Nucleolin Discriminates Drastically between Long-Loop and Short-Loop Quadruplexes

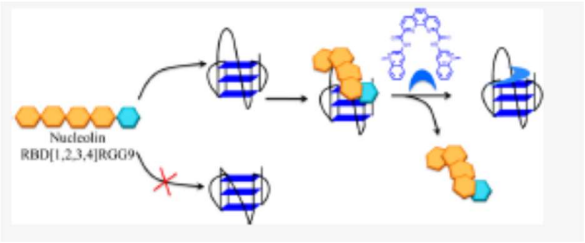
Abhijit Saha, Patricia Duchambon, Vanessa Masson, Damarys Loew, Sophie Bombard, Marie-Paule Teulade-Fichou

Biochemistry, 2020, 59, 1261-1272

Nucleolin Discriminates Drastically between Long-Loop and Short Loop Quadruplexes

Abhijit Saha, Patricia Duchambon, Vanessa Masson, Damarys Loew, Sophie Bombard,[4] and Marie-Paule Teulade-Fichou*

ABSTRACT:
We investigate herein the interaction between nucleolin (NCL) and a set of G4 sequences derived from the CEB25 human minisatellite that adopt a parallel topology while differing in the length of the central loop (from nine nucleotides to one nucleotide). It is revealed that NCL strongly binds to long- loop (five to nine nucleotides) G4 while interacting weakly with the shorter variants (loop with fewer than three nucleotides). Photo-cross-linking experiments using 5-bromo-2′-deoxyuridine (BrU)-modified sequences further confirmed the loop-length dependency, thereby indicating that the WT-CEB25-L191 (nine- nucleotide loop) is the best G4 substrate. Quantitative proteomic analysis (LC-MS/MS) of the product(s) obtained by photo-cross- linking NCL to this sequence enabled the identification of one contact site corresponding to a 15-amino acid fragment located in helix α2 of RNA binding domain 2 (RBD2), which sheds light on the role of this structural element in G4-loop recognition. Then, the ability of a panel of benchmark G4 ligands to prevent the NCL–G4 interaction was explored. It was found that only the most potent ligand PhenDC3 can inhibit NCL binding, thereby suggesting that the terminal guanine quartet is also a strong determinant of G4 recognition, putatively through interaction with the RGG domain. This study describes the molecular mechanism by which NCL recognizes G4-containing long loops and leads to the proposal of a model implying a concerted action of RBD2 and RGG domains to achieve specific G4 recognition via a dual loop−quartet interaction.



INTRODUCTION

G-Quadruplexes (G4) are noncanonical secondary stru tures that arise in G-rich sequences of DNA and RNA. G4 result from the π-stacking of two or more sets of four guanines (G-quartets) self-associated in a square planar arrangement through Hoogsteen H-bonding and further stabilized by coordination to K+ cations.[1−3] The subject of quadruplexes has been extensively reviewed in the literature over the past two decades, and currently, a large body of evidence supports the formation of these structures in various regions of the genome and the transcriptome. However, there remain many questions about the prevalence of G4 structures, their transient/dynamic nature, and their implications in multiple biological processes. In particular, the number of putative quadruplex-forming sequences is a matter of debate as considerable variations have been observed depending on the studies and the methods used. Indeed, up to 700000 potential G4-forming sequences (PQS) in the human genome have been predicted by computational analyses and in vitro sequencing,[4−6] whereas only 10000 G4 structures have been identified using G4-chromatin immunoprecipitation sequencing (G4- ChIP-seq) with G4 antibodies.[7] In the same vein, using a functional assay in yeast based on G4-induced genomic instability, Nicolas et al. have shown that only highly stable G4 structures with short loops are likely to form in vivo, thus restricting de

facto the PQS number to ~15000.[8,9] These discrepancies point to the difficulties in determining the parameters that regulate the existence of G4 in a cellular context. Therefore, how G4 are recognized and regulated by endogenous protein partners is a prerequisite for shedding light on their occurrence and functional roles. Initially, G4 were mainly viewed as steric blocks for proteins involved in replication and transcription, implying recognition and unwinding by helicases to maintain genetic stability. [10]

Subsequently, together with our ever-improving knowledge of the biology of G4, it appeared that these structures might also act as signaling elements that can recruit proteins for regulation of DNA/RNA-related events. [11−13]

The molecular basis for the recognition of G4 by proteins is still poorly understood, although this field is evolving at a rapid pace. For instance, although G4 helicases are known to load on a single-stranded tail close to G4 motifs, [10,11] two recent X-ray studies demonstrated that direct contacts with residues that constitute of the G4 core are established. This is the case of RHAU that was shown to interact both by π-stacking on an external quartet using its N-terminal G4 binding domain (DSM domain)[14−16] and by direct contact between with the G4-phosphodiester backbone and its OB domain.[17] It also was found that the bacterial RecQ helicase operates via flipping out the 3′ guanine of the external quartet using a specific guanine binding pocket (GSP).[18] This study nicely complements the initial finding that RecQ helicases possess a conserved G4 binding domain (RQC).[19] Besides helicases, a wide range of G4 binding proteins have been identified so far, which are roughly comprised of the telomeric proteins, transcription factors, and RNA binding proteins (RBP).[20−23] The latter, in particular, exhibit G4 substrate promiscuity as they bind both G4 DNA and G4 RNA. Again, structural data are still scarce, but several current studies have converged on the fact that

81 interactions with G4 are mediated by the glycine/arginine (RGG) domain, a typical feature of most RBP.[24−27] In a number of cases, the RGG motif was found to interact with the phosphate backbone and/or the base residues of the G4 loops, whereas in others, it was proposed to contribute to π-stacking interactions with the external quartets.28 Therefore, although RGG has been viewed as a critical element in recent studies,[27,29] it is poorly structured and should act in conjunction with other RNA binding domains (RBD) to achieve G4 structure specificity.[30,31] Taken together, there are still many questions unanswered about protein domains involved in the recognition of G4 substrates. In addition, G4s have topological diversity based on strand directionality and the length and sequence of the three loops connecting the G-tracts that induce significant differences in their three-dimensional (3D) structure and hydrodynamic size,[21,28] and it is still unclear how proteins cope with these structural features and their subtle differences.[32,33] In particular, interaction with G4 loops has been discussed in only a few cases. For instance, the central-loop length seems to be crucial in the case of the Ewing sarcoma protein binding [24,32,33] whereas this is more ambiguous for nucleolin (NCL) that was initially shown to prefer the short (three-nucleotide) central-loop pattern of the c-Myc G4-sequence (variant Pu47)[34,35] but was recently found5 to bind better the long (six-nucleotide) loop of the modified c- Myc G4 sequence (variant 1245-14T).[36] The long-loop preference of NCL was also observed with parallel and antiparallel hybrid G4 originating from telomeric sequence variants.37 We were thus wondering if the loop length is a strong determinant of G4 − NCL interaction (irrespective of the topology) or if this loop dependency is related to the G4 sequence context and/or to the nature of the recombinant NCL employed (full-length or polypeptide fragment, MPB vs His tag, etc.) that may eventually impact the binding. NCL was one of the first G4 interactive proteins identified and has been shown to interact with G4 DNA, G4 RNA, and G4 aptamers. [34,38,39] NCL plays a multifunctional role in cells ranging from ribosome biogenesis to chromatin remodeling transcription regulation and apoptosis [40] and has been shown to strongly repress transcription by binding G4 structures in oncogenes [41] and viral promoters.[42] Additionally, some of us recently reported that NCL regulates the translation of the Epstein-Barr viral antigen (EBNA1) through binding to its G4 mRNA. [43] To explore the interaction of NCL with G4 loops in particular, we choose to use a panel of G4-forming sequences derived from the human minisatellite CEB25. [44] The native (WT) CEB25 sequence forms a G4 structure of parallel topology harboring a central long loop of nine nucleotides and two short (1T) loops (termed CEB25-L191), which has been fully characterized by nuclear magnetic resonance (NMR) (see Table 1). In addition, systematic variation of the central-loop t1 length [from nine nucleotides to one nucleotide (Table 1)] has been performed to generate a CEB25 variant series used in a yeast functional assay

developed by Nicolas et al. This assay demonstrated that only the short-loop variants (fewer than three nucleotides) can cause genomic instability (G4-mediated response) that strongly correlates with their high thermody- namic stability. 9 It was concluded that only short-loop CEB25 variants form G4 in vivo whereas the long-loop variants do not. Thus, capitalizing on the loop-length-dependent behavior of the CEB25 series observed in a cellular context, it appeared to be ideally suited to investigating the interaction of NCL in vitro with this sequence panel. Interestingly, in the course of our study, it has been reported that the yeast NCL analogue Nsr1 may play a role in G4-mediated genome instability, [45] which gives us additional impetus for exploring interaction of NCL with the CEB25 sequences. The interaction of NCL with the G4 CEB25 series was explored using a combination of gel retardation (EMSA) and photo-cross-linking. First, we found that NCL binds the long-loop variant CEB25-L191 with a higher affinity than the short-loop variant sequences (loop with fewer than three nucleotides) and that the affinity is directly correlated with the loop length. Using a photo-cross-linking approach based on systematic replacement of loop thymines by BrU, it was shown that the central loop is indeed a strong determinant of NCLrecognition. Subsequently, proteomic analysis revealed one putative cross-linking site located in the RBD2 domain of NCL, which definitely demonstrated the specific role of the central loop in the recognition of the G4 structure. Finally, competition experiments with a panel of benchmark G4 interactive small compounds (G4 ligands) showed that only very high affinity ligands (e.g., PhenDC3) can outcompete the protein, thereby suggesting that NCL－G4 binding also involves interaction with the external quartets that might be mediated by the RGG domain. Altogether, our data led us to propose an interaction model based on concerted action between the two domains of NCL for achieving dual loop－ quartet recognition.

■ MATERIALS AND METHODS

Preparation of Oligonucleotides.

All oligonucleotides were purchased from Eurogentec (HPLC-purified grade) and used without further purification.

Preparation of Nucleolin

Recombinant NCL was produced using pET28 having a 6His-Sumo human NCL RBD-1,2,3,4-RGG expression vector (platform protein production and purification, Institut Curie, Orsay, France). This vector was generated by HIFI Gibson Assembly from the original pMal2 Nuc-1,2,3,4-RGG vector (a kind gift from N. Maizels, University of Washington, Seattle, WA). Escherichia coli BL21 DE3 pISO Dscb were transformed and grown at 20 °C in 7 L of Terrific Broth and induced with 0.5% arabinose and 0.5 mM IPTG at an optical density of 2, during 22 h in a Labfors bioreactor. The cells were collected in PBS (pH 7.4), 350 mM NaCl, 5 mM β-mercaptoethanol, and 1× Complete EDTA-free protease inhibitor cocktail (Roche), and the suspension was disrupted by being passed through a T75 cell disruptor (Constant Systems). The resulting cell lysate was centrifuged at 43000g for 1 h at 4 °C. The supernatant was applied to a HisTrap FF crude 5 mL column (GE Healthcare) and washed thoroughly, and the proteins were eluted in elution buffer [PBS (pH 7.4), 350 mM NaCl, 5 mM β- mercaptoethanol, and 250 mM imidazole]. 6His-Sumo was cleaved using Sumo protease (10 μg/1 mg of protein) and dialyzed overnight against 20 mM Tris-HCl (pH 7.4), 100 mMNaCl, 10% glycerol, 0.5 mM EDTA, and 5 mM β-mercaptoethanol. The eluate was loaded onto a HiTrapHeparin column (GE Healthcare) for elution with a continuous gradient of NaCl (from 0.1 to 1 M) in the same buffer. Fractions containing NCL were dialyzed against 20 mM Tris-HCl (pH 8.00), 400 mM NaCl, 0.5 mM EDTA, and 10% glycerol. The purity of the protein was checked by sodium dodecyl sulfate－ polyacrylamide gel electrophoresis (SDS－ PAGE) in a 4 to 20% acrylamide gel, and the protein concentration was determined by measuring the absorbance at 280 nm [size of 46.1 kDa (Figure S1)].

Electrophoretic Mobility Shift Assay (EMSA).

The samples were prepared by mixing 5′ -radiolabeled (γ-32P) G4 DNA (100000 cpm) with the corresponding nonradiolabeled G4 DNA (100 nM) in K-100 buffer [which consisted of 10 mM LiAsO2Me2 and 100 mM KCl (pH 7.2)]. After mixing, the sample was heated at 95 °C for 5 min followed by overnight cooling. For protein binding, the preannealed sample G4 DNA was diluted at a final concentration of 20 nM in the 1× binding buffer [which consisted of 100 mM Tris-HCl (pH 7.5), 10 mM EDTA, 1 M KCl, 1 mM DTT, 50% glycerol, and 0.10 mg/mL BSA]. NCL was added in the desired concentration range of 0－2 μM and incubated at room temperature for 1 h. The protein－ DNA complexes were esolved on a pre-electrophoresed 10% polyacrylamide native gel

(37.5:1 acrylamide:bis ratio) in 0.5× TBE buffer at room temperature for 2 h at 120 V. After electrophoresis, the gel was analyzed by phosphor-imaging, by being scanned on a Typhoon Trio Variable mode imager. The quantitative gel was fitted in OriginPro 8.6 by the Hill equation, y = Bmax[P]n/ (KD 226
n + [P]n), where y is the fraction bound, Bmax is the fraction bound at which the data plateau, [P] is the concentration of the protein, KD is the equilibrium dissociation constant, and n is the number of cooperative binding sites; the equation was fitted in a 1:1 binding mode. Photo-Cross-Linking. The samples were prepared by annealing 100 nM 5-radiolabeled G4 DNA in K-100 buffer.The annealed DNA was then diluted to 20 nM with 1× binding buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM KCl, 0.1 mM DTT, 5% glycerol, and 0.01 mg/mL BSA]. The annealed DNA was incubated with 0.4－1 μM NCL for 1 h at room temperature and irradiated for 10－60 min with ultraviolet (UV) light (xenon light source MAX-303, 300 W, 300 nm monochromatic light) by keeping a distance of 8 cm between the light source and the sample in a 1.5 mL centrifuge tube placed on ice (irradiance of 7.9 mW cm 241 －2). After irradiation, 10 μL of the loading dye [2 M urea, 1% SDS, 250 mM Tris-HCl (pH 6.8), 40% glycerol, 0.01% bromophenol, and 0.01% xylene blue] was added and the mixture heated at 95 °C for 5 min before being subjected to 10% SDS－PAGE.

Proteomics and Mass Spectrometry Analysis.

Samples were prepared by mixing 125 pmol of preannealed G4 DNA and 7.3 μg of NCL in 1× binding buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM KCl, 0.1 mM DTT, and 5% glycerol]. After being incubated at room temperature for 1 h, the samples were irradiated with UV light (300 nm); 2 μg of protein samples was digested with 0.2 μg of trypsin/LysC (Promega) overnight in 20 μL of 25 mM NH4HCO3. The sample was then loaded onto a homemade C18 StageTips for desalting. Peptides were eluted using a 40:60 MeCN/H2Omixture with 0.1% formic acid and vacuum concentrated to dryness. Online chromatography was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled online to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). Peptides were trapped on a C18 column (75 μm inner diameter × 2 cm; nanoViper Acclaim PepMap 100, Thermo Scientific) with buffer A (2:98 MeCN/H2O in 0.1% formic acid) at a flow rate of 4.0 μL/min over 4 min. Separation was performed on a 50 cm × 75 μm C18 column (nanoViper Acclaim PepMap RSLC, 2 μm, 100 Å, Thermo Scientific) regulated to a temperature of 55 °C with alinear gradient of 5% to 25% buffer B (100% MeCN in 0.1% formic acid) at a flow rate of 300 nL/min over 100 min. Full- scan MS was performed in the Orbitrap analyzer with a resolution set to 120000, and ions from each full scan were HCD fragmented and analyzed in the linear ion trap. For identification, the data were searched against the Human (UP000005640) UniProt or the NCL construct database using Mascot 2.5. Enzyme specificity was set to trypsin, and a maximum of two missed cleavage sites was allowed. Oxidized methionine was set as the variable modification. The maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.6 Da for MS/MS peaks. The resulting files were further processed using myProMS (https://github.com/bioinfo-pf-curie/myproms)
version 3.6: the maximum false discovery rate (FDR) was set to 1% at the peptide level (Mascot score).46 To quantify the modified peptides, we extracted from the MS survey of nano-LC-MS/MS raw files the extracted ion chromatogram (XIC) signal of the well-characterized NCL tryptic peptide ions using Skyline (version 4.1) (MacCoss Lab Software, Seattle, WA; https://skyline.ms/
project/home/sofware/Skyline/begin.view). The peptide XIC areas were log2 transformed, and the mean log2－ area was normalized by the mean area of nonmodified peptide NDLAVVDVR using R version 3.1.0. A linear model was built using all peptide areas to estimate the mean and the 5% confidence interval of each condition. The associated ratio and p value were computed via a two-sided t test. The mass spectrometry proteomics data have been deposited with the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD016529 (username reviewer88634@ebi.ac.uk, password swD9eGSb).[46,47]

FRET Melting Assay.

Stabilization of compounds with CEB25-L191 was monitored via the FRET melting assay using the doubly labeled sequence 5′-Fam-AAGGGTGGGT-GTAAGTGTGGGTGGGT-Tamra that was prefolded by heating at 90 °C for 5 min and left to cool at 4 °C overnight. The assay was performed in 96-well plates on a real time PCR apparatus 7900HT Fast Real-Time PCR System as follows: 5 min at 25 °C and then increase of 0.5 °C every minute until 95 °C. Each

experimental condition was tested in duplicate in a volume of 25 µL for each sample. To determine ligand G4 versus duplex selectivity, an unlabeled DNA competitor (ds26) was used. Final concentrations are as follows: 0.2 µM oligonucleotide, 1 µM ligand, and 3 or 10 µM duplex competitor. Measurements were taken with excitation at 492 nm and detection at 516 nm in a buffer of 10 mM lithium cacodylate (pH 7.2), 99 mM LiCl, and 1 mM KCl. The ligand-induced stabilization (ΔT1/2) was calculated from the temperature at half-denaturation of the G4 in the absence and presence of the ligand.

Table 1: G-Quadruplex sequences used in this study

Acronym	Sequence (5′ → 3′)	G4-Topology
CEB25-L191 (WT)	AAGGGTGGGTGTAAGTGTGGGTGGGT	parallel [44] (PDB: 2lpw)
CEB25-L151	AAGGGTGGGAGTGTGGGTGGGT	parallel [9]
CEB25-L131	AAGGGTGGGTGTGGGTGGGT	parallel [9]
CEB25-L121	AAGGGTGGGTTGGGTGGGT	parallel [9]
CEB25-L111	AAGGGTGGGTGGGTGGGT	parallel [9]
T95-2T	TTGGGTGGGTGGGTGGGT	parallel [48] (PDB: 2LK7)
Bcl2mid	GGGCGCGGGAGGAATTGGGCGGG	Hybrid [49] (PDB: 2F8U)
G-Mut	AAGCGTCGCTGTAAGTGTGCGTCGCT	ss DNA

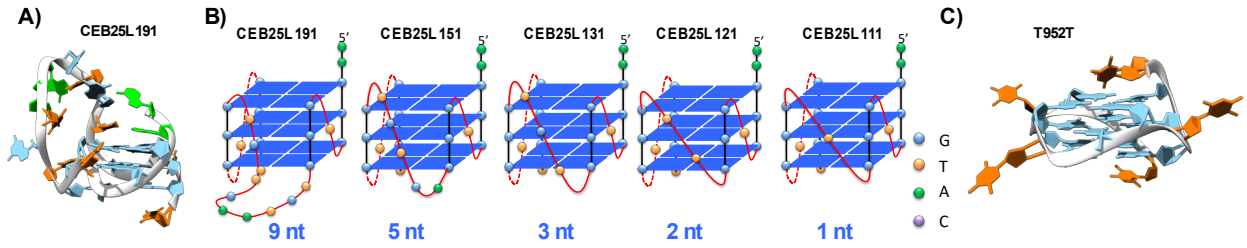


Figure 1: A) NMR Structure of the WT CEB25-L191 (PDB: 2lpw) [44], G blue, A green, T orange B) Schematic representation of the G4-parallel topology adopted by the loop-variant series a previously shown by CD [9]; blue square stands for G-quartet, red ribbon stands for phosphodiester backbone, color spheres stand for loop bases C) NMR structure of T95-2T (PDB: 2LK7) [48], similar structure is likely to be adopted by CEB25-L111 harboring 2A instead of 2T in 5′.

■ RESULTS

NCL Binds Preferentially to Long-Loop CEB25 Substrates.

To study the interaction between NCL and G4 DNA, we used the C-terminal domain of NCL containing four RBD-[1,2,3,4]-domains and the RGG, which has been shown previously to bind c-Myc G4. 322 41 NCL constructs reported for studying G4 interactions contain a maltose binding protein (MBP) fragment in most cases. 324 34,38 Considering that the MBP tag may impact the G4 binding behavior of NCL, [37] the expression protocol was modified to produce the protein construct with a His tag. Finally, the His tag was cleaved by Sumo protease to produce nucleolin containing the RBD-[1,2,3,4]RGG (see Materials and Methods). Then we evaluated the binding affinity of NCL for the CEB25 series comprised of five sequences all exhibiting two short side loops (1T) and one central loop varying from nine nucleotides to one nucleotide (Table 1). The NMR structures of CEB25- L191 as well as that of a variant sequence (T95-2T) differing from CEB25-L111 by only two nucleotides at the 5 end (see Table 1) are shown in Figure 1. All G4 are stable at room temperature under our conditions (see the Tm in Table S1). The binding was monitored by a gel electrophoresis retardation assay (EMSA) using 32P-labeled oligonucleotides. As shown from Figure 2A, binding of NCL to CEB25-L191 indicated a concentration-dependent association that rapidly leveled off. Fitting of the data with a 1:1 model allowed us to determine a KD value in the nanomolar range [KD = 241 nM, 1:1 fitting (Figure 2F)]. This binding affinity is close to that previously reported with parallel G4 sequences such as c-Myc and PDGF-A.34 Next, we modified the CEB25 sequence in reducing the number of nucleotides of the central loop from nine to five, three, and two (Table 1). Remarkably, these modifications greatly affected the association of NCL, which experienced a gradual decrease as the loop size was reduced (Figure 2B − D). Accordingly, a 3 − 4-fold decrease in the apparent affinity constant was determined as a function of loop length (KD = 859 nM for CEB25-L151; KD = 1020 nM for CEB25-L131; CEB25-L121 binding too low to allow KD to be determined). In a striking manner, the interaction appeared to be completely abolished when the central-loop length was reduced to one nucleotide [CEB25-L111 (Figure 2E)]. This dramatic loop-length dependence indicates that at least three nucleotides in the central loop are needed for a significant interaction of NCL. Subsequently, the experiment was reproduced with a single-stranded scrambled mutant sequence (G-Mut) that cannot form G4 structure. As expected, much less binding of NCL was observed with a large smearing fraction typical of unstable species (Figure S2), thereby revealing significantly weaker interaction compared to that of CEB25-L191. This suggests that the observed binding preference does not result simply from the ability of NCL to interact with single-stranded nucleic acids but that the protein recognizes a single-stranded loop in the context of a G- quadruplex. This is further supported by the fact that NCL binding does not induce significant changes in the G4 substrate topology and thermodynamic stability as shown by CD and FRET melting measurement (Figure S3). To strengthen our observation with the CEB25 series, we evaluated the binding of NCL with the Bcl2mid sequence that contains a central-loop length of seven nucleotides. Interestingly, this sequence forms a G4 hybrid structure harboring the seven-nucleotide central

379 loop in a lateral arrangement (Table 1 and Figure S4).[49] Remarkably, EMSA analysis indicated strong binding of NCL (KD = 558 nM), which is consistent with the behavior observed with the two long-loop (nine and five nucleotides) CEB25 sequences (Figure S4). Taken together, our results suggest that a long loop (at least five nucleotides) positioned in a G4 structure is a strong determinant of the interaction with NCL independently of G4 topology.

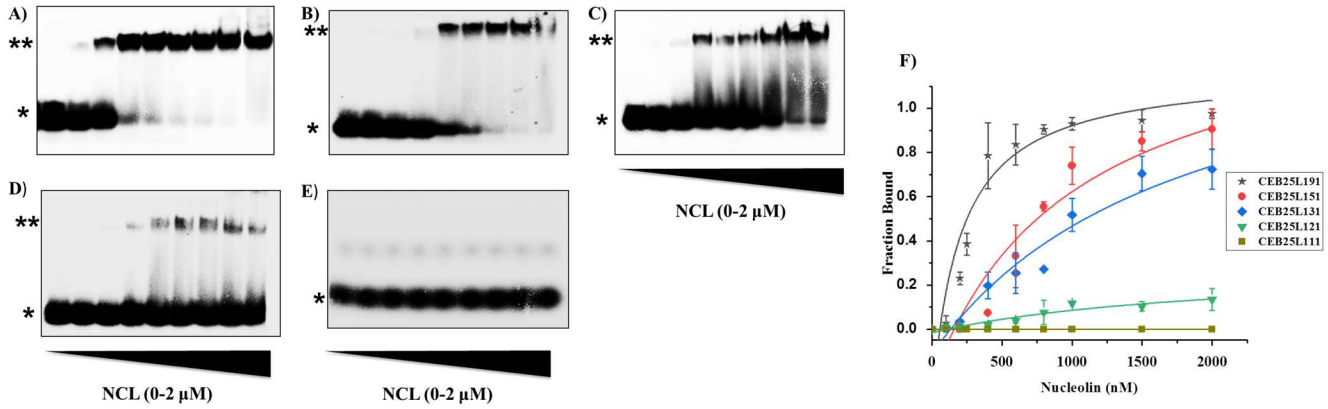


Figure 2: Gel electrophoresis (EMSA) analysis for determination of binding of NCL to CEB25 sequences A) CEB25-L191 B) CEB25-L151 C) CEB25-L131 D) CEB25-L121 E) CEB25-L111. [DNA]= 20nM, [NCL]= 100-2000 nM, corresponding to 0.1, 0.2, 0.5, 0.7, 0.9, 1.1, 1.7, 2.3 µg of recombinant protein respectively, (*) free DNA and (**) DNA /NCL complex (F) Binding curves obtained from quantification of bound NCL as function of concentration (average of triplicate experiments), solid lines represent the fitted curves.

Exploring the Loop Influence by a Photo-Cross-Linking Approach.

Photo-cross-linking based on incorporation of photoactivatable bromo or iodouracil in place of thymine residues is a powerful method for assessing interactions in nucleic acid − protein complexes and also for the identification of close contact points. Usually, UVB irradiation produces covalent adducts linking the two partners with variable yields highly dependent on the relative reactivity of the amino acids (mostly aromatic electron-rich amino acids) situated in the proximity to the uracil site generating the short-lived reactive radical species. 397 50,51 Therefore, the CEB25 series is ideally suited for this photo-cross-linking approach as all sequences contain thymine residues in their three loops (Table 1). First, a single BrU residue was introduced into the central loop of each CEB25 sequence (Figure 3A). As previously reported, T to BrU replacement does not affect significantly the G4 structure; 403 52 it was nonetheless verified by CD measurement of the BrU-modified CEB25-L191 sequence (Figure S5). On the basis of the EMSA titration, the protein:DNA ratio was fixed at 50:1 for photoreaction to ensure formation of a significant amount of complexes while avoiding saturation that could minimize differences between the sequences (see Figure 2F). After irradiation of the set of BrU-modified sequences, PAGE analysis of the samples revealed the formation of a major retarded band corresponding to a cross-linking product (Figure 3A and Figure S6). The yield of cross-linking is high for the long loop (L191-1) (~49%) and decreases gradually as the loop length is shortened [~36%, ~27%, and ~11% for L151-1, L131-1, and L121-1, respectively (Figure 3B)]. Remarkably, no adduct was observed for the shortest one-nucleotide loop G4-L111-1 (Figure 3A). These results are in full agreement with the EMSA results, confirming unambiguously that NCL interacts with the central loop of the G4 structure with a threshold fixed around two or three nucleotides. It is noteworthy that a second slower mobility band was observed on the gel (Figure 3A), which might correspond to higher-order species (2DNA:NCL)

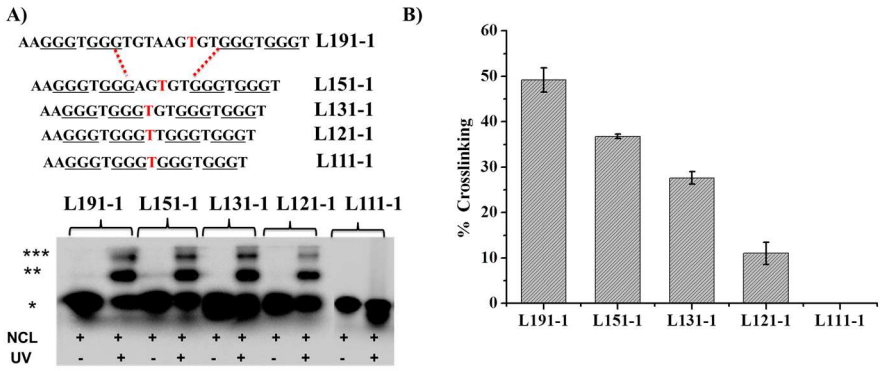


Figure 3 A) SDS PAGE analysis of photo-crosslinking of NCL with single BrU substituted CEB25 sequence variants named CEB25-L1X1-1, X indicates central loop length X= 9,5,3,2,1, -1 stands for 1 BrU residue. (*) represents unmodified DNA, (**) represents DNA-NCL crosslink products and (***) represents putative (DNA)2-NCL cross-linked products. B) Quantification of the crosslink products [DNA] = 20 nM, [NCL] = 1 μM, UV irradiation time 10 min, Exc: 300 nm.

Identifying Contact Site(s) of NCL with the G4 Central Loop

Using Proteomic Analysis. Having identified the central nine-nucleotide loop as the strong determinant of the interaction, we were then interested in identifying if other sites of the native CEB25-L191 sequence can produce cross- linking of NCL. Therefore, we designed a panel of six oligonucleotides having a single BrU replacing each T at all possible locations, meaning that the three loops are labeled with a photoactive moiety (Figure 4). After UV irradiation, gel analysis showed clearly that the oligonucleotides containing a BrU residue in the central loop (oligonucleotides 10BrU, 12BrU, 16BrU, and 18BrU) are the most reactive ones as compared to the other sequences with BrU in the two side loops (oligonucleotides 6BrU and 22BrU). Quantification of the photo-cross-linking products indicates that the yield is almost in the same range (38 − 45%) for oligonucleotides labeled in the central loop (10BrU, 12BrU, 16BrU, and 18BrU), whereas it is <10% for oligonucleotides labeled in the side loops (6BrU and 22BrU). The striking difference in reactivity indicates clearly that strong interactions are established between NCL and the central nine-nucleotide loop, while the protein seems to interact poorly or in a weaker manner with the one-nucleotide side loops. Examination of the 3D NMR structure of the CEB25-L191 sequence showing the four BrU residues in the nine-nucleotide central loop (Figure 4C) suggests that the protein may surround this single- stranded structured domain that is protruding from the G4 core, using it as a hook to interact with the G4 structure. Altogether, these data indicate that NCL establishes tight contact with nearly all loop nucleotides. In an attempt to identify amino acid residues or peptidic domains involved in the interaction between NCL and its long- loop substrate, we decided to carry out quantitative proteomic analysis (LC-MS/MS) of the cross-linked product(s). Because the 16BrU-modified sequence gives the highest yield of cross-linking suggesting intimate contact with the protein, this sequence was used for analysis. To this end, the experiment described above (Figure 4) was reproduced four times. In the first experiment, the classical protocol was applied to extract the cross-linked products from the gel and make them suitable for LC-MS/MS analysis (i.e., elution from the gel, nuclease and phosphatase treatment, and purification by SDS − PAGE followed by enzymatic digestion with trypsin/LysC). However, under these conditions, LC-MS/MS analysis showed broad profiles from which no relevant information could be extracted due to the remaining DNA. Therefore, an alternative approach was applied, which is based on a well-established workflow for analyzing RNA − protein complexes directly from solution

samples. [53,54] Then direct trypsinization without DNA digestion should generate bulky DNA−peptide adducts unlikely to be detected by LC-MS/MS. Under these conditions, comparison of the ratio of all peptides in the UV-irradiated solution versus the nonirradiated control solution should show specific peak depletion corresponding to cross-linked peptides. 478 [54,55] Following this protocol (shown schematically in Figure S7), we were able to identify a highly significantly reduced peptide in the UV-treated sample by applying a 1% false discovery rate (FDR) for the peptide validation and an intensity threshold in skyline and a p value of <0.05 as the statistical cutoff for the ratio significance (TableS2). The MS analysis indicated that this is a 15-amino acid peptide located in RBD2 of NCL corresponding to the fragment from the Gly at amino acid 429 to the Lys at amino acid 444 (Figure 5), which suggests a direct interaction of this domain with the nine-nucleotide loop of the G4 substrate. Interestingly, the RBD1 and RBD2 structures of the human NCL have been characterized by NMR56 and are known to be involved in the recognition of the G4 aptamer AS1411.[56] Remarkably, examination of the protein sequence indicates that the depleted peptide is located at the junction between strand β3 and helix α2 of RBD2,56 suggesting these elements might be strong determinants of G4-loop binding. Finally, although RBD1 has not been detected in our analysis, it may act in tandem with RBD2 for G4-loop recognition as has been shown for NCL−RNA hairpin-loop complexes.[57] In addition, other domains not detected by cross-linking might be involved in the interaction. For instance, the RGG domain of NCL has been reported to contribute to G4 binding24−29 via either loop or quartet contacts. Cross-linking within this domain was not detected in this case, which could indicate the absence of contact or weak contact between the RGG domain and the central loop and/or result from the low level of electron-rich aromatic residues that are the most reactive with regard to BrU photo-cross-linking.[50,51] In addition, arginine residues strongly promote trypsinization, generating small fragments that can hardly be detected by LC- MS/MS. Altogether, this makes the detection of the RGG domain using our approach difficult.

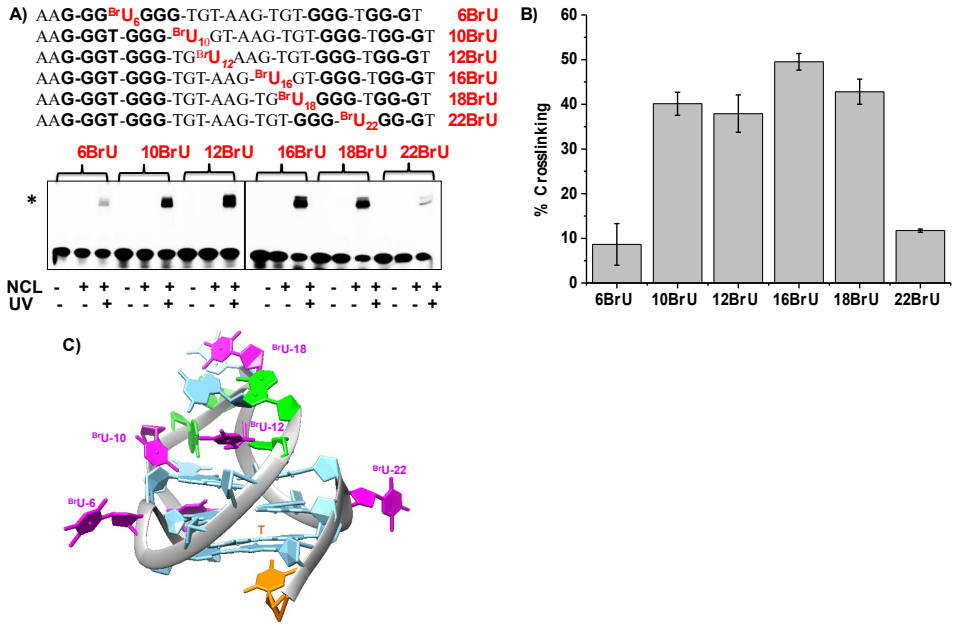


Figure 4. A) SDS PAGE analysis of photo-crosslinking of NCL with ^{32}P labelled single BrU substituted CEB25-L191 sequences (oligonucleotides 6BrU-22BrU). [DNA]=20nM, [NCL]= 400 nM, UV irradiation time 10 min. B) Quantification of the photo-adduct. Single asterisk (*) represents unmodified DNA, double asterisks represent (**) DNA-NCL crosslink product C) Representation of CEB25-L191 structure showing positions of the BrU residues (pink) in the three loops, G in blue, A in green, T in orange.

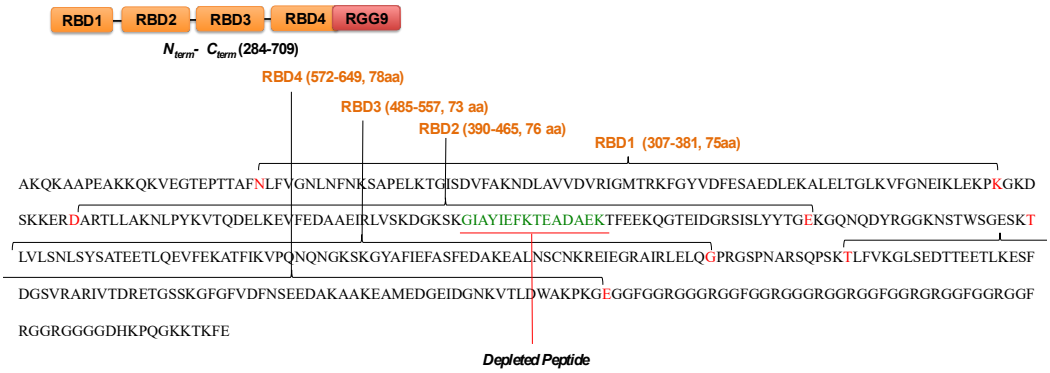


Figure 5. *N*-terminal sequence of NCL (284 to 709) showing the four RBD domains and RGG9 domain. The depleted peptide is shown in green.

Competition between NCL and G4 Binding Ligands.

Therefore, to gain a more global picture of the NCL interaction and eventually probe other binding sites in particular, with the G4 quartets, we next carried out binding competition between NCL and a set of G4-interactive compounds that are mostly accommodated on the top and bottom external quartets of G4.[58] G-Quartet interactions have been shown for a number of proteins and peptide fragments[14−17] and are likely to be mediated by stacking forces between peptidic aromatic residues and guanines.[24−29] In addition, G4 ligands such as BRACO-19 and CX-3543 have been shown to compete with G4 binding proteins containing RGG motifs, including NCL.[59,60] In addition, we recently reported on the very efficient disruption of NCL−G4 RNA association using PhenDC3 and N-acyl hydrazone derivatives in the context of Epstein-Barr virus. [43,61] Thus, to explore the interference between NCL and the external quartets of CEB25- L191, the competition assay was performed with the most potent benchmark ligands reported so far, namely, PhenDC3, Pyridostatin (PDS), PDC-(360A), RHPS4, and Pt-ttpy (Figure 6), together with a negative control (BisDC3) exhibiting weak affinity for G4. [62] The binding experiment with NCL and CEB25-L191 was reproduced as described above but at a lower protein:DNA ratio (10:1) to obtain 50% binding (Figure 2), and then ligands were added at increasing concentrations. Remarkably, a rapid and gradual decrease in the intensity of the retarded band corresponding to the NCL−G4 complex was observed when using PhenDC3 (Figure 7A and Figure S8). The complete disappearance of the bound fraction was induced at a relatively low PhenDC3:NCL ratio (≤2 molar equivalents), therebyindicating that this compound is highly efficient in preventing NCL−G4 interaction. In stark contrast, all of the other G4 binding ligands appeared to be unable to compete with NCL. Indeed, like the negative control (Figure 7B and Figure S7), the fraction bound remained unchanged regardless of the ligand concentration (Figure 7C−F), although it was applied at a 2-fold higher range than PhenDC3 (0−1000 nM). Though these ligands are assumed to interact strongly with G4, their incapacity to impact NCL binding may suggest two possibilities. These ligands have either a different binding site or a lower affinity than PhenDC3 for CEB25-L191. Although PhenDC3, PDC, and PDS are thought to be high-affinity binders with roughly nanomolar KD, their G4 affinity may differ slightly (or substantially) depending on the G4 substrate, and few studies have systematically and quantitatively compared G4 ligands.[63,64] Thus, in an attempt to understand the strikingly different behavior of the used G4 ligands, we examined their binding to CEB25-L191 using the classical FRET melting assay. The melting measurements indicate that all ligands stabilize

CEB25-L191, but the ΔT1/2 values greatly differ, showing clearly that PhenDC3 is by far the stronger binder of the series [ΔT1/2 values of 24.6 °C for PhenDC3, 16.4 °C for PDS, 13.6 °C for PDC, and 13.5 °C for Pt-ttpy (Figure S9)]. This implies that ligands with ΔT1/2 values of ≤ 18 °C display overly high KD values relative to that of NCL and thus cannot act as a competitor for G4 binding. It is noteworthy that RHPS4 cannot be properly evaluated by this method due to its intrinsic fluorescence that interferes with the FRET signal (Figure S9),65 but the KD value reported in the literature for telomeric G4 lies in the micromolar range, i.e., 100-fold lower than that found for PhenDC3 for the same G4 sequence.66Alternatively, the distribution of the ligand on allosteric sites (loop or groove) cannot be excluded, which may eventually induce loop reorganization and affect (or not) protein binding. This has been reported for carboxyPDS and 7ODT (hexaoxazole analogue of telomestatin) that were shown to bind G4 without affecting antibody or protein G4 recognition through formation of ternary complexes[25,67] On the contrary in the case of PhenDC3, the strong and exclusive stacking interaction with external quartets that has been structurally well-characterized 583 58,68 is more likely to block access of the protein to these sites, thereby causing competitive binding inhibition. Taken together, these findings indicate that G4 ligands considered as benchmarks may differ significantly in their binding strength for a given G4 in vitro, which explain why they may display significant biological effects in vivo. This aspect that is often overlooked should be considered with care in particular when competition with proteins is involved. [64] In summary, our data indicate that the external quartet of CEB25-L191 is also a site of interaction for NCL and that the occupation of this site by a G4 ligand of high affinity like PhenDC3 can disrupt or prevent protein association. This suggests that the G-quartet surface is a strong determinant of the NCL－G4 interaction that may involve RGG or eventually other domains. However, although this interaction is a prerequisite for the stability of the complex, it is not sufficient in itself to ensure strong binding of the protein as indicated by the absence of binding to the short-loop CEB25 variants that harbor external quartets of high accessibility.

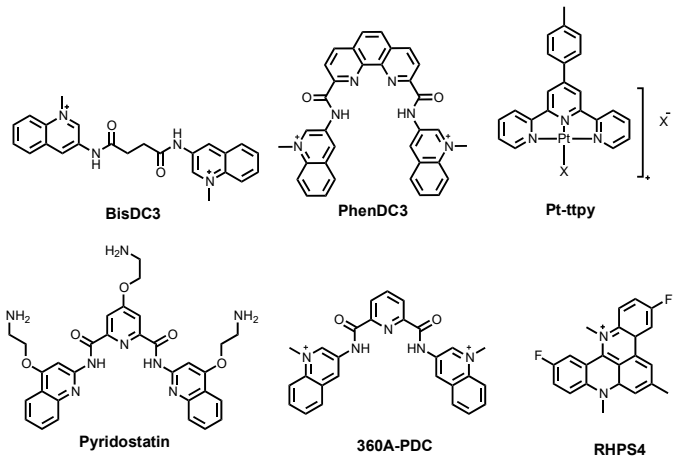


Figure 6. Chemical structure of G4 ligands used in this study.

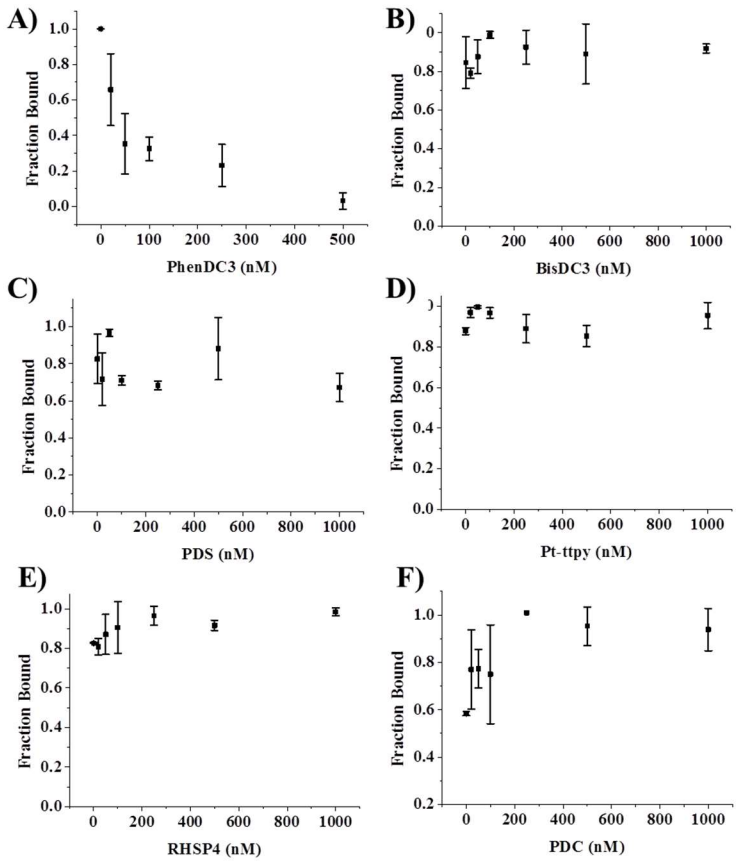


Figure 7. Displacement of NCL by small molecules monitored by EMSA (A) PhenDC3 (B) BisDC3 (C) PDS (D) Pt-ttpy (E) RHSP4 (F) PDC. Fraction bound was calculated from band intensity quantification of EMSA gels (three replicates) shown in figure S8.

■ DISCUSSION AND CONCLUSION

Exploring recognition of G4 by proteins is key to deciphering the functional roles of these noncanonical structures. Nucleolin (NCL) is well-known for its capacity to bind a variety of secondary structures and is one of the first G4 binding proteins to have been identified. Nonetheless, NCL−G4 interaction is not yet fully elucidated at the molecular level, as the structural determinants of such association have not been completely identified from the perspectives of both the protein and the G4 substrate. NCL−G4 binding has been investigated extensively with the sequences originating from the c-Myc promoter because NCL is a ubiquitous protein likely to interact at many G4 DNA and G4 RNA loci in the genome, and it was worth investigating its interaction with other sequences like the CEB25 minisatellite variants for which a structure G4 phenotype relationship has been well established. [9] We report herein that the NCL construct (RBD-1,2,3,4-RGG) devoid of a tag recognizes G4-CEB25 containing a central long loop (four to nine bases), whereas short-loop variants (fewer than three bases) are not bound as shown from EMSA experiments. This first result is in line with the recent studies making use of full-length NCL. 623 37 The discrepancy with the initial reports of NCL−G4 binding might result from differences in the structure of the recombinant protein used and that of the G4 partners. 626 34 First, a non-tagged NCL protein is more akin to its native analogue. Second, a well-characterized monomorphic G4 structure like the CEB25 series is more likely to reflect one binding event rather than polymorphic sequences like Pu27/Pu47 that may form a variety of G4 populations. [69] Subsequently, we used a photo-cross-linking approach based on the systematic replacement of thymines by BrU in the three loops of the quadruplex, which confirmed that

the NCL－G4 interaction occurs mainly in the central long nine-nucleotide loop of the native sequence CEB25-L191. Finally, quantitative proteomic analysis showed that this recognition involves the RBD2 domain that is strongly cross-linked to the central loop of the G4 DNA substrate via a 15-amino acid fragment located at the β3－α2 junction of RBD2, thereby identifying this structural element as a strong determinant of the protein－G4 interaction. Finally, to gain a more global picture of the interaction, we conducted competition with G4 interactive small molecules that are known to bind strongly to the external quartet(s) of quadruplexes via a combination of π-stacking mode and electrostatic forces. The striking inhibition of NCL binding observed with PhenDC3, the best G4 binding compound of the series, clearly suggests competition for the same binding site indicating that NCL－G-quartet interaction is a specific requirement for the stability of the interaction. Taken together, the collected data demonstrate clearly that the RBD2 domain interacts with the single-stranded central loop of the native CEB25-L191 while another domain may bind the terminal guanine quartet, both interactions being absolutely required. The second NCL domain involved in the interaction could be the RGG domain, which is supported by the early observation of Maizels et al. that the RGG domain of NCL recognizes G4 DNA independently of sequence context.[70] The RGG (arginine-glycine-glycine) box present in most RNA binding proteins (RBP) is a highly disordered domain known to establish electrostatic interactions with the phosphate backbone of nucleic acids. Although its specificity is questionable, it is nonetheless known to mediate the recognition of a large variety of nucleic acid secondary structures.[71] Currently, the role of the RGG domain in G4－protein interactions is intensively scrutinized, and a number of RGG-containing G4-interactive proteins have been identified. In most cases, deletion of the RGG domain abolishes or drastically weakens the binding to G4, pointing out the key role of this domain.[41] RGG domains have been reported to stack on terminal guanine quartets putatively through aromatic/hydrophobic residues isolated between the Gly-Arg repeats, but the nature of the interactions taking place remains unclear.[27] Other studies suggested interactions of RGG with G4 loops that may stabilize or destabilize the G4 structure;[24,72] alternatively, both loop and quartet RGG interactions have been reported.[27] Therefore, in view of this variety of interaction modes and contrasting conclusions, the mode of binding of RGG to G4 is still an open question. In addition, the binding mode may vary from one quadruplex to another in view of the flexibility of the RGG domain that provides high adaptability as illustrated by its interaction at the G4-duplex junction.[73] Finally, the implication of another RBD domain 684 (RBD1, RBD3, or RBD4) can also be hypothesized in view of the NMR structure reported for NCL bound to RNA hairpin loops, which implies at least two RBD domains. On the whole, our findings led us to propose a model interaction based on a concerted action between at least two protein domains to achieve long-loop G4 recognition possibly via a dual loop－quartet interaction. Thus, our study shed light on the molecular mechanism by which NCL recognizes G4s, thereby adding a piece to the puzzle of protein－G4 interactions. From the biology perspective, NCL is a protein with high promiscuity that underlies the variety of functions it achieves in cells. Its large substrate promiscuity is likely to result from the modular feature of NCL that can use a combination of its various domains to adapt to different substrates. However, the study presented here indicates that NCL can discriminate in a drastic way between long-loop and short-loop G4 DNA structures. Therefore, this observation opens doors to whether this loop-dependent behavior has functional significance and more importantly the possible consequences for NCL functions. In particular, it raises the question of correlation between the loop dependence observed for inducing genomic instability within the CEB25 series and that observed for NCL binding. The G4-mediated instability of CEB25 minisatellites is driven by persistent G4 with a short central loop (no more than three nucleotides), whereas long central-loop G4 (no fewer than three nucleotides), less likely to

form in vivo, are not able to trigger dysfunction. [9,74] Strikingly, the same three-nucleotide loop threshold is observed for NCL binding, which suggests that NCL could play a role in the genetic instability of the CEB25 series and other G-rich tandem sequences. [74] Therefore, we can hypothesize that binding of NCL to long-loop G4 could play a protective role in helping to unfold G4 and/or participating in the recruitment of G4 helicases, [75] thereby contributing to the suppression of G4-induced genetic instability in line with its role in genome maintenance. Interestingly, from a broader perspective, a number of oncogene promoter sequences forming parallel G4 feature a central loop with at least three nucleotides and two side loops of one nucleotide (G4-L1X1) as mentioned in recent studies by Yang et al. [36,76] It thus could be of great interest to verify the binding proteins with these promoter sequences and their loop-length variants with the aim of establishing if the G4- L1X1 ($X \geq 3$) represents a consensus sequence and a general rule for recognition by NCL or other regulatory proteins (transcription factors, etc.). Finally, the incapacity of NCL to bind the very compact and highly stable short-loop G4 (G4-L1X1, where X = 1) interrogates once again the roles of G4 in cells and is fully consistent with the numerous studies showing that all G4 motifs are not equivalent and that mostly short-loop G4 are at risk for genome integrity [9,74] due to their high thermodynamic stability and slow unfolding kinetics, making them long-lived motifs that cannot be processed by proteins. This aspect has been fully documented for proteins with enzymatic activity (helicases and polymerases) [6,75,77] but is demonstrated for the first time herein with a protein more likely to act on G4 as a chaperone or signaling intermediate, pointing one more time to the fascinating but complex interplay between G4 structures and proteins. ■ **ASSOCIATED CONTENT**

***SI Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.9b01094. SDS−PAGE to check the purity of recombinant NCL (Figure S1), EMSA for monitoring binding of NCL with G4Mut (Figure S2), CD spectra of CEB25-L191 (3 μM) after incubation with NCL (Figure S3), EMSA for monitoring binding of NCL with Bcl2mid (Figure S4), CD spectra of WT-CEB25-L191 (Figure S5), SDS−PAGE analysis of the control photo-cross-linking experiment (Figure S6), schematic representation of the LC-MS/MS analysis (Figure S7), displacement of NCL from CEB25-L191 (Figure S8), stabilization of G4-CEB25-191 by ligands measured by FRET melting (Figure S9), thermal stability of the CEB25 variants measured by UV melting (Table S1), and a list of peptides identified following the workflow of Figure S5 with corresponding UV+/UV− ratios of extracted areas and p values (Table S2) (PDF)

Accession Codes

Nucleolin, P19338. RBD-1,2,3,4-RGG domain

■ **AUTHOR INFORMATION**

Corresponding Author

Marie-Paule Teulade-Fichou − CNRS UMR9187, INSERM U1196, Institut Curie, PSL Research University, 91405 Orsay, France; CNRS UMR9187, INSERM U1196, Universite´ Paris Sud, Universite´ Paris Saclay, 91405 Orsay, France; Email: mp.teulade-fichou@curie.fr*

Authors

Abhijit Saha − CNRS UMR9187, INSERM U1196, Institut Curie, PSL Research University, 91405 Orsay, France; CNRS UMR9187, INSERM U1196, Universite´ Paris Sud, Universite´Paris Saclay, 91405 Orsay, France Patricia Duchambon − CNRS UMR9187, INSERM U1196,Institut Curie, PSL Research University, 91405 Orsay, France; CNRS UMR9187, INSERM U1196, Universite´ Paris Sud,Universite´ Paris Saclay, 91405 Orsay, France

Vanessa Masson − Institut Curie, PSL Research University, Centre de Recherche, Laboratoire

de Spectrometrie de Masse Proteomique, Paris 75248, France

Damarys Loew − Institut Curie, PSL Research University,Centre de Recherche, Laboratoire de Spectrometrie de Masse Proteomique, Paris 75248, France

Sophie Bombard − CNRS UMR9187, INSERM U1196, Institut Curie, PSL Research University, 91405 Orsay, France;CNRS UMR9187, INSERM U1196, Universite´ Paris Sud, Universite´ Paris Saclay, 91405 Orsay, France; orcid.org/790 0000-0002-6543-4497

Complete contact information is available at

Author Contributions

A.S., S.B., and M.-P.T.-F. designed this study. V.M. carried out the MS experimental work, and D.L. supervised MS and data analysis. P.D. produced recombinant NCL and optimized the protocol. A.S., S.B., and M.-P.T.-F. wrote the manuscript.

Funding

This work was supported by "Re´gion Ile-de-France" and Fondation pour la Recherche Me´dicale grants (to D.L.) and by the Agence Nationale de la Recherche (Grant-in-Aid ANR-14- CE35-0003-02, postdoctoral fellowship to A.S.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Prof. Nancy Maizels (University of Washington, Seattle, WA) for the generous gift of the plasmid expression vector for nucleolin, Dr. Alain Nicolas and Dr. Arturo Londono (UMR-CNRS 3244, Institut Curie) for the design of the CEB25 sequences, long term collaboration, and fruitful discussions, Dr. Anton Granzhan, Dr. Daniela Verga, and Dr. Claire Beauvineau (UMR9187-U1196, Institut Curie) for their help with the figures and CD and FRET melting
814 measurements.

■ ABBREVIATIONS

G4, G-quadruplex; NCL, nucleolin; RBD, RNA binding domains; RGG, arginine-glycine-glycine; BrdU, 5-bromo-2-deoxyuridine; PDB, Protein Data Bank.

■ REFERENCES

(1) Neidle, S., Balasubramanian, S. *Quadruplex Nucleic Acids*, Royal Society of Chemistry, 2007.
(2) Society of Chemistry, 2007.
(3) Rhodes, D., Lipps, H. J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res. 43*, 8627-8637.
(4) Chen. Y., Yang, D. (2012) Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. *Curr. Protoc. Nucleic Acid Chem.* Chapter 17:Unit17.5.
(5) Bedrat, A., Lacroix, L., Mergny, J. L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746-1759.
(6) Huppert, J. L., Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 33, 2908−2916.
(7) Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33, 877–881.
(8) Hansel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., Tannahill, D., Balasubramanian, S. (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* 48, 1267−1272.
(9) Piazza, A., Boulé, J. B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M. P., Nicolas, A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acids Res.* 38, 4337-4348.
(10) Piazza, A., Adrian, M., Samazan, F., Heddi, B., Hamon, F., Serero, A., Lopes, J., Teulade-Fichou, M. -P., Phan, A. T., Nicolas, A. (2015) Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-

forming minisatellites. *EMBO J.* 34, 1718-1734.
(11) Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M. Jr., Mergny, J. L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.* 44, 1989–2006.
(12) Bharti, S. K., Sommers, J. A., George, F., Kuper, J., Hamon, F., Shin-ya, K., Teulade-Fichou, M. P., Kisker, C., Brosh, R. M. Jr. (2013) Specialization among iron-sulfur cluster helicases to resolve G-quadruplex DNA structures that threaten genomic stability. *J. Biol. Chem.* 288, 28217–28229.
(13) Valton, A. L., Prioleau, M. N. (2016) G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet.* 32, 697-706.
(14) Cammas, A., Millevoi, S. (2017) RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* 45, 1584-1595.
(15) Heddi, B., Cheong, V. V., Martadinata, H., Phan, A. T. (2015) Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proc. Natl. Acad. Sci. USA.* 112, 9608-9613.
(16) You, H., Lattmann, S., Rhodes, D. & Yan, J. (2017) RHAU helicase stabilizes G4 in its nucleotide-free state and destabilizes G4 upon ATP hydrolysis. *Nucleic Acids Res.* 45, 206–214.
(17) Lattmann, S., Giri, B., Vaughn, J. P., Akman, S. A., Nagamine, Y. (2010) Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.* 38, 6219-6233.
(18) Chen, M. C., Tippana, R., Demeshkina, N. A., Murat, P., Balasubramanian, S., Myong, S., Ferré-D'Amaré, A. R. (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. Nature. 558, 465-469.
(19) Voter, A. F., Qiu, Y., Tippana, R., Myong, S., Keck, J. L. (2018) A guanine-flipping and sequestration mechanism for G-quadruplex unwinding by RecQ helicases. *Nat. Commun.* 9, 4201.
(20) Huber, M. D., Duquette, M. L., Shiels, J. C., Maizels, N. (2006) A conserved G4 DNA binding domain in RecQ family helicases. *J. Mol. Biol.* 358, 1071-1080.
(21) Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., Lipps, H. J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* 12, 847-854.
(22) Sissi, C., Gatto, B., Palumbo, M., (2011) The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie.* 93, 1219-1230.
(23) Thandapani, P., Song, J., Gandin, V., Cai, Y., Rouleau, S. G., Garant, J. M., Boisvert, F. M., Yu, Z., Perreault, J. P., Topisirovic, I., Richard, S., (2015) Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. *Elife.* 4.
(24) Cogoi, S., Paramasivam, M., Spolaore, B., 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.
(25) Takahama, K., Sugimoto, C., Arai, S., Kurokawa, R., Oyoshi, T., (2011) Loop lengths of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's sarcoma. *Biochemistry* 50, 5369-5378.
(26) Liu, X., Ishizuka, T., Bao, H. L., Wada, K., Takeda, Y., Iida, K., Nagasawa, K., Yang, D., Xu, Y., (2017) Structure-Dependent Binding of hnRNPA1 to Telomere RNA *J. Am. Chem. Soc.* 139, 7533-7539.
(27) Phan, A. T., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R. B., Patel, D. J. (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* 18, 796-804.
(28) Huang, Z. L., Dai, J., Luo, W. H., Wang, X. G., Tan, J. H., Chen, S. B., Huang, Z. S. (2018) Identification of G-Quadruplex-Binding Protein from the Exploration of RGG

Motif/G-Quadruplex Interactions. *J. Am. Chem. Soc.* 140, 17945-17955.
(29) McRae, E. K. S., Booy, E. P., Padilla-Meier, G. P., McKenna, S. A. (2017) On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids. *J. Nucleic Acids.* 2017:9675348.
(30) Herdy, B., Mayer, C., Varshney, D., Marsico, G., Murat, P., Taylor, C., D'Santos, C., Tannahill, D., Balasubramanian, S. (2018) Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res.* 46, 11592-11604.
(31) Williams, P., Li, L., Dong, X., Wang, Y. (2017) Identification of SLIRP as a G Quadruplex-Binding Protein. *J. Am. Chem. Soc.* 139, 12426-12429.
(32) Federici, L., Arcovito, A., Scaglione, G. L., Scaloni, F., Lo Sterzo, C., Di Matteo, A., Falini, B., Giardina, B., Brunori, M. (2010) Nucleophosmin C-terminal leukemia-associated domain interacts with G-rich quadruplex forming DNA. *J. Biol. Chem.* 285, 37138-37149.
(33) Ariyo, E. O., Booy, E. P., Dzananovic, E., McRae, E. K., Meier, M., McEleney, K., Stetefeld, J., McKenna, S. A. (2017) Impact of G-quadruplex loop conformation in the PITX1 mRNA on protein and small molecule interaction. *Biochem. Biophys. Res. Commun.* 487, 274-280.
(34) Scholz, O., Hansen, S., Plückthun, A., (2014) G-quadruplexes are specifically recognized and distinguished by selected designed ankyrin repeat proteins. *Nucleic Acids Res.* 42, 9182-9194.
(35) González, V., Guo, K., Hurley, L., Sun, D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* 284, 23622-23635.
(36) González, V., Hurley, L. H. (2010) The c-MYC NHE III(1): function and regulation. *Annu. Rev. Pharmacol. Toxicol.* 50, 111-129.
(37) Dickerhoff, J., Onel, B., Chen, L., Chen, Y., Yang, D. (2019) Solution Structure of a MYC Promoter G-Quadruplex with 1:6:1 Loop Length. *ACS Omega* 4, 2533-2539.
(38) Lago, S., Tosoni, E., Nadai, M., Palumbo, M., Richter, S. N. (2017) The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochim. Biophys. Acta* 1861, 1371-1381.
(39) Hanakahi, L. A., Sun, H. & Maizels, N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.* 274, 15908–15912.
(40) Bates, P. J., Kahlon, J. B., Thomas, S. D., Trent, J. O., Miller, D. M. (1999) Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. *J. Biol. Chem.* 274, 26369-26377.
(41) Ginisty, H., Sicard, H., Roger, B., and Bouvet, P. (1999) Structure and functions of nucleolin. *J. Cell Sci.* 112, 761–772.
(42) González, V., Hurley, L. H. (2010) The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry 49*, 9706-9714.
(43) Tosoni, E., Frasson, I., Scalabrin, M., Perrone, R., Butovskaya, E., Nadai, M., Palu, G., Fabris, D., Richter, S. N. (2015) Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* 43, 8884–8897.
(44) Lista, M. J., Martins, R. P., Billant, O., Contesse, M. A., Findakly, S., Pochard, P., Daskalogianni, C., Beauvineau, C., Guetta, C., Jamin, C., Teulade-Fichou, M. P., Fåhraeus, R., Voisset, C., Blondel, M. (2017) Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. *Nat. Commun.* 8:16043.
(45) Amrane, S., Adrian, M., Heddi, B., Serero, A., Nicolas, A., Mergny, J. L., Phan, A. T. (2012) Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite. *J. Am. Chem. Soc.* 134: 5807–5816.
(46) Singh, S., Berroyer, A., Kim, N. The RGG Domain of the Yeast Nucleolin Nsr1 Is Required for the Genome Instability Associated with Co-transcriptionally Formed G4 DNA bioRxiv 802876; doi: https://doi.org/10.1101/802876
(47) Poullet, P., Carpentier, S., Barillot, E. (2007) myProMS, a web server for management

and validation of mass spectrometry-based proteomic data. *Proteomics* 7, 2553–2556.

(48) Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., Hermjakob, H. (2016) 2016 update of the PRIDE database and related tools. *Nucleic Acids Res.* 44(D1): D447-D456

(49) Do, N. Q., Phan, A. T. (2012) Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chem.- Eur. J.* 18, 14752–14759.

(50) Dai, J., Chen, D., Jones, R. A., Hurley, L. H., Yang, D. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* 34, 5133-5144.

(51) Willis, M.C., Hicke, B.J., Uhlenbeck, O.C., Cech, T.R., Koch, T.H. (1993) Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science* 262, 1255-1257.

(52) Dietz, T. M., Koch, T. H. (1987) Photochemical coupling of 5-bromouracil to tryptophan, tyrosine and histidine, peptide-like derivatives in aqueous fluid solution. *Photochem. Photobiol.* 46, 971-978.

(53) Saha, A., Bombard, S., Granzhan, A., Teulade-Fichou, M. P. (2018) Probing of G-Quadruplex Structures via Ligand-Sensitized Photochemical Reactions in BrU-Substituted DNA. *Sci. Rep.* 8, 1–14.

(54) Schmidt, C., Kramer, K., Urlaub, H. (2012) Investigation of protein-RNA interactions by mass spectrometry—techniques and applications. *J. Proteomics* 75, 3478–3494

(55) Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K. L., Hentze, M. W., Kohlbacher, O., Urlaub, H. (2014) Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* 11, 1064–1070.

(56) Kellermann, G., Dingli, F., Masson, V., Dauzonne, D., Ségal-Bendirdjian, E., Teulade-Fichou, M. P., Loew, D., Bombard, S. (2017) Exploring the mechanism of inhibition of human telomerase by cysteine-reactive compounds. *FEBS Lett.* 591, 863-874.

(57) Arumugam, S., Miller, M.C., Maliekal, Bates, P.J., Trent, J.O. Lane A.N. (2010) Solution structure of the RBD1,2 domains from human nucleolin *J. Biomol. NMR*, 47, 79-83.

(58) Johansson, C., Finger, L. D., Trantirek, L., Mueller, T. D., Kim, S., Laird-Offringa, I. A., Feigon J. (2004) Solution Structure of the complex formed by the two N-terminal RNA-binding domains of Nucleolin and a pre-rRNA target *J.Mol.Biol*. 337, 799-816.

(58) Neidle, S. (2016), Quadruplex Nucleic Acids as novel therapeutic targets - *J.Med.Chem*. 59, 5987-6011.

(59) Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C. B., Proffitt, C., Trent, K., Whitten, J. P., Lim, J. K., Von Hoff, D., Anderes, K., Rice, W. G. (2009) Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* 69, 7653-7661.

(60) Norseen, J., Johnson, F. B., Lieberman, P. M. (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.* 83, 10336-10346.

(61) Reznichenko, O., Quillévéré, A., Martins, R. P., Loaëc, N., Kang, H., Lista, M. J., Beauvineau, C., González-García, J., Guillot, R., Voisset, C., Daskalogianni, C., Fåhraeus, R., Teulade-Fichou, M. P., Blondel, M., Granzhan, A. (2019) Novel cationic bis(acylhydrazones) as modulators of Epstein-Barr virus immune evasion acting through disruption of interaction between nucleolin and G-quadruplexes of EBNA1 mRNA. *Eur. J. Med. Chem.* 178, 13-29.

(62) Beauvineau, C., Guetta, C., Teulade-Fichou, M. P., Mahuteau-Betzer, F. (2017) PhenDV, a turn-off fluorescent quadruplex DNA probe for improving the sensitivity of drug screening assays. *Org. Biomol. Chem.* 15, 7117-7121.

(63) De Rache, A., Mergny, J. L. (2015) Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay, *Biochimie.* 115, 194-202.

(64) Maleki, P., Mustafa, G., Gyawali, P., Budhathoki, J. B., Ma, Y., Nagasawa, K.,

Balci, H. (2019) Quantifying the impact of small molecule ligands on G-quadruplex stability against Bloom helicase. *Nucleic Acids Res.* 47, 10744-10753.

(65) De Cian, A., Guittat, L., Kaiser, M., Saccà, B., Amrane, S., Bourdoncle, A., Alberti, P., Teulade-Fichou, M. P., Lacroix, L., Mergny, J. L. (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods* 42, 183-195.

(66) Rizzo, A., Iachettini, S., Zizza, P., Cingolani, C., Porru, M., Artuso, S., Stevens, M., Hummersone, M., Biroccio, A., Salvati, E., Leonetti, C., (2014) Identification of novel RHPS4-derivative ligands with improved toxicological profiles and telomere-targeting activities. *J. Exp. Clin. Cancer Res.* 33, 81.

(67) Yangyuoru, P. M., Di Antonio, M., Ghimire, C., Biffi, G., Balasubramanian, S., Mao, H. (2015) Dual Binding of an Antibody and a Small Molecule Increases the Stability of TERRA G-Quadruplex. *Angew Chem Int. Ed.* 127, 924-927

(68) Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P., Phan, A. T. (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC3. *Angew. Chem. Int. Ed. Engl.* 53, 999-1002.

(69) Le, H. T., Miller, M. C., Buscaglia, R., Dean, W. L., Holt, P. A., Chaires, J. B., Trent, J. O. (2012) Not all G-quadruplexes are created equally: an investigation of the structural polymorphism of the c-Myc G-quadruplex-forming sequence and its interaction with the porphyrin TMPyP4. *Org. Biomol. Chem.* 10, 9393-9404. Le, H. T.,

(70) Hanakahi, L.A.; Sun, H.; Maizels, N. (1999),High affinity Interactions of Nucleolin with G-G paired rDNA *J.Biol.Chem*.274, 22, 15908-15913.

(71) Ozdilek, B. A., Thompson, V. F., Ahmed, N. S., White, C. I., Batey, R. T., Schwartz, J. C. (2017) Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.* 45, 7984-7996.

(72) Ghosh, M., Singh, M. (2018) RGG-box in hnRNPA1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain *Nucleic Acids Res.* 46, 10246-10261.

(73) Vasilyev, N., Polonskaia, A., Darnell, J. C., Darnell, R. B., Patel, D. J., Serganov, A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci.* 112, E5391–E5400.

(74) Puig Lombardi, E; Verga, D., Holmes, A., Teulade-Fichou, M.-P., Nicolas, A., Londono-Vallejo, A. (2019) Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species *Nucleic Acids Res.* 47, 12, 6098-6113.

(75) Mendoza,O.; Bourdoncle, A.; Boulé, J.-B.; Brosh Jr. R.M. Merngy J.-L. (2016) G-quadruplexes and Helicases *Nucleic Acids Res.* 44, 5, 1989-2006.

(76) Agrawal,,P.,Hatzakis, E.;Guo, K.; Carver, M. Yang, D. (2013) 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.* 41, 22,10584-10592.

(77) Wang,L. Wang, Q.M. Wang, Y.R.; Xi, X.G.; Hou, X.M. (2018) DNA-unwinding activity of Saccharomyces Cerevisiae Pif1 is modulated by thermal stability , folding conformation and loop length of G-quadruplex DNA *J.Biol.Chem*. 293,18504-18515.