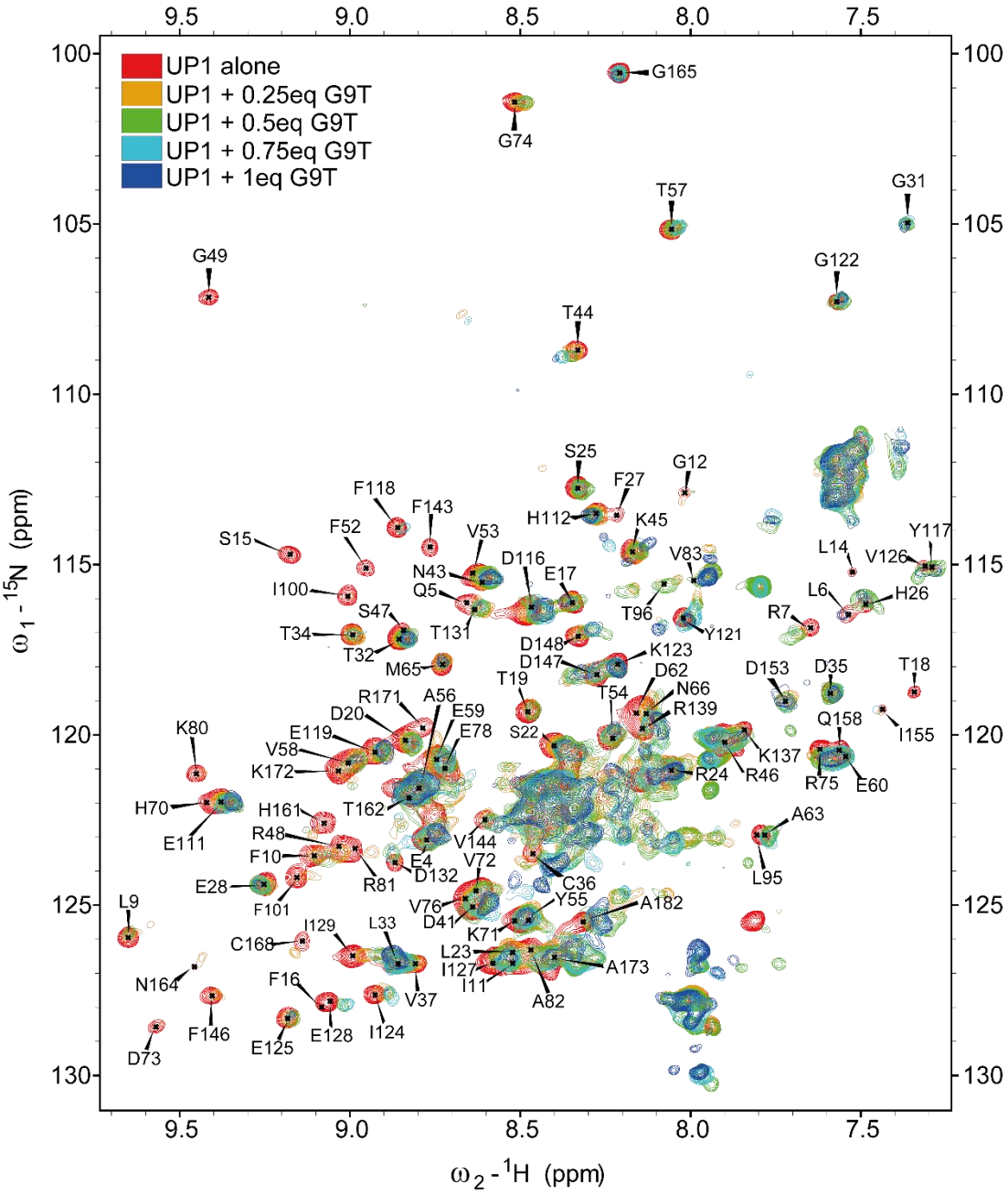


Figure 53. Superimposition of ^{15}N-^{1}H NMR HSQC spectra of UP1 showing each residue with NH bond of the backbone, measured alone and with increasing amount of KRAS32 G9T G-quadruplex

We performed titrations using isotopically labelled protein and followed with 2D ^{1}H-^{15}N HSQC NMR spectroscopy allowing us to follow the evolution of each protein residue upon the addition of DNA. We could extract chemical shift deviations (Δδ / ppm) for the amide group for the most affected amino acids. We measured HSQC spectrum without G4 sequences and

then we measured again after each addition of DNA. Finally, I superimposed all the spectra to assess shift of the most implicated residues in the interaction **(Figure 53 and 54)**.

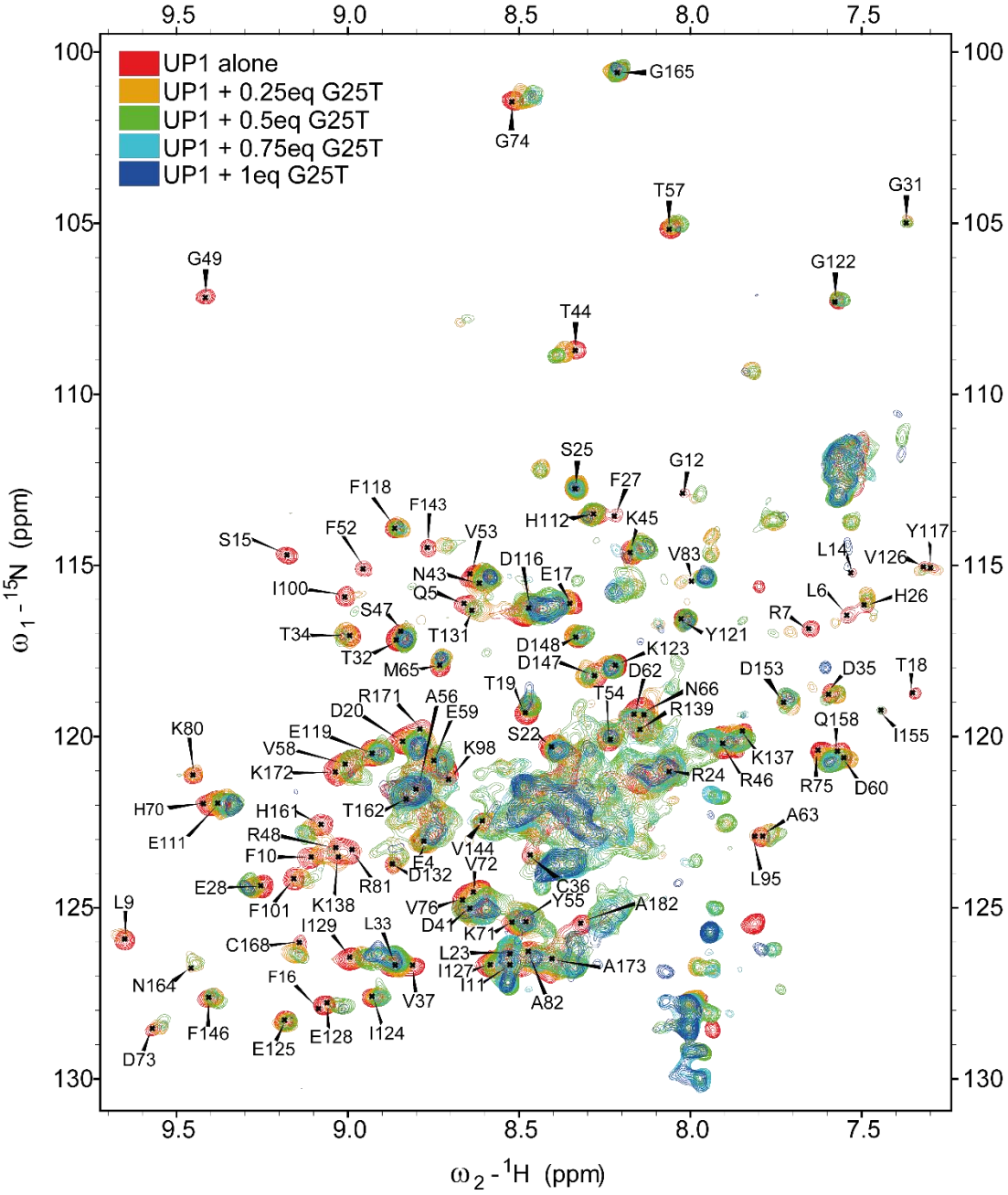


Figure 54. Superimposition of 15N-1H NMR HSQC spectra of UP1 showing each residue with NH bond of the backbone, measured alone and with increasing amount of KRAS32 G25T G-quadruplex

The spectrum of UP1 alone showed peaks which were separated and well resolved. Upon addition of DNA G-quadruplexes it became more complex and we cannot clearly see all peaks

especially in a central region around 8.5 ppm for ^{1}H dimension and around 122.5 ppm for ^{13}C dimension. Among the remaining peaks that could be analysed, some peaks do not shift or disappear. The first ones are probably not involved in the interaction but for the second set of peaks, it is difficult to conclude about their role. Then, when possible, we measured shifts in each dimension for peaks and calculated a global shift with an equation described in Material and methods. All shifts where plotted with the corresponding residue in a graph in **Figure 55**.

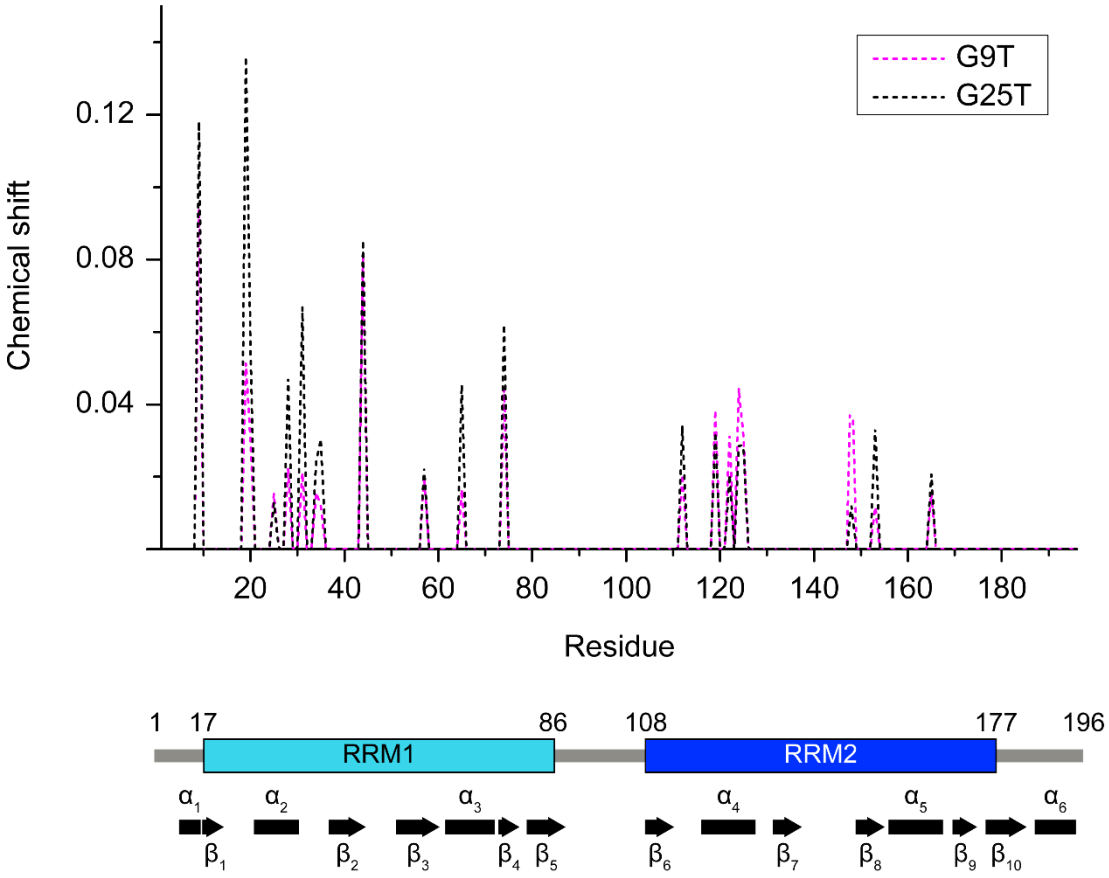


Figure 55. Plotted chemical shifts with their corresponding residue with a schematic view of UP1 structure to identify regions implicated in the interaction with KRAS32R G9T and G25T G-quadruplexes

Despite the fact that we missed several residues shifts due to the overlapping of peaks in the spectra, we had enough data to observe that the main regions interacting with G9T and G25T G4s corresponded to the two RRM domains of UP1. This result is not surprising considering that RRM domain is known to interact with a variety of nucleic acids sequences, and for the case of G9T and G25T with almost a perfect match between them. There is only one exception in the peaks we identified in the interaction which corresponds to Leucine 9. Indeed, this

residue does not belong to RRM domains but it seems to be highly implicated in UP1 binding and can be coloured- visualized in UP1 structure in **Figure 56**.

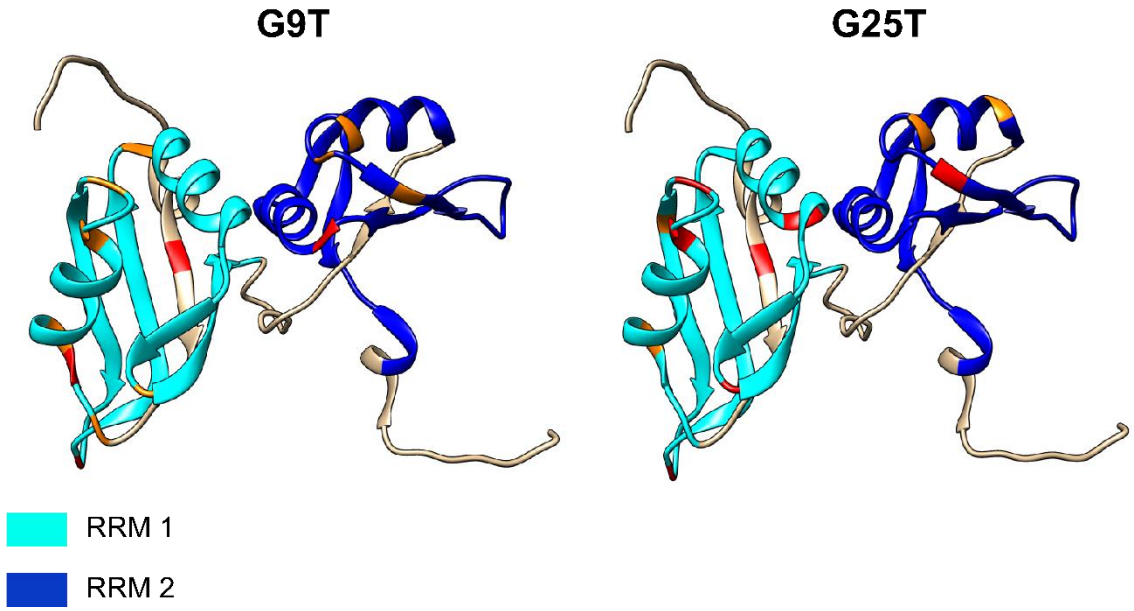


Figure 56. UP1 structure with the two RRM domains with the most shifted residues in NMR colored (red for strong shifts and orange for medium shifts)

By looking at the UP1 interactions identified with G9T and G25T, we could see that in both cases most of the residues that belong to the RRM domains but the residues involved are not exactly the same. Concerning the peak shift observed for Leucine 9, even it is not part of RRM domain 1, it is structurally close to this motif. Although we obtained some structural information about the interaction, especially in UP1 structure, this experiment does not provide direct observation that UP1 unfold the G4. For this end, we have to perform the same experiments looking at the G-quadruplexes structure instead of UP1, using labelled DNA sequences. Concerning the mechanism of interaction, if we considered a 1:1 stoichiometry, it means that two RRM domains from one UP1 molecule are needed which has to form a kind of "sandwich" with the G-quadruplex. In the case of a 2:1 stoichiometry then four RRM domains from two UP1 molecules are necessary for the interaction thereby for the unfolding process.

V.4. NMR preliminary studies of the unfolding by UP1

As we did not have any information about the unfolding process of G9T and G25T G4 by UP1 we tried to follow the event by 1D NMR experiments, similarly to the experiments we

performed in NMR competition studies with ligands. We looked at the imino profile of G9T and G25T G-quadruplexes before and after adding 1 equivalent of UP1 protein. The NMR spectra were acquired at different time periods **(Figure 57)** and the samples were kept at 37°C for a week.

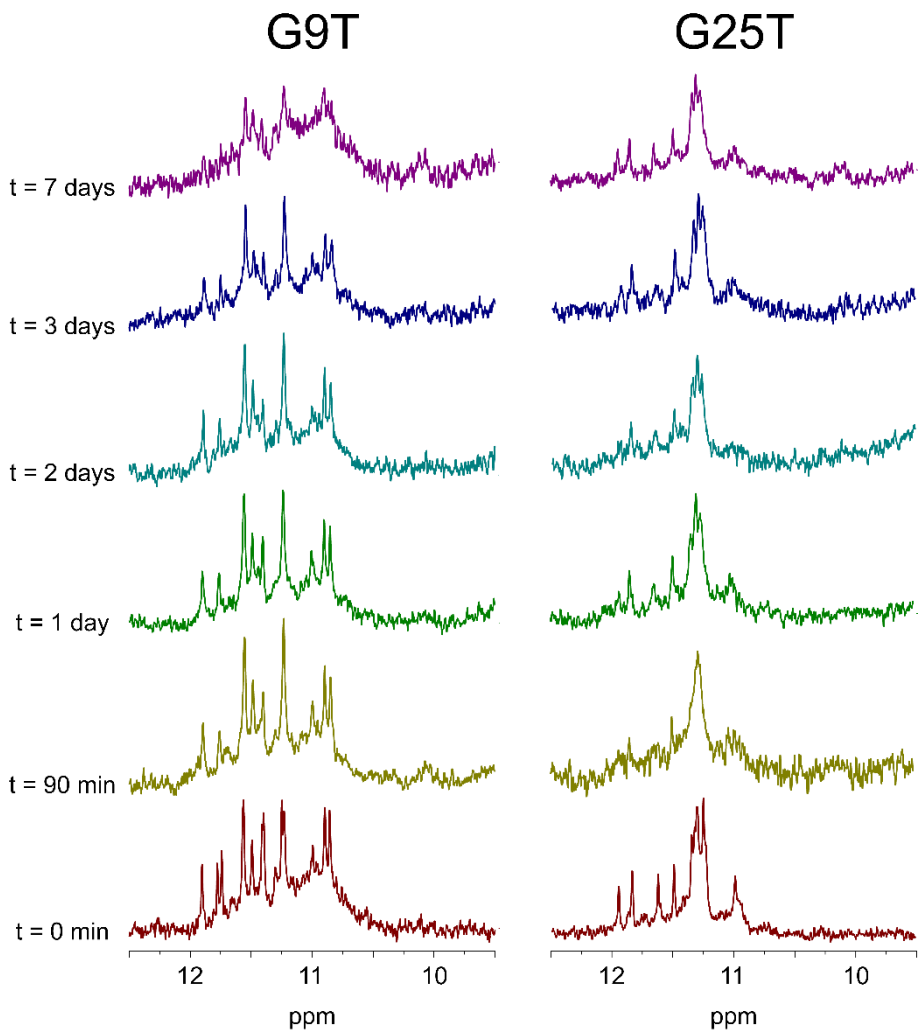


Figure 57. 32R G9T and G25T G-quadruplexes imino region after the addition of 1 equivalent of UP1 at different time periods.

The unfolding process of UP1 is not clear. Ninety minutes after the addition of UP1, we have the impression that very mild or almost no unfolding process did occur with G9T sample. In the case of G25, almost all G4 peaks disappeared, with the exception of the peaks of the central tetrad around 11.3 ppm. Visible changes were 24h after including UP1 in samples. In this case, the G9T spectra remain virtually the same but some peaks in G25T seemed to have reappeared. Even the peaks around 11.3 ppm look more resolved. After several days of

incubation at 37°C both G9T and G25T samples looks to start losing peak intensity and resolution. These results indicate that further unfolding is occurring and at same time samples are precipitating. The precipitation can be induced by both protein and DNA unfolding. I did not explore the conditions further due to a lack of UP1. For example, I would like to repeat the same experiment but in conditions where UP1 would be in excess (2:1) molar ration of oligonucleotide. Nevertheless, I was more encouraged with these experiments because G25T in presence of UP1 really looks that was partially unfolded. Since I was expecting a refolding back event it is not surprising the reappearance of some imino peaks after a certain amount of time.

VI. Preliminary results from ligands studies in UP1 interaction

Even without conclusive results about the unfolding role of UP1, assessed the action of ligands against the complex formed between KRAS G4s and UP1. We were obtained preliminary results with native polyacrylamide gel experiments. In these experiments I chose to use PhenDC3 and C8 ligands because they were tested with 22RT and 32R G4s giving good overall results. I included 2 equivalents of each UP1 and ligand **(Figure 58)**. As described before, I observed that a complex is formed with 22RT upon addition of UP1. Upon addition of PhenDC3, the band corresponding to this complex becomes less intense and new bands appeared with higher MW. This means that the ligand induced formation of a high-order complex with the protein and is also capable to protect the G4 by interacting with it. Similar results were obtained with 32R WT and the conformer G9T. For G9T we observed a band with a lower mobility compared to the absence of ligand meaning that PhenDC3 is interacting with G9T. For G25T, it seems that PhenDC3 is not capable of protecting the G4 from the interaction with UP1. In the case of the mutated 32R sequence, PhenDC3 was also capable of partially avoid the interaction with UP1 and looking at the new band corresponding to the sequence in presence of PhenDC3, it is also capable to induce the formation of a more compact structure.

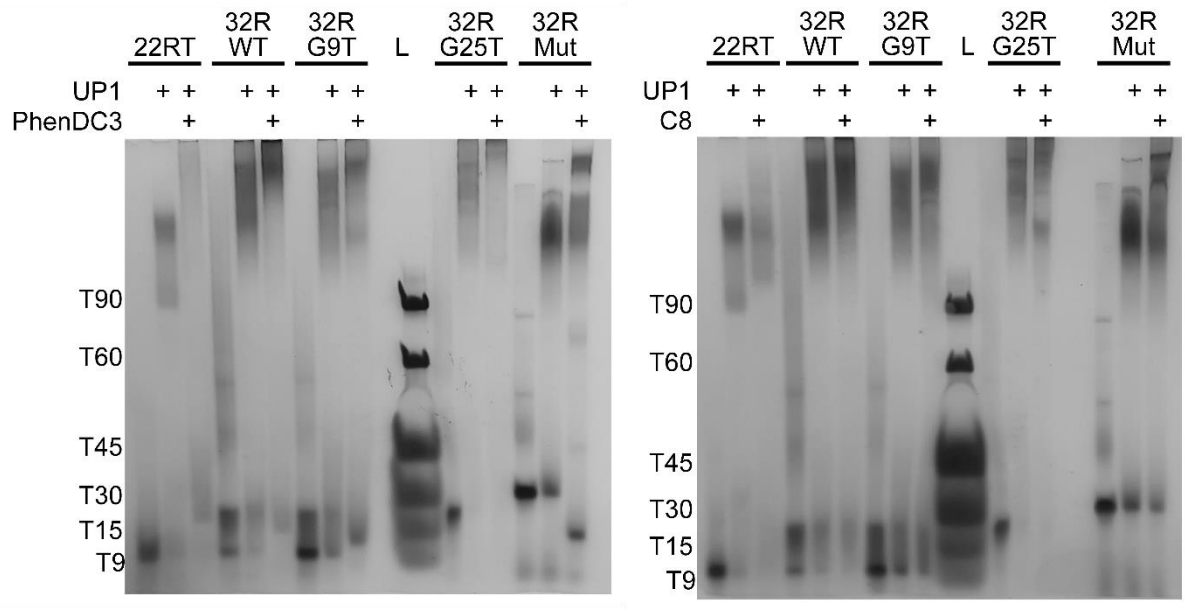


Figure 58. Native gel experiments to study the role of PhenDC3 and C8 ligands in UP1 interaction. We used KRAS 22RT, 32RWT, 32RG9T, 32RG25T and 32RMut and added 2 equivalents of each UP1 first and the ligand after. A ladder has been added containing several polythymine sequences T9, T15, T30, T45, T60 and T90

For C8 ligand, I observed that the addition of ligand did not have any effects with oligo/UP1 interaction. However, the ligand seems to enhance the interaction between UP1 and G-quadruplexes probably by interacting and stabilizing the formed complex. In conclusion, PhenDC3 seems to be a good ligand to inhibit UP1 interaction whereas C8 is not a good candidate in this end. We also tested the effect of PhenDC3 on the UP1 interaction by Circular Dichroism **(Figure 59)**. I first recorded the spectrum of 32R WT and the two conformers G9T and G25T. I obtained a parallel signal for all sequences. I then added 2 equivalents of UP1. I observed that the signal corresponding to the protein which appears as a negative signal around 225 nm. However, there were no sign of the G-quadruplex unfolding since the parallel signal remained unchanged. We finally mixed 2 equivalents of PhenDC3. G-quadruplex and the protein signals were severely modified with the disappearance of protein CD signature and the reduction of the G4 signal (about one third of the signal intensity). This observation could mean that the ligand has probably an effect on G4-UP1 interaction by interacting and modifying G-quadruplex structure.

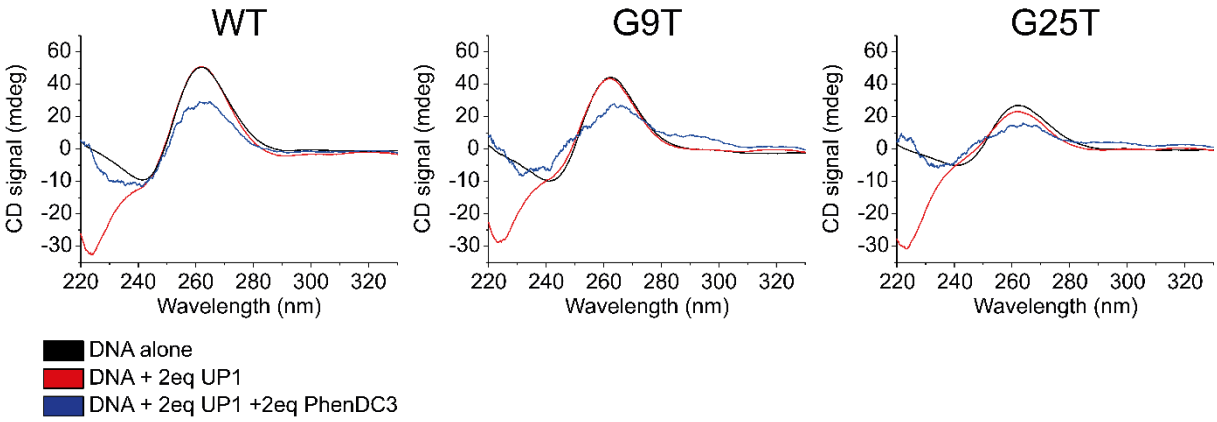


Figure 59. CD signal of KRAS32R WT, G9T and G25T upon the addition of 2 equivalents of UP1 and then PhenDC3. All G-quadruplexes showed parallel signal. Signal corresponding to the protein appeared around 225 nm

VII. Conclusions

In this part of the study, we tested a panel of different ligands from several chemical families in order to find good candidates for G-quadruplexes stabilization and prevent UP1 interaction through competition. Among all the ligands we tested, there is no ligand with desired properties. Several of them gave good results in term of thermal stabilization, protection from the complementary strand or UP1 protein, cytotoxicity or cellular uptake. We have tested several interesting ligands, among those PhenDC3 which can efficiently stabilize KRAS G-quadruplex, gives a good IC50 in KRAS-driven cancer cells and can inhibit UP1 interaction. Another one was TriPropil revealed to be a good stabilizer but it was too toxic. C8 is probably our best ligands regarding its stabilization effect and its capacity to protect KRAS G4s against complementary strand. Unfortunately, it seems that it cannot avoid UP1 interaction, nevertheless it has good cytotoxicity results with good IC_{50} but not enough specific for cancer cells. Although AG is not the best ligand, we decided to performed advanced structural studies with 22RT because it was the best candidate to this end. We wanted to obtain structural information of the interaction to know which part of the ligand is important for the interaction. We obtained a list of cross-peaks giving distance restrains that we could use in molecular dynamics between 22RT and AG, and finally determine the structure of the complex. For 32R G-quadruplexes, the best candidates for these studies would be G9T conformer with PhenDC3 looking at the spectrum we obtained in NMR titrations. Concerning

the characterization of UP1 interaction with KRAS G4s, we confirmed that this protein is capable of interacting with them and we obtained some structural information about UP1 binding and G25T partial unfolding. Our preliminary studies with ligands showed that it is possible to avoid interaction with UP1 as demonstrated with PhenDC3. However, the mechanism of its action remains unclear and additional experiments such as 2D NMR or ITC need to be performed.

Part 3: Discussion and Perspectives

The overall aim of this thesis was to understand how G-quadruplex from the NHE promoter region of KRAS organizes and folds into a stable 3D structure. Knowing about the 3D high-resolution of structures fits the global objective, understanding how molecular machines work. In addition, my project also aims identifying and characterizing new targets related to cancer, where numerous private and public laboratories do huge efforts to develop new drugs against cancer. Since no FDA-approval drugs that directly bind G4s are yet available, G-quadruplexes still need to be validate as important targets. So a global effort must be pursued to assess the relevance of nucleic acid unusual structures and at same time work at the level of drug development since both studies complement each other and help to achieve a steady-progress in the field. The goal is to design or find existing molecules that are capable of increasing the effect of G-quadruplexes which are known to reduce the efficiency of the transcription process. As G-quadruplexes have multiple putative roles in transcription by directly disturbing the process acting as a roadblock or indirectly by recruiting effectors (activators or repressors), we have two major possibilities. The first one is to find molecules that can bind and stabilize G-quadruplexes, the second one is to find molecules capable of inhibiting interaction between G-quadruplexes and effectors, with ligand that can bind to G-quadruplexes or to effectors. In this work, we focused on KRAS G-quadruplexes and especially the sequence called KRAS32R looking for molecules that can be used as candidates for cancers therapies. We were identified good stabilizers of KRAS32R and inhibitors of interaction with UP1, a hnRNP A1 derivative known to unfold KRAS G-quadruplexes (21R and 32R). Previous studies have identified ligands against KRAS G-quadruplexes, however, there is no structural information concerning this interaction and only few information concerning KRAS G-quadruplexes structure alone. In order to give new insights in the field of KRAS G-quadruplexes, we determined the major scaffold adopted in KRAS32R and we studied the interaction between this G-quadruplex and ligands by also assessing the interaction with UP1.

I. Determination of KRAS32R structure G-quadruplex

In this work, we successfully provided new structural insights by determining two structures of KRAS32R G-quadruplexes at physiological temperature. Indeed, we rapidly saw and confirmed by NMR that KRAS32R is polymorphic adopting several different conformations. After performing several mutations to understand the G-quadruplexes formation, we proposed a model with two major conformations in equilibrium along with uncharacterized

minor conformations (<5%). These conformations, named KRAS32R G9T and KRAS32R G25T, are the two structures we determined. Before using mutated sequences, we tried to isolate one conformation in the wild type sequence by influencing the equilibrium. We tried to change the experiment conditions (temperature, pH or ionic strength). In addition, we performed ligand and protein binding interactions to see if we could induce the formation of a preferred conformation. Unfortunately, I was not able to shift the equilibrium far enough by playing with the experimental conditions. A future work to be done to better understand the exchange between conformers of 32R and characterize a possible intermediate could be to perform temperature-induced folding and unfolding events using 2D ^{1}H-^{1}H NOESY type spectra. By looking at the appearance and disappearance of peaks, we could in principle decipher the transition between both states. But those experiments would require to analyze the extremely complex 32R WT in addition to G9T and G25T at different conditions. It would require a steady access to a high field an instrument > 900 MHz. Another possible issue has also to be addressed. In our model, we proposed the two major structures that also coexist with minor confirmations. I identified a sliding in the last tract of guanines in G25T conformation that could lead to the existence of one of these minor conformations. The implication of guanines in minor confirmations is not clear as we cannot observe any other base involved in KRAS32R G-quadruplexes except the ones we already observed and characterized in the G4 core and triad. This observation means that the base exchange leading to minor conformations could be too slow to be observed in our experiments. Thanks to these new insights we have now structural information that could be used to study interaction of both ligands and transcription factors such as the truncated version of hnRNP A1 (UP1). We also studied the structure of KRAS22RT G-quadruplex which corresponds to the major scaffold adopted by KRAS21R. This sequence is a part of KRAS32R and has been shown, as 32R, to have an effect on transcription when it is deleted and it is able to interact with hnRNP A1. In a previous work, KRAS22RT structure has been determined at 20°C and we then studied this structure at 37°C. In addition to the help in the more heavily studies about KRAS32R. The shorter G4 structure of KRAS could be used as structural base to study the effect of ligands and UP1. In this sequence of studies, I also think that now it will be possible to further extend the KRAS sequence and try to understand the G-quadruplex structures that could be formed in longer sequences (5’ AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGGCAGCGAGCGCCG 3’) called KRAS44R, which is also known to form G-quadruplexes. When compared to KRAS32R,

the additional residues, contain five additional guanines but not in a tract meaning that they will not be implicated in the G-quadruplex formation, and five additional cytosines which could be implicated in the formation of a duplex segment. Thus duplex stem could further stabilize the G-quadruplex formation and influence the equilibrium between G9T and G25T species. Finally, another point that need to be discussed is related to the comparison between the structures I determined *in vitro* and how these results could be expanded *in vivo* or *ex vivo* conditions. A major difference is the staggering amount of possible partners and soft interactions that could emerge in complex media that can be found in cells. The G4 structure and G4 folding could be modulated by proteins such as transcription factors, histones, zinc-fingers, polymerase and so on. Despite the fact that they could have an effect on G-quadruplex formation and structure, their quantity will also lead to an increased viscosity in the cell which is far from our NMR conditions. One way to simulate this condition is to use polymer such as PEG or Ficoll to increase viscosity in our sample and looked at the effect on our G-quadruplex structure. So, one important aspect that must be investigated is the effect of high viscosity environment in the G4 equilibrium. A similar approach was studied by the group of Prof. N. Sugimoto (Konan University Japan). They studied the effect of (poly)ethylene glycols (PEG) with different lengths on G-quadruplex formation. And determined G4 structures in presence of different cosolutes[290]. They found that the G4 core remained unchanged in presence of PEG but they observed several changes especially in loops. Indeed, their results showed that PEGs are capable of interacting with G-quadruplexes mainly due to dehydration. Moreover. Heddi and colleagues determined the structure of a tetrameric G-quadruplex in presence of PEG[291]. They showed that this G4, which was known as polymorphic, could formed a monomorphic structure in crowed solution. It is therefore reasonable to consider that viscosity may affect our equilibria between G9T and G25T.

Structure determination of KRAS32R conformers, as well as KRAS22RT G-quadruplex at physiological temperature brought new structural features that could be used in interaction study with ligands or proteins. Indeed, we could specifically identify moieties implicated in interaction with both ligands and proteins, which could allow for example the design of better ligands or competitors that ultimately would interfere with G4 metabolism in the cell.

II. UP1 interaction with KRAS G-quadruplexes

In this work, we tried to assess the interaction that takes place between KRAS G-quadruplexes and the hnRNP A1 derivative UP1. As demonstrated by the group of L. Xodo (Univ. Udine – Italy), UP1 is capable of interacting and unfolding KRAS21R and 32R G4s. We decided to test the activity of UP1 on the KRAS G4 forming sequences KRAS22RT, 32R and both G9T and G25T conformers. We showed that UP1 could interact with each of the four sequences to form a molecular complex of different oligomerization states. Nevertheless, it is also capable of interacting with a mutated sequence which cannot form a G4 structure. These results were not so surprising considering that UP1 RRM domains are not very selective in terms of DNA binding sequence. In collaboration with Anne Bourdoncle and Samir Amrane, we performed ITC experiments and showed that UP1 had very similar binding affinities for all four G4 sequences and an unfolded control-sequence. We also found that two proteins, or four RRM domains, are involved in each G4-UP1 complex. In addition, HSQC experiments showed that the interaction with G9T and G25T G4s and isotopically labelled UP1 resulted in very similar results in terms of binding. We were not able to decipher exactly which nucleotides interact with UP1, but in terms of protein surface the results were very similar. Further experiments need to be performed with a non-forming G-quadruplex sequence such as the one we used in ITC. Concerning the unfolding role of UP1, our preliminary studies did not confirm the hypothesis according to which UP1 unfolds G4 structures. However, we only used one equivalent of UP1, and it seems that two proteins may be required to form G4-UP1 complex. Even with one equivalent of UP1 we saw that G-quadruplexes started to be partially unfolded which could support the unfolding role of UP1. However, these results have to be carefully taken because our sample may aggregate. These experiments need to be repeated with an increased amount of UP1, different incubation times and amounts of K^+ salt in order to decrease re-folding process of the G4 once unfolded by UP1. Although we had only proven that the UP1 is capable of interacting with KRAS G-quadruplexes, it is enough to start looking for ligands which can inhibit this interaction, which is one of the main objectives of my research work.

III. Ligands binding studies with KRAS G-quadruplexes

My research time was also focused on the screening of ligands with high affinity for the KRAS G4 sequences. We first focused on 22RT since we completed the structure calculations on the smaller G4 from KRAS and we then started our investigations on 32R WT, G9T and G25T

sequences. In the initial screens for ligands against 22RT, we tested a pool of compounds representing a large spectra of chemical families. Compounds were selected from FRET melting results and the best compounds in terms of ΔTm were subjected to different biophysical experiments in order to assess not only the G4 stabilization, but also the capacity to protect G4 from the complementary strand, the cytotoxicity and cellular localization. We also investigated peak dispersion in 1D and 2D NMR spectra to continue characterize the structure of a drug-G4 complex. Some biophysical methods were not well suited for ligand binding studies due to many reason: among those methods SPR and ITC were the most deceiving. Nevertheless, I identified candidates showing good results in several of the tests, which allow to proceed to structural studies. For example, C8, an acridine derivative, was able to stabilize both 22RT and 32R (with 40°C ΔTm at 2 equivalents for 22RT and 20°C ΔTm at 5 equivalents for G25T for example), was also good to protect the G4 against duplex hybridization, and was also capable of entering the nucleus and killing KRAS-driven addicted cancer cells. Unfortunately, it exhibited a high degree of cytotoxicity in non-malignant cells. TriPropil, from (Naphthalene DiImide) NDI, showed very good results in stabilization and cellular penetration (being localized in the cell nucleus), but not a good capacity to protect 22RT G4 from complementary strand hybridization. It was also found to be cytotoxic in normal cells. Regarding the NMR titrations results, both ligands showed a decent peak separation that would render them capable for further structural studies in order to determine the complex with 22RT G4. One of the best ligands that showed an excellent profile for NMR structural studies was a Ni-Salphen derivative (AG). This ligand only showed average results in stabilization and protection from complementary strand hybridization and it was not efficient in entering the nucleus, leading to poor cytotoxicity properties for both normal and cancer cells. However, we started to study this ligand to identify chemical moieties involved in the G4 interaction. Based on these results, we could modify the ligand to remove the useless parts for the interaction and include chemical moieties that would for example, improve passage through the membrane and localization to the nucleus. The same experiments need to be repeated with the longer KRAS sequences. This global strategy needs to be used for other ligands such as C8 and TriPropil to increase their selectivity against KRAS G-quadruplexes. This is the first point I want to discuss. Indeed, most of the actual studies are focused on inspecting ligands for their stabilization and cytotoxicity capacity, and modified them depending on the results obtained. These modifications comprise often the addition of aliphatic arms or the

modification of the metal cation but without structure-based strategy. Within all studies that have been performed with ligands against G-quadruplexes (KRAS or not), several good candidates have been identified. It is necessary to performed structural studies with the best identified candidates to improve their properties or design new derivatives to be close to be the "perfect" ligand. This approach "rational structure-based drug design" is well stablished in other segments of drug development such as in the case of kinases where several block buster drugs were design such as Imatinib or the antiviral Zanamivir. There are still too many studies performed so far looking for good ligands against G4s structures focused primarily in stabilization studies obtained from screening large libraries with a big "fishing net". The results have not been good at all and I assume that it is not the proper way to tackle the problem. I understand that this strategy is also the result of lack of proper numbers and quality of structures of G4s. Fortunately there is also some excellent research that uses ligands in original and relevant ways such as using a mixture of several molecules in order to maximize the effect against tumors. The main idea is to select several good candidates and improve their effects thanks to structural studies to maximize their potential. I think that in the future we will be find very good G4 ligands that will have the potential to be included in a "drug-cocktail" in order to improve the synergy between drugs and maximize the chance of tumor-growth inhibition.

IV. Global conclusion

We provided novel structural insights concerning the organization of G-quadruplexes in the NHE region of KRAS. I solved the structure of three G-quadruplexes by NMR methods, including a small and two large G4s that are in equilibrium under *in vi*tro conditions. The structure determination of 32R was considered as a huge challenge for years among our peers. We also used the previous structure of KRAS22RT at 20°C to translate this structure at physiological temperature also. Any of these three structures have the potential to be used for rational drug design. Now we have a target we could focus on to develop ligands with increased affinity for KRAS G4s. I believe this is a major step for studying the complexes between these G-quadruplexes and ligands that have been identified as good candidates in order to extract information concerning the implication of the different moieties, and improve ligand selectivity. We started to perform studies in order to find good ligands and obtain some structural information with a complex formed between KRAS22RT and ligands (AG, C8). The

next steps of this work will be to continue looking to the complexes formed with KRAS G-quadruplexes and good ligands at atomic level and improve the salphen and the acridine derivatives. For KRAS32, more ligands need to be tested especially with NMR titrations to identify good candidates for further NMR structural studies. This work can be used in drug design applied to KRAS G-quadruplexes and can help to solve the major societal problem represented by pancreatic, lung or colorectal cancers. These cancers are among the deadliest cancers in the world and this work could be an additional step to fight their cancer growth and tumor spreading.

Annexes

Article 1: High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation

JBC ARTICLE

High-resolution three-dimensional NMR structure of the *KRAS* proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation

Received for publication, February 17, 2017, and in revised form, March 15, 2017 Published, Papers in Press, March 22, 2017, DOI 10.1074/jbc.M117.781906

Abdelaziz Kerkour[‡], Julien Marquevielle[‡], Stefaniia Ivashchenko[‡], Liliya A. Yatsunyk[‡§], Jean-Louis Mergny[‡1], and Gilmar F. Salgado[‡2]

From the [‡]Université Bordeaux, INSERM, CNRS, ARNA laboratory, European Institute of Chemistry and Biology, U1212, UMR 5320, 2 Rue Robert Escarpit, 33000 Pessac, France and [§]Department of Chemistry and Biochemistry, Swarthmore College, Swarthmore, Pennsylvania 19081

Edited by Wolfgang Peti

Non-canonical base pairing within guanine-rich DNA and RNA sequences can produce G-quartets, whose stacking leads to the formation of a G-quadruplex (G4). G4s can coexist with canonical duplex DNA in the human genome and have been suggested to suppress gene transcription, and much attention has therefore focused on studying G4s in promotor regions of disease-related genes. For example, the human *KRAS* proto-oncogene contains a nuclease-hypersensitive element located upstream of the major transcription start site. The KRAS nuclease-hypersensitive element (NHE) region contains a G-rich element (22RT; 5′-AGGGCGGTGTGGGAATAGG-GAA-3′) and encompasses a Myc-associated zinc finger-binding site that regulates *KRAS* transcription. The NEH region therefore has been proposed as a target for new drugs that control *KRAS* transcription, which requires detailed knowledge of the NHE structure. In this study, we report a high-resolution NMR structure of the G-rich element within the KRAS NHE. We found that the G-rich element forms a parallel structure with three G-quartets connected by a four-nucleotide loop and two short one-nucleotide double-chain reversal loops. In addition, a thymine bulge is found between G8 and G9. The loops of different lengths and the presence of a bulge between the G-quartets are structural elements that potentially can be targeted by small chemical ligands that would further stabilize the structure and interfere or block transcriptional regulators such as Myc-associated zinc finger from accessing their binding sites on the KRAS promoter. In conclusion, our work suggests a possible new route for the development of anticancer agents that could suppress KRAS expression.

Non-canonical base pairing within guanine-rich DNA and RNA sequences can produce G-quartets stabilized via eight hydrogen bonds involving both the "Watson-Crick" and "Hoogsteen" edges of each guanine. Stacking of the planar G-quartets (also called a G-tetrad) leads to the formation of a G-quadruplex (G4).[3] G4s are maintained by the presence of cations such as K^+ and to a lesser degree Na^+ and NH_4^+. The stacked G-quartets constitute the nearly invariant core of all G4 structures (1–3). This core is stabilized by cooperation between three key factors: hydrogen-bonding dipole interactions, metal coordination, and π-π stacking. The orientations of the loop regions within G4 structures are tightly related to strand directionality and give rise to the heterogeneity of G4 structures (4–6). G4 structures interfere with replication (7–9), transcription (10–13), and recombination (14–16). Bioinformatics analyses have provided evidence that sequences with potential to adopt G4 are not randomly localized within genomes but are specifically enriched in particular regions such as telomeres and promoters of genes (17–20). Proto-oncogenes are particularly enriched with G4 motifs, whereas tumor suppressor genes are not (21, 22). The formation of intramolecular G4s has been studied *in vitro* for motifs found in different human promoters regions, including c-*myc* (23–25), c-*kit* (26), and *bcl2* (27). G-rich elements found in other proto-oncogenes such as *KRAS* have received less attention. *KRAS* is located on chromosome 12 and encodes a GTP/GDP-binding protein. Previous studies showed that mutant alleles of *KRAS* are prevalent in pancreatic, biliary tract, colorectal, and lung carcinomas (28–33). Mutations in the *KRAS* promoter are found in about 30% of these cases. It is thought that the *KRAS* oncogene promotes glycolysis through the activation of downstream signaling pathways (34) to sustain the energy requirements for uncontrolled cellular proliferation, thus contributing to survival of cancer cells. The very high affinity of the RAS GTP/GDP-binding site (picomolar range) (35, 36) has made it difficult to synthesize molecules that effectively compete with GTP at millimolar range inside cells to block KRAS activity (37). It is no surprise that after many

This work was supported in part by Fondation Association pour la Recherche sur le Cancer (ARC), Ligue Régionale contre le Cancer-Gironde, and Agence National de Recherche grant "Oligoswitch," and an Aquitaine-Midi Pyrénées inter-regional grant. The authors declare that they have no conflicts of interest with the contents of this article.

This article contains supplemental Figs. S1–S5 and Tables S1–S3.

The atomic coordinates and structure factors (code 5I2V) have been deposited in the Protein Data Bank (http://wwpdb.org/).

[1] Recipient of a Chaire d'accueil from the Région Aquitaine.

[2] Benefits from a Chaire Mixte INSERM-Université Bordeaux position. To whom correspondence should be addressed. Tel.: 33-540002224; Fax: 33-540003004; E-mail: g.salgado@iecb.u-bordeaux.fr.

[3] The abbreviations used are: G4, G-quadruplex; NHE, nuclease-hypersensitive element; MAZ, Myc-associated zinc finger; DMS, dimethyl sulfate; DOSY, diffusion-ordered spectroscopy; r.m.s.d., root mean square deviation; dABz, N^6-benzoyl-2′-deoxyadenosine; dCAc, N^4-acetyl-2′-deoxycytidine; dGiBu, N^2-isobutyryl-2′-deoxyguanosine; HSQC, heteronuclear single quantum coherence; TOCSY, total correlation spectroscopy; HMBC, heteronuclear multiple bond correlation.

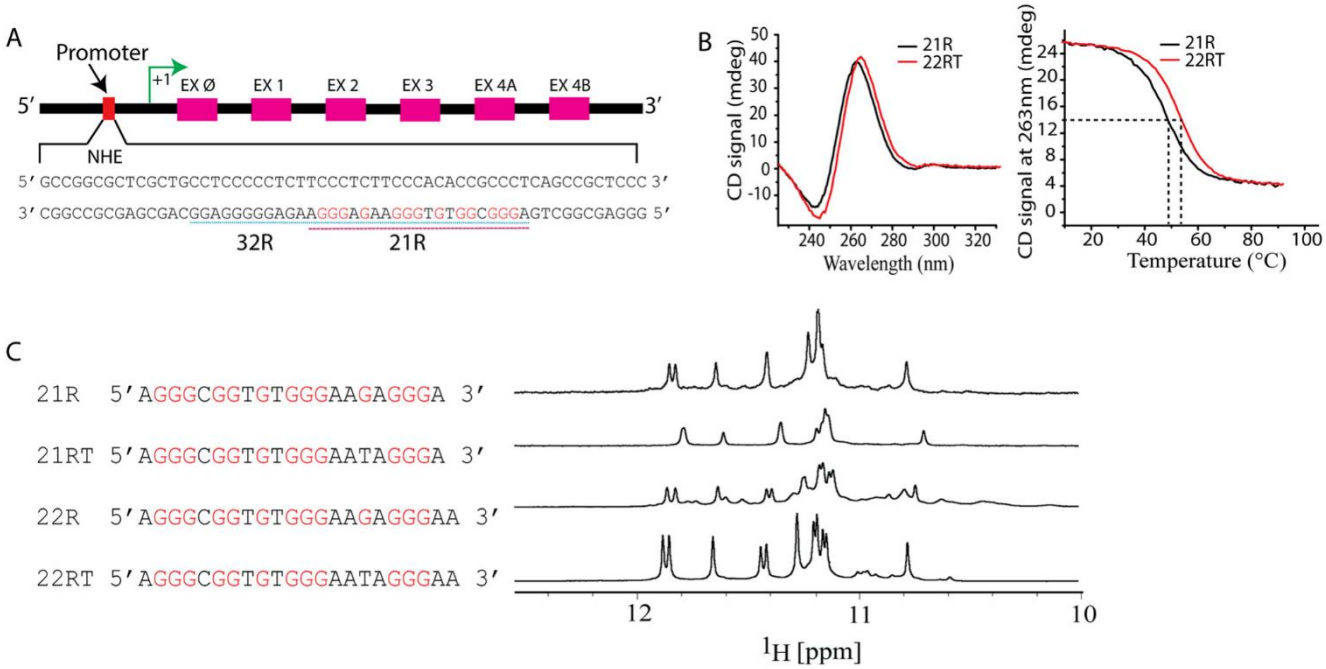


Figure 1. *A*, representation of human *KRAS* gene showing sequences of the NHE (13) and 32R and 21R(46). *EX*, exon. *B*, CD spectra of 21R and 22RT *KRAS* sequences at 5 μM (*left*) and CD melting profiles at 3 μM. *C*, sequences of *KRAS* NHE fragments (*left*) and corresponding imino region 1D NMR spectra obtained at 1 mM strand concentration at 20 °C in a pH 6.5 buffer containing 20 mM potassium phosphate supplemented with 70 mM KCl. *mdeg*, millidegrees.

decades of unsuccessfully battling against the RAS proteins new strategies in the field of drug design have emerged. Among those, some target alternative binding sites in the GTPase domain (38), others target the altered metabolic pathways (39), whereas some target the mRNA with antisense oligonucleotides (40). Alternatively, several strategies, including ours, target the promotor region in an effort to block *KRAS* expression (41). The *KRAS* promoter contains a polypurine nuclease-hypersensitive element (42) that plays an essential role in transcription. Its deletion results in a significant down-regulation of *KRAS* transcription (12, 43, 44). The promoter region of the *KRAS* gene comprises more than 500 bp and is susceptible to digestion by nucleases such as DNase I, micrococcal nuclease, and other endogenous nucleases (45). A particular sequence inside the NHE, between positions −327 and −296 nucleotides upstream of the main transcription initiation site, is particularly rich in guanines (sequence 32R; Fig. 1*A*). 32R contains six guanine stretches and is able to form several G4 conformations. Among those stretches of guanines, two regions overlap and includes the Myc-associated zinc finger (MAZ)-binding sites, the recognition sequence for a transcription factor that recognizes GGGCGG and GGGAGG sequences (46). Recent biophysical studies using circular dichroism and DMS footprinting (43, 47) suggest that oligonucleotides corresponding to 32R are able to adopt different intramolecular G-quadruplex topologies depending on which G-runs are included. Some of the topologies were tested as decoys for sequestration of MAZ (48). In addition, certain G4 structures in this region are stabilized by G-quadruplex-interacting ligands (12, 13, 43, 49) such as guanidine-modified phthalocyanines that interfere with *KRAS* transcription by competing with MAZ and poly(ADP-ribose) polymerase 1 proteins. Here we studied the conformations adopted by different stretches within the 32R sequence using NMR spectroscopy. The G4 conformation revealed by our studies provides a model that could potentially be used for *in silico* drug screening for ligands that stabilize the G4 structure in the *KRAS* promoter. The approach targeting unusual motifs present in genomic DNA is actively being pursued and can be seen as a new alternative strategy with promising results (1, 28, 29, 41).

Results and discussion

We began our study with circular dichroism (CD) and NMR analyses of oligonucleotides within the sequence NHE (Fig. 1*A*). To make spectral assignments possible, different oligonucleotides were evaluated (supplemental Table S2) with the objective of identifying a sequence that formed a single G4 conformer based on the dispersion and intensities of imino peaks observed in the NMR spectra. The sequence 21R and three other related sequences display a 1D imino peak pattern that corresponds to a single conformer as shown by 1D ^{1}H NMR spectroscopy (Fig. 1*C*). The similar imino signatures suggest the presence of a predominant conformer within the human NHE of *KRAS* gene. The sequence 22RT with a G16 to T16 mutation displayed a better resolved imino peak pattern and a slightly better stabilization observed by the CD melting studies (Fig. 1*B*). In addition, DMS footprinting (50) and our ^{15}N-filtered 1D NMR experiments (results not shown) demonstrated that G16 did not participate in the tetrad formation. Oligonucleotides of the native sequence with four G-tracts, 21R and 22R, and those with single G16 to T (16G→T) mutations, 21RT and 22RT, respectively, appeared to adopt a predominant conformation based on analysis of the imino proton region from 10 to 12 ppm. The sequence and the respective ^{1}H 1D NMR spectra are presented in Fig. 1*C*. Remarkably, 22RT showed a better resolved peak pattern in both imino and aromatic regions (not shown) than

did 21R or 22R. For stability purposes, we have included an additional A at the 22RT 3′-end. These modifications resulted in better imino and aromatic peak resolution when compared with 22R and 21R. The CD spectral signatures are similar between 21R and 22RT. The spectrum of each includes a positive band at 263 nm and a negative band at 243 nm suggestive of a parallel G4 fold (Fig. 1*B*, *left*). The thermal stability of 21R and 22RT was determined through CD melting experiments. The melting temperatures (T_m) in 90 mM K^+ were 49.2 ± 0.2 and 51.8 ± 0.3 °C for 21R and 22RT, respectively (Fig. 1*B*, *right*). The molecularity of 21R and 22RT was assessed by inspecting the UV-visible melting curves and by diffusion NMR experiments (supplemental Figs. S2 and S3, respectively). The results demonstrate that both 21R and 22RT fold into a monomeric structure, inferred from the reversible and superimposable cooling *versus* heating UV-visible curves in the concentration range that spans an order of magnitude from ≈5 to ≈50 μM. These experiments showed that melting transitions are reversible and independent of DNA concentration, demonstrating that the G4 structures formed by both 21R and 22RT are unimolecular. As the melting process was reversible, model-dependent van't Hoff enthalpies of folding could be calculated. The ΔH^0_{37} values for 21R and 22RT were 164 ± 4 and 191 ± 7 kJ/mol, respectively. Diffusion NMR spectroscopy was used to determine the diffusion coefficient value for 22RT. From diffusion-ordered spectroscopy (DOSY) experiments, we obtained a diffusion coefficient of 1.54×10^{-10} m^2 s^{-1} ($\log D \approx -9.81$), which is in the range of those found for monomeric G4 oligonucleotides of similar size such as the human telomeric sequence (22AG) observed elsewhere (23, 51, 52). The results support a model where 22RT is monomeric under the experimental conditions probed in this work. For reference, we report a diffusion value of 8.9×10^{-11} m^2 s^{-1} ($\log D \approx -10.05$) obtained for the oligonucleotide KRAS 44R from the NHE region that contains the sequence 32R at the 5′-end (supplemental Table S2). The imino proton spectrum of 22RT is characterized by 10 individually well resolved and sharp peaks in the 10.5-12-ppm region (Fig. 1*C*) plus one additional broad peak that was later identified as G19 and G7 overlapped imino peaks. The imino pattern in this region is often used as a fingerprint for G4 structures. The pattern observed suggests the formation of three G-quartets, each involving four imino protons. Based on the specific intraquartet characteristic guanine H1–H8 NOE correlations (Fig. 2*B*), the folding pattern of the 22RT G4 involves three G-quartets: G2·G6·G11·G18, G3·G7·G12·G19, and G4·G9·G13·G20 (Fig. 2). For clarity, we have selected the six lowest-energy structures after refinement of the best 20 structures with a heavy atom r.m.s.d. value of ≈1.5 Å (Fig. 3). When depicted against the calculated mass-weighting principal axis, the tetrad core is placed with its averaged planes almost in a perpendicular fashion (Fig. 3). The G3·G7·G12·G19 residues form the central tetrad of the quadruplex core as the imino protons on these residues are better protected from water/deuterium exchange than are those of other guanines of 22RT (supplemental Fig. S4). In addition and as expected, the guanines in a central G-quartet have by far the strongest NOE inter-residue connectivities between exchangeable protons (*e.g.* NH2/H1), supporting the increased protection of these protons from exchange with water. Interestingly, G3 is the only base that has NOE cross-peaks to both amino-exchangeable protons (NH21/NH22). These protons are probably protected by the single-nucleotide chain reversal C5 loop, which bridges and completely blocks the groove between G3 and G7. The 22RT G4 has an ~30° helical twist on average and a rise of 3.4 Å for each G-tetrad step. On average, the four grooves are of medium size with similar widths in the range of 12 ± 2 Å as defined by the distances between phosphates of opposing guanines in the structure. The orientations of the aromatic bases toward the sugars are determined by the conformation of the glycosidic bond angle, which is determined by the intraresidue NOE correlation intensities between the H8 aromatic proton and H1′ sugar proton. All the guanine glycosidic torsion angles are in the *anti* conformation as reflected by the medium/low intraguanine NOE cross-peaks observed between H8 and H1′ protons (Fig. 2*A*). These glycosidic torsion angle conformations are expected for a parallel G4 as suggested by the CD spectra of 22RT (Fig. 1*B*). Our CD and NMR results are consistent and indicate that 22RT adopts a parallel G4 as shown schematically in Fig. 2*C*. The three G-tetrads are connected by four linkers: two single-residue loops (C5 and T10), a bulge (T8), and one four-nucleotide loop (A14, A15, T16, and A17). C5 and T10 each form double-chain reversal loops that allow these single residues to bridge three G-tetrad blocks. Inspection of r.m.s.d. values and conformation diversity for C5 indicates hingelike motions parallel to the mass-weighting principal axis. The T10 base is oriented toward the 3′-end of the oligonucleotides, and fewer distinct conformers are observed. Interestingly, T8 forms a bulge projected out of the G-tetrad core. Although over 700,000 G-quadruplexes with single or multiple bulges may exist in the human genome (53, 54), only a few G-quadruplex structures have been deposited in the Protein Data Bank, and besides our model, only another deposited structure (Protein Data Bank code 2M4P) has a bulge between G-tetrads. This unusual structural feature within the 3D fold may be attractive to design ligands specific for *KRAS* G4 *in silico*. The fourth linker is composed of A14, A15, T16, and A17 and forms a medium-size propeller loop that crosses all three tetrads. The size of the loop allows several water molecules to fit between the G4 core surface and the loop residues. At both oligonucleotide extremities, two adenine residues, A1 and A22, cap the 5′- and 3′-ends of 22RT, respectively (Fig. 3). A1 interacts with A17, and both are tilted inward, capping the G4 core surface at the 5′-end. At the 3′-end, A21 and A22 interact through π-π stacking and are tilted toward the tetrad surface, which is much more exposed to the solvent than the opposite end (Fig. 3*b*). In most of the lowest energy conformers, A21 partially blocks one of the grooves. T10 shows a profile slightly different from T8 and T16. T10 methyl protons do not make cross-correlations with any other proton except with T10 H2′/H2″, and the cross-peak with its own H1′ is very weak, indicating free rotational motion around the C1′ and N1 of the pyrimidine base without any appreciable out-of-axis torsion of the base. For T8, we observe low-intensity NOEs, which indicate that the T8 methyl slightly interacts with G4 and G9; there is a low-intensity cross-correlation between the T8 methyl and the G9 H8 and a very weak correlation from the methyl of T8 to G4 H1, indicating that the

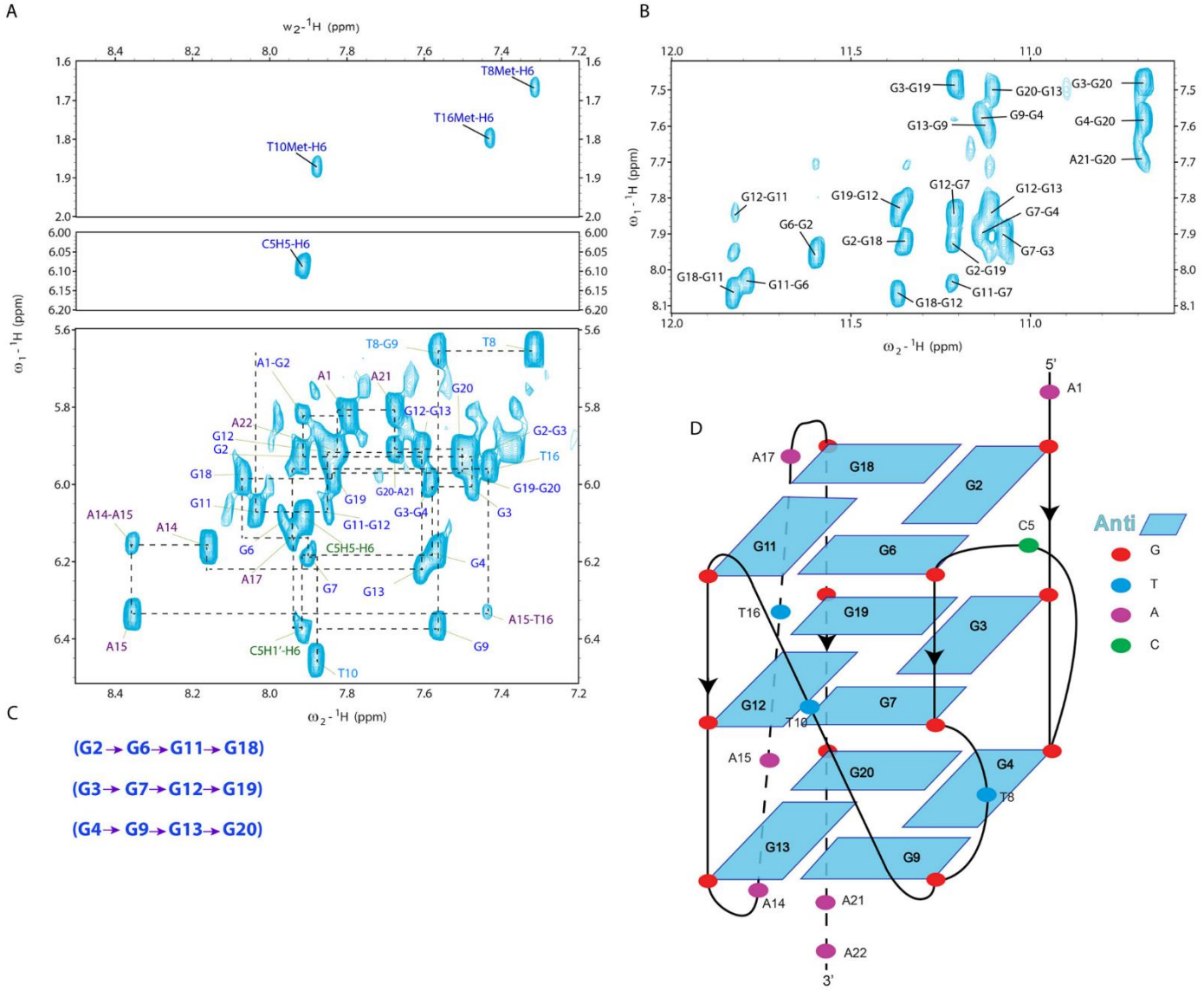


Figure 2. *A*, 2D ^{1}H NOESY spectra at 300-ms mixing time in $^{1}H_2O$ showing H6 to thymine methyl (*top*) and H5-H6 cytosine (*middle*) correlations from 2D ^{1}H TOCSY spectrum and H8/H6-H1′ sequential walk assignments (*bottom*). *B*, H1/H8 correlations from 2D ^{1}H NOESY spectrum at 300-ms mixing time. *C*, sequential imino-H8 intertetrad NOE connectivities observed by NMR. *D*, schematic representation of the G-quartets of 22RT. Guanines are numbered based on position in the 22RT sequence. All experiments were performed at 20 °C in a buffer containing 20 mM potassium phosphate, pH 6.5, supplemented with 70 mM KCl and ~2 mM DNA. Characteristic guanine imino-H8 NOE connectivities were observed for the following tetrads: G2·G6·G11·G18, G3·G7·G12·G19, and G4·G9·G13·G20.

methyl may be positioned in the edge of the top tetrad. Finally, both sugar CH2 protons of T16 have strong cross-peaks with the A17 H6 and H8 protons, not observed in any other of the two thymines. Overall, we observe a more restricted mobility of T16 compared with the other two thymines. Fig. 3*a* shows the ensemble of structures chosen by the lowest total energy criterion. Structures were refined in water and deposited under Protein Data Bank code 5I2V. Of the 11 structures deposited, only six are shown in Fig. 3 for clarity. Two K^+ counterions are expected per conformer coordinated between three G-quartet planes. In the ribbon diagrams shown in Fig. 3*b*, the ribbon thickness is proportional to the all-atom r.m.s.d. A1 and T8 have the highest r.m.s.d. values (≈3.5 Å) of all nucleotides in the molecule. This was expected as these two residues have fewer inter-residue correlations in the NOESY spectra. The propeller-type loops are also characterized by above average r.m.s.d. fluctuations. G-quadruplexes are highly polymorphic, and sequences with G4-forming potential are suggested to be formed in different key genomic regions, mainly in telomeres and gene promoters. Recently, high-resolution models of several different folding topologies have been reported, and some were highlighted in the aim to develop interesting pharmacological compounds that could be exploited as new anticancer drugs (47, 55–64). The G4 structure presented here is in agreement with the topology calculated for 21R (50) using DMS footprinting and may serve as a template to target and design new drugs that may diminish or inhibit *KRAS* expression. Interestingly, our structure contains a four-residue loop that covers one of the four grooves. It is well documented that the stability of G4 structures somewhat decreases as bulge and loop sizes increase (65, 66). We observed a more important and expected structural heterogeneity in the region of the four-nucleotide loop

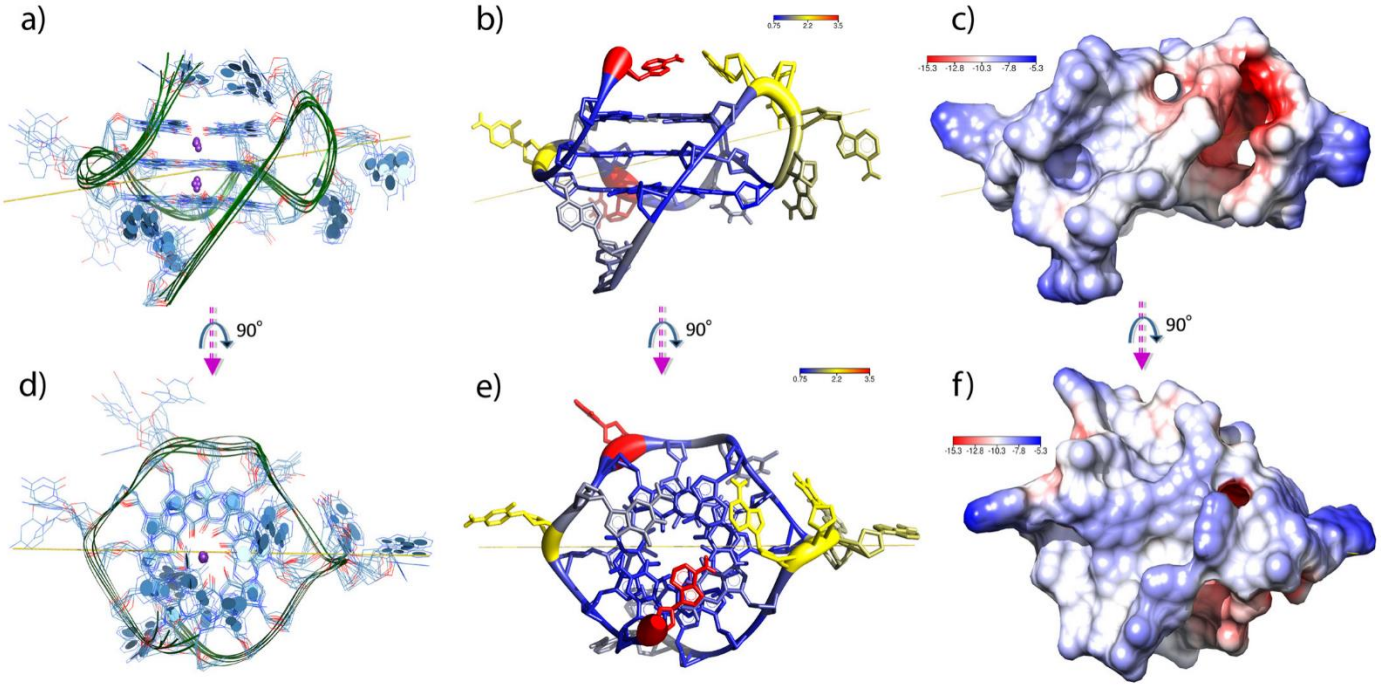


Figure 3. *a*, depiction of the ensemble of the six lowest-total-energy refined structures of 22RT. All guanines are in the *anti* conformation. K^+ counterions are depicted in *purple*, and the *yellow line* represents the principal axis of the six averaged all-atom mass-weighting principal inertial axes. On average, each conformer structure principal axis spans nearly 26 × 44 Å. The average groove width is 12 ± 2 Å. In the depiction on the *right*, the 5′- and 3′-ends of the oligonucleotide are at the *top* and *bottom* of the image, respectively. *b*, the average structure calculated from the six lowest-total-energy conformers. The *ribbon thickness* is proportional to the all-atom r.m.s.d. numerically represented by the color (see *key* for code in Å). A1 and T8 in *red* have the largest r.m.s.d. values (≈3.5 Å) of all residues. *c*, electrostatic surface of 22RT calculated using the Adaptive Poisson-Boltzmann Solver (APBS). The map was calculated using multiple Debye-Hückel boundary conditions and a nonlinear Poisson-Boltzmann equation with 40 points as the surface density at 293 K. The scale is reported in dimensionless units. The electronegative tunnel/cavity (*red*) represents a unique feature not usually found in canonical nucleic acid structures. In the *bottom panels* (*d–f*), the structure has been rotated (90°) so that the view is down onto the G-quartet plane that contains the 5′-most G.

with a significant degree of plasticity as indicated by the low number of inter-residue NOEs from the loop nucleotides. The four-base loop blocks the groove but creates an additional narrow backbone-backbone interface in the region where A17 links to G18 (Fig. 3, *c* and *f*, *red* colored surface, and supplemental Fig. S5). The conformation seems unique to this G4 fold and could be an interesting site to target small ligands. In our view, this medium-size loop presents in terms of drug design another interesting feature, *i.e.* together with the G4 core surface it makes a sort of tunnel or cavity with negative electrostatic potential values significantly lower than those found on the rest of the surface (*blue*). The cavity could be targeted by ligands with non-bulky side arms usually hanging out from the aromatic core that conventionally targets the more exposed tetrad. This cavity/tunnel is similar to the one found in the structure of a G4 adopted by a G-rich region of c-*kit2* (Protein Data Bank code 2KQG) (67). As opposed to KRAS 22RT, c-*kit2* G4 coexists in two parallel-stranded propeller-type folds that are similar but dynamically distinct substates.

Conclusion

In summary, G-quadruplex DNA structural motifs can behave as transcription repressors (12, 68), and *KRAS* is an important target against many forms of cancer that so far have very poor responses to standard therapies (38). We have determined a 3D high-resolution NMR structure of a stable G-quadruplex from a sequence located in the nuclease-hypersensitive element of the promoter region of the *KRAS* proto-oncogene. The structure may be one among others potentially available in the promotor region and may offer new, interesting possibilities for selective recognition by small ligands (69–73). In that sense, developing inhibitors of KRAS that target G-quadruplex motifs may be an interesting alternative route in the fight against certain types of cancers, as important as others currently being pursued against the intricate signaling network involved in oncogenic KRAS activation (74). Current existing strategies involve targeting post-translational modifications to prevent membrane association (75), small GTPase-targeting peptides (76), SOS-mediated nucleotide exchange (77), and inhibitors that allosterically control GTP affinity (38).

Materials and methods

DNA oligonucleotides

The unlabeled DNA oligonucleotides used in this work were purchased from both Eurogentec (Belgium) and Integrated DNA Technologies. They were synthesized on a 200-nmol or 1-μmol scale and then purified by reverse-phase HPLC. The sequences were supplied lyophilized. The 5% ^{15}N,^{13}C site-specifically labeled 22RT used in this study was synthesized in our laboratory (INSERM U1212, Bordeaux, France) on an automated Expedite 8909 DNA synthesizer at a 1-μmol scale on a 1000-Å primer support (Link Technologies SynBase CPG). All the standard phosphoramidites (dABz, dT, dGiBu, and dCAc), reagents, and solvents used during the synthesis were purchased from Glen Research. The dGiBu phosphoramidite

(U-$^{13}C_{10}$, 98%; U-$^{15}N_5$, 98%; CP, 95%) was purchased from Cambridge Isotope Laboratories. After the synthesis, the oligonucleotides were cleaved from the support, and the nucleobases were deprotected with ammonium hydroxide at 55 °C for 16 h and then lyophilized. All the sequences used for NMR were prepared in potassium NMR buffer (20 mM K_2HPO_4/KH_2PO_4, 70 mM KCl, pH 6.5, 10% D_2O). For CD and UV-visible studies, samples were prepared in KP_i buffer (20 mM K_2HPO_4/KH_2PO_4, 70 mM KCl, pH 6.5). After dissolving in buffers, the oligonucleotides were heated for 5 min at 95 °C and chilled on ice several times. After the last annealing cycle, they were refrigerated for 24 h before use. Supplemental Table S2 lists some characteristics of the sequences used in this work.

CD spectroscopy

CD experiments were performed on a Jasco J-815 spectrometer using Spectra Manager software. Each DNA sample was prepared at 3–5 μM in KP_i buffer and annealed at 90 °C for 5 min, cooled slowly by turning off the heating block (about 3–5 h), and then incubated overnight at 4 °C. The CD spectra were measured in the region between 220 and 330 nm using a scan speed of 50–100 nm/min and a response time of 1 s. Three scans were collected and averaged. Data were processed as described elsewhere (78). CD melting studies were performed on ≈3 μM DNA samples using either a full-wavelength or a single-wavelength mode. In the former case, the data were collected in the wavelength range 330–220 nm with 0.5-s averaging time, 2-nm bandwidth, 100-nm/min scan speed, two accumulations, and 0.2-nm step. The temperature was raised from 4 to 90 °C with 1 °C intervals and a 0.4 °C/min rate. These parameters lead to the overall acquisition time of 1 h/10 °C temperature change. All data collected in this manner were examined for the presence of possible intermediates during the melting process. However, the overall shape of the CD signature remained unchanged for every sample examined. Thus, after completion of the full-wavelength scan for each sample, additional melting data were collected, monitoring at 264 (characteristic for a parallel G4) and 330 nm (used as a reference to factor out instrument fluctuations) with an averaging time of 32 s and a bandwidth of 2 nm. The temperature was raised from 4 to 95 °C and then cooled to 4 °C at a rate of 1 °C/min. This set of experiments allowed us to test the reversibility of the melting process and to obtained thermodynamic parameters. Melting data were processed as described in a previous report (78).

NMR spectroscopy

NMR spectra were recorded on Bruker Avance 700- and 800-MHz instruments equipped with cryogenically cooled probes. Experiments were performed at 20 °C. For solution NMR, standard 3- or 5-mm NMR tubes were used. The samples were prepared in potassium NMR buffer. The concentrations of oligonucleotide were between 1 and 5 mM depending on the experiment requirements. Most of the ^1H 1D spectra were recorded using the 1-1 echo pulse sequence (79) (based on the use of "double pulsed field gradient spin-echo"), which selectively removes resonance due to water without affecting other resonances, including those that are in fast exchange with water. The gradient pulse was a smoothed square shape (SMSQ10).100. Resonance assignments were made using 5% ^{15}N,^{13}C site-specific low-enrichment labeling for imino protons and ^1H-^{13}C HSQC for aromatic H8 protons. To correlate imino and H8 protons of the same guanine via the $^{13}C5$ carbon at natural abundance through-bond correlations, the ^1H-^{13}C HMBC experiment were performed at natural abundance. Thymines were unambiguously identified using the following mutants: 22RT T8 to C8 and 22RT T10 to C10 (supplemental Table S2). The remaining resonances were identified using TOCSY and COSY experiments and were independently verified using NOESY experiments.

DOSY experiment

A reference 1D ^1H spectrum was recorded before the DOSY experiment. The pulse program stebpgp1s191d was used for the 1D DOSY spectra, and stebpgp1sp9pr was used for the 2D DOSY spectra. The sequences used simulated echo with a bipolar gradient pulse pair and one spoiled gradient, and 3-9-19 WATERGATE (80) solvent suppression was also applied. Sixty-four scans were recorded for the 1D DOSY experiment, and 1024 were recorded for 2D DOSY. A relaxation delay of 2 ms was applied, and a 20-μs delay was used for water suppression. The time domain was fixed to 8000 points for the F2 dimension and 32 points for the F1 dimension. The diffusion time (Δ) was 150 ms, the gradient length (δ) was 1 ms, and the recovery delay after gradient was fixed to 200 μs. The gradient strengths applied were set between 5 and 95%, and the gradient strength change was set to linear. The data processing was performed using Bruker-designed DOSY software. The following equation was applied to fit the curve of diffusion,

$$I = I_0 e^{-D\gamma^2 g^2 \delta^2 (\Delta - \delta/3 - \tau/2)} \quad \text{(Eq. 1)}$$

where I is the observed intensity, I_0 is the reference intensity (unattenuated signal intensity), D is the diffusion coefficient, γ is the gyromagnetic ratio of the observed proton, g is the gradient strength, δ is the length of the gradient, and τ is the diffusion time. The diffusion coefficient D for a given molecule is described by the Stokes-Einstein equation,

$$D = \frac{kT}{6\pi\eta R_S} \quad \text{(Eq. 2)}$$

where k is the Boltzmann constant, T is the temperature, η is the viscosity of the liquid, and R_S is the (hydrodynamic) radius of the molecule.

UV-visible spectroscopy

Thermal difference spectra—Thermal difference spectra were obtained in KP_i buffer by collecting UV-visible wavelength scans from 220 to 350 nm at two temperatures, one well below (usually 4 °C) and another well above (usually 95 °C) the melting temperature of the DNA secondary structure. The difference spectra were obtained by subtracting the data at 4 °C from the data at 95 °C. Both 21R and 22RT showed thermal difference spectral signatures characteristic of G4 structures (81).

UV-visible melting of 21R and 22RT—Concentration dependences of the melting transitions of 21R (5′-AGGGCGGTGTGGGAAGAGGGA-3′) and 22RT (5′-AGGGCGGT-

GTGGGAATAGGGAA-3′) were determined in UV-visible melting experiments by monitoring the signals at 295 and 335 nm using an Uvikon XL spectrophotometer. The former wavelength is sensitive to the G4 folding state, and the latter wavelength was used as a reference to monitor instrument performance. The extinction coefficient of the DNA at 335 nm is negligible. Five separate samples with concentration ranging from 5.0 to 50 μM were annealed in KP_i buffer as described above and equilibrated at 4 °C overnight. Samples were placed in cuvettes with 1.0- or 0.2-cm path lengths depending on strand concentration. The temperature was measured with the temperature sensor inserted in the cuvette holder next to the DNA sample. The temperature was changed at a rate of 0.2 °C/min, and the averaging time was 0.3–0.5 s. Each experiment included two temperature ramps from 95 to 0.5 °C with a 15-min hold and from 0.5 back to 95 °C. The experiments were repeated twice. The value of signal at 335 nm was subtracted from each data set. The data suggest that the folding/unfolding of both 21R and 22RT is reversible as the melting and cooling data are nearly superimposable, consistent with our CD melting study. The melting curves were analyzed assuming a two-state model with temperature-independent enthalpy, ΔH^0 (82). Starting and final baselines were assumed to be linear, and melting temperature and enthalpy of unfolding were adjusted to get the best fit. Data were also analyzed assuming non-zero heat capacity. This analysis included an additional parameter but did not lead to significant improvement of the fit. Melting temperatures and ΔH^0 obtained from CD and UV-visible data are in good agreement with each other.

NMR structural calculations based on NOE distance restraints and simulated annealing

NOE-derived distance restraints were calculated from spectral densities obtained from different ^{1}H-^{1}H NOESY spectra at various mixing times (50, 200, 300, and 400 ms). In the final structure calculations, only data from the 300-ms mixing time was used. All NMR restraints were obtained from spectra collected at 293 K unless otherwise stated. The peak volumes were classified as weak (4.0–6.5 Å), medium (2.5–4.5 Å), and strong (1.8–3 Å). Planarity restraints (20 kcal/mol/Å^2) were introduced for the following tetrad architecture: G2·G6·G11·G18, G3·G7·G12·G19, and G4·G9·G13·G20. Hydrogen-bond and planarity restraints were defined between 1.9 and 2.1 Å and between 2.9 and 3.1 Å for the bonds established between H1 and O6 and between H21 and N7, respectively, and were only applied to the guanine bases involved in tetrad formation. The *anti* conformation was defined by the glycosidic torsion angles (χ) determined from the H1′-H8 intrabase distances. They were restrained to be in the range of $-130 \pm 40°$. Altogether, the hydrogen bonds and the artificial planarity restraints kept the G-quartets in their quasiplanar conformation during the first steps of the ARIA-CNS calculations; these restraints were removed during the refinement process. The integration of NOE volumes, calibration of distances with a relaxation matrix spin diffusion correction, and setting of lower and upper bounds were done by ARIA2.3/CNS1.2. Two distinct steps were used to calculate the final assembly of 20 structures. First, eight iterations of calculations were carried out using 200 structures per iteration in ARIA2.3/CNS1.2 (83, 84) with mixed Cartesian and torsion angle dynamics during the simulated annealing runs. The protocol contains four stages: (*a*) an initial high-temperature torsion angle simulated annealing of 50,000 steps at 10,000 K with 27 fs for each step, (*b*) a torsion angle dynamic cooling stage of 10,000 steps from 10,000 to 2000 K, (*c*) a Cartesian dynamics cooling stage of 10,000 steps from 2000 to 1000 K, and finally (*d*) a Cartesian dynamics cooling stage of 20,000 steps from 1000 to 50 K with 3 fs per step. For all bonds, angles, and improper dihedral energy terms of the force field, the standard CNS dna-rna-allatom topology and parameter files were used with uniform energy constants. For distances and hydrogen bonds, 10 kcal mol^{-1}Å^{-2} was applied during the initial stage of dynamics and was increased up to 50 kcal mol^{-1}Å^{-2} for the remaining steps of the dynamics. For dihedral restraints, energy constants applied were 5, 25, 200, and 200 kcal mol^{-1}Å^{-2} for the phases a, b, c, and d, respectively. An energy constant of 25 kcal mol^{-1}Å^{-2} was applied for planarity restraints. Distance restraints together with G-tetrad hydrogen-bonding distance restraints, glycosidic angle restraints, and planarity restraints were used during this calculation step. In the second step, we performed the necessary refinement of the t20 best structures in explicit water molecules as solvent. For that purpose, the SANDER module of Amber 12 (University of California, San Francisco) was used. The calculation was performed with the AMBER force field FF12SB, which contains the AMBER force field for nucleic acids and Barcelona changes (85). Two K^+ ions were included between the G-tetrads, and 21 additional K^+ ions were included to counter the negative charge of the DNA. The 20 structures were solvated by a truncated octahedral box of TIP3P water molecules (86). The structures were energy-minimized using harmonic position restraints of 25 kcal/mol/Å^2. First, 1000 steps of minimization were carried out holding the system fixed and minimizing just the water box, including ions with 500 steps of steepest descent minimization followed by 500 steps of conjugate gradient minimization. Then, 2500 steps of minimization were performed for the entire system with 1000 steps of steepest descent minimization followed by 1500 steps of conjugate gradient minimization. Afterward, 20 ps of simulated annealing was acquired with heating from 0 to 300 K during 5 ps under a constant volume while maintaining the position restraints at 25 kcal/mol/Å^2. Finally, we performed a cooling step to 100 K during 13 s; during this step, the time constant for heat bath coupling was varied in the range of 0.05–0.5 ps. A final cooling stage was also performed with more rapid cooling (0.1–0.05 ps) to bring the system to 0 K. The weight of distance restraints was increased gradually during this simulated annealing from 0.1 to 1 in the first 3 seconds and was then kept at 1 for the rest of the annealing procedure.

NMR assignments

Using site-specific low (5%)-enrichment [^{13}C,^{15}N]guanine-labeled samples, the imino H1 and the aromatic H8 protons for each guanine were unambiguously assigned (supplemental Fig. S1*A*). The guanine H8 aromatic protons were assigned using classical ^{13}C-^{1}H HSQC spectral analysis for each guanine separately (supplemental Fig. S1*B*). The assignments were also

confirmed by natural abundance through-bond correlations using a jump and return (JR) HMBC experiment, which correlates guanine imino protons with H8 aromatic protons through ^{13}C5 (supplemental Fig. S1*C*). The complete spectral assignment was achieved by combining through-bond (TOCSY and COSY) and through-space (NOESY) experiments. An ^{1}H-^{1}H TOCSY experiment allowed unambiguous correlation of H5-H6 protons of the cytosine at position 5 and the H6 methyl protons of thymines at positions 8, 10, and 17 (Fig. 2*A*). The remaining proton assignments such as those of sugar protons (H3′, H4′, H2′/H2″, and H5′/H5″) were determined as described previously (5, 6).

Restraints used in structure calculations

NMR restraints used for the calculations are listed in supplemental Table S1.

Data deposition

Water-refined structures of KRAS 22RT G-quadruplex were deposited in the Protein Data Bank under Protein Data Bank code 5I2V.

Author contributions—A. K. prepared samples, acquired NMR data, performed analysis and structure determination, and helped prepare some figures. J. M. prepared oligonucleotides samples, acquired CD spectra, and helped prepare some figures. S. I. prepared samples for CD and UV-visible melting experiments and analyzed the CD spectra. L. A. Y. supervised and acquired data from CD and UV-visible experiments and participated in manuscript conception. J.-L. M. directed the research subject and participated in manuscript conception. G. F. S. acquired NMR spectra, analyzed and interpreted various data, performed atomic structure determination, directed the research subject, wrote the article, and prepared the figures.

Acknowledgments—This work benefited from the facilities and expertise of UMS3033/US001. Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged. We thank B. Vialet for synthesis of isotopically labeled oligonucleotides and the Olifans team for useful discussions.

References

1. Balasubramanian, S., and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chemical Biol.* **13,** 345–353
2. Düchler, M. (2012) G-quadruplexes: targets and tools in anticancer drug design. *J. Drug Target.* **20,** 389–400
3. Phan, A. T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **277,** 1107–1117
4. Wong, A., Ida, R., Spindler, L., and Wu, G. (2005) Disodium guanosine 5′-monophosphate self-associates into nanoscale cylinders at pH 8: a combined diffusion NMR spectroscopy and dynamic light scattering study. *J. Am. Chem. Soc.* **127,** 6990–6998
5. 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
6. Marusic, M., Sket, P., Bauer, L., Viglasky, V., and Plavec, J. (2012) Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Res.* **40,** 6946–6956
7. Madireddy, A., Purushothaman, P., Loosbroock, C. P., Robertson, E. S., Schildkraut, C. L., and Verma, S. C. (2016) G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. *Nucleic Acids Res.* **44,** 3675–3694
8. Valton, A. L., and Prioleau, M. N. (2016) G-quadruplexes in DNA replication: a problem or a necessity? *Trends Genet.* **32,** 697–706
9. Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M. P., Foiani, M., and Nicolas, A. (2011) G-quadruplex-induced instability during leading-strand replication. *EMBO J.* **30,** 4033–4046
10. Du, Z., Kong, P., Gao, Y., and Li, N. (2007) Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochem. Biophys. Res. Commun.* **354,** 1067–1070
11. Yan, J., Zhao, X., Liu, B., Yuan, Y., and Guan, Y. (2016) An intramolecular G-quadruplex structure formed in the human MET promoter region and its biological relevance. *Mol. Carcinog.* **55,** 897–909
12. 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
13. Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K. K., and Xodo, L. E. (2010) The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J. Biol. Chem.* **285,** 22003–22016
14. Stanton, A., Harris, L. M., Graham, G., and Merrick, C. J. (2016) Recombination events among virulence genes in malaria parasites are associated with G-quadruplex-forming DNA motifs. *BMC Genomics* **17,** 859
15. Piekna-Przybylska, D., Sullivan, M. A., Sharma, G., and Bambara, R. A. (2014) U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry* **53,** 2581–2593
16. De Nicola, B., Lech, C. J., Heddi, B., Regmi, S., Frasson, I., Perrone, R., Richter, S. N., and Phan, A. T. (2016) Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic Acids Res.* **44,** 6442–6451
17. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33,** 2908–2916
18. Wong, H. M., Stegle, O., Rodgers, S., and Huppert, J. L. (2010) A toolbox for predicting G-quadruplex formation and stability. *J. Nucleic Acids* **2010,** 564946
19. Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33,** 2901–2907
20. Stegle, O., Payet, L., Mergny, J. L., MacKay, D. J., and Leon, J. H. (2009) Predicting and understanding the stability of G-quadruplexes. *Bioinformatics* **25,** i374–382
21. Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35,** 406–413
22. Eddy, J., and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34,** 3887–3896
23. Ambrus, A., Chen, D., Dai, J., Jones, R. A., and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44,** 2048–2058
24. 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
25. Phan, A. T., Modi, Y. S., and Patel, D. J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **126,** 8710–8716
26. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.* **129,** 4386–4392
27. Dai, J., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* **128,** 1096–1098
28. Lavrado, J., Brito, H., Borralho, P. M., Ohnmacht, S. A., Kim, N. S., Leitão, C., Pisco, S., Gunaratnam, M., Rodrigues, C. M., Moreira, R., Neidle, S., and Paulo, A. (2015) KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines. *Sci. Rep.* **5,** 9696
29. Brito, H., Martins, A. C., Lavrado, J., Mendes, E., Francisco, A. P., Santos, S. A., Ohnmacht, S. A., Kim, N. S., Rodrigues, C. M., Moreira, R., Neidle, S., Borralho, P. M., and Paulo, A. (2015) Targeting KRAS oncogene in colon

cancer cells with 7-carboxylate indolo[3,2-b]quinoline tri-alkylamine derivatives. *PLoS One* **10,** e0126891
30. Forbes, S. A., Bindal, N., Bamford, S., Cole, C., Kok, C. Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., Teague, J. W., Campbell, P. J., Stratton, M. R., and Futreal, P. A. (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945–D950
31. Dent, P. (2013) Multi-kinase modulation for colon cancer therapy. *Cancer Biol. Ther.* **14,** 877–878
32. Krens, L. L., Baas, J. M., Gelderblom, H., and Guchelaar, H. J. (2010) Therapeutic modulation of k-ras signaling in colorectal cancer. *Drug Discov. Today* **15,** 502–516
33. Andreyev, H. J., Ross, P. J., Cunningham, D., and Clarke, P. A. (2001) Antisense treatment directed against mutated Ki-ras in human colorectal adenocarcinoma. *Gut* **48,** 230–237
34. Ying, H., Kimmelman, A. C., Lyssiotis, C. A., Hua, S., Chu, G. C., Fletcher-Sananikone, E., Locasale, J. W., Son, J., Zhang, H., Coloff, J. L., Yan, H., Wang, W., Chen, S., Viale, A., Zheng, H., *et al.* (2012) Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell* **149,** 656–670
35. John, J., Sohmen, R., Feuerstein, J., Linke, R., Wittinghofer, A., and Goody, R. S. (1990) Kinetics of interaction of nucleotides with nucleotide-free H-ras p21. *Biochemistry* **29,** 6058–6065
36. McCormick, F. (2016) K-Ras protein as a drug target. *J. Mol. Med.* **94,** 253–258
37. Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J., and Der, C. J. (2014) Drugging the undruggable RAS: mission possible? *Nat. Rev. Drug Discov.* **13,** 828–851
38. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A., and Shokat, K. M. (2013) K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503,** 548–551
39. Zhu, Z., Golay, H. G., and Barbie, D. A. (2014) Targeting pathways downstream of KRAS in lung adenocarcinoma. *Pharmacogenomics* **15,** 1507–1518
40. Wang, J. H., Newbury, L. J., Knisely, A. S., Monia, B., Hendry, B. M., and Sharpe, C. C. (2012) Antisense knockdown of Kras inhibits fibrosis in a rat model of unilateral ureteric obstruction. *Am. J. Pathol.* **180,** 82–90
41. Cogoi, S., and Xodo, L. E. (2016) G4 DNA in ras genes and its potential in cancer therapy. *Biochim. Biophys. Acta* **1859,** 663–674
42. Rokney, A., Shagan, M., Kessel, M., Smith, Y., Rosenshine, I., and Oppenheim, A. B. (2009) *E. coli* transports aggregated proteins to the poles by a specific and energy-dependent process. *J. Mol. Biol.* **392,** 589–601
43. Cogoi, S., Paramasivam, M., Spolaore, B., and Xodo, L. E. (2008) Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Res.* **36,** 3765–3780
44. Paramasivam, M., Membrino, A., Cogoi, S., Fukuda, H., Nakagama, H., and Xodo, L. E. (2009) Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription. *Nucleic Acids Res.* **37,** 2841–2853
45. Jordano, J., and Perucho, M. (1986) Chromatin structure of the promoter region of the human c-K-ras gene. *Nucleic Acids Res.* **14,** 7361–7378
46. Cogoi, S., Zorzet, S., Rapozzi, V., Géci, I., Pedersen, E. B., and Xodo, L. E. (2013) MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRAS in pancreatic cancer cells and delays tumor growth in mice. *Nucleic Acids Res.* **41,** 4049–4064
47. Cogoi, S., Paramasivam, M., Filichev, V., Géci, I., Pedersen, E. B., and Xodo, L. E. (2009) Identification of a new G-quadruplex motif in the KRAS promoter and design of pyrene-modified G4-decoys with antiproliferative activity in pancreatic cancer cells. *J. Med. Chem.* **52,** 564–568
48. Podbevšek, P., and Plavec, J. (2016) KRAS promoter oligonucleotide with decoy activity dimerizes into a unique topology consisting of two G-quadruplex units. *Nucleic Acids Res.* **44,** 917–925
49. Xodo, L., Paramasivam, M., Membrino, A., and Cogoi, S. (2008) Protein hnRNPA1 binds to a critical G-rich element of KRAS and unwinds G-quadruplex structures: implications in transcription. *Nucleic Acids Symp. Ser.* **52,** 159–160
50. Paramasivam, M., Cogoi, S., and Xodo, L. E. (2011) Primer extension reactions as a tool to uncover folding motifs within complex G-rich sequences: analysis of the human KRAS NHE. *Chem. Commun.* **47,** 4965–4967
51. Kerkour, A., Mergny, J. L., and Salgado, G. F. (2016) NMR based model of human telomeric repeat G-quadruplex in complex with 2,4,6-triarylpyridine family ligand. *Biochim. Biophys. Acta* 10.1016/j.bbagen.2016.12.016
52. Groves, P., and Webba da Silva, M. (2010) Rapid stoichiometric analysis of G-quadruplexes in solution. *Chemistry* **16,** 6451–6453
53. Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33,** 877–881
54. Bedrat, A., Lacroix, L., and Mergny, J. L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44,** 1746–1759
55. Merle, P., Gueugneau, M., Teulade-Fichou, M. P., Müller-Barthélémy, M., Amiard, S., Chautard, E., Guetta, C., Dedieu, V., Communal, Y., Mergny, J. L., Gallego, M., White, C., Verrelle, P., and Tchirkov, A. (2015) Highly efficient radiosensitization of human glioblastoma and lung cancer cells by a G-quadruplex DNA binding compound. *Sci. Rep.* **5,** 16255
56. Tomar, J. S. (2015) *In-silico* modeling studies of G-quadruplex with soy isoflavones having anticancerous activity. *J. Mol. Model.* **21,** 193
57. Zhou, J. K., Yang, D. Y., and Sheu, S. Y. (2015) The molecular mechanism of ligand unbinding from the human telomeric G-quadruplex by steered molecular dynamics and umbrella sampling simulations. *Phys. Chem. Chem. Phys.* **17,** 12857–12869
58. Lavrado, J., Ohnmacht, S. A., Correia, I., Leitão, C., Pisco, S., Gunaratnam, M., Moreira, R., Neidle, S., Santos, D. J., and Paulo, A. (2015) Indolo [3,2-c]quinoline G-quadruplex stabilizers: a structural analysis of binding to the human telomeric G-quadruplex. *ChemMedChem* **10,** 836–849
59. Ebrahimi, M., Khayamian, T., Hadadzadeh, H., Sayed Tabatabaei, B. E., Jannesari, Z., and Khaksar, G. (2015) Spectroscopic, biological, and molecular modeling studies on the interactions of [Fe(III)-meloxicam] with G-quadruplex DNA and investigation of its release from bovine serum albumin (BSA) nanoparticles. *J. Biomol. Struct. Dyn.* **33,** 2316–2329
60. Qin, X. Y., Liu, Y. N., Yu, Q. Q., Yang, L. C., Liu, Y., Zhou, Y. H., and Liu, J. (2014) Mixed-ligand mononuclear copper(II) complex: crystal structure and anticancer activity. *ChemMedChem* **9,** 1665–1671
61. Mancini, J., Rousseau, P., Castor, K. J., Sleiman, H. F., and Autexier, C. (2016) Platinum(II) phenanthroimidazole G-quadruplex ligand induces selective telomere shortening in A549 cancer cells. *Biochimie* **121,** 287–297
62. Shin, Y. J., Kumarasamy, V., Camacho, D., and Sun, D. (2015) Involvement of G-quadruplex structures in regulation of human RET gene expression by small molecules in human medullary thyroid carcinoma TT cells. *Oncogene* **34,** 1292–1299
63. Ali, A., Bansal, M., and Bhattacharya, S. (2015) Ligand 5,10,15,20-tetra(N-methyl-4-pyridyl)porphine (TMPyP4) prefers the parallel propeller-type human telomeric G-quadruplex DNA over its other polymorphs. *J. Phys. Chem. B* **119,** 5–14
64. Gupta, P., Rastede, E. E., and Appella, D. H. (2015) Multivalent LKγ-PNA oligomers bind to a human telomere DNA G-rich sequence to form quadruplexes. *Bioorg. Med. Chem. Lett.* **25,** 4757–4760
65. Guédin, A., Gros, J., Alberti, P., and Mergny, J. L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **38,** 7858–7868
66. Mukundan, V. T., and Phan, A. T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.* **135,** 5017–5028
67. Hsu, S. T., Varnai, P., Bugaut, A., Reszka, A. P., Neidle, S., and Balasubramanian, S. (2009) A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Am. Chem. Soc.* **131,** 13399–13409
68. David, A. P., Margarit, E., Domizi, P., Banchio, C., Armas, P., and Calcaterra, N. B. (2016) G-quadruplexes as novel cis-elements controlling transcription during embryonic development. *Nucleic Acids Res.* **44,** 4163–4173
69. Carvalho, J., Nottelet, P., Mergny, J. L., Queiroz, J. A., Salgado, G. F., and Cruz, C. (2017) Study of the interaction between indole-based compounds and biologically relevant G-quadruplexes. *Biochimie* **135,** 186–195

70. Zimmer, J., Tacconi, E. M., Folio, C., Badie, S., Porru, M., Klare, K., Tumiati, M., Markkanen, E., Halder, S., Ryan, A., Jackson, S. P., Ramadan, K., Kuznetsov, S. G., Biroccio, A., Sale, J. E., *et al.* (2016) Targeting BRCA1 and BRCA2 deficiencies with G-quadruplex-interacting compounds. *Mol. Cell* **61,** 449–460
71. 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.* **43,** 10376–10386
72. Di Leva, F. S., Novellino, E., Cavalli, A., Parrinello, M., and Limongelli, V. (2014) Mechanistic insight into ligand binding to G-quadruplex DNA. *Nucleic Acids Res.* **42,** 5447–5455
73. Laguerre, A., Stefan, L., Larrouy, M., Genest, D., Novotna, J., Pirrotta, M., and Monchaud, D. (2014) A twice-as-smart synthetic G-quartet: PyroTASQ is both a smart quadruplex ligand and a smart fluorescent probe. *J. Am. Chem. Soc.* **136,** 12406–12414
74. Eser, S., Schnieke, A., Schneider, G., and Saur, D. (2014) Oncogenic KRAS signalling in pancreatic cancer. *Br. J. Cancer* **111,** 817–822
75. Berndt, N., Hamilton, A. D., and Sebti, S. M. (2011) Targeting protein prenylation for cancer therapy. *Nat. Rev. Cancer* **11,** 775–791
76. Cromm, P. M., Schaubach, S., Spiegel, J., Fürstner, A., Grossmann, T. N., and Waldmann, H. (2016) Orthogonal ring-closing alkyne and olefin metathesis for the synthesis of small GTPase-targeting bicyclic peptides. *Nat. Commun.* **7,** 11300
77. Leshchiner, E. S., Parkhitko, A., Bird, G. H., Luccarelli, J., Bellairs, J. A., Escudero, S., Opoku-Nsiah, K., Godes, M., Perrimon, N., and Walensky, L. D. (2015) Direct inhibition of oncogenic KRAS by hydrocarbon-stapled SOS1 helices. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 1761–1766
78. Nicoludis, J. M., Barrett, S. P., Mergny, J. L., and Yatsunyk, L. A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.* **40,** 5432–5447
79. Sklenar, V., and Bax, A. (1987) Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson.* **74,** 469–479
80. Piotto, M., Saudek, V., and Sklenár, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2,** 661–665
81. Saccà, B., Lacroix, L., and Mergny, J. L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* **33,** 1182–1192
82. Ramsay, G. D., and Eftink, M. R. (1994) Analysis of multidimensional spectroscopic data to monitor unfolding of proteins. *Methods Enzymol.* **240,** 615–645
83. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921
84. Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2,** 2728–2733
85. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* **92,** 3817–3829
86. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935

Article 2: Fluorescent light-up acridine orange derivatives bind and stabilize KRAS-22RT G-quadruplex

Biochimie 144 (2018) 144–152

Contents lists available at ScienceDirect

Biochimie

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

Research paper

Fluorescent light-up acridine orange derivatives bind and stabilize KRAS-22RT G-quadruplex

Josué Carvalho [a], Edgar Pereira [b], Julien Marquevielle [c], Maria P.C. Campello [b], Jean-Louis Mergny [c, d], António Paulo [b, *], Gilmar F. Salgado [c], João A. Queiroz [a], Carla Cruz [a, **]

[a] CICS-UBI - Centro de Investigação em Ciências da Saúde, Universidade da Beira Interior, Av. Infante D. Henrique, 6200-506, Covilhã, Portugal
[b] Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico, Universidade de Lisboa, Estrada Nacional 10 (km 139,7), 2695-066, Bobadela LRS, Portugal
[c] Univ. Bordeaux, ARNA Laboratory, INSERM, U1212, CNRS UMR 5320, IECB, F-33600, Pessac, France
[d] Institute of Biophysics, AS CR, v.v.i. Kralovopolska 135, 612 65, Brno, Czech Republic

ARTICLE INFO

Article history:
Received 23 July 2017
Accepted 6 November 2017
Available online 10 November 2017

Keywords:
G-quadruplex
KRAS promoter
Acridine orange ligands
NMR spectroscopy
Fluorescent probes

ABSTRACT

KRAS is often found mutated in lethal cancers and should be an important target for anticancer drug development. However, no effective inhibitor has been reported so far, prompting the scientific community to describe the RAS proteins as nearly "undruggable". Recent approaches developed to modulate KRAS protein expression comprises the targeting of G-quadruplex (G4) structures formed within the nuclease hypersensitive element of KRAS promoter region, by designing small and specific ligands to stabilize the tertiary fold and reduce gene expression. In this work, we report *in vitro* and *in silico* studies of novel acridine orange (AO) derivatives (C_3-C_8), developed as G4 stabilizing agents. The results show that the ligands bind with high affinity and stabilize KRAS22-RT G4 with modest specificity over duplex DNA. The most promising ligand C_8 stabilizes the structure by $\approx$ 40 °C. Molecular docking using NMR-derived distance restraints reveal atomic details about the ligand structural features in the interaction with KRAS22-RT G4. In vitro studies with HeLa cells show that the ligands are cytotoxic with IC50 values between 0.9 μM and 5.7 μM. Moreover, the ligands tend to localize in the nucleus as shown by confocal fluorescence microscopy. Overall, these results show that the reported AO ligands display favourable properties as G4 ligands and this study provides structural detail for the development of lead KRAS G4 ligands.

1. Introduction

The KRAS proto-oncogene encodes for a GTPase ($p21^{KRAS}$) which cycles between GTP-bound active and GDP-inactive states, influencing key signalling pathways related to cell proliferation and apoptosis [1,2]. Due to activating mutations occurring in 30% of all human cancers, the mutated $p21^{KRAS}$ protein remains in activated state leading to increased cell proliferation, anti-apoptotic effects and consequently to cancer development [2,3]. KRAS mutations are found in a wide variety of cancers such as pancreatic, colorectal, lung and cervical cancers with high incidence in the top-ranked lethal cancers [1,3]. Moreover, $p21^{KRAS}$ is associated with the epidermal growth factor receptor (EGFR) and its mutation has been shown to increase the resistance to anti-EGFR therapies and conventional chemotherapy [3,4]. Therefore, KRAS is an attractive target for cancer therapeutics, despite no effective KRAS inhibitor has been approved thus far. The reported approaches focus on the development of mutated KRAS protein inhibitors, blocking downstream or upstream effectors, and on gene silencing by miRNA or siRNA. However, these strategies failed in achieving a suitable clinical response due to cancerous cell adaption and resistance to inhibitors and inefficient delivery of gene suppressors [1].

* Corresponding author.
** Corresponding author.
E-mail addresses: apaulo@ctn.tecnico.ulisboa.pt (A. Paulo), carlacruz@fcsaude.ubi.pt (C. Cruz).

https://doi.org/10.1016/j.biochi.2017.11.004
0300-9084/

One alternative approach to modulate KRAS protein expression relies on the targeting of G4 structures formed in the nuclease hypersensitive element (NHE) of KRAS promoter. Those structures were shown to have a regulatory role in the expression of KRAS [2,5,6]. Quadruplexes modulate important mechanisms such as gene transcription, translation and telomere elongation, and are appealing therapeutic targets for the development of specific small molecules that bind and stabilize the G4 structures [7–10]. The KRAS NHE region comprises a particular 32-mer sequence (32R) which contains 6 guanine tracts prone to G4 formation, whose structure was already shown to be highly polymorphic adopting different intramolecular topologies [11]. A smaller 21-mer sequence (21R) adopts the major G4 conformation of 32R. Recently, Salgado and co-workers determined the NMR structure of a 22-mer sequence (22RT) within the NHE which folds into a stable parallel-stranded G4 with 3 G-tetrads [12]. This structure (PDB code: 5I2V) shows unique structural elements such as the presence of a bulge and a 4-nucleotide loop, which can be used for G4-targeted ligands design. Different families of ligands have been synthesized and evaluated in the past decade, among those are porphyrins, acridines, anthraquinones, phenanthrolines, perylenes together with natural compounds such as telomestatin and berberine [13,14]. These compounds interact with the G4 structure through a variety of modes namely external stacking and groove binding [15]. The interest in acridine derivatives is due to its favourable characteristics such as a large planar moiety prone to π-stacking with the tetrads of the G4 structure, possibility of a cationic central ring nitrogen atom, intrinsic fluorescence and cellular uptake with known nucleus penetrating ability [16]. The compound BRACO-19 for instance, is an acridine derivative that binds telomeric G4s specifically over duplex DNA, inhibits telomerase and induces DNA damage with significant *in vivo* anticancer activity in tumour xenografts [17].

In this study, we have evaluated a series of Acridine Orange (AO) derivatives (Fig. 1) as a new class of G4 ligands. To tackle this goal, we studied their ability to bind and stabilize the quadruplex found in the KRAS promoter sequence (KRAS-22RT) using spectroscopic techniques such as circular dichroism (CD), fluorescence and nuclear magnetic resonance (NMR). Moreover, restrain-assisted molecular docking simulations using Autodock 4 [18] were performed in order to understand how this family of ligands interact with KRAS-22RT G4. Finally, we have assessed their inhibition properties against cancer cell proliferation using HeLa cell line and normal fibroblasts. These cytotoxicity tests were performed by MTT assays. Moreover, the intracellular localization of the compounds was studied by confocal fluorescence microscopy.

2. Materials and methods

2.1. Oligonucleotides and AO derivatives

All oligonucleotides were obtained from Eurofins MWG Operon (Ebersberg, Germany) or STAB VIDA Genomins Lab (Lisbon, Portugal) with HPLC-grade purification and further purified using Vivaspin® centricons. The following oligonucleotide sequences were used: KRAS-22RT: 5′-AGGGCGGTGTGGGAATAGGGAA-3'; c-MYC: 5′-TGAGGGTGGGTAGGGTGGGTAA-3'; 22AG: 5′-AGGGTTAGGGTTAGGGTTAGGG-3′ and ds26: d(5′- CAATCGGATCGAATTCGATCCGATTG-3′). Stock solutions of approximately 500 μM were prepared using Milli-Q water and stored at −20 °C until used.

The synthesis and purification of 10-(3-(4-iodobenzamide)propyl)-3,6-bis(dimethylamine) acridinium iodide (C_3), 10-(5-(4-iodobenzamide)pentyl)-3,6-bis(dimethylamine) acridinium iodide (C_5) and 10-(8-(4-iodobenzamide)octyl))-3,6-bis(dimethylamine) acridinium iodide (C_8) were performed as previously described [19]. Stock solutions of the compounds were prepared as 10 mM in DMSO. The following dilutions were done using Milli-Q water.

2.2. Circular dichroism assays

CD spectra were recorded with a Jasco J-815 spectropolarimeter equipped with a temperature control Peltier system. All measurements were made at 25 °C using previously annealed KRAS-22RT G4 (heating to 95 °C for 10 min, followed by cooling in ice) in 10 mM K-phosphate buffer pH 7.0 containing 50 mM K^+ or in physiologically relevant conditions buffer mimicking the intracellular salt environment (25 mM HEPES (pH = 7.2), 10.5 mM NaCl, 140 mM KCl, 130 nM $CaCl_2$, 2 mM $MgCl_2$). Test compounds were dissolved in Milli-Q water at 1 mM stock solution and the titrations performed by adding the required amount directly in a 1 mm path-length quartz cell containing 10 μM G4 DNA. Spectra were collected between 200 and 340 nm using an instrument scanning speed of

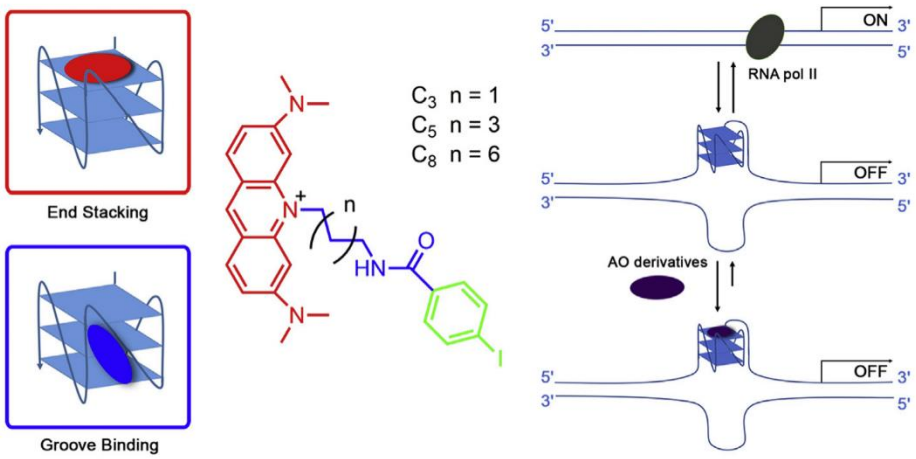


Fig. 1. Schematic diagram showing the structures of C_3, C_5 and C_8 and how the ligands stabilize KRAS G4 structure hence inhibiting the expression of KRAS protein. The acridine moiety interact with the G-tetrads surface through end stacking interactions (red). The alkylamide substituents interact through groove binding and a positively charged nitrogen in the acridine ring is able to occupy a position similar to that of the G-tetrad central ion (blue). The iodobenzene is expected to enhance the cellular uptake while providing additional interactions with the G4 grooves/loop interface (green). Interaction of the AO derivatives with the G4 structure is expected to maintain the "gene-off" state induced by the conversion of the double stranded "gene-on" form to a G4 "gene-off" form (model for the regulation of KRAS gene transcription on the right).

100 nm/min, 1 nm bandwidth, 2 s response time and averaged over three scans. All spectra were corrected by subtracting a buffer spectrum. For CD-melting experiments, KRAS22-RT G4 in the absence and presence of the compounds was heated between 20 and 110 °C (temperature set by the Peltier system), with a heating rate of 2 °C/min, while monitoring the ellipticity at 263 nm. Competition assays were performed by adding an excess of ds26 to the cuvette. Additionally, c-MYC and 22AG G4s were tested while monitoring the ellipticity at 263, 290 and 295 nm for parallel, hybrid-1 and antiparallel topologies. For these, buffer conditions were adjusted to favor different topologies and so that the G4s would melt around 50 °C: c-MYC in 100 mM LiCl; 22AG hybrid-1 in 10 mM KCl and 90 mM LiCl; 22AG antiparallel in 100 mM NaCl. Annealing was performed as described above. Data was converted into fraction folded (θ) plots using the following equation:

$$\theta = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}} \quad (1)$$

where CD is the ellipticity of the monitored wavelength at each temperature and CD^{min} and CD^{max} are the lowest and highest ellipticity, respectively. Data was fitted to a Boltzmann distribution (OriginPro 8) and the melting temperatures (T_m) determined.

2.3. Fluorescence binding studies

Fluorescence spectra were recorded with a Horiba FloroMax4 fluorometer equipped with a temperature control system. Ligands were excited at 498 nm and fluorescence emission collected between 530 and 700 nm in 10 mM K-phosphate buffer pH 7.0 containing 50 mM KCl at 25 °C using a quartz cuvette with a path length of 1 cm. Similar experiment was conducted using physiologically relevant buffers as described above. Excitation and emission slits were fixed at 2 nm. The association between G4 or duplex DNA and ligands was assessed by titrating the oligonucleotides and measuring the change in fluorescence. The titration was performed by adding previously annealed G4 solution or duplex DNA, followed by 3-min equilibration time. Data was converted into fraction of bound ligand (α) plots using the following equation:

$$\alpha = \frac{I - I_{\lambda}^{free}}{I_{\lambda}^{bound} - I_{\lambda}^{free}} \quad (2)$$

where I is the fluorescence intensity at 530 nm at each ligand:DNA ratio and I_{free} and I_{bound} are the fluorescence intensity of the free and fully bound ligand, respectively. Data points were then fitted to a hyperbolic function (OriginPro 8) and K_D values were determined from the following saturation binding model [20]:

$$\alpha = \frac{[DNA]}{K_D + [DNA]} \quad (3)$$

where α is the fraction of ligand bound and [DNA] is the concentration of the DNA. When required, the saturation binding Hill slope model was used [21]:

$$\alpha = \frac{[DNA]^h}{K_D + [DNA]^h} \quad (4)$$

where h is the Hill constant which describes cooperativity of ligand binding.

The Job plot analysis of ligand C_8 was performed by continuous variation of the ligand/DNA molar fraction while keeping a constant total (DNA + ligand) concentration of 1.5 μM. The excitation wavelength was set at 498 nm and the emission fluorescence recorded at 530 nm.

2.4. Nuclear magnetic resonance spectroscopy

^{1}H NMR spectra were recorded at temperature of 310 K on a Bruker Avance III 700 MHz spectrometer equipped with a liquid TXI ^{1}H/^{13}C/^{15}N/^{2}H with Z-gradient probe. Oligonucleotide concentration was typically 0.5 mM in 3 mm tubes with a total volume of 200 μL. G4 samples were annealed as described above in 20 mM K-phosphate buffer pH 6.5 containing 70 mM KCl and supplemented with 10% D_2O. NMR titration was performed by adding increasing amounts of ligand to the G4 solution.

2D ^{1}H-^{1}H NOESY and 1D ^{1}H-^{15}N HSQC spectra NMR spectra were recorded in 3 mm tubes at 37 °C unless otherwise stated. Usually the samples were prepared in K-phosphate NMR buffer (20 mM K_2HPO_4/KH_2PO_4, 70 mM KCl, pH 6.5, 10% D_2O). The concentrations of oligonucleotide were close to 1 mM. Water suppression was achieved with double pulse-field gradient spin echo (DPFGSE) methods. All spectra were processed with the software Topspin 3.2 and analyzed with MestReNova. Chemical shifts (δ) were measured in ppm.

2.5. Cell viability assay

Normal human dermal fibroblasts (NHDF) and human cervical cancer cells (HeLa) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, and 1% antibiotic/antimycotic (DMEM/F-12 was used for NHDF cells). Cultures were maintained at 37 °C in a humidified atmosphere containing 5% CO_2. Cells were seeded in 48-well plates (1×10^4 cells/well) and after 24 h these were incubated with the compounds at seven different concentrations ranging from 50 nM to 100 μM for 24 h. Wells containing untreated cells were used as control, as well as vehicle control (DMSO) and a common cytotoxic agent 5-fluorouracil (5-FU). At the end of incubation, the media was replaced with fresh media containing 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide salt (MTT) and further incubated at 37 °C for 2 h. Finally, MTT containing media was removed, formazan crystals were dissolved in DMSO and absorbance was recorded in a Bio-Rad xMark™ microplate reader at 570 nm. Cell viability data was expressed as mean ± SEM from at least three different experiments in comparison with control cells. IC50 values were determined using GraphPad Prism 6.

2.6. Confocal fluorescence microscopy

HeLa cells were cultured in glass bottom culture dish coated with poly-D-lysine (1×10^5 cells/dish) for 37 °C in a humidified atmosphere containing 5% CO_2. After growth, cells were incubated in the dark with 10 μM of compound at different times ranging from 5 min to 24 h. At the end of incubation, the excess of probe was washed off by rinsing with PBS three times. For the co-localization studies, cells were further treated with 1 μM nuclear probe Hoechst 33342 for 15 min. Prior to visualization, excess probe was washed off by rinsing in PBS three times. Cells were observed using a Zeiss AxioObserver LSM 710 microscope with 405, 488 and 514 nm laser excitation for the 4,6-diamidino-2-phenylindole (DAPI), fluorescein isothiocyanate (FITC) and rhodamine channels, respectively. Appropriate emission bands were selected for the two fluorescent channels, taking into account the spectral overlap and potential bleed-through between both channels. Images were processed with Zeiss ZEN software.

3. Results

3.1. AO ligands design

Most of the G4 ligands described in the literature share the same characteristics such as a large aromatic core suitable for π-π stacking interaction with the external G-tetrads and side chains enable to establish additional interactions with the grooves/loops of the G4 structure which confer selectivity over duplex DNA. Herein, three acridine orange derivatives with alkylamide side arms containing an iodobenzene substituent were synthesized and evaluated for their G4 binding ability (Fig. 1). These ligands were introduced originally by Paulo and collaborators who also reported the synthesis of their radioiodinated counterparts [19]. Interestingly, the possibility of obtaining the radioiodinated congeners might allow to speed-up the preclinical evaluation of this class of compounds, through biodistribution and *in vivo* stability/metabolism studies based on quantitative gamma-counting measurements and/or nuclear imaging techniques [19,22]. The planar acridine chromophore was already shown to be an interesting moiety for G4 ligands in the synthesis of 3,6,9-trisubstituted acridine BRACO-19 [17]. The ability of acridine to bind G4s is in part attributed to the fact that it is protonated at physiological pH being able to take advantage of the negative electrostatic potential of the G-tetrads central carbonyl groups [13,23]. The addition of the alkylamide side chains could provide additional interactions between the ligands and the grooves/loops and/or phosphate backbone of the G4 by H-bonding through the amide groups. Moreover, as reported previously, the inclusion of a halogenated substituent such as iodine may increase the binding affinity and stabilization degree of the ligands as well as the cellular uptake [24–26]. This reasoning led us to consider that the AO derivatives presented in Fig. 1 would present the structural requisites necessary to act as potent and selective G4-binders.

3.2. Binding and stabilization of KRAS-22RT G-quadruplex

Circular dichroism (CD) studies were performed to evaluate the ligands ability to bind and stabilize KRAS-22RT G4 structure. CD spectroscopy is an interesting technique to study the conformation of a G4 structure regarding its strands orientation and to assess the binding potential of G4 ligands [27]. KRAS-22RT adopts a parallel topology in the presence of K^+ cations showing the typical CD spectrum of such conformation (Fig. S1), with a maximum positive band around 263 nm and a weak minimum negative band around 240 nm. Upon addition of the compounds, only a slight decrease in ellipticity was observed for all ligands, which suggests the preferential binding of the pre-folded parallel topology without disrupting or converting the G4 conformation (Fig. S1). The ligand ability to induce thermal stabilization of KRAS-22RT G4 structure was then evaluated by CD melting experiments. The melting temperature (T_m) of the G4 in the absence and presence of 2 molar equivalents of ligands was determined by monitoring the ellipticity at 263 nm and increasing temperature. In the experimental conditions used, KRAS-22RT G4 yielded a melting temperature of about 51 °C. Upon addition of the ligands, the stability of KRAS-22RT was clearly enhanced as denoted by the higher T_m values (Fig. 2, Table 1).

C_3 promoted an increase in the T_m of 17.8 °C, while C_5 and C_8 stabilized the G4 structure by 29.0 and 40.7 °C, respectively. The ΔT_m value of near 40 °C obtained for the ligand C_8 was also corroborated by 1D NMR spectroscopy (Fig. S2). The increase in the length of the alkylamide side chain seems to promote a higher thermal stabilization degree as ligand C_8 was the best stabilizer. This feature was studied with more detail by molecular docking

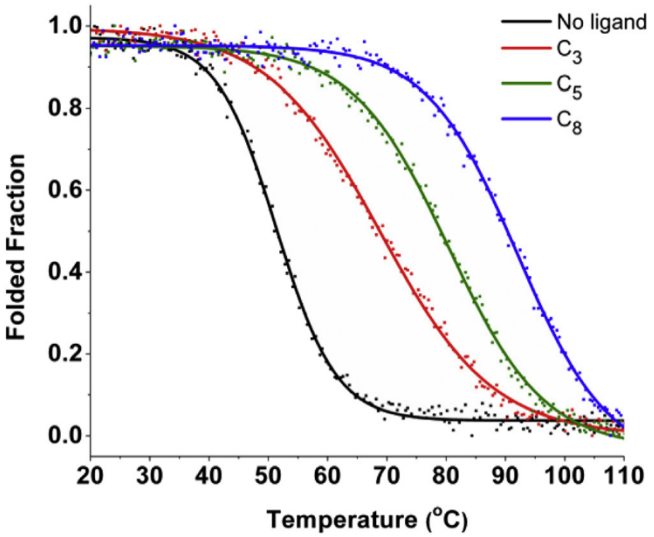


Fig. 2. CD melting curves of KRAS-22RT G-quadruplex at 5 μM in the absence and presence of 2 molar equivalents of ligands, in 10 mM K-phosphate buffer and 50 mM KCl. Data points were recorded at 263 nm. The Boltzmann curve fitting of the data points is shown. Temperature corresponds to the temperature set by the Peltier system.

studies (see Supporting Information). The obtained stabilization values for C_5 and C_8 are higher than those previously reported for other KRAS G4 ligands such as porphyrin TMPyP4 [21,28]. In addition, a competition assay was conducted with the self-complementary DNA sequence ds26, in order to evaluate the specificity of the ligands towards G4 DNA over duplex DNA (Fig. S3). Competition melting studies have been used before to assess whether ligands can bind preferentially to G4 structures compared to duplex DNA [28,29]. The ratio of the measured ΔT_m values in the presence and absence of competitor ds26 was used as a measure for specificity (S). In the presence of an excess of ds26 the ΔT_m induced by the ligands was almost unchanged as shown by the S values around 0.9. These results suggest a reasonable specificity of the ligands for KRAS-22RT G4 over duplex DNA.

To assess the selectivity of the ligands towards KRAS-22RT G4, we performed a similar CD melting study with other known G4 structures such as the promoter c-MYC and telomeric sequence 22AG. Promoter sequence c-MYC folds into a parallel G4 conformation while the telomeric sequence 22AG can form several distinct conformations depending on the cation in solution, namely hybrid-1 and antiparallel, in the presence K^+ or Na^+, respectively [7]. CD melting curves of the three different G4 conformations in the presence and absence of 2 molar equivalents of ligand were acquired (Figs. S4-S6). Despite using different type of cations, the ionic strength was kept equal for all G4 samples used. Moreover, KRAS-22RT G4 has a very robust and stable conformation even at a KCl concentration 10 times higher than the oligonucleotide concentration and at different pH in the range of 6.0–7.0 (checked by NMR, Figs. S7-S8). Therefore, the salt conditions used should affect the binding of the AO derivatives in a lesser degree, being the results comparable. The ΔT_m values for each G4 are summarized in Table 1. The results indicate that the ligands are able to stabilize other G4 structures, despite the lower degree of stabilization observed. A similar trend to that of KRAS-22RT is observed as C_8 promotes the higher stabilization while C_3 is the less effective stabilizer, for all G4 structures. The ligands seem to have a binding preference towards the parallel G4s as the experiments with c-MYC shown the higher ΔT_m values ranging between 18.5 and 33.9 °C, when compared to the hybrid-1 and antiparallel structures of 22AG. The results suggest the

Table 1
Ligand-induced thermal stabilization of various G4s measured by CD melting experiments in the absence and presence of ligands.

	ΔT_m (°C)[a]			
Ligand	KRAS-22RT	c-MYC	22AG-h[b]	22AG-a[c]
C_3	17.8 ± 0.2	18.5 ± 0.3	6.4 ± 0.1	1.5 ± 0.2
C_5	29.0 ± 0.2	22.0 ± 0.2	7.6 ± 0.1	2.7 ± 0.2
C_8	40.7 ± 0.3	33.9 ± 0.7	13.2 ± 0.1	6.8 ± 0.2

[a] ΔT_m represents difference in melting temperature [$\Delta T_m = T_m$ (DNA + 2 molar eq. ligand) − T_m (DNA)]. The buffer was 10 mM K-phosphate buffer, pH 7.0. The T_m values are 51.2 ± 0.1 °C [KRAS-22RT (5 μM) in 50 mM KCl]; 51.2 ± 0.2 °C [c-MYC (5 μM) in 100 mM LiCl]; 54.4 ± 0.1 °C [22AG hybrid-1 (5 μM) in 10 mM KCl + 90 mM LiCl]; 56.3 ± 0.2 °C [22AG antiparallel (5 μM) in 100 mM NaCl]. Values reported are average of three measurements with the estimated standard deviation.
[b] Hybrid-1 topology.
[c] Antiparallel topology.

order KRAS-22RT > c-MYC > 22AG-h > 22AG-a in terms of thermal stabilization by the ligands.

It is also important to investigate the stabilization effect of the ligands in physiological buffer conditions [30]. Therefore, the CD melting studies (Fig. S9) with KRAS-22RT G4 were reconducted in a buffer mimicking the intracellular salt environment (25 mM HEPES (pH = 7.2), 10.5 mM NaCl, 140 mM KCl, 130 nM $CaCl_2$, 2 mM $MgCl_2$) [30]. Under these buffer conditions, the T_m of KRAS-22RT G4 structure is 57.9 ± 0.1 °C, which is higher than that obtained previously (51.2 ± 0.2 °C) possibly due to the ionic strength. The obtained ΔT_m values were 9.5 ± 0.2, 16.4 ± 0.2 and 31.5 ± 0.3 °C for C_3, C_5 and C_8, respectively. Despite a slightly lower stabilization (in terms of relative ΔT_m) as compared to the results shown in Table 1, the absolute T_m values of KRAS-22RT in the presence of the ligands were roughly the same (68.9, 80.4 and 92.1 °C *vs* 67.4, 74.3 and 89.4 °C for C_3, C_5 and C_8, respectively). The difference in the ΔT_m could be a result of the higher T_m value of the G4 structure alone due to the ionic strength.

3.3. Fluorescence binding studies with KRAS-22RT G4

The interaction of the ligands with KRAS-22RT G4 was further studied using fluorescence spectroscopy. The ligands excitation at 498 nm produces the emission of light with a broad band centered

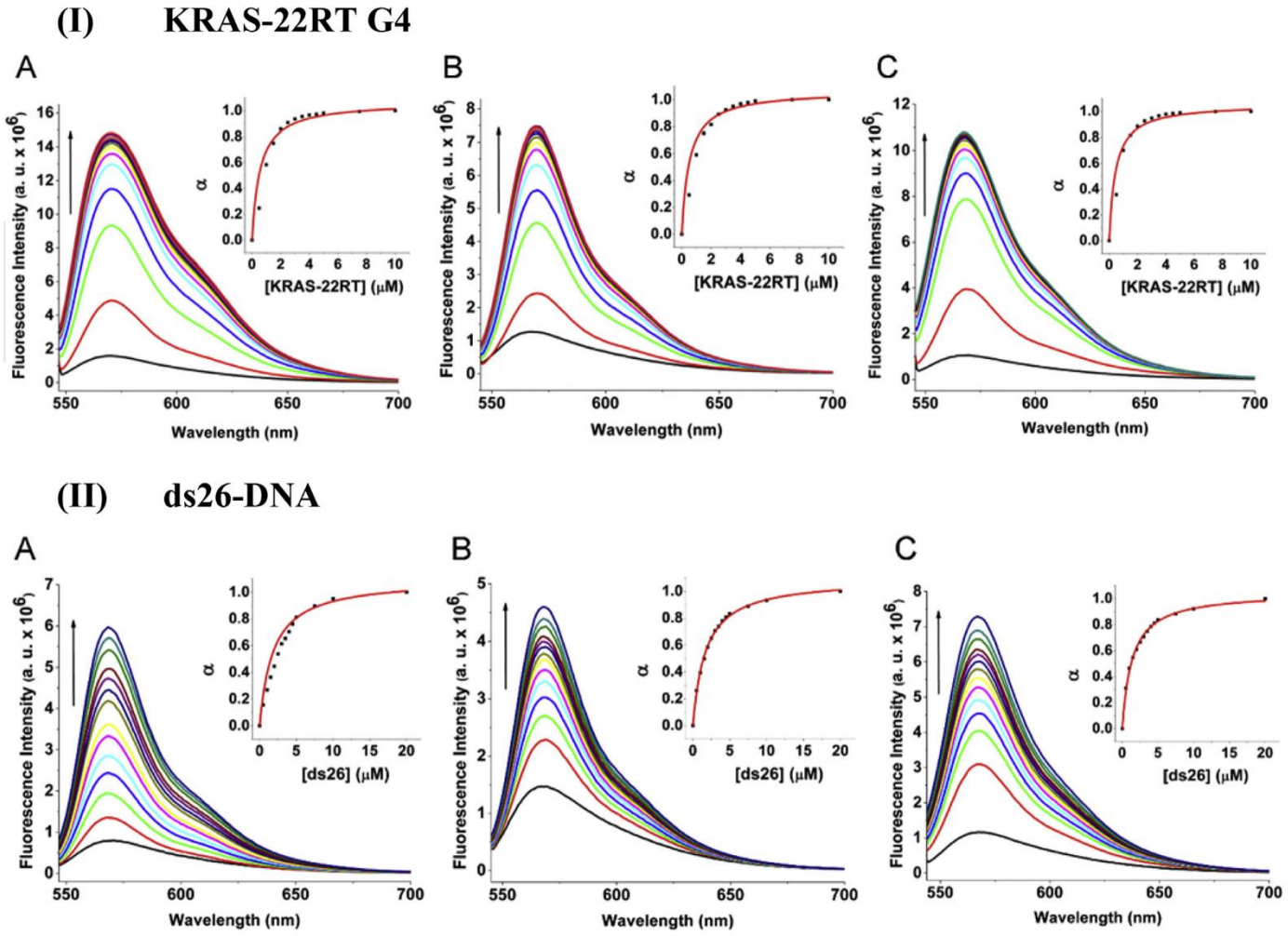


Fig. 3. Fluorescence emission spectra of (A) C_3, (B) C_5 and (C) C_8 with increasing concentration of (I) KRAS-22RT G4 (0.5–10 μM) and (II) DNA ds26 (0.5–20 μM) in 10 mM K-phosphate buffer and 50 mM KCl. Insets: fraction of ligand bound plots fitted to the saturation binding equation.

Table 2
Apparent dissociation constants (K_D) obtained from fluorescence titration experiments.

	K_D (M)[a]			
Ligand	KRAS-22RT	c-MYC	22AG	ds26
C_3	5.5×10^{-7}	1.8×10^{-7}	5.8×10^{-7}	1.9×10^{-6}
C_5	5.1×10^{-7}	2.7×10^{-7}	5.8×10^{-7}	1.7×10^{-6}
C_8	4.4×10^{-7}	7.4×10^{-7}	6.4×10^{-7}	1.4×10^{-6}

[a] Determined from fitting the fluorometric titration points to Equation (3) or 4 in the case of c-MYC. The buffer used was 10 mM K-phosphate buffer, pH 7.0. SD $\leq 1.2 \times 10^{-7}$.

at 568 nm. Spectrofluorometric titrations were performed with G4 and duplex DNA and the apparent dissociation constants (K_D) determined for the different ligands (Fig. 3 and Table 2) in 10 mM K-phosphate buffer and 50 mM KCl. Upon titration of pre-annealed G4 solution, up to ≈10-fold fluorescence enhancement was observed (Fig. 3-I). The K_D values are presented in Table 2. The fluorescence titrations were also conducted in a buffer mimicking physiological conditions (same buffer used in CD) and the fitting of the data points to the saturation binding model are presented in Fig. S10 of Supporting Information. The K_D values obtained were 3.3×10^{-7}, 1.4×10^{-7} and 2.6×10^{-7} for C_3, C_5 and C_8, respectively. The values are similar to those reported in Table 2 and both are indicative of high affinity in the sub-micromolar range, similar to those of potent ligands such as pyridostatin [31].

In the case of ds26 (Fig. 3-II), the K_D values obtained are about 3.5-fold higher than those obtained for KRAS-22RT G4, indicating weaker binding and modest specificity (both by S and K_D values). Although the specificity was less than that obtained for telomestatin or Phen-DC_3 [32], these results are in line with compounds such as quinacridines (MMQs) [33], indicating that the AO derivatives are interesting scaffolds for further development of G4 specific compounds. For all the complexes formed with G4 and ds26, no variation in the emission wavelength maximum was observed. This suggests a similar chemical environment for the complex formation, although different structures are formed in each complex. However, the fluorescence enhancement upon binding G4 DNA is higher than that observed for ds26 (≈3–7 fold). The varying extent of fluorescence enhancement caused by different DNAs suggests a stronger interaction between the ligands and G4 DNA [34].

These results for ds26 are also in agreement with the CD-melting competition assays and indicate some binding preference of the ligands towards G4 structures.

The K_D values for the interaction of the AO derivatives with c-MYC and 22AG G4s were also determined and are presented in Table 2 and Figs. S11 and S12. Intriguingly, the binding curves of the ligands with c-MYC G4 show a sigmoidal rather than hyperbolic shape which may indicate a cooperative mechanism of binding (Fig. S11). Therefore, the data points were fitted to Hill equation assuming cooperative binding sites oppositely to that of KRAS-22RT which assumes independent and equivalent binding sites [21]. C_3 has a higher affinity towards c-MYC and the ΔT_m was higher when compared with other G4s; C_8 has lower affinity and ΔT_m towards c-MYC G4 when compared to KRAS-22RT. The Hill constants (h) were greater than 1 for all ligands (1.9, 1.8 and 3.5) which denotes positively cooperative binding, *i.e.* the binding of one ligand increases the affinity of other ligand molecules [21]. In the case of 22AG, the affinity was lower for all ligands when compared to the other G4s studied (Table 2), which is in agreement with the lower degree of stabilization observed in CD melting studies (Fig. S12).

An increase in the alkylamide chain length seems to promote more ligand plasticity, further increasing the binding contacts between the ligands and the G4 structures. This fact may be the explanation for the different ΔT_m values obtained for the ligands since the binding affinities are quite similar between ligands. Furthermore, between different G4 structures, similar K_D values in the order of 10^{-7} M are also observed, indicating that the binding modes are different between the different ligands and the various G4 architectures used.

3.4. NMR studies with KRAS-22RT G4

NMR spectroscopy was used to study the complex formation between the AO ligands and KRAS-22RT G4 (Fig. 4, Figs. S13 and S14). The imino region of KRAS-22RT ^{1}H NMR spectrum shows 12 well resolved peaks corresponding to a single G4 conformation with three G-tetrads [12]. Titration of KRAS-22RT G4 with the strongest ligand C_8 is shown in Fig. 4.

Upon addition of the ligand, the imino peaks became significantly broader at low ligand concentration. At 2:1 ratio, the imino peaks coalesce and become almost undetectable. This behavior is observed by many different ligands probed against the same oligonucleotide sequence (results not shown), and indicate the formation of higher order complexes with shared ligand molecules which slowly tumble in the liquid state optimal time scale (μs-ps) [35]. In addition, the occurrence of multiple binding modes and

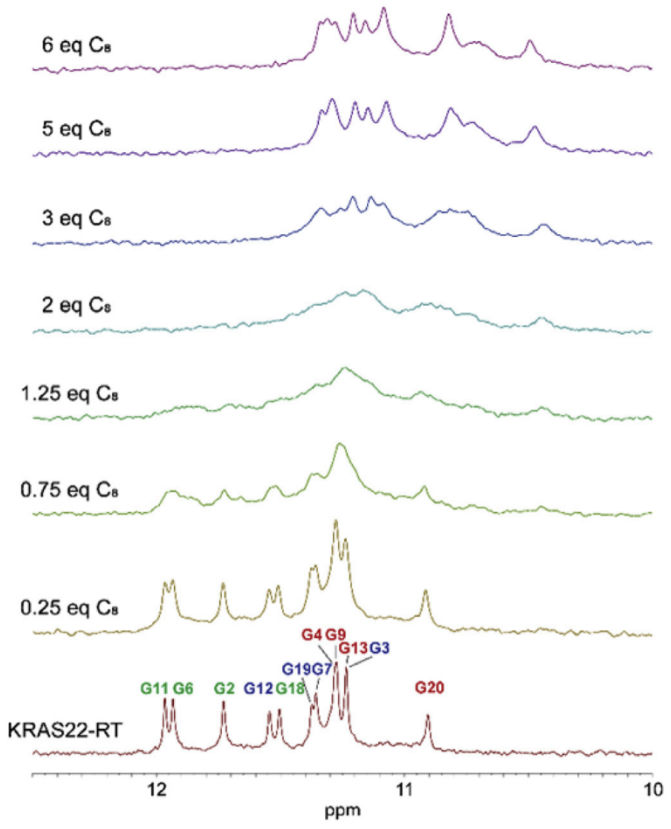


Fig. 4. Imino region of KRAS22-RT G4 ^{1}H spectra upon titration with compound C_8 at 310 K. The assignment of free KRAS22-RT is shown: top 5′ G-quartet protons are labelled in green, middle tetrad in blue and bottom 3′ tetrad in red.

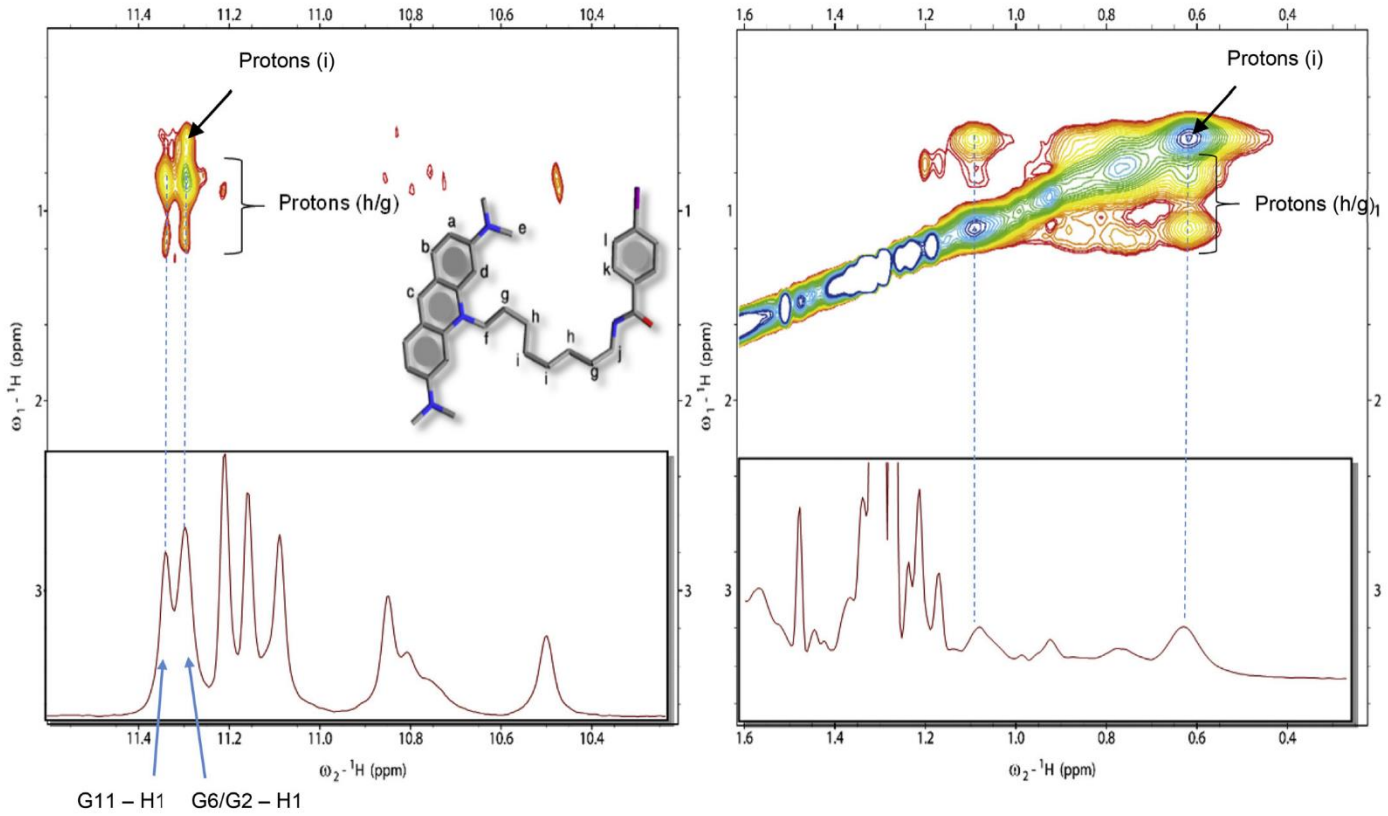


Fig. 5. Expanded region of the 2D-NOESY spectrum showing the intermolecular NOEs between the G2, G6 and G11 H1 protons (imino) and ligand C_8, namely the alkylamide chain aliphatic protons of the ligand (inset). KRAS-22RT G-quadruplex at 350 μM in presence of 4 molar equivalents of C_8.

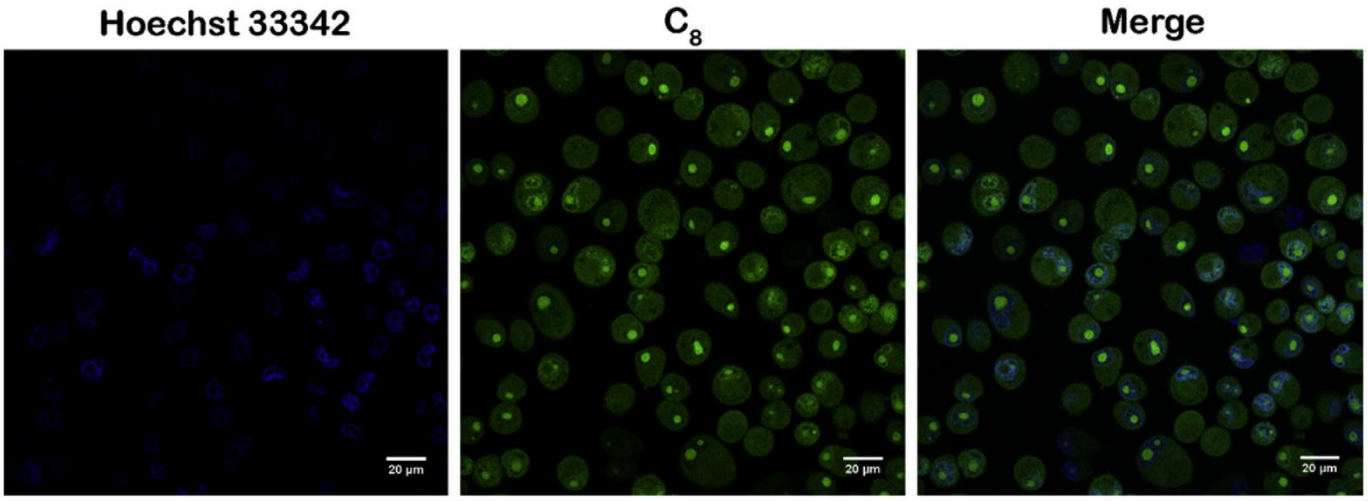


Fig. 6. Confocal microscopy images of HeLa cells treated for 1 h with 10 μM C_8 (middle). Nuclear probe Hoechst 33342 (left) was used for co-localization (merge, right). Scale bar is 20 μm. C_3 and C_5 shown similar results.

conformations in the ensemble, means that one nucleus can have multiple peaks; hence, in the intermediate timescale the averaged NMR signal will be broadened and may disappear. Finally, at high ligand concentrations, the signals re-appear, indicating the saturation of the binding sites. At a ligand:DNA ratio above 4:1 the imino protons signals did not change significantly. These results suggest a 4:1 ligand:DNA binding ratio. In order to confirm the binding stoichiometry of the C_8 ligand, Job plot analysis was performed using the fluorescence titration data with KRAS-22RT (Fig. S15). The molar fraction (χ) of 0.2 was determined for C_8 by the intersection of the linear regressions for high and low molar fractions, which suggests a 4:1 binding ratio. This supports the stoichiometry observed in the ^{1}H NMR titrations.

The imino peaks were coherently and significantly shifted upfield, which is described as being associated with π-π stacking interactions between ligands and the G-tetrads [36,37]. Such interaction mode is expected with the large aromatic acridine moiety of the ligands. Following the determination of the 4:1 stoichiometry for C_8, the CD melting experiment was carried using 4 molar equivalents of ligand which further increased the KRAS G4

thermal stabilization (>45 °C). However, the melting temperature was not possible to be determined due to the temperature range limitation. Regarding C_3 and C_5, the analysis of the ^{1}H NMR titration (Figs. S13 and S14) suggest the same ligand:DNA binding stoichiometry of 4:1. The imino protons signals did not change significantly at higher ligand concentrations, suggesting the saturation of the binding sites. Additional ^{1}H-^{1}H 2D-NOESY experiments at 37 °C were performed using the most promising ligand C_8 at a ligand:DNA ratio of 4:1. Despite the overlapping between the signals of KRAS22-RT G4 and C_8, several NOE cross-peaks are clearly observed between the ligand and the DNA, namely with the alkylamide chain aliphatic protons of C_8 (Fig. 5). Using ^{15}N-labelled G4 samples it was possible to assign unambiguously G11. Two other imino-protons (G2 and G6) are overlapped and were assigned to the same chemical shift ($\approx$11.95 ppm). In addition, the cross-peaks show almost a two-fold peak volume compared to G11 imino, suggesting that both iminos could participate in the cross peak with aliphatic protons i, g and h. Since G2, G6, G11 and G18 belong to the same tetrad it is reasonable to think that the top 5′ tetrad is the principal binding site, although other binding sites are expected due to the 4:1 binding stoichiometry. However, the overlapping of signals volume hinders the unambiguous assignment and analysis of other NOE coming from the 4 ligands bound. Nevertheless, the assigned peaks denote a strong interaction and are expected to correspond to the principal binding site.

A deeper insight on the binding mode between ligand C_8 and the G4 structure was obtained by molecular docking studies using the NOE distance restraints obtained in a 250 ms mixing time homonuclear NOESY NMR experiments. The NMR structure of KRAS-22RT (PDB code: 5I2V) was used as docking target for simulation using Autodock 4.2. The results depict the interactions involved in the recognition of the apparent principal binding site, according to the NOE signals observed (see Supporting Information for further discussion).

3.5. Cell studies: cytotoxicity and intracellular localization

The effect of the ligands on cell proliferation was assessed on a cervical cancer cell line (HeLa) and non-malignant human cells (NHDF) by the MTT assay. KRAS gene appears frequently mutated in cervical cancer [38]. The anticancer drug 5-fluorouracil (5-FU) was also used as a control. All the compounds had an effect on the viability of the cancer cells (Fig. S16). The concentrations inducing 50% proliferation inhibition (IC50) were remarkably lower for the AO ligands when compared with the reference agent 5-FU. C_3 had an IC50 value of 5.7 μM, while C_5 and C_8 had a lower value of 2.4 and 0.9 μM, respectively. Notably, 5-FU presented an IC50 value > 100 μM which is >100-fold greater than the value obtained for C_8. As seen previously, the increase in length of the alkylamide chain seems to be important for the activity of the ligand as there is a relationship between higher stabilization degree, affinity and cytotoxicity, and side chain length. Regrettably, the compounds induced a similar anti-proliferative effect on NHDF cells showing non-selective cytotoxic effect. The IC50 values for NHDF cells were 3.7, 2.5 and 0.7 for C_3, C_5 and C_8, respectively.

Taking advantage of the intrinsic fluorescence of the ligands, confocal fluorescence microscopy was employed to assess their intracellular localization. All the ligands emit green fluorescence upon excitation at 488 nm. After treating the cells with the compounds, the acridine orange derivatives seem to preferentially localize within the nucleus and particularly in the nucleolus, despite some localization in the cytoplasm (Fig. 6). Nucleolus accumulation of ligands often means G4 RNA binding, thus being possible that the ligand binds these types of structures intracellularly as well [39]. The cellular uptake seems to be fast as considerable green fluorescence was observed even at 5 min incubation time of the ligand (Fig. S17). The intracellular localization is similar for the three ligands. Acridine orange is known for penetrating the nucleus of cells, thus this aromatic moiety may be responsible for the ligands behavior [40]. The intracellular uptake and nucleus/nucleolus accumulation of the ligands suggest that these compounds may actually reach their intended intracellular target.

4. Conclusions

KRAS plays an important role in the pathogenesis of several types of cancer and is therefore an attractive target for anticancer drugs. Targeting G4 structures formed within KRAS promoter to block transcription and expression has attracted the scientific community in the past years despite no effective suppressor has yet been found [2]. Thus, the selection of molecules capable of binding to KRAS G4s is still on demand. Overall, this study shows the potential use of AO-based ligands for anticancer drug development based on KRAS oncogene promoter targeting and modulation. Acridine derivatives have been shown already to successfully recognize G4 DNA with antitumor effect, in particular BRACO-19 [17].

The AO moiety together with the alkylamide side chains led to high affinity and reasonable selectivity towards KRAS-22RT G4, promoting a thermal stabilization of about 40 °C for the best ligand C_8. The observed degree of stabilization for C_8 is competitive with other ligands reported for KRAS G4 such as triarylpyridines, indoloquinolines or TMPyP4 [21,28]. Additionally, the *S* values of around 0.9 and 3.5-fold higher K_D of the ligand in the presence of duplex DNA shown a modest but significant G4 specificity. The ligands were shown to improve the thermal stability of other known G4 structures but to a lesser extent. Studies using physiological buffers indicate that the ability of the ligands to bind and stabilize G4 structures in a cellular environment is maintained.

The ligands showed high toxicity against HeLa cancer cells; however the toxicity against non-malignant cells should be improved. To overcome these issues, we plan to introduce additional side chain functionalities in order to increase selectivity towards G4, for instance, by substituting the iodobenzene group with other promising moieties containing fluorine or chloride (work in progress). Another strategy could be to increase the size of the aromatic moiety to limit its access to the smaller surface area of AT and GC base pairs of a typical duplex DNA. Moreover, the nuclear penetration ability and the fluorescence properties of the ligands make them attractive for the development of potential probes for G4 labelling in cellular environment. In brief, the *in silico* and the structural results that we provide in this work will greatly help in designing ligands with better selectivity for KRAS G4 and with promise for G4 DNA based therapeutics.

Acknowledgments

Josué Carvalho acknowledges the doctoral fellowship grant from FCT - Foundation for Science and Technology ref. SFRH/BD/122953/2016. This work was supported by project “Ações Integradas Luso-Francesas” ref. TC-15/17 entitled “Aptameros de G-quadruplex para tratamento do cancro do colo do útero”, FCT project “Projeto de Investigação Exploratória” ref. IF/00959/2015 entitled “Nucleolin targeting by G-quadruplex aptamers for cervical cancer therapy” financed by Fundo Social Europeu e Programa Operacional Potencial Humano (IF/00959/2015). This work was also supported by FEDER funds through the POCI - COMPETE 2020 - Operational Programme Competitiveness and Internationalisation in Axis I - Strengthening research, technological development and innovation

(Project POCI-01-0145-FEDER-007491), by National Funds by FCT (Project UID/Multi/00709/2013), by Conseil régional d'Aquitaine (to J.L.M.) Ligue régionale contre le Cancer (to G.F.S) and by the SYMBIT project (reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477) financed by the ERDF (CZ.02.1.01/0.0/0.0/15_003/0000477).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.biochi.2017.11.004.

References

[1] A.D. Cox, S.W. Fesik, A.C. Kimmelman, J. Luo, C.J. Der, Drugging the undruggable RAS: mission possible? Nat. Rev. Drug Discov. 13 (2014) 828–851, https://doi.org/10.1038/nrd4389.
[2] S. Cogoi, L.E. Xodo, G4 DNA in ras genes and its potential in cancer therapy, Biochim. Biophys. Acta - Gene Regul. Mech. 1859 (2016) 663–674, https://doi.org/10.1016/j.bbagrm.2016.02.002.
[3] R. Chetty, D. Govender, Gene of the month: KRAS, J. Clin. Pathol. 66 (2013) 548–550, https://doi.org/10.1136/jclinpath-2015-203207.
[4] A. Casadei Gardini, L. Capelli, P. Ulivi, M. Giannini, E. Freier, S. Tamberi, E. Scarpi, A. Passardi, W. Zoli, A. Ragazzini, D. Amadori, G.L. Frassinet, KRAS, BRAF and PIK3CA status in squamous cell anal carcinoma (SCAC), PLoS One 9 (2014) e92071, https://doi.org/10.1371/journal.pone.0092071.
[5] M. Paramasivam, S. Cogoi, L.E. Xodo, Primer extension reactions as a tool to uncover folding motifs within complex G-rich sequences: analysis of the human KRAS NHE, Chem. Commun. (Camb) 47 (2011) 4965–4967, https://doi.org/10.1039/c1cc10269a.
[6] M. Paramasivam, A. Membrino, S. Cogoi, H. Fukuda, H. Nakagama, L.E. Xodo, Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription, Nucleic Acids Res. 37 (2009) 2841–2853, https://doi.org/10.1093/nar/gkp138.
[7] G. Collie, G. Parkinson, The application of DNA and RNA G-quadruplexes to therapeutic medicines, Chem. Soc. Rev. 40 (2011) 5867–5892, https://doi.org/10.1039/c1cs15067g.
[8] S. Balasubramanian, L.H. Hurley, S. Neidle, Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nat. Rev. Drug Discov. 10 (2011) 261–275, https://doi.org/10.1038/nrd3428.
[9] S. Neidle, Quadruplex nucleic acids as targets for anticancer therapeutics, Nat. Rev. Chem. 1 (2017) 41, https://doi.org/10.1038/s41570-017-0041.
[10] S. Neidle, Quadruplex nucleic acids as novel therapeutic targets, J. Med. Chem. 59 (2016) 5987–6011, https://doi.org/10.1021/acs.jmedchem.5b01835.
[11] S. Cogoi, M. Paramasivam, B. Spolaore, L.E. Xodo, Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins, Nucleic Acids Res. 36 (2008) 3765–3780, https://doi.org/10.1093/nar/gkn120.
[12] A. Kerkour, J. Marquevielle, S. Ivashchenko, L.A. Yatsunyk, J.-L. Mergny, G.F. Salgado, High-resolution 3D NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation, J. Biol. Chem. 292 (2017) 8082–8091, https://doi.org/10.1074/jbc.M117.781906.
[13] D. Monchaud, M.-P. Teulade-Fichou, A hitchhiker's guide to G-quadruplex ligands, Org. Biomol. Chem. 6 (2008) 627–636, https://doi.org/10.1039/b714772b.
[14] M. Franceschin, L. Rossetti, A. D'Ambrosio, S. Schirripa, A. Bianco, G. Ortaggi, M. Savino, C. Schultes, S. Neidle, Natural and synthetic G-quadruplex interactive berberine derivatives, Bioorg. Med. Chem. Lett. 16 (2006) 1707–1711, https://doi.org/10.1016/j.bmcl.2005.12.001.
[15] S. Zhang, Y. Wu, W. Zhang, G-quadruplex structures and their interaction diversity with ligands, ChemMedChem. 9 (2014) 899–911, https://doi.org/10.1002/cmdc.201300566.
[16] J. Debray, W. Zeghida, M. Jourdan, D. Monchaud, M.-L. Dheu-Andries, P. Dumy, M.-P. Teulade-Fichou, M. Demeunynck, Synthesis and evaluation of fused bispyrimidinoacridines as novel pentacyclic analogues of quadruplex-binder BRACO-19, Org. Biomol. Chem. 7 (2009) 5219–5228, https://doi.org/10.1039/b912716j.
[17] A.M. Burger, F. Dai, C.M. Schultes, A.P. Reszka, M.J. Moore, J.A. Double, S. Neidle, The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function, Cancer Res. 65 (2005) 1489–1496, https://doi.org/10.1158/0008-5472.CAN-04-2910.
[18] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791, https://doi.org/10.1002/jcc.21256.
[19] E. Pereira, L. Quental, E. Palma, M.C. Oliveira, F. Mendes, P. Raposinho, I. Correia, J. Lavrado, S. Di Maria, A. Belchior, P. Vaz, I. Santos, A. Paulo, Evaluation of acridine orange derivatives as DNA-targeted radiopharmaceuticals for auger Therapy: influence of the radionuclide and distance to DNA, Sci. Rep. 10 (2017) 42544, https://doi.org/10.1038/srep42544.
[20] T.D. Pollard, A guide to simple and informative binding assays, Mol. Biol. Cell. 21 (2010) 4061–4067, https://doi.org/10.1091/mbc.E10-08-0683.
[21] J. Lavrado, H. Brito, P.M. Borralho, S. a. Ohnmacht, N.-S. Kim, C. Leitão, S. Pisco, M. Gunaratnam, C.M.P. Rodrigues, R. Moreira, S. Neidle, A. Paulo, KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines, Sci. Rep. 5 (2015) 9696, https://doi.org/10.1038/srep09696.
[22] A.P. Kiess, I. Minn, Y. Chen, R. Hobbs, G. Sgouros, R.C. Mease, M. Pullambhatla, C.J. Shen, C.A. Foss, M.G. Pomper, Auger radiopharmaceutical therapy targeting prostate-specific membrane antigen, J. Nucl. Med. 56 (2015) 1401–1407, https://doi.org/10.2967/jnumed.115.155929.
[23] M. Read, R.J. Harrison, B. Romagnoli, F.A. Tanious, S.H. Gowan, A.P. Reszka, W.D. Wilson, L.R. Kelland, S. Neidle, Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 4844–4849, https://doi.org/10.1073/pnas.081560598.
[24] T. Meng, S.-F. Tang, Q.-P. Qin, Y.-L. Liang, C.-X. Wu, C.-Y. Wang, H.-T. Yan, J.-X. Dong, Y.-C. Liu, Evaluation of the effect of iodine substitution of 8-hydroxyquinoline on its platinum (ii) complex: cytotoxicity, cell apoptosis and telomerase inhibition, Med. Chem. Commun. 7 (2016) 1802–1811, https://doi.org/10.1039/C6MD00201C.
[25] R. Nanjunda, E. a. Owens, L. Mickelson, T.L. Dost, E.M. Stroeva, H.T. Huynh, M.W. Germann, M.M. Henary, W.D. Wilson, Selective G-quadruplex DNA recognition by a new class of designed cyanines, Molecules 18 (2013) 13588–13607, https://doi.org/10.3390/molecules181113588.
[26] A. Sidibe, F. Hamon, E. Largy, D. Gomez, M.-P. Teulade-Fichou, C. Trentesaux, J.F. Riou, Effects of a halogenated G-quadruplex ligand from the pyridine dicarboxamide series on the terminal sequence of XpYp telomere in HT1080 cells, Biochimie 94 (2012) 2559–2568, https://doi.org/10.1016/j.biochi.2012.07.003.
[27] S. Paramasivan, I. Rujan, P.H. Bolton, Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding, Methods 43 (2007) 324–331, https://doi.org/10.1016/j.ymeth.2007.02.009.
[28] N.M. Smith, G. Labrunie, B. Corry, P.L.T. Tran, M. Norret, M. Djavaheri-Mergny, C.L. Raston, J.-L. Mergny, Unraveling the relationship between structure and stabilization of triarylpyridines as G-quadruplex binding ligands, Org. Biomol. Chem. 9 (2011) 6154–6162, https://doi.org/10.1039/c1ob05560g.
[29] A. De Rache, N.M. Gueddouda, A. Bourdoncle, P. Hommes, H.-U. Reissig, J.-L. Mergny, A flexible terpyridine derivative interacts specifically with G-quadruplexes, Chem. - A Eur. J. 22 (2016) 12651–12654, https://doi.org/10.1002/chem.20160302.
[30] R. Hansel, S. Foldynová-Trantírková, V. Dotsch, L. Trantirek, Investigation of quadruplex structure under physiological conditions using in-cell NMR, Top. Curr. Chem. 330 (2013) 47–66, https://doi.org/10.1007/128.
[31] D. Koirala, S. Dhakal, B. Ashbridge, Y. Sannohe, R. Rodriguez, H. Sugiyama, S. Balasubramanian, H. Mao, A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands, Nat. Chem. 3 (2011) 782–787, https://doi.org/10.1038/nchem.1126.
[32] M. Bončina, Č. Podlipnik, I. Piantanida, J. Eilmes, M.-P. Teulade-Fichou, G. Vesnaver, J. Lah, Thermodynamic fingerprints of ligand binding to human telomeric G-quadruplexes, Nucleic Acids Res. (2015), https://doi.org/10.1093/nar/gkv1167 gkv1167.
[33] C. Hounsou, L. Guittat, D. Monchaud, M. Jourdan, N. Saettel, J.-L. Mergny, M.-P. Teulade-Fichou, G-quadruplex recognition by quinacridines: a SAR, NMR, and biological study, ChemMedChem. 2 (2007) 655–666, https://doi.org/10.1002/cmdc.200600286.
[34] X. Zhang, Y. Wei, T. Bing, X. Liu, N. Zhang, J. Wang, J. He, B. Jin, D. Shangguan, Development of squaraine based G-quadruplex ligands using click chemistry, Sci. Rep. 7 (2017) 4766, https://doi.org/10.1038/s41598-017-04344-x.
[35] A.G. Palmer, NMR characterization of the dynamics of biomacromolecules, Chem. Rev. 104 (2004) 3623–3640, https://doi.org/10.1021/cr030413t.
[36] A.T. Phan, V. Kuryavyi, H.Y. Gaw, D.J. Patel, Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter, Nat. Chem. Biol. 1 (2005) 167–173, https://doi.org/10.1038/nchembio723.
[37] V. Dhamodharan, S. Harikrishna, A.C. Bhasikuttan, P.I. Pradeepkumar, Topology specific stabilization of promoter over telomeric g-quadruplex DNAs by bisbenzimidazole carboxamide derivatives, ACS Chem. Biol. 10 (2015) 812–833, https://doi.org/10.1021/cb5008597.
[38] A.A. Wright, B.E. Howitt, A.P. Myers, S.E. Dahlberg, E. Palescandolo, P. Van Hummelen, L.E. MacConaill, M. Shoni, N. Wagle, R.T. Jones, C.M. Quick, A. Laury, I.T. Katz, W.C. Hahn, U.A. Matulonis, M.S. Hirsch, Oncogenic mutations in cervical cancer: genomic differences between adenocarcinomas and squamous cell carcinomas of the cervix, Cancer 119 (2013) 3776–3783, https://doi.org/10.1016/j.pestbp.2011.02.012.
[39] D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C.B. Ho, C. Proffitt, K. Trent, J.P. Whitten, J.K.C. Lim, D. Von Hoff, K. Anderes, W.G. Rice, Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis, Cancer Res. 69 (2009) 7653–7661, https://doi.org/10.1158/0008-5472.CAN-09-1304.
[40] N. Agorastos, L. Borsig, A. Renard, P. Antoni, G. Viola, B. Spingler, P. Kurz, R. Alberto, Cell-specific and nuclear targeting with [M(CO)3]+ (M=99mTc, Re)-Based complexes conjugated to acridine orange and bombesin, Chem. - A Eur. J. 13 (2007) 3842–3852, https://doi.org/10.1002/chem.200700031.

Supplementary Figures

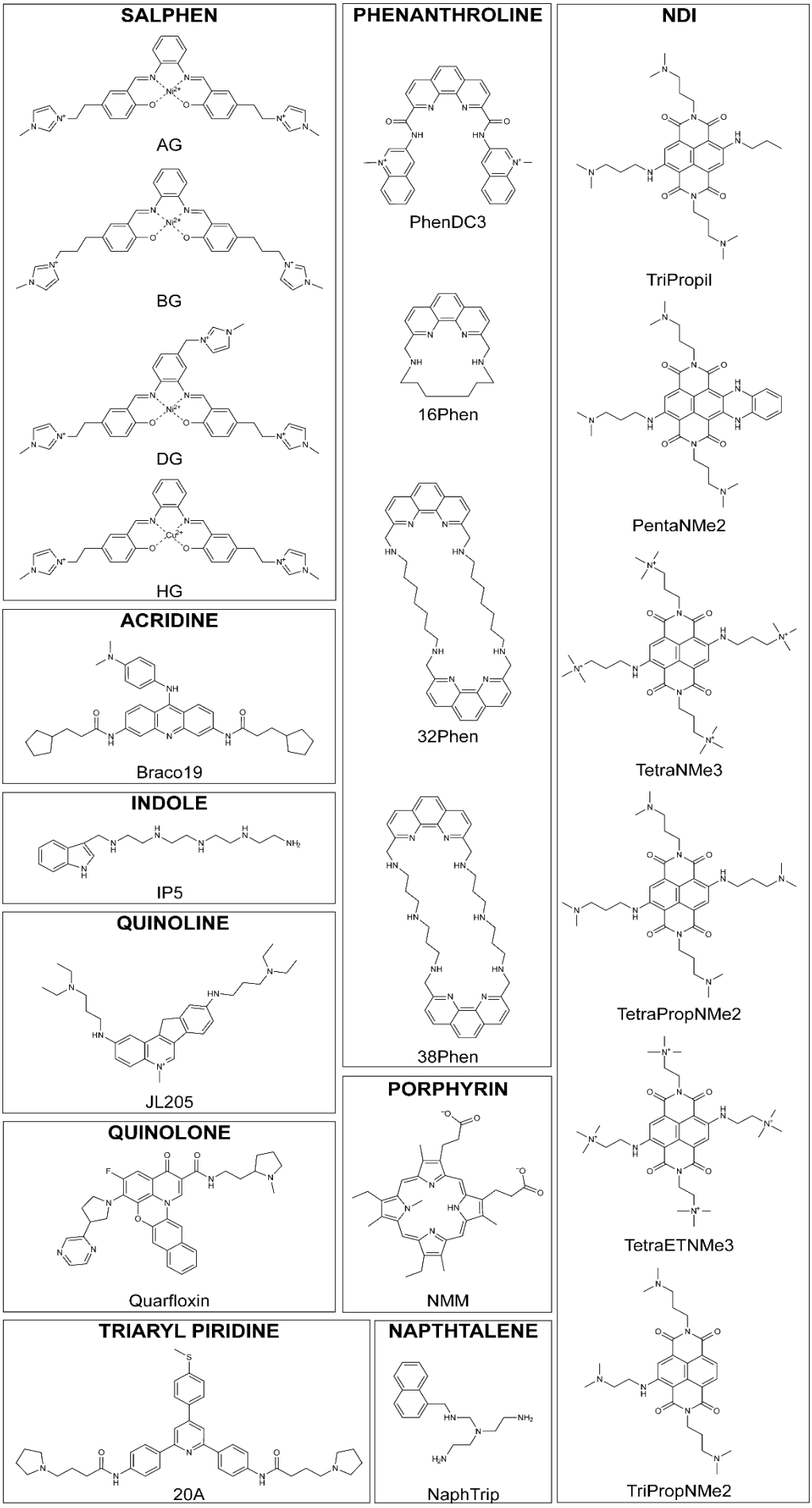


Figure A1. Chemical structures of the ligands used for KRAS G-quadruplexes interaction with their corresponding chemical families

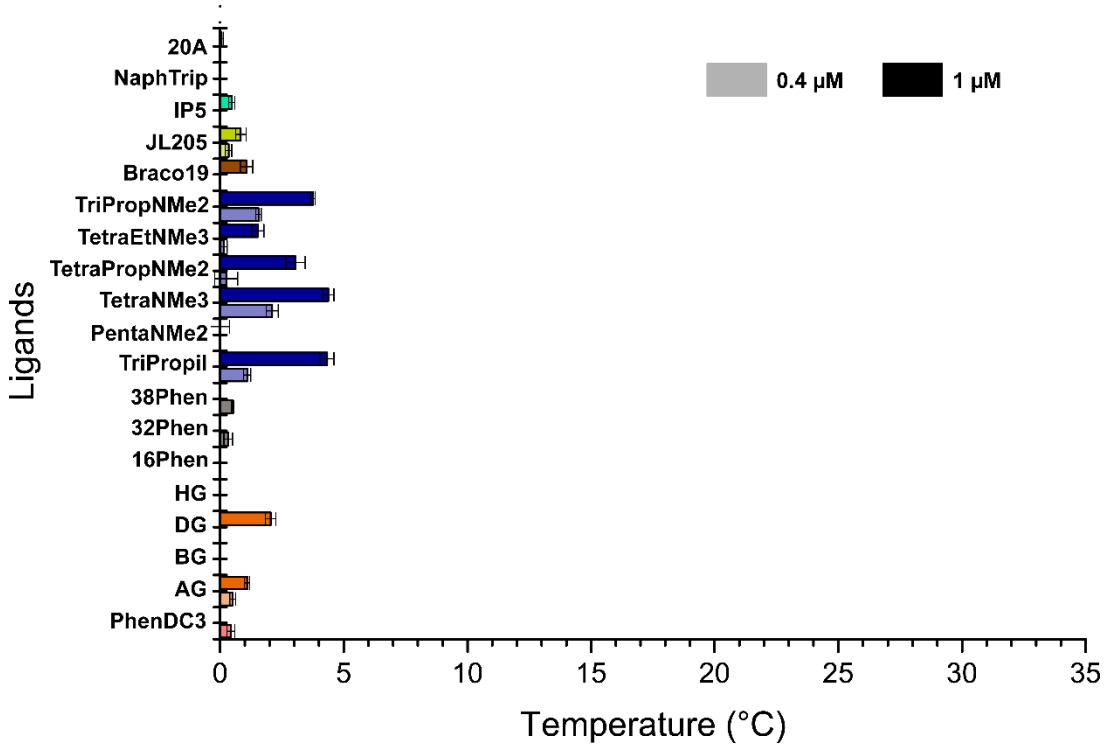


Figure A2. FRET melting experiments with a panel of ligands against DNA duplex. All ligands have been tested at 2 and 5 equivalent of ligands. ΔT_M have been obtained by subtracting T_M of negative control in T_M of each well.

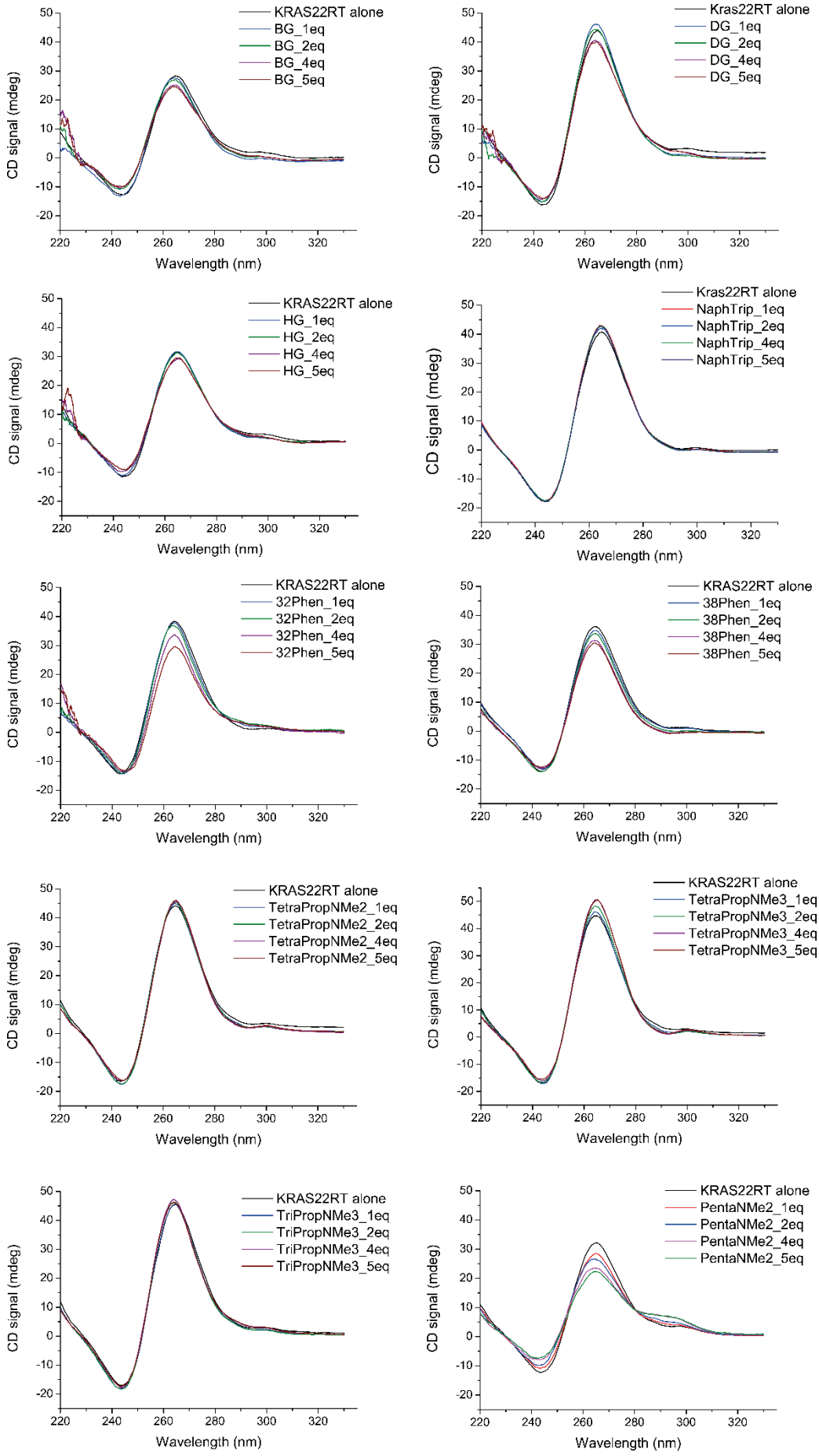


Figure A3. All CD titrations with the tested ligands against KRAS22RT parallel G-quadruplex performed at 37°C

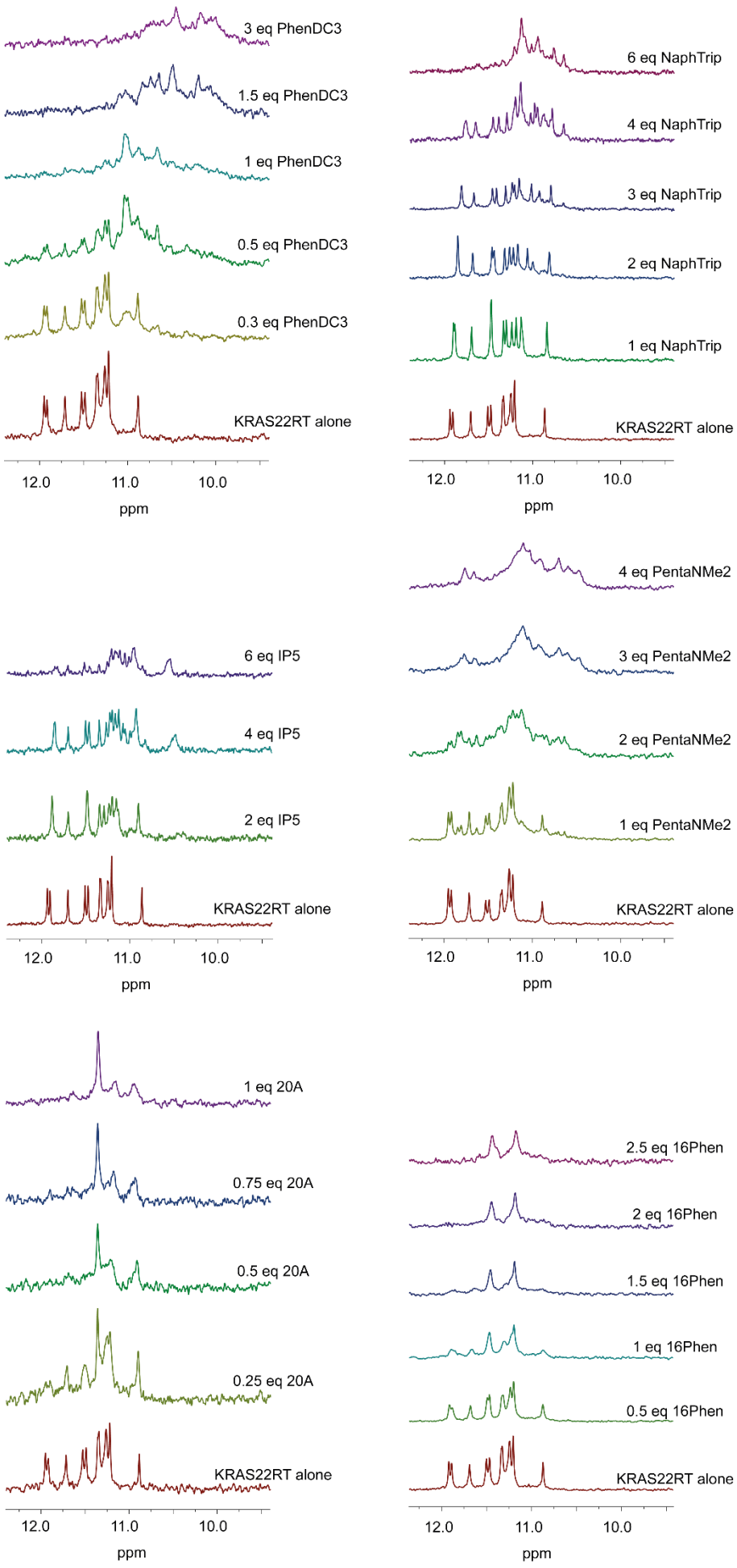


Figure A4. All NMR titrations with the tested ligands with KRAS22RT parallel G-quadruplex performed at 37°C with increased addition of ligands

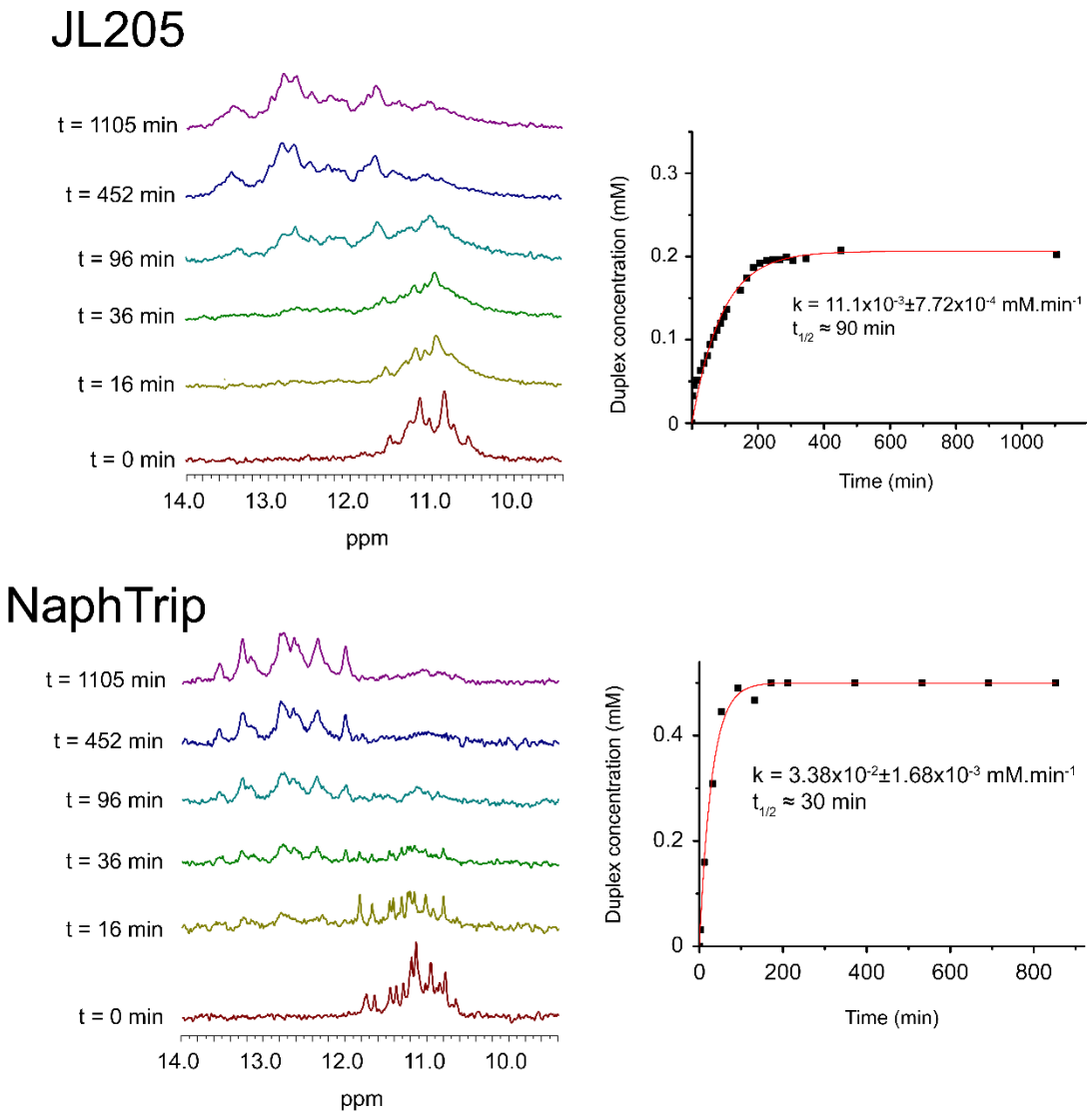


Figure A5. All NMR competitions with the other tested ligands with KRAS22RT parallel G-quadruplex performed at 37°C in function of time

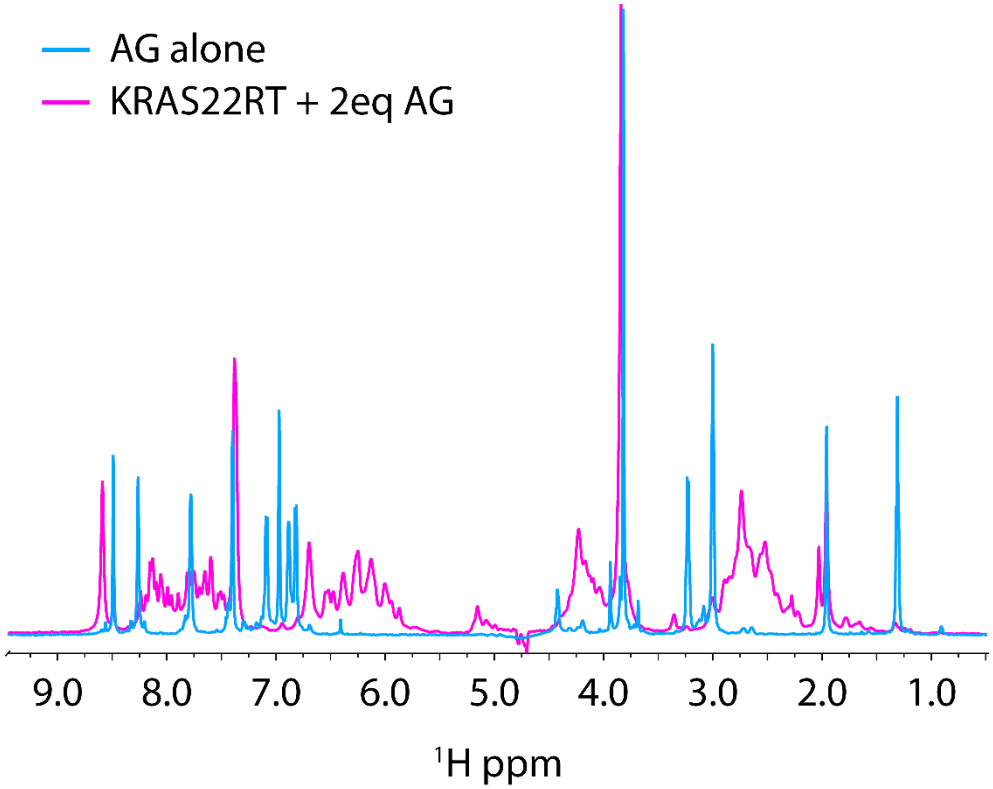


Figure A6. Spectra superimposition of AG alone (blue) and KRAS22RT in presence of 2 equivalents AG (pink) showing that peaks from AG are different when mixed with G-quadruplex

References

1. Dahm, R., *Friedrich Miescher and the discovery of DNA.* Dev Biol, 2005. **278**(2): p. 274-88.
2. Dahm, R., *Discovering DNA: Friedrich Miescher and the early years of nucleic acid research.* Hum Genet, 2008. **122**(6): p. 565-81.
3. Dahm, R., *From discovering to understanding. Friedrich Miescher's attempts to uncover the function of DNA.* EMBO Rep, 2010. **11**(3): p. 153-60.
4. Xie, C. and J.P. O'Leary, *DNA and the brains behind its discovery.* Am Surg, 1996. **62**(11): p. 979-80.
5. Levene, P.A., *THE STRUCTURE OF YEAST NUCLEIC ACID: IV. AMMONIA HYDROLYSIS.* The Journal of Biological Chemistry, 1919. **40**: p. 415-424.
6. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii.* J Exp Med, 1944. **79**(2): p. 137-58.
7. Chargaff, E., *Chemical specificity of nucleic acids and mechanism of their enzymatic degradation.* Experientia, 1950. **6**(6): p. 201-9.
8. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
9. Shakked, Z., et al., *The conformation of the DNA double helix in the crystal is dependent on its environment.* Nature, 1989. **342**(6248): p. 456-60.
10. Hoogsteen, K., *The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine.* acta Crystallogr D Biol Crystallogr, 1959. **12**: p. 822-823.
11. Courtois, Y., P. Fromageot, and W. Guschlbauer, *Protonated polynucleotide structures. 3. An optical rotatory dispersion study of the protonation of DNA.* Eur J Biochem, 1968. **6**(4): p. 493-501.
12. Swadling, J.B., et al., *Origins of biological function in DNA and RNA hairpin loop motifs from replica exchange molecular dynamics simulation.* Phys Chem Chem Phys, 2018. **20**(5): p. 2990-3001.
13. Paulus, M., M. Haslbeck, and M. Watzele, *RNA stem-loop enhanced expression of previously non-expressible genes.* Nucleic Acids Res, 2004. **32**(9): p. e78.
14. Darfeuille, F., et al., *An antisense RNA inhibits translation by competing with standby ribosomes.* Mol Cell, 2007. **26**(3): p. 381-92.
15. Balbontin, R., et al., *Expression of IroN, the salmochelin siderophore receptor, requires mRNA activation by RyhB small RNA homologues.* Mol Microbiol, 2016. **100**(1): p. 139-55.
16. Staple, D.W. and S.E. Butcher, *Pseudoknots: RNA structures with diverse functions.* PLoS Biol, 2005. **3**(6): p. e213.
17. Michiels, P.J., et al., *Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting.* J Mol Biol, 2001. **310**(5): p. 1109-23.
18. Nixon, P.L., et al., *Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot.* J Mol Biol, 2002. **322**(3): p. 621-33.
19. Brunel, C., et al., *RNA loop-loop interactions as dynamic functional motifs.* Biochimie, 2002. **84**(9): p. 925-44.
20. Shen, L.X. and I. Tinoco, Jr., *The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus.* J Mol Biol, 1995. **247**(5): p. 963-78.

21. Paillart, J.C., et al., *A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA.* Proc Natl Acad Sci U S A, 1996. **93**(11): p. 5572-7.
22. Ennifar, E., et al., *Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site.* Nat Struct Biol, 2001. **8**(12): p. 1064-8.
23. Panganiban, A.T. and D. Fiore, *Ordered interstrand and intrastrand DNA transfer during reverse transcription.* Science, 1988. **241**(4869): p. 1064-9.
24. Felsenfeld, G. and A. Rich, *Studies on the formation of two- and three-stranded polyribonucleotides.* Biochim Biophys Acta, 1957. **26**(3): p. 457-68.
25. Dervan, P.B., *Design of sequence-specific DNA-binding molecules.* Science, 1986. **232**(4749): p. 464-71.
26. Reither, S. and A. Jeltsch, *Specificity of DNA triple helix formation analyzed by a FRET assay.* BMC Biochem, 2002. **3**: p. 27.
27. Demidov, V.V. and M.D. Frank-Kamenetskii, *Sequence-specific targeting of duplex DNA by peptide nucleic acids via triplex strand invasion.* Methods, 2001. **23**(2): p. 108-22.
28. Vasquez, K.M. and P.M. Glazer, *Triplex-forming oligonucleotides: principles and applications.* Q Rev Biophys, 2002. **35**(1): p. 89-107.
29. Simon, P., et al., *Targeting DNA with triplex-forming oligonucleotides to modify gene sequence.* Biochimie, 2008. **90**(8): p. 1109-16.
30. Sugimoto, N., et al., *pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes.* Biochemistry, 2001. **40**(31): p. 9396-405.
31. Plum, G.E., et al., *Thermodynamic characterization of the stability and the melting behavior of a DNA triplex: a spectroscopic and calorimetric study.* Proc Natl Acad Sci U S A, 1990. **87**(23): p. 9436-40.
32. Wu, Q., et al., *High-affinity triplex-forming oligonucleotide target sequences in mammalian genomes.* Mol Carcinog, 2007. **46**(1): p. 15-23.
33. Gaddis, S.S., et al., *A web-based search engine for triplex-forming oligonucleotide target sequences.* Oligonucleotides, 2006. **16**(2): p. 196-201.
34. Gehring, K., J.L. Leroy, and M. Gueron, *A tetrameric DNA structure with protonated cytosine.cytosine base pairs.* Nature, 1993. **363**(6429): p. 561-5.
35. Akinrimisi, E.O., C. Sander, and P.O. Ts'O, *Properties of helical polycytidylic acid.* Biochemistry, 1963. **2**: p. 340-4.
36. Langridge, R. and A. Rich, *Molecular structure of helical polycytidylic acid.* Nature, 1963. **198**: p. 725-8.
37. Gueron, M. and J.L. Leroy, *The i-motif in nucleic acids.* Curr Opin Struct Biol, 2000. **10**(3): p. 326-31.
38. Zhou, J., et al., *Formation of i-motif structure at neutral and slightly alkaline pH.* Mol Biosyst, 2010. **6**(3): p. 580-6.
39. Leroy, J.L., et al., *Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif.* Nucleic Acids Res, 1994. **22**(9): p. 1600-6.
40. Mergny, J.L. and L. Lacroix, *Kinetics and thermodynamics of i-DNA formation: phosphodiester versus modified oligodeoxynucleotides.* Nucleic Acids Res, 1998. **26**(21): p. 4797-803.
41. Fojtik, P. and M. Vorlickova, *The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH.* Nucleic Acids Res, 2001. **29**(22): p. 4684-90.
42. Leroy, J.L. and M. Gueron, *Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons.* Structure, 1995. **3**(1): p. 101-20.

43. Nonin, S., A.T. Phan, and J.L. Leroy, *Solution structure and base pair opening kinetics of the i-motif dimer of d(5mCCTTTACC): a noncanonical structure with possible roles in chromosome stability.* Structure, 1997. **5**(9): p. 1231-46.
44. Phan, A.T. and J.L. Leroy, *Intramolecular i-Motif Structures of Telomeric DNA.* J Biomol Struct Dyn, 2000. **17 Suppl 1**: p. 245-51.
45. Collin, D., Gehring, K., *Stability of Chimeric DNA/RNA Cytosine Tetrads: Implications for i-Motif Formation by RNA.* J Am Chem Soc, 1998. **120**: p. 4069-4072.
46. Brooks, T.A., S. Kendrick, and L. Hurley, *Making sense of G-quadruplex and i-motif functions in oncogene promoters.* FEBS J, 2010. **277**(17): p. 3459-69.
47. Catasti, P., et al., *Cystosine-rich strands of the insulin minisatellite adopt hairpins with intercalated cytosine+.cytosine pairs.* J Mol Biol, 1997. **272**(3): p. 369-82.
48. Gallego, J., S.H. Chou, and B.R. Reid, *Centromeric pyrimidine strands fold into an intercalated motif by forming a double hairpin with a novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer.* J Mol Biol, 1997. **273**(4): p. 840-56.
49. Takimoto, M., et al., *Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro.* J Biol Chem, 1993. **268**(24): p. 18249-58.
50. Bang, I., *Untersuchungen über die Guanyksäure.* Biochem Z., 1910. **26**: p. 293-231.
51. Gellert, M., M.N. Lipsett, and D.R. Davies, *Helix formation by guanylic acid.* Proc Natl Acad Sci U S A, 1962. **48**: p. 2013-8.
52. Zimmerman, S.B., G.H. Cohen, and D.R. Davies, *X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid.* J Mol Biol, 1975. **92**(2): p. 181-92.
53. Howard, F.B., J. Frazier, and H.T. Miles, *Stable and metastable forms of poly(G).* Biopolymers, 1977. **16**(4): p. 791-809.
54. Cang, X., J. Sponer, and T.E. Cheatham, 3rd, *Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes.* Nucleic Acids Res, 2011. **39**(10): p. 4499-512.
55. Bochman, M.L., K. Paeschke, and V.A. Zakian, *DNA secondary structures: stability and function of G-quadruplex structures.* Nat Rev Genet, 2012. **13**(11): p. 770-80.
56. Cevec, M. and J. Plavec, *Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes.* Biochemistry, 2005. **44**(46): p. 15238-46.
57. Risitano, A. and K.R. Fox, *Influence of loop size on the stability of intramolecular DNA quadruplexes.* Nucleic Acids Res, 2004. **32**(8): p. 2598-606.
58. Guedin, A., et al., *How long is too long? Effects of loop size on G-quadruplex stability.* Nucleic Acids Res, 2010. **38**(21): p. 7858-68.
59. Karsisiotis, A.I. and M. Webba da Silva, *Structural probes in quadruplex nucleic acid structure determination by NMR.* Molecules, 2012. **17**(11): p. 13073-86.
60. Smargiasso, N., et al., *G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation.* J Am Chem Soc, 2008. **130**(31): p. 10208-16.
61. Li, T., E. Wang, and S. Dong, *Potassium-lead-switched G-quadruplexes: a new class of DNA logic gates.* J Am Chem Soc, 2009. **131**(42): p. 15082-3.
62. Dai, J., et al., *Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence.* Nucleic Acids Res, 2007. **35**(15): p. 4927-40.
63. Chen, Y. and D. Yang, *Sequence, stability, and structure of G-quadruplexes and their interactions with drugs.* Curr Protoc Nucleic Acid Chem, 2012. **Chapter 17**: p. Unit17 5.

64. Singh, A. and S. Kukreti, *A triple stranded G-quadruplex formation in the promoter region of human myosin beta(Myh7) gene.* J Biomol Struct Dyn, 2018. **36**(11): p. 2773-2786.
65. Hansel, R., et al., *The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions.* Nucleic Acids Res, 2011. **39**(13): p. 5768-75.
66. Yang, C., et al., *Visualization of Parallel G-Quadruplexes in Cells with a Series of New Developed Bis(4-aminobenzylidene)acetone Derivatives.* ACS Omega, 2018. **3**(9): p. 10487-10492.
67. Geng, Y., et al., *The crystal structure of an antiparallel chair-type G-quadruplex formed by Bromo-substituted human telomeric DNA.* Nucleic Acids Res, 2019. **47**(10): p. 5395-5404.
68. Lim, K.W., et al., *Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers.* J Am Chem Soc, 2009. **131**(12): p. 4301-9.
69. Sengar, A., et al., *Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter.* Nucleic Acids Res, 2019. **47**(3): p. 1564-1572.
70. Cang, X., J. Sponer, and T.E. Cheatham, 3rd, *Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis.* J Am Chem Soc, 2011. **133**(36): p. 14270-9.
71. Havrila, M., et al., *Structural dynamics of propeller loop: towards folding of RNA G-quadruplex.* Nucleic Acids Res, 2018. **46**(17): p. 8754-8771.
72. Peng, P., et al., *Probing the propeller-like loops of DNA G-quadruplexes with looped-out 2-aminopurine for label-free switchable molecular sensing.* Analyst, 2018. **143**(16): p. 3814-3820.
73. Fadrna, E., et al., *Molecular dynamics simulations of Guanine quadruplex loops: advances and force field limitations.* Biophys J, 2004. **87**(1): p. 227-42.
74. Lech, C.J., B. Heddi, and A.T. Phan, *Guanine base stacking in G-quadruplex nucleic acids.* Nucleic Acids Res, 2013. **41**(3): p. 2034-46.
75. Lavery, R., et al., *A comprehensive classification of nucleic acid structural families based on strand direction and base pairing.* Nucleic Acids Res, 1992. **20**(19): p. 5011-6.
76. Kuryavyi, V., et al., *A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G(syn)-G(syn)-G(anti)-G(anti) tetrads flanked by a G-(T-T) Triad and a T-T-T triple.* J Mol Biol, 2001. **310**(1): p. 181-94.
77. Pan, B., et al., *Crystal structure of a bulged RNA tetraplex at 1.1 a resolution: implications for a novel binding site in RNA tetraplex.* Structure, 2003. **11**(11): p. 1423-30.
78. Mukundan, V.T. and A.T. Phan, *Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences.* J Am Chem Soc, 2013. **135**(13): p. 5017-28.
79. Yang, C., E. Kim, and Y. Pak, *Free energy landscape and transition pathways from Watson-Crick to Hoogsteen base pairing in free duplex DNA.* Nucleic Acids Res, 2015. **43**(16): p. 7769-78.
80. Malgowska, M., et al., *Overview of the RNA G-quadruplex structures.* Acta Biochim Pol, 2016. **63**(4): p. 609-621.
81. Ruggiero, E. and S.N. Richter, *G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy.* Nucleic Acids Res, 2018. **46**(7): p. 3270-3283.

82. Wlodarczyk, A., et al., *Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies.* J Phys Chem B, 2005. **109**(8): p. 3594-605.
83. Bhattacharyya, D., G. Mirihana Arachchilage, and S. Basu, *Metal Cations in G-Quadruplex Folding and Stability.* Front Chem, 2016. **4**: p. 38.
84. Hardin, C.C., et al., *Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG).* Biochemistry, 1992. **31**(3): p. 833-41.
85. Venczel, E.A. and D. Sen, *Parallel and antiparallel G-DNA structures from a complex telomeric sequence.* Biochemistry, 1993. **32**(24): p. 6220-8.
86. Nagesh, N. and D. Chatterji, *Ammonium ion at low concentration stabilizes the G-quadruplex formation by telomeric sequence.* J Biochem Biophys Methods, 1995. **30**(1): p. 1-8.
87. Miyoshi, D., et al., *Effect of divalent cations on antiparallel G-quartet structure of d(G4T4G4).* FEBS Lett, 2001. **496**(2-3): p. 128-33.
88. Zaccaria, F., G. Paragi, and C. Fonseca Guerra, *The role of alkali metal cations in the stabilization of guanine quadruplexes: why K(+) is the best.* Phys Chem Chem Phys, 2016. **18**(31): p. 20895-904.
89. Wang, Y. and D.J. Patel, *Solution structure of a parallel-stranded G-quadruplex DNA.* J Mol Biol, 1993. **234**(4): p. 1171-83.
90. Gray, R.D. and J.B. Chaires, *Kinetics and mechanism of K+- and Na+-induced folding of models of human telomeric DNA into G-quadruplex structures.* Nucleic Acids Res, 2008. **36**(12): p. 4191-203.
91. Bose, K., et al., *High-resolution AFM structure of DNA G-wires in aqueous solution.* Nat Commun, 2018. **9**(1): p. 1959.
92. Marsh, T.C. and E. Henderson, *G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures.* Biochemistry, 1994. **33**(35): p. 10718-24.
93. Kolesnikova, S., et al., *GTP-Dependent Formation of Multimeric G-Quadruplexes.* ACS Chem Biol, 2019. **14**(9): p. 1951-1963.
94. Kolesnikova, S. and E.A. Curtis, *Structure and Function of Multimeric G-Quadruplexes.* Molecules, 2019. **24**(17).
95. Chaires, J.B., et al., *An improved model for the hTERT promoter quadruplex.* PLoS One, 2014. **9**(12): p. e115580.
96. Do, N.Q., et al., *Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity.* Nucleic Acids Res, 2011. **39**(21): p. 9448-57.
97. Kato, Y., et al., *Dynamics and thermodynamics of dimerization of parallel G-quadruplexed DNA through pi-pi stacking interaction.* Nucleic Acids Symp Ser (Oxf), 2005(49): p. 247-8.
98. Haider, S., G.N. Parkinson, and S. Neidle, *Molecular dynamics and principal components analysis of human telomeric quadruplex multimers.* Biophys J, 2008. **95**(1): p. 296-311.
99. Lane, A.N., et al., *Stability and kinetics of G-quadruplex structures.* Nucleic Acids Res, 2008. **36**(17): p. 5482-515.
100. Stegle, O., et al., *Predicting and understanding the stability of G-quadruplexes.* Bioinformatics, 2009. **25**(12): p. i374-82.
101. Todd, A.K., M. Johnston, and S. Neidle, *Highly prevalent putative quadruplex sequence motifs in human DNA.* Nucleic Acids Res, 2005. **33**(9): p. 2901-7.

102. Huppert, J.L. and S. Balasubramanian, *Prevalence of quadruplexes in the human genome.* Nucleic Acids Res, 2005. **33**(9): p. 2908-16.
103. Huppert, J.L., *Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes.* Chem Soc Rev, 2008. **37**(7): p. 1375-84.
104. Huppert, J.L., *Structure, location and interactions of G-quadruplexes.* FEBS J, 2010. **277**(17): p. 3452-8.
105. Maizels, N., *G4 motifs in human genes.* Ann N Y Acad Sci, 2012. **1267**: p. 53-60.
106. Williamson, J.R., M.K. Raghuraman, and T.R. Cech, *Monovalent cation-induced structure of telomeric DNA: the G-quartet model.* Cell, 1989. **59**(5): p. 871-80.
107. Huppert, J.L., *Hunting G-quadruplexes.* Biochimie, 2008. **90**(8): p. 1140-8.
108. Huppert, J.L., et al., *G-quadruplexes: the beginning and end of UTRs.* Nucleic Acids Res, 2008. **36**(19): p. 6260-8.
109. Capra, J.A., et al., *G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae.* PLoS Comput Biol, 2010. **6**(7): p. e1000861.
110. Hershman, S.G., et al., *Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae.* Nucleic Acids Res, 2008. **36**(1): p. 144-56.
111. Blackburn, E.H., et al., *Molecular manifestations and molecular determinants of telomere capping.* Cold Spring Harb Symp Quant Biol, 2000. **65**: p. 253-63.
112. Zakian, V.A., *Telomeres: the beginnings and ends of eukaryotic chromosomes.* Exp Cell Res, 2012. **318**(12): p. 1456-60.
113. Blackburn, E.H., *Telomere states and cell fates.* Nature, 2000. **408**(6808): p. 53-6.
114. Hackett, J.A., D.M. Feldser, and C.W. Greider, *Telomere dysfunction increases mutation rate and genomic instability.* Cell, 2001. **106**(3): p. 275-86.
115. Wright, W.E., et al., *Normal human chromosomes have long G-rich telomeric overhangs at one end.* Genes Dev, 1997. **11**(21): p. 2801-9.
116. Henderson, E., et al., *Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs.* Cell, 1987. **51**(6): p. 899-908.
117. Harley, C.B., *Telomere loss: mitotic clock or genetic time bomb?* Mutat Res, 1991. **256**(2-6): p. 271-82.
118. Harley, C.B. and B. Villeponteau, *Telomeres and telomerase in aging and cancer.* Curr Opin Genet Dev, 1995. **5**(2): p. 249-55.
119. Harley, C.B., A.B. Futcher, and C.W. Greider, *Telomeres shorten during ageing of human fibroblasts.* Nature, 1990. **345**(6274): p. 458-60.
120. Zvereva, M.I., D.M. Shcherbakova, and O.A. Dontsova, *Telomerase: structure, functions, and activity regulation.* Biochemistry (Mosc), 2010. **75**(13): p. 1563-83.
121. Shay, J.W. and W.E. Wright, *Role of telomeres and telomerase in cancer.* Semin Cancer Biol, 2011. **21**(6): p. 349-53.
122. 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, 2011. **39**(14): p. 6229-37.
123. Oganesian, L., et al., *Telomerase recognizes G-quadruplex and linear DNA as distinct substrates.* Biochemistry, 2007. **46**(40): p. 11279-90.
124. Zaug, A.J., E.R. Podell, and T.R. Cech, *Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro.* Proc Natl Acad Sci U S A, 2005. **102**(31): p. 10864-9.

125. He, Q., et al., *G-quadruplex-mediated regulation of telomere binding protein POT1 gene expression.* Biochim Biophys Acta, 2014. **1840**(7): p. 2222-33.
126. Parkinson, G.N., M.P. Lee, and S. Neidle, *Crystal structure of parallel quadruplexes from human telomeric DNA.* Nature, 2002. **417**(6891): p. 876-80.
127. Ambrus, A., et al., *Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution.* Nucleic Acids Res, 2006. **34**(9): p. 2723-35.
128. Dai, J., et al., *Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation.* Nucleic Acids Res, 2007. **35**(7): p. 2440-50.
129. Huppert, J.L. and S. Balasubramanian, *G-quadruplexes in promoters throughout the human genome.* Nucleic Acids Res, 2007. **35**(2): p. 406-13.
130. Yadav, V.K., et al., *QuadBase: genome-wide database of G4 DNA--occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes.* Nucleic Acids Res, 2008. **36**(Database issue): p. D381-5.
131. Mullen, M.A., et al., *RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles.* Nucleic Acids Res, 2010. **38**(22): p. 8149-63.
132. Sun, D. and L.H. Hurley, *The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression.* J Med Chem, 2009. **52**(9): p. 2863-74.
133. Kouzine, F., et al., *The functional response of upstream DNA to dynamic supercoiling in vivo.* Nat Struct Mol Biol, 2008. **15**(2): p. 146-54.
134. Qin, Y. and L.H. Hurley, *Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions.* Biochimie, 2008. **90**(8): p. 1149-71.
135. Paeschke, K., et al., *Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo.* Nat Struct Mol Biol, 2005. **12**(10): p. 847-54.
136. Giraldo, R. and D. Rhodes, *The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA.* EMBO J, 1994. **13**(10): p. 2411-20.
137. David, A.P., et al., *CNBP controls transcription by unfolding DNA G-quadruplex structures.* Nucleic Acids Res, 2019. **47**(15): p. 7901-7913.
138. Niu, K., et al., *Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates.* Nucleic Acids Res, 2019. **47**(14): p. 7306-7320.
139. Brooks, T.A. and L.H. Hurley, *Targeting MYC Expression through G-Quadruplexes.* Genes Cancer, 2010. **1**(6): p. 641-649.
140. Dang, C.V., *MYC on the path to cancer.* Cell, 2012. **149**(1): p. 22-35.
141. Hsieh, A.L., et al., *MYC and metabolism on the path to cancer.* Semin Cell Dev Biol, 2015. **43**: p. 11-21.
142. Ambrus, A., et al., *Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization.* Biochemistry, 2005. **44**(6): p. 2048-58.
143. Mathad, R.I., et al., *c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability.* Nucleic Acids Res, 2011. **39**(20): p. 9023-33.

144. Dickerhoff, J., et al., *Solution Structure of a MYC Promoter G-Quadruplex with 1:6:1 Loop Length.* ACS Omega, 2019. **4**(2): p. 2533-2539.
145. Siddiqui-Jain, A., et al., *Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11593-8.
146. De Armond, R., et al., *Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter.* Biochemistry, 2005. **44**(49): p. 16341-50.
147. Agrawal, P., et al., *Solution structure of the major G-quadruplex formed in the human VEGF promoter in K+: insights into loop interactions of the parallel G-quadruplexes.* Nucleic Acids Res, 2013. **41**(22): p. 10584-92.
148. Dai, J., et al., *NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region.* Nucleic Acids Res, 2006. **34**(18): p. 5133-44.
149. Dixon, S.J. and B.R. Stockwell, *Identifying druggable disease-modifying gene products.* Curr Opin Chem Biol, 2009. **13**(5-6): p. 549-55.
150. Fauman, E.B., B.K. Rai, and E.S. Huang, *Structure-based druggability assessment--identifying suitable targets for small molecule therapeutics.* Curr Opin Chem Biol, 2011. **15**(4): p. 463-8.
151. Lin, C., Y. Liu, and H. Yan, *Designer DNA nanoarchitectures.* Biochemistry, 2009. **48**(8): p. 1663-74.
152. Krishnan, Y. and F.C. Simmel, *Nucleic acid based molecular devices.* Angew Chem Int Ed Engl, 2011. **50**(14): p. 3124-56.
153. Lieberman, J., *Tapping the RNA world for therapeutics.* Nat Struct Mol Biol, 2018. **25**(5): p. 357-364.
154. Kole, R., A.R. Krainer, and S. Altman, *RNA therapeutics: beyond RNA interference and antisense oligonucleotides.* Nat Rev Drug Discov, 2012. **11**(2): p. 125-40.
155. Goodchild, J., *Therapeutic oligonucleotides.* Methods Mol Biol, 2011. **764**: p. 1-15.
156. Crooke, S.T., *Molecular Mechanisms of Antisense Oligonucleotides.* Nucleic Acid Ther, 2017. **27**(2): p. 70-77.
157. Crooke, S.T., *Vitravene--another piece in the mosaic.* Antisense Nucleic Acid Drug Dev, 1998. **8**(4): p. vii-viii.
158. Raal, F.J., et al., *Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial.* Lancet, 2010. **375**(9719): p. 998-1006.
159. Santos, R.D., et al., *Mipomersen, an antisense oligonucleotide to apolipoprotein B-100, reduces lipoprotein(a) in various populations with hypercholesterolemia: results of 4 phase III trials.* Arterioscler Thromb Vasc Biol, 2015. **35**(3): p. 689-99.
160. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.
161. de Fougerolles, A., et al., *Interfering with disease: a progress report on siRNA-based therapeutics.* Nat Rev Drug Discov, 2007. **6**(6): p. 443-53.
162. Elbashir, S.M., et al., *Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.* Nature, 2001. **411**(6836): p. 494-8.
163. Silva, J.M., et al., *Profiling essential genes in human mammary cells by multiplex RNAi screening.* Science, 2008. **319**(5863): p. 617-20.

164. Novina, C.D., et al., *siRNA-directed inhibition of HIV-1 infection.* Nat Med, 2002. **8**(7): p. 681-6.
165. Song, E., et al., *RNA interference targeting Fas protects mice from fulminant hepatitis.* Nat Med, 2003. **9**(3): p. 347-51.
166. Nair, J.K., et al., *Multivalent N-acetylgalactosamine-conjugated siRNA localizes in hepatocytes and elicits robust RNAi-mediated gene silencing.* J Am Chem Soc, 2014. **136**(49): p. 16958-61.
167. Matsuda, S., et al., *siRNA conjugates carrying sequentially assembled trivalent N-acetylgalactosamine linked through nucleosides elicit robust gene silencing in vivo in hepatocytes.* ACS Chem Biol, 2015. **10**(5): p. 1181-7.
168. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
169. Marraffini, L.A., *CRISPR-Cas immunity in prokaryotes.* Nature, 2015. **526**(7571): p. 55-61.
170. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* J Bacteriol, 1987. **169**(12): p. 5429-33.
171. Jiang, W. and L.A. Marraffini, *CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems.* Annu Rev Microbiol, 2015. **69**: p. 209-28.
172. Mirza, Z. and S. Karim, *Advancements in CRISPR/Cas9 technology-Focusing on cancer therapeutics and beyond.* Semin Cell Dev Biol, 2019.
173. Kaushik, I., S. Ramachandran, and S.K. Srivastava, *CRISPR-Cas9: A multifaceted therapeutic strategy for cancer treatment.* Semin Cell Dev Biol, 2019.
174. Jain, A., G. Wang, and K.M. Vasquez, *DNA triple helices: biological consequences and therapeutic potential.* Biochimie, 2008. **90**(8): p. 1117-30.
175. Rao, T.S., et al., *Incorporation of 2'-deoxy-6-thioguanosine into G-rich oligodeoxyribonucleotides inhibits G-tetrad formation and facilitates triplex formation.* Biochemistry, 1995. **34**(3): p. 765-72.
176. Cooney, M., et al., *Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro.* Science, 1988. **241**(4864): p. 456-9.
177. Christensen, L.A., et al., *Targeting oncogenes to improve breast cancer chemotherapy.* Cancer Res, 2006. **66**(8): p. 4089-94.
178. Vasquez, K.M., et al., *High-efficiency triple-helix-mediated photo-cross-linking at a targeted site within a selectable mammalian gene.* Biochemistry, 1996. **35**(33): p. 10712-9.
179. Perkins, B.D., et al., *Triplex targets in the human rhodopsin gene.* Biochemistry, 1998. **37**(32): p. 11315-22.
180. Cheng, A.J. and M.W. Van Dyke, *Monovalent cation effects on intermolecular purine-purine-pyrimidine triple-helix formation.* Nucleic Acids Res, 1993. **21**(24): p. 5630-5.
181. Olivas, W.M. and L.J. Maher, 3rd, *Overcoming potassium-mediated triplex inhibition.* Nucleic Acids Res, 1995. **23**(11): p. 1936-41.
182. Balasubramanian, S., L.H. Hurley, and S. Neidle, *Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?* Nat Rev Drug Discov, 2011. **10**(4): p. 261-75.
183. Nakajima, A., et al., *Telomerase inhibition enhances apoptosis in human acute leukemia cells: possibility of antitelomerase therapy.* Leukemia, 2003. **17**(3): p. 560-7.
184. Ivancich, M., et al., *Treating Cancer by Targeting Telomeres and Telomerase.* Antioxidants (Basel), 2017. **6**(1).

185. D'Urso, A., et al., *Stabilization vs. destabilization of G-quadruplex superstructures: the role of the porphyrin derivative having spermine arms.* Phys Chem Chem Phys, 2017. **19**(26): p. 17404-17410.
186. Mendoza, O., et al., *G-quadruplexes and helicases.* Nucleic Acids Res, 2016. **44**(5): p. 1989-2006.
187. Arola, A. and R. Vilar, *Stabilisation of G-quadruplex DNA by small molecules.* Curr Top Med Chem, 2008. **8**(15): p. 1405-15.
188. Yan, Y., et al., *DNA G-quadruplex binders: a patent review.* Expert Opin Ther Pat, 2013. **23**(11): p. 1495-509.
189. Shan, C., et al., *Chemical intervention of the NM23-H2 transcriptional programme on c-MYC via a novel small molecule.* Nucleic Acids Res, 2015. **43**(14): p. 6677-91.
190. Shan, C., et al., *Design, Synthesis, and Evaluation of Isaindigotone Derivatives To Downregulate c-myc Transcription via Disrupting the Interaction of NM23-H2 with G-Quadruplex.* J Med Chem, 2017. **60**(4): p. 1292-1308.
191. Steeg, P.S., et al., *Evidence for a novel gene associated with low tumor metastatic potential.* J Natl Cancer Inst, 1988. **80**(3): p. 200-4.
192. Wang, Y.Q., et al., *Design, Synthesis, and Evaluation of New Selective NM23-H2 Binders as c-MYC Transcription Inhibitors via Disruption of the NM23-H2/G-Quadruplex Interaction.* J Med Chem, 2017. **60**(16): p. 6924-6941.
193. Machireddy, B., et al., *Probing the Binding Pathway of BRACO19 to a Parallel-Stranded Human Telomeric G-Quadruplex Using Molecular Dynamics Binding Simulation with AMBER DNA OL15 and Ligand GAFF2 Force Fields.* J Chem Inf Model, 2017. **57**(11): p. 2846-2864.
194. Gowan, S.M., et al., *A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity.* Mol Pharmacol, 2002. **61**(5): p. 1154-62.
195. Burger, A.M., et al., *The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function.* Cancer Res, 2005. **65**(4): p. 1489-96.
196. Gunaratnam, M., et al., *Mechanism of acridine-based telomerase inhibition and telomere shortening.* Biochem Pharmacol, 2007. **74**(5): p. 679-89.
197. Perrone, R., et al., *Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19.* J Antimicrob Chemother, 2014. **69**(12): p. 3248-58.
198. Piekna-Przybylska, D. and S.B. Maggirwar, *CD4+ memory T cells infected with latent HIV-1 are susceptible to drugs targeting telomeres.* Cell Cycle, 2018. **17**(17): p. 2187-2203.
199. Taetz, S., et al., *Biopharmaceutical characterization of the telomerase inhibitor BRACO19.* Pharm Res, 2006. **23**(5): p. 1031-7.
200. Izbicka, E., et al., *Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells.* Cancer Res, 1999. **59**(3): p. 639-44.
201. Freyer, M.W., et al., *Biophysical studies of the c-MYC NHE III1 promoter: model quadruplex interactions with a cationic porphyrin.* Biophys J, 2007. **92**(6): p. 2007-15.
202. Mikami-Terao, Y., et al., *Antitumor activity of TMPyP4 interacting G-quadruplex in retinoblastoma cell lines.* Exp Eye Res, 2009. **89**(2): p. 200-8.
203. Mikami-Terao, Y., et al., *Antitumor activity of G-quadruplex-interactive agent TMPyP4 in K562 leukemic cells.* Cancer Lett, 2008. **261**(2): p. 226-34.

204. Rodriguez, R., et al., *A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres.* J Am Chem Soc, 2008. **130**(47): p. 15758-9.
205. Zimmer, J., et al., *Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds.* Mol Cell, 2016. **61**(3): p. 449-460.
206. Drygin, D., et al., *Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis.* Cancer Res, 2009. **69**(19): p. 7653-61.
207. Xu, H., et al., *CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours.* Nat Commun, 2017. **8**: p. 14432.
208. Wennerberg, K., K.L. Rossman, and C.J. Der, *The Ras superfamily at a glance.* J Cell Sci, 2005. **118**(Pt 5): p. 843-6.
209. Bourne, H.R., D.A. Sanders, and F. McCormick, *The GTPase superfamily: a conserved switch for diverse cell functions.* Nature, 1990. **348**(6297): p. 125-32.
210. Lowy, D.R. and B.M. Willumsen, *Function and regulation of ras.* Annu Rev Biochem, 1993. **62**: p. 851-91.
211. Fiorucci, G. and A. Hall, *All three human ras genes are expressed in a wide range of tissues.* Biochim Biophys Acta, 1988. **950**(1): p. 81-3.
212. Barbacid, M., *ras genes.* Annu Rev Biochem, 1987. **56**: p. 779-827.
213. Jancik, S., et al., *Clinical relevance of KRAS in human cancers.* J Biomed Biotechnol, 2010. **2010**: p. 150960.
214. Adjei, A.A., *Ras signaling pathway proteins as therapeutic targets.* Curr Pharm Des, 2001. **7**(16): p. 1581-94.
215. Pylayeva-Gupta, Y., E. Grabocka, and D. Bar-Sagi, *RAS oncogenes: weaving a tumorigenic web.* Nat Rev Cancer, 2011. **11**(11): p. 761-74.
216. Hancock, J.F., *Ras proteins: different signals from different locations.* Nat Rev Mol Cell Biol, 2003. **4**(5): p. 373-84.
217. Chavan, T.S., et al., *High-Affinity Interaction of the K-Ras4B Hypervariable Region with the Ras Active Site.* Biophys J, 2015. **109**(12): p. 2602-2613.
218. Ahearn, I.M., et al., *Regulating the regulator: post-translational modification of RAS.* Nat Rev Mol Cell Biol, 2011. **13**(1): p. 39-51.
219. Reuther, G.W. and C.J. Der, *The Ras branch of small GTPases: Ras family members don't fall far from the tree.* Curr Opin Cell Biol, 2000. **12**(2): p. 157-65.
220. Bos, J.L., H. Rehmann, and A. Wittinghofer, *GEFs and GAPs: critical elements in the control of small G proteins.* Cell, 2007. **129**(5): p. 865-77.
221. Buday, L. and J. Downward, *Epidermal growth factor regulates the exchange rate of guanine nucleotides on p21ras in fibroblasts.* Mol Cell Biol, 1993. **13**(3): p. 1903-10.
222. Buday, L. and J. Downward, *Epidermal growth factor regulates p21ras through the formation of a complex of receptor, Grb2 adapter protein, and Sos nucleotide exchange factor.* Cell, 1993. **73**(3): p. 611-20.
223. Cox, A.D. and C.J. Der, *The dark side of Ras: regulation of apoptosis.* Oncogene, 2003. **22**(56): p. 8999-9006.
224. Lavoie, H. and M. Therrien, *Regulation of RAF protein kinases in ERK signalling.* Nat Rev Mol Cell Biol, 2015. **16**(5): p. 281-98.
225. Lavoie, H., et al., *Inhibitors that stabilize a closed RAF kinase domain conformation induce dimerization.* Nat Chem Biol, 2013. **9**(7): p. 428-36.
226. Wang, H.L., et al., *KRAS mutation testing in human cancers: The pathologist's role in the era of personalized medicine.* Adv Anat Pathol, 2010. **17**(1): p. 23-32.

227. Riely, G.J., et al., *Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma.* Clin Cancer Res, 2008. **14**(18): p. 5731-4.
228. Perkins, G., et al., *Beyond KRAS status and response to anti-EGFR therapy in metastatic colorectal cancer.* Pharmacogenomics, 2014. **15**(7): p. 1043-52.
229. Tyner, J.W., et al., *High-throughput sequencing screen reveals novel, transforming RAS mutations in myeloid leukemia patients.* Blood, 2009. **113**(8): p. 1749-55.
230. Keohavong, P., et al., *Detection of K-ras mutations in lung carcinomas: relationship to prognosis.* Clin Cancer Res, 1996. **2**(2): p. 411-8.
231. Young, A., D. Lou, and F. McCormick, *Oncogenic and wild-type Ras play divergent roles in the regulation of mitogen-activated protein kinase signaling.* Cancer Discov, 2013. **3**(1): p. 112-23.
232. di Magliano, M.P. and C.D. Logsdon, *Roles for KRAS in pancreatic tumor development and progression.* Gastroenterology, 2013. **144**(6): p. 1220-9.
233. Viale, P.H., A. Fung, and L. Zitella, *Advanced colorectal cancer: current treatment and nursing management with economic considerations.* Clin J Oncol Nurs, 2005. **9**(5): p. 541-52.
234. Venook, A., *Critical evaluation of current treatments in metastatic colorectal cancer.* Oncologist, 2005. **10**(4): p. 250-61.
235. Ostrem, J.M. and K.M. Shokat, *Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design.* Nat Rev Drug Discov, 2016. **15**(11): p. 771-785.
236. Patricelli, M.P., et al., *Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State.* Cancer Discov, 2016. **6**(3): p. 316-29.
237. Janes, M.R., et al., *Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor.* Cell, 2018. **172**(3): p. 578-589 e17.
238. Sakamoto, K., et al., *K-Ras(G12D)-selective inhibitory peptides generated by random peptide T7 phage display technology.* Biochem Biophys Res Commun, 2017. **484**(3): p. 605-611.
239. Haluska, P., G.K. Dy, and A.A. Adjei, *Farnesyl transferase inhibitors as anticancer agents.* Eur J Cancer, 2002. **38**(13): p. 1685-700.
240. Cox, A.D. and C.J. Der, *Farnesyltransferase inhibitors and cancer treatment: targeting simply Ras?* Biochim Biophys Acta, 1997. **1333**(1): p. F51-71.
241. Berndt, N., A.D. Hamilton, and S.M. Sebti, *Targeting protein prenylation for cancer therapy.* Nat Rev Cancer, 2011. **11**(11): p. 775-91.
242. Liu, M., et al., *Targeting the protein prenyltransferases efficiently reduces tumor development in mice with K-RAS-induced lung cancer.* Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6471-6.
243. Friday, B.B. and A.A. Adjei, *K-ras as a target for cancer therapy.* Biochim Biophys Acta, 2005. **1756**(2): p. 127-44.
244. Chandra, A., et al., *The GDI-like solubilizing factor PDEdelta sustains the spatial organization and signalling of Ras family proteins.* Nat Cell Biol, 2011. **14**(2): p. 148-58.
245. Zimmermann, G., et al., *Small molecule inhibition of the KRAS-PDEdelta interaction impairs oncogenic KRAS signalling.* Nature, 2013. **497**(7451): p. 638-42.
246. Papke, B., et al., *Identification of pyrazolopyridazinones as PDEdelta inhibitors.* Nat Commun, 2016. **7**: p. 11360.
247. Yuneva, M.O., et al., *The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type.* Cell Metab, 2012. **15**(2): p. 157-70.

248. Bryant, K.L., et al., *KRAS: feeding pancreatic cancer proliferation.* Trends Biochem Sci, 2014. **39**(2): p. 91-100.
249. Hutton, J.E., et al., *Oncogenic KRAS and BRAF Drive Metabolic Reprogramming in Colorectal Cancer.* Mol Cell Proteomics, 2016. **15**(9): p. 2924-38.
250. Son, J., et al., *Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway.* Nature, 2013. **496**(7443): p. 101-5.
251. Padanad, M.S., et al., *Fatty Acid Oxidation Mediated by Acyl-CoA Synthetase Long Chain 3 Is Required for Mutant KRAS Lung Tumorigenesis.* Cell Rep, 2016. **16**(6): p. 1614-1628.
252. Singh, A., et al., *De novo lipogenesis represents a therapeutic target in mutant Kras non-small cell lung cancer.* FASEB J, 2018: p. fj201800204.
253. Hartwell, L.H., et al., *Integrating genetic approaches into the discovery of anticancer drugs.* Science, 1997. **278**(5340): p. 1064-8.
254. Kaelin, W.G., Jr., *The concept of synthetic lethality in the context of anticancer therapy.* Nat Rev Cancer, 2005. **5**(9): p. 689-98.
255. Guarente, L., *Synthetic enhancement in gene interaction: a genetic tool come of age.* Trends Genet, 1993. **9**(10): p. 362-6.
256. Dietlein, F., et al., *A Synergistic Interaction between Chk1- and MK2 Inhibitors in KRAS-Mutant Cancer.* Cell, 2015. **162**(1): p. 146-59.
257. Corcoran, R.B., et al., *Synthetic lethal interaction of combined BCL-XL and MEK inhibition promotes tumor regressions in KRAS mutant cancer models.* Cancer Cell, 2013. **23**(1): p. 121-8.
258. Barbie, D.A., et al., *Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1.* Nature, 2009. **462**(7269): p. 108-12.
259. Tutt, A., et al., *Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial.* Lancet, 2010. **376**(9737): p. 235-44.
260. Saif, M.W., *Colorectal cancer in review: the role of the EGFR pathway.* Expert Opin Investig Drugs, 2010. **19**(3): p. 357-69.
261. Mendelsohn, J. and J. Baselga, *The EGF receptor family as targets for cancer therapy.* Oncogene, 2000. **19**(56): p. 6550-65.
262. Hagan, S., M.C. Orr, and B. Doyle, *Targeted therapies in colorectal cancer-an integrative view by PPPM.* EPMA J, 2013. **4**(1): p. 3.
263. Messersmith, W.A. and M. Hidalgo, *Panitumumab, a monoclonal anti epidermal growth factor receptor antibody in colorectal cancer: another one or the one?* Clin Cancer Res, 2007. **13**(16): p. 4664-6.
264. Jonker, D.J., et al., *Cetuximab for the treatment of colorectal cancer.* N Engl J Med, 2007. **357**(20): p. 2040-8.
265. Douillard, J.Y., et al., *Final results from PRIME: randomized phase III study of panitumumab with FOLFOX4 for first-line treatment of metastatic colorectal cancer.* Ann Oncol, 2014. **25**(7): p. 1346-55.
266. Dong, Z.Y., et al., *Potential Predictive Value of TP53 and KRAS Mutation Status for Response to PD-1 Blockade Immunotherapy in Lung Adenocarcinoma.* Clin Cancer Res, 2017. **23**(12): p. 3012-3024.
267. Cong, J., et al., *Dysfunction of Natural Killer Cells by FBP1-Induced Inhibition of Glycolysis during Lung Cancer Progression.* Cell Metab, 2018. **28**(2): p. 243-255 e5.

268. Yamamoto, F. and M. Perucho, *Characterization of the human c-K-ras gene promoter.* Oncogene Res, 1988. **3**(2): p. 125-30.
269. Hoffman, E.K., et al., *An S1 nuclease-sensitive homopurine/homopyrimidine domain in the c-Ki-ras promoter interacts with a nuclear factor.* Proc Natl Acad Sci U S A, 1990. **87**(7): p. 2705-9.
270. Cogoi, S., et al., *Anti-gene effect in live cells of AG motif triplex-forming oligonucleotides containing an increasing number of phosphorothioate linkages.* Biochemistry, 2001. **40**(5): p. 1135-43.
271. Cogoi, S., et al., *Antiproliferative activity of a triplex-forming oligonucleotide recognizing a Ki-ras polypurine/polypyrimidine motif correlates with protein binding.* Cancer Gene Ther, 2004. **11**(7): p. 465-76.
272. Morgan, R.K., et al., *Identification and characterization of a new G-quadruplex forming region within the kRAS promoter as a transcriptional regulator.* Biochim Biophys Acta, 2016. **1859**(2): p. 235-45.
273. Cogoi, S. and L.E. Xodo, *G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription.* Nucleic Acids Res, 2006. **34**(9): p. 2536-49.
274. Lavrado, J., et al., *KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines.* Sci Rep, 2015. **5**: p. 9696.
275. Cogoi, S., et al., *Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins.* Nucleic Acids Res, 2008. **36**(11): p. 3765-80.
276. Podbevsek, P. and J. Plavec, *KRAS promoter oligonucleotide with decoy activity dimerizes into a unique topology consisting of two G-quadruplex units.* Nucleic Acids Res, 2016. **44**(2): p. 917-25.
277. Cogoi, S., et al., *MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRAS in pancreatic cancer cells and delays tumor growth in mice.* Nucleic Acids Res, 2013. **41**(7): p. 4049-64.
278. Bedrat, A., L. Lacroix, and J.L. Mergny, *Re-evaluation of G-quadruplex propensity with G4Hunter.* Nucleic Acids Res, 2016. **44**(4): p. 1746-59.
279. Paramasivam, M., et al., *Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription.* Nucleic Acids Res, 2009. **37**(9): p. 2841-53.
280. Hudson, J.S., et al., *Recognition and binding of human telomeric G-quadruplex DNA by unfolding protein 1.* Biochemistry, 2014. **53**(20): p. 3347-56.
281. Cogoi, S., et al., *Critical role of hnRNP A1 in activating KRAS transcription in pancreatic cancer cells: A molecular mechanism involving G4 DNA.* Biochim Biophys Acta Gen Subj, 2017. **1861**(5 Pt B): p. 1389-1398.
282. Vitali, J., et al., *Correlated alternative side chain conformations in the RNA-recognition motif of heterogeneous nuclear ribonucleoprotein A1.* Nucleic Acids Res, 2002. **30**(7): p. 1531-8.
283. Lecarme, L., et al., *Interaction of polycationic Ni(II)-salophen complexes with G-quadruplex DNA.* Inorg Chem, 2014. **53**(23): p. 12519-31.
284. Nadai, M., et al., *Naphthalene diimide scaffolds with dual reversible and covalent interaction properties towards G-quadruplex.* Biochimie, 2011. **93**(8): p. 1328-40.
285. Zuffo, M., et al., *G-quadruplex fluorescence sensing by core-extended naphthalene diimides.* Biochim Biophys Acta Gen Subj, 2017. **1861**(5 Pt B): p. 1303-1311.

286. Muoio, D., et al., *Naphthalene diimide-derivatives G-quadruplex ligands induce cell proliferation inhibition, mild telomeric dysfunction and cell cycle perturbation in U251MG glioma cells.* FEBS J, 2018. **285**(20): p. 3769-3785.
287. Calabrese, D.R., et al., *Characterization of clinically used oral antiseptics as quadruplex-binding ligands.* Nucleic Acids Res, 2018. **46**(6): p. 2722-2732.
288. De Cian, A., et al., *Highly efficient G-quadruplex recognition by bisquinolinium compounds.* J Am Chem Soc, 2007. **129**(7): p. 1856-7.
289. 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, 2008. **90**(8): p. 1207-23.
290. Trajkovski, M., et al., *Pursuing origins of (poly)ethylene glycol-induced G-quadruplex structural modulations.* Nucleic Acids Res, 2018. **46**(8): p. 4301-4315.
291. Heddi, B. and A.T. Phan, *Structure of human telomeric DNA in crowded solution.* J Am Chem Soc, 2011. **133**(25): p. 9824-33.

Titre : Etudes structurales par RMN des acides nucléiques G-quadruplexes de KRAS et leur interaction avec des ligands.

Résumé : L'oncogène KRAS code pour une protéine GTPasique hautement mutée qui agit comme un « interrupteur » entre des états actifs et inactifs, un mécanisme important dans les processus comme la réplication ou la prolifération cellulaire. Quand ils sont dérégulés, ces processus sont à l'origine des cancers. Les mutations de KRAS sont particulièrement impliquées dans les cancers des poumons (30%), colorectaux (44%) et pancréatiques (97%). Malgré le fait que ces mutations soient bien connues, aucune molécule ne cible KRAS car toutes les stratégies actuelles ne sont pas assez efficaces pour les thérapies contre le cancer. C'est pourquoi de nouvelles stratégies ont émergé il y a quelques années visant directement la région promotrice de KRAS et plus précisément des structures appelées G-quadruplexes (G4). Même si le phénomène n'est pas encore parfaitement compris, de nombreux exemples dans la littérature montre que ces structures peuvent se former in vitro et dans les conditions cellulaires. Il a été montré que les G4 formés dans la région promotrice de KRAS peuvent lier des facteurs de transcription et perturber le processus en agissant comme un bloc lorsque l'enzyme vient lire la séquence. La stabilisation ou la destruction des G4, en utilisant de petits ligands chimiques par exemple, pourrait devenir une nouvelle voie de thérapie. Ce travail se concentre sur une séquence de 32 résidus (KRAS32R) qui peut former un G4 et correspond également au domaine minimal d'interaction de certains facteurs de transcription comme MAZ ou hnRNP A1. Cette dernière est capable de lier les G4 de KRAS32R et de les défaire favorisant ainsi la transcription de KRAS. Ce projet vise à comprendre la formation du G4 de KRAS32R au niveau atomique ainsi que son interaction avec de petites molécules organiques qui pourraient agir sur la transcription

Mots clés : G-quadruplexe, RMN, Ligands, Cancer, KRAS

Title: NMR structural studies of G-quadruplexes nucleic acids from KRAS and their interaction with ligands.

Abstract: KRAS gene codes for a highly mutated GTPase protein acting as a « switch » between an active and an inactive state, a mechanism found to be important in biological processes such as cell replication and proliferation. When misregulated, these processes are found to be at the origin many types of cancer. KRAS mutations are particularly implicated in lung (30%), colorectal (44%) and pancreatic (97%) cancers. Despite the fact that those mutations are well known, KRAS is still an undruggable target because all the actual strategies (RAS activator inhibitors, membrane association inhibitors, and so on) are not efficient enough as cancer therapies. That is why new strategies have emerged recently, such as directly targeting the KRAS promoter region and especially some specific structures called G-quadruplexes (G4). Although we do not understand well the phenomena, there are plenty of evidence in the literature that these structures can assemble both in vitro and in cellular conditions. It was shown that G4 within KRAS promoter region can bind transcription related proteins and disturb transcription process acting as a block mechanism when transcription machinery is reading the genetic sequence. Stabilization of these structures, using small chemical ligands for example, could become a new area of therapy. In my thesis work, I am focused on a 32 nucleotide sequence (KRAS32R) which can form G4 and also corresponds to the minimal interaction domain of transcription proteins such as MAZ or hnRNP1. This last protein is capable of binding to KRAS32R G-quadruplexes and possibly unfolding it, favoring the transcription of KRAS. This project aims to understand the folding of this KRAS32R G-quadruplex at atomic level and its interaction with small organic molecules that would have some effect on transcription process.

Keywords: G-quadruplex, NMR, Ligands, Cancer, KRAS

Unité de recherche

[ARNA, INSERM U1212, IECB, 2 rue Robert Escarpit 33607 Pessac]