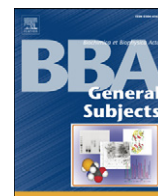




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids<sup>☆</sup>



Sara Lago<sup>a</sup>, Elena Tosoni<sup>a</sup>, Matteo Nadai<sup>a</sup>, Manlio Palumbo<sup>b</sup>, Sara N. Richter<sup>a,\*</sup>

<sup>a</sup> Department of Molecular Medicine, University of Padua, via Gabelli 63, 35121 Padua, Italy

<sup>b</sup> Department of Pharmaceutical Sciences, University of Padua, via Marzolo 5, 35131 Padua, Italy

## ARTICLE INFO

### Article history:

Received 16 October 2016

Received in revised form 25 November 2016

Accepted 26 November 2016

Available online 30 November 2016

### Keywords:

G-quadruplexes

Nucleolin

Loops

Conformation

Nucleic acids

Protein-nucleic acid binding

## ABSTRACT

**Background:** G-quadruplexes (G4s) are four-stranded nucleic acid structures that form in G-rich sequences. Nucleolin (NCL) is a cellular protein reported for its functions upon G4 recognition, such as induction of neurodegenerative diseases, tumor and virus mechanisms activation. We here aimed at defining NCL/G4 binding determinants.

**Methods:** Electrophoresis mobility shift assay was used to detect NCL/G4 binding; circular dichroism to assess G4 folding, topology and stability; dimethylsulfate footprinting to detect G bases involved in G4 folding.

**Results:** The purified full-length human NCL was initially tested on telomeric G4 target sequences to allow for modulation of loop, conformation, length, G-tract number, stability. G4s in promoter regions with more complex sequences were next employed. We found that NCL binding to G4s heavily relies on G4 loop length, independently of the conformation and oligonucleotide/loop sequence. Low stability G4s are preferred. When alternative G4 conformations are possible, those with longer loops are preferred upon binding to NCL, even if G-tracts need to be spared from G4 folding.

**Conclusions:** Our data provide insight into how G4s and the associated proteins may control the ON/OFF molecular switch to several pathological processes, including neurodegeneration, tumor and virus activation. Understanding these regulatory determinants is the first step towards the development of targeted therapies.

**General significance:** The indication that NCL binding preferentially stimulates and induces folding of G4s containing long loops suggests NCL ability to modify the overall structure and steric hindrance of the involved nucleic acid regions. This protein-induced modification of the G4 structure may represent a cellular mechanosensor mechanism to molecular signaling and disease pathogenesis. This article is part of a Special Issue entitled "G-quadruplex" Guest Editor: Dr. Concetta Giancola and Dr. Daniela Montesarchio.

© 2016 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

G-quadruplexes (G4s) are unique four-stranded nucleic acid structures that may form in guanine-rich sequences [1]. Four Gs bind via Hoogsteen-type hydrogen bonds to yield G-quartets that in turn stack on top of each other to form the G4. The presence of K<sup>+</sup> cations specifically supports G4 formation and stability. G4s are highly polymorphic, in terms of strand stoichiometry (forming both inter- and intramolecular structures), strand orientation/topology and loop sequence, length and position [2]. G4s have been shown in humans and other eukaryotes [3,4], in prokaryotes [5–8] and in viruses [9,10]. From a functional point of view, in humans G4s play key regulatory and pathological roles,

including transcriptional regulation of gene promoters and enhancers, translation, chromatin epigenetic regulation, DNA recombination [11–13]. Human G4 DNA motifs have been reported to be associated with recombination prone regions [14] and to show mutational patterns that preserved the potential to form G4 DNA structures [15]. G4 misregulation has been associated with a number of relevant human diseases, such as the amyotrophic lateral sclerosis [16], Alzheimer [17] and fragile X syndrome [18], in which expansion of G4-forming regions has been reported. In viruses, G4s have been proposed both as virus cycle modulators and antiviral targets [10]. In the human immunodeficiency virus-1 (HIV-1), the etiologic agent of the acquired immune deficiency syndrome (AIDS), functionally significant G4s in the Nef coding region [19] and in the unique long terminal repeat (LTR) promoter [20–22] have been identified. These studies have shown that G4 folding at the LTR promoter decreases viral transcription with an effect that is augmented by G4 ligands [23,24]. G4s have also been reported in the herpes simplex virus 1 [25], Epstein–Barr virus [26,27], SARS coronavirus [28], human papilloma, hepatitis C, Zika and Ebola virus genomes [29,30,31,32].

**Abbreviations:** G-quadruplexes, G4s; nucleolin, NCL; circular dichroism, CD; electrophoresis mobility shift assay, EMSA; temperature of melting, T<sub>m</sub>; dimethylsulfate, DMS; polyethylene glycol, PEG.

<sup>☆</sup> This article is part of a Special Issue entitled "G-quadruplex" Guest Editor: Dr. Concetta Giancola and Dr. Daniela Montesarchio.

\* Corresponding author.

E-mail address: [sara.richter@unipd.it](mailto:sara.richter@unipd.it) (S.N. Richter).

**Table 1**  
Sequences of oligonucleotides used in this study.

Name	Length (nts)	Sequence 5'–3'
(GGGTTA) <sub>4</sub>	21	GGGTTAGGGTTAGGGTTAGGG
(GGGTTA) <sub>6</sub>	36	GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
(GGGTTA) <sub>8</sub>	54	TCGAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAC
(GGGTTT) <sub>4</sub>	21	GGGTTTGGGTTTGGGTTTGGG
(GGGTTT) <sub>6</sub>	36	GGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTT
(GGGTTT) <sub>8</sub>	48	GGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTT
(GGGTTT) <sub>5</sub>	27	GGGTTTGGGTTTGGGTTTGGGTTTGGG
n5M1	27	TTTTTTGGGTTTGGGTTTGGGTTTGGG
n5M2	27	GGGTTTTTTTTTGGGTTTGGGTTTGGG
n5M3	27	GGGTTTGGGTTTTTTTTGGGTTTGGG
n5M4	27	GGGTTTGGGTTTGGGTTTTTTTTTTGGG
n5M5	27	GGGTTTGGGTTTGGGTTTGGGTTTTTTT
M43	43	TTTTTGGAGTCGTGTCGCGTGTGAGCGTGTGTAGTGGTTTTT
M30	30	TTTTTTTTTTTTTTTTTTTTTTTTTTTTT
M29i	29	AAAACAGACAGCTCACTGCGCTCACAAA
M29ii	29	TTTTGTGAGCGCGTGAGGTGTGTGTTTT
M26	26	CCTCCTCCTCTCTCCTCCTCCTCC
M21	21	TTTTTTTTTTTTTTTTTTTTTTTTT
M14	14	CGGCATGCGTGGCT
111	15	GGGTGGGTGGGTGGG
131	17	GGGTGGGTTTGGGTGGG
161	20	GGGTGGGTTTTTTGGGTGGG
313	19	GGGTTTGGGTGGGTTTGGG
333	21	GGGTTTGGGTTTGGGTTTGGG
363	24	GGGTTTGGGTTTTTTGGGTTTGGG
616	25	GGGTTTTTTGGGTGGGTTTTTTGGG
636	27	GGGTTTTTTGGGTTTGGGTTTTTTGGG
666	30	GGGTTTTTTGGGTTTTTTGGGTTTTTTGGG
T111T	17	TGGGTGGGTGGGTGGGT
T131T	19	TGGGTGGGTTTGGGTGGGT
T161T	22	TGGGTGGGTTTTTTGGGTGGGT
c-kit1	21	AGGGAGGGCGCTGGGAGGAGG
bcl-2	39	AGGGGCGGCGCGGGAGGAAGGGGCGGGAGCGGGGCTG
c-myc	27	TGGGGAGGGTGGGGAGGGTGGGGAAGG
hTert 1–4	20	AGGGGAGGGGCTGGGAGGGG
hTert 5–12	46	AGGGGGCTGGGCCGGGACCCGGGAGGGGTCGGGACGGGGCGGGGT
hTert 1–12	70	AGGGGAGGGGCTGGGAGGGCCCGGAGGGGCTGGGCCGGGACCCGGGAGGGGTCGGGACGGGGCGGGGT

The biological effects of G4s are often modulated by the interaction with proteins [33]: the shelterin complex proteins are involved in telomere homeostasis [34], and other G4 interacting proteins either unfold or stabilize the G4 conformation [35,36]. Mutations in some of these G4-interacting proteins have been linked to genetic diseases, such as the Bloom's and Werner's syndromes, Fanconi anemia [37,38], and cancer [39].

Among G4 interacting proteins, nucleolin (NCL) is the protein most often reported for its biological functions upon G4 recognition. It is widely believed that NCL plays a chaperone role by helping the correct folding of complex nucleic acids structures. Indeed, NCL has been shown to display a marked preference for both endogenous and exogenous G-rich sequences that can fold into G4 [40]. It has been reported that binding of NCL to the endogenous (GGGGCC)<sub>n</sub> hexanucleotide repeat expansion (HRE) in *C9orf72* is responsible for the initiation of molecular cascades that lead to neurodegenerative diseases [16]. At the promoter level, binding of NCL to G4 structures augments the basal effect of the folded conformation [41–44]. For example, in the case of HIV-1, it induces potent inhibition of viral transcription by stabilization of the LTR promoter [45]. NCL is mainly present in the nucleolus, but it can be found also in cell membranes and, upon stress stimuli, in the nucleoplasm and cytoplasm to some extent [46,47]. NCL has been initially described as a RNA binding protein, but recent evidence indicates the relevance, and in some case the preference, of its binding to DNA structures [45]. Because of NCL role in viral mechanisms, such as HIV cell attachment/entry and budding [48–50] and virus replication and transcription [51–56], NCL inhibitors have been developed [57–59]. Interestingly, one of these is a G4-forming oligonucleotide, which displays interesting HIV activity [60,61].

Despite the involvement of NCL/G4 complex in several human diseases, a deep understanding of the G4 structural features that drive binding to NCL is lacking. A certain degree of selectivity towards nucleic acids has been reported [41,45]. For instance, in the case of the HIV-1 LTR promoter, where three G4s can form (LTR-II, LTR-III and LTR-IV), NCL binds with high affinity to LTR-III and with intermediate affinity to LTR-II, whereas LTR-IV is not recognized [45].

To understand the structural requirements that stimulate NCL binding to G4s, we tested NCL binding towards different sets of G4 oligonucleotides, considering loop length, topology, sequence, number of G-runs, stability. We found that loop length was always the prevalent determinant of NCL binding to G4s, independently of the conformation and oligonucleotide/loop sequence. These findings both allow a deeper understanding of NCL/G4-regulated physiological and pathological mechanisms, and the possibility to develop new target-specific drugs.

## 2. Materials and methods

### 2.1. Oligonucleotides and proteins

Desalted oligonucleotides were purchased from Sigma-Aldrich (Milan, Italy). Oligonucleotides were denatured for 5 min at 95 °C and gradually cooled to room temperature to achieve proper G4 folding. The human recombinant nucleolin (NCL) is from Origene Technologies, Rockville, USA.

### 2.2. Circular dichroism (CD)

CD spectroscopy was used to probe G4 folding and conformation of the tested DNA oligonucleotides. Oligonucleotides were diluted to a

final concentration of 4  $\mu\text{M}$  in lithium cacodylate buffer (10 mM, pH 7.4), KCl or NaCl (100 mM) and 20% of PEG200 where indicated. After heat denaturation samples were folded at room temperature overnight. CD spectra were recorded on a Chirascan-Plus (applied Photophysics, Leatherhead, UK) equipped with a Peltier temperature controller using a quartz cell of 5 mm optical path length and an instrument scanning speed of 50 nm/min. The temperature was set at 20 °C and the spectra were recorded over a wavelength range of 230–320 nm. For  $T_m$  determination, spectra were recorded over a temperature range of 20–95 °C, with temperature increase of 5 °C/min.  $T_m$  values were calculated according to the van't Hoff equation, applied for a two-state transition from a folded to unfolded state. Acquired spectra were baseline corrected for signal contribution due to the buffer and the observed ellipticities converted to mean residue ellipticity ( $\theta$ ) =  $\text{deg} \times \text{cm}^2 \times \text{dmol}^{-1}$  (mol ellip).

### 2.3. Electrophoresis Mobility Shift Assay (EMSA)

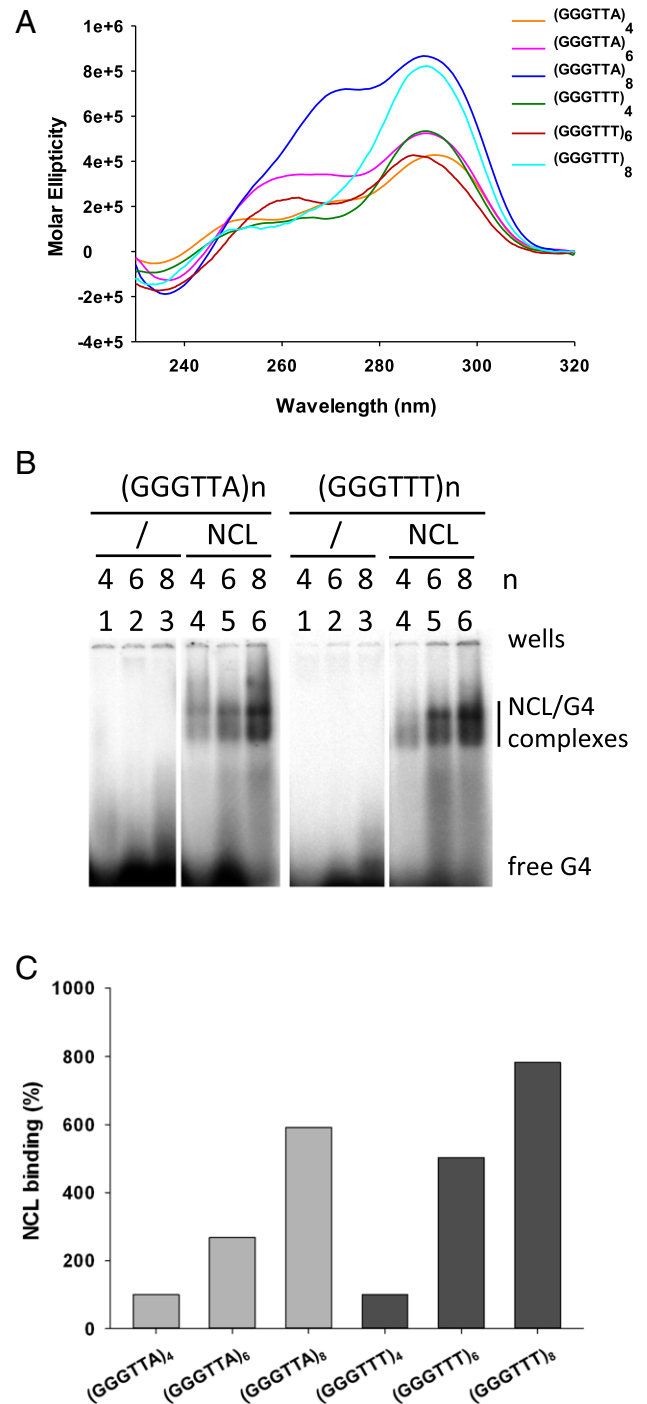
EMSA was used to detect G4-NCL interaction and the stoichiometry of G4 folding. Oligonucleotides were 5'-end labeled with [ $\gamma$ - $^{32}\text{P}$ ]ATP by T4 polynucleotide kinase after 30 min incubation at 37 °C. After DNA precipitation, labeled species were resuspended in lithium cacodylate buffer (10 mM, pH 7.4, KCl 100 mM or NaCl 100 mM and 20% (w/v) PEG200 where indicated), heat denatured and folded at room temperature.

For the determination of G4 stoichiometry samples were run in 20% non-denaturing polyacrylamide gels for ~24 h at 60 V. Gels were fresh impressed and visualized by phosphorimaging (Typhoon FLA 9000, GE Healthcare Europe, Milan, Italy).

For NCL binding to DNA, reactions were performed in binding buffer (Tris-HCl 20 mM pH 8, KCl 30 mM,  $\text{MgCl}_2$  1.5 mM, DTT 1 mM, glycerol 8%, protease inhibitor cocktail – Sigma Aldrich, Milan, Italy – 1%, NaF 5 mM,  $\text{Na}_3\text{VO}_4$  1 mM), labeled oligonucleotides (40 nM), human recombinant NCL (300 ng) and incubated for 1 h at 37 °C. Binding reactions were supplemented with 20% (w/v) PEG200 where indicated. To load the same amount of DNA in each reaction, counts per second emitted by the radiolabeled oligonucleotides were measured by a Geiger counter (Mini 900 Scintillation Monitor, Thermo Scientific™, Monza, Italy). Samples were loaded on 8% non-denaturing polyacrylamide gels for ~15 h at 40 V. Gels were dried using a gel dryer (Bio Rad Laboratories, Milan, Italy) and visualized by phosphorimaging (Typhoon FLA 9000, GE Healthcare Europe, Milan, Italy). Quantification of bound DNA molecules was performed by ImageQuant TL Software (GE Healthcare Europe, Milan, Italy). Each gel was performed twice to confirm quantification results; only results consistent throughout the repeated gels were retained and reported.

### 2.4. Dimethylsulfate footprinting (DMS)

The DNA substrate of interest was gel-purified before use, 5'-end-labeled with [ $\gamma$ - $^{32}\text{P}$ ]ATP by T4 polynucleotide kinase, purified using MicroSpin G-25 columns (GE Healthcare Europe, Milan, Italy), resuspended in lithium cacodylate buffer 10 mM pH 7.4 with or without KCl 100 mM, heat denatured and folded. Sample solutions were then treated with dimethylsulfate (DMS, 0.5% in ethanol) for 5 min and stopped by addition of 10% glycerol and  $\beta$ -mercaptoethanol. Samples were loaded onto a 16% native polyacrylamide gel and run until the desired resolution was obtained. DNA bands were localized via autoradiography, excised and eluted overnight. The supernatants were recovered, ethanol-precipitated and treated with piperidine 1 M for 30 min at 90 °C. Samples were dried in a speed-vac, washed with water, dried again, and resuspended in formamide gel loading buffer. Reaction products were analyzed on 20% denaturing polyacrylamide gels, visualized by phosphorimaging analysis, and quantified by ImageQuant TL software (GE Healthcare Europe, Milan, Italy).



**Fig. 1.** NCL binding to G4s with increasing number of G-tracts. (A) CD spectra of telomeric sequences (GGGTTA)<sub>n</sub> and (GGGTTT)<sub>n</sub>, where n = 4, 6 and 8. (B) EMSA analysis of NCL binding to telomeric sequences (GGGTTA)<sub>n</sub> - lane 4, 5, 6 left - and (GGGTTT)<sub>n</sub> - lane 4, 5, 6 right -. Lanes 1–4 are control reactions without NCL. (C) Quantification of NCL/G4 complex bands obtained in the EMSA shown in panel (B). Binding of lanes 4–6 left and 4–6 right was normalized for (GGGTTA)<sub>4</sub> and (GGGTTT)<sub>4</sub> NCL/G4 band intensity, respectively.

## 3. Results

### 3.1. Nucleolin binding to G-quadruplexes directly correlates with the number of G-tracts

To investigate the determinants required for efficient NCL binding, we initially tested the human telomeric sequence, formed by GGGTTA repeats, and a simplified sequence formed by GGGTTT repeats, which is

found in telomeres of ciliates of the genus *Paramecium* [62]. The interaction of NCL and its four RNA binding domains (RBD) and the Arg–Gly–Gly (RGG) domain to telomeric G4s has been previously reported [41,63,64]. Telomeric sequences are rather simple compared to the G4 sequences in oncogene and HIV-1 promoters and were thus chosen to allow assessment of oligonucleotide length, loops, repeats, conformation and sequence contribution to NCL binding. NCL/oligonucleotide binding was tested by electrophoresis mobility shift assays (EMSA) with  $^{32}$ P-labelled nucleic acids and the purified full-length human NCL protein [45].

We initially tested telomeric sequences  $((GGGTTA)_n$  and  $((GGGTTT)_n$  with increasing number of G-tracts, i.e.  $n = 4, 6$  and 8 (Table 1).

All sequences folded into mixed type G4s as assessed by circular dichroism (CD) analysis: in fact, all oligonucleotides displayed positive peaks around 290 nm and 260 nm, which are signatures of hybrid/mixed type G4 topologies (Fig. 1A). NCL binding to the G4 oligonucleotides was visualized as two bands running slower than the free unbound DNA. The presence of the two bands has been previously demonstrated to be due to the N-terminus self-cleaving activity of NCL [45]: thus the two bands were ascribed to the binding of the full-length (upper) and the N-terminus-minus (lower) NCL protein to the G4s (Fig. 1B). In these conditions, NCL binding was directly proportional to the length of the telomeric sequences and the difference in loop sequence between human and ciliate G4s did not modify this trend (Fig. 1C).

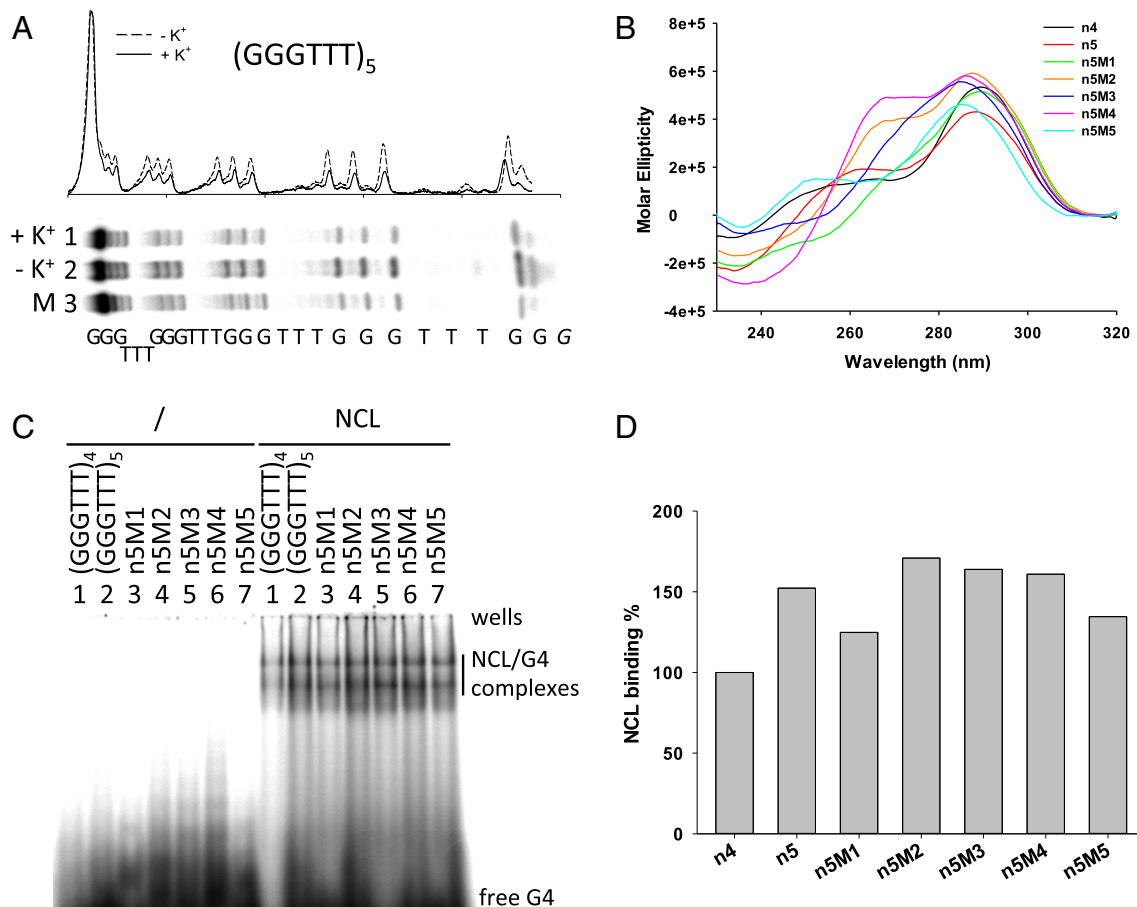
**Table 2**

Stability ( $T_m$  values) of oligonucleotides of the  $(GGGTTT)_5$  series with G-tracts replaced with Ts.  $T_m$  were measured by CD thermal unfolding and reported at the positive CD peaks, signatures of the parallel (265 nm) and antiparallel (285 nm) G4 conformation contribution.

Oligonucleotide	$T_m$ (265 nm) ( $^{\circ}$ C)	$T_m$ (285 nm) ( $^{\circ}$ C)
$(GGGTTT)_4$	$66.6 \pm 0.4$	$66.1 \pm 0.4$
$(GGGTTT)_5$	$65.5 \pm 1.1$	$64.5 \pm 0.5$
n5M1	$58.0 \pm 0.8$	$55.7 \pm 0.5$
n5M2	$53.4 \pm 0.5$	$54.5 \pm 0.2$
n5M3	$56.7 \pm 0.3$	$56.6 \pm 0.1$
n5M4	$52.0 \pm 0.3$	$53.2 \pm 0.1$
n5M5	/	$60.4 \pm 0.4$

### 3.2. Long loops favor NCL binding to G4s

While the sequences with 4 and 8 G-tracts can fold into 1 and 2 stacked G4s, respectively, sequences with 5 and 6 G-tracts can fold into one G4 structure leaving one or two spare G-tracts. To test whether there was a preferred involvement of G-tracts in these conformationally dynamic sequences, oligonucleotide  $(GGGTTT)_5$  (Table 1) was folded in the presence and absence of  $K^+$  and treated with DMS to highlight G bases protected from DMS-induced methylation when in the G4 folded state. The analysis showed that all Gs were equally protected during folding (Fig. 2A) indicating the equilibrium of multiple G4 structures



**Fig. 2.** Analysis of NCL binding to alternative G4 structures formed by the  $(GGGTTT)_5$  telomeric sequence (A) DMS protection analysis of  $(GGGTTT)_5$  folded in the presence or absence of  $K^+$ . M is a marker lane obtained with the Maxam and Gilbert sequencing protocol. Oligonucleotide sequence is shown below the gel image. The densitogram above the gel shows quantification of cleaved bands intensity. (B) CD spectra of  $(GGGTTT)_5$  and mutant derivatives where each G-tract was systematically replaced by Ts. (C) EMSA analysis of NCL binding to  $(GGGTTT)_4$ ,  $(GGGTTT)_5$  and mutant oligonucleotides. Oligonucleotide identity is shown above each lane in the gel image. (D) Quantification of NCL/G4 complex bands obtained in the EMSA in panel (C). n4 and n5 stand for  $(GGGTTT)_4$  and  $(GGGTTT)_5$ , respectively.

that involved alternative G-tracts. Similar results have been previously reported [65]. In particular, in the case of the (GGGTTT)<sub>5</sub> oligonucleotide, 5 alternative G4 structures, which comprise G-tracts 1-2-3-4, 2-3-4-5, 1-3-4-5, 1-2-4-5, 1-2-3-5, may be envisaged. To test NCL binding to each of these sequences, mutant oligonucleotides where each G-tract was replaced by Ts were employed (Table 1). The (GGGTTT)<sub>4</sub> oligonucleotide was used as a control since it displays the same sequence and number of G-tracts as the mutant oligonucleotides, but lacks extra T bases in the loops or at the 5'- or 3'-end (Table 1). All mutant sequences were shown to maintain the G4 folding by CD analysis: oligonucleotides displayed mainly a hybrid/mixed type topology with different levels of parallel/antiparallel contributions (Fig. 2B). The stability of the mutant sequences forming internal long loops was slightly lower than that of the wild-type sequence, as gathered by T melting ( $T_m$ ) values measured by CD thermal unfolding (Table 2).

EMSA analysis of NCL binding to the mutant sequences showed mild preferential binding towards oligonucleotides where internal G-tracts were mutated (Fig. 2C and D). Binding to sequences with external Ts was consistently lower, but higher than that to the (GGGTTT)<sub>4</sub> control. The (GGGTTT)<sub>5</sub> oligonucleotide displayed an average binding to all other mutant sequences. These data indicate that NCL preferentially binds G4s with at least one loop longer than 3 nts (e.g. 9 nts, compare n5M2, n5M3 and n5M4 binding, lanes 4–6 NCL, to n5M1 and n5M5, lanes 3 and 7 NCL, Fig. 2C) and with free nts at the 5' or 3' termini (compare n5M1, lane 3 NCL, and n5M5, lane 7 NCL, to (GGGTTT)<sub>4</sub>, lane 1 NCL, Fig. 2C).

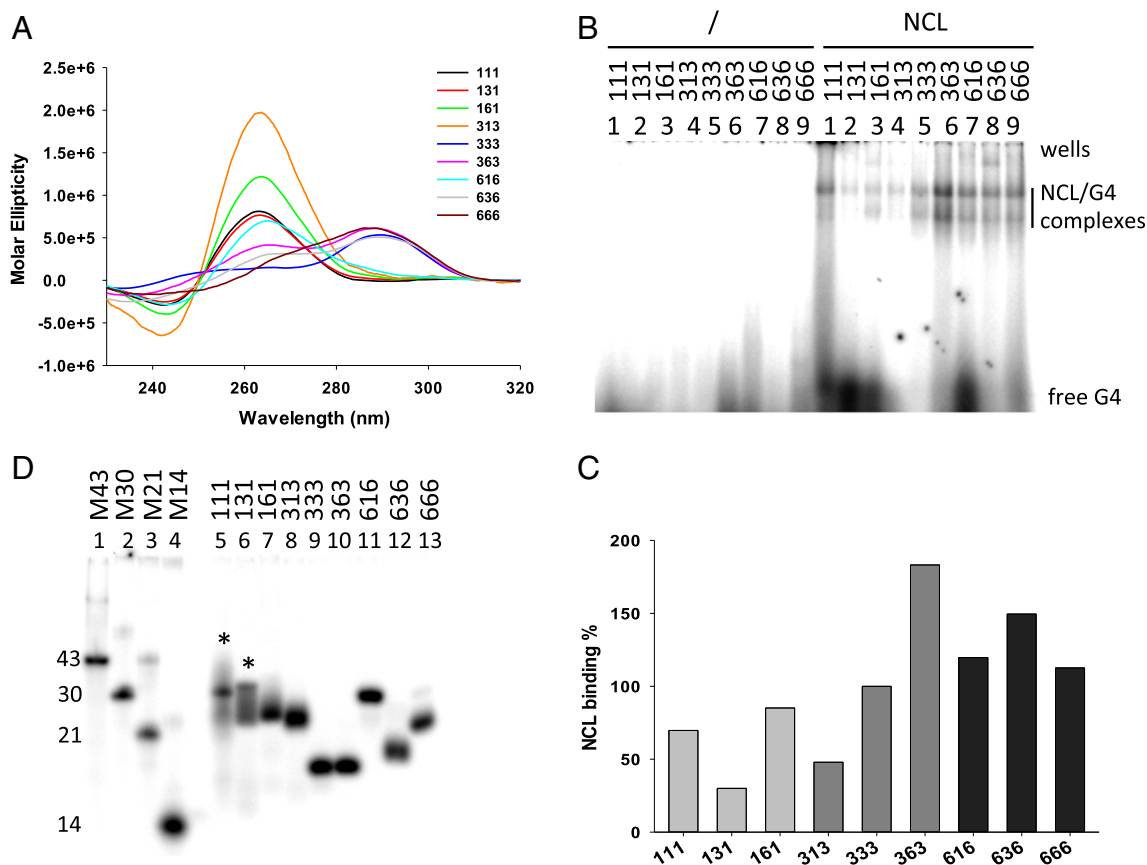
**Table 3**

Stability ( $T_m$  values) of oligonucleotides of the (GGGTTT)<sub>4</sub> series with variable-length loops.  $T_m$  were measured by CD thermal unfolding and reported at the positive CD peaks, signatures of the parallel (265 nm) and antiparallel (285 nm) G4 conformation contribution.

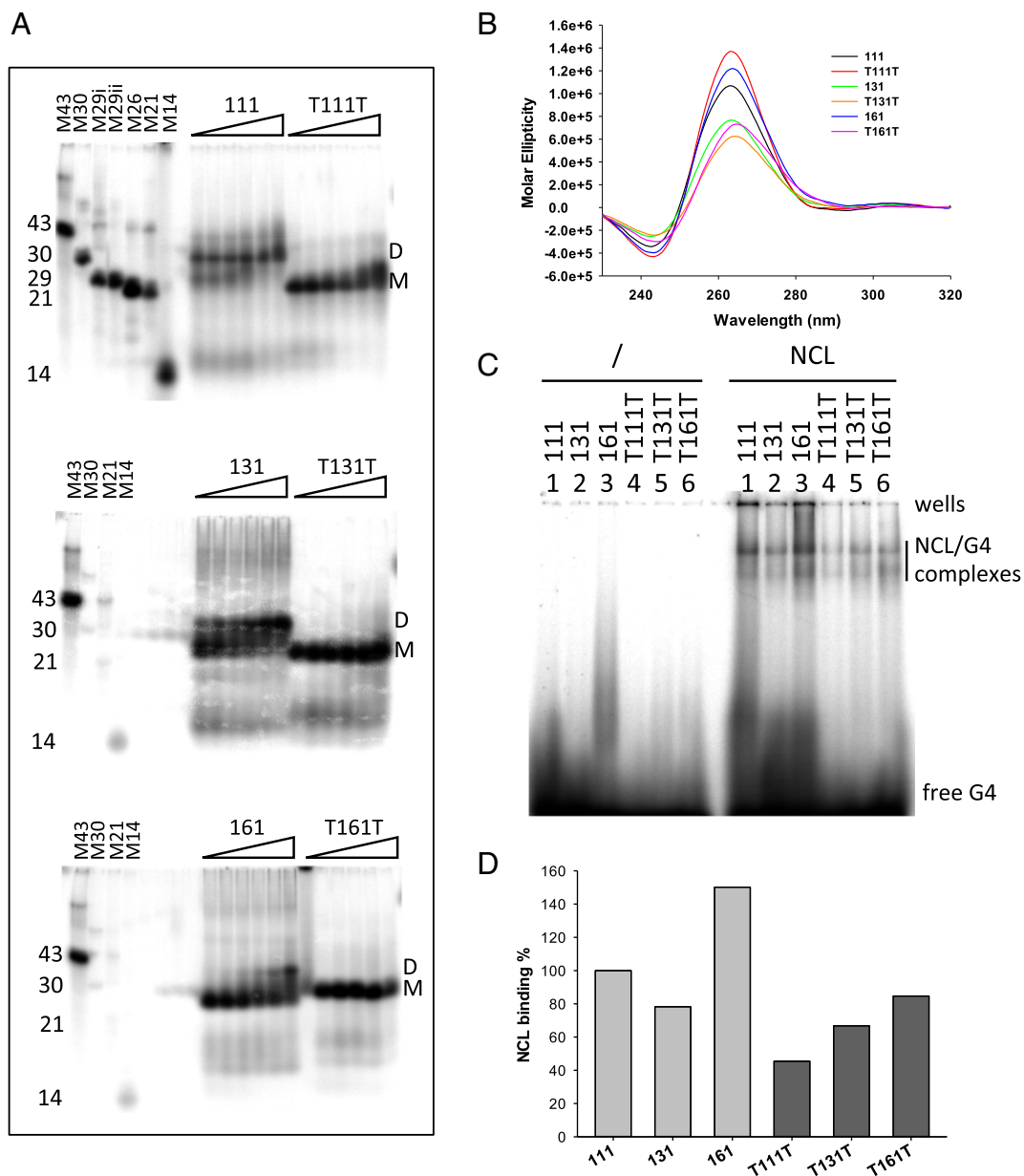
Oligonucleotide	$T_m$ (265 nm) (°C)	$T_m$ (285 nm) (°C)
111	>90	/
131	>90	/
161	72.1 ± 0.3	/
313	66.1 ± 0.4	/
333	66.6 ± 0.4	66.1 ± 0.4
363	63.0 ± 0.2	64.0 ± 0.2
616	52.7 ± 0.1	/
636	52.9 ± 0.3	53.0 ± 0.1
666	45.3 ± 0.3	44.4 ± 0.1

### 3.3. More than two long or extremely short loops destabilize NCL interaction with G4s

To explore the impact of loop length on NCL binding to G4s, three series of oligonucleotides with different loop-length combinations were tested: in each series, the two lateral loops were kept constant (1, 3 or 6 nts in each series) and the central loop was modified (1, 3 or 6 nts) (Table 1). All oligonucleotides folded into G4: in this case some sequences displayed a neat parallel topology (111, 131, 161, 313, 616, Fig. 3A), while other a mixed/hybrid type one (333, 363, 636, 666, Fig. 3A), in keeping with the notion that G4s with 1-nt loops are constrained



**Fig. 3.** NCL binding to G4s with systematic elongation of loop length. (A) CD spectra of oligonucleotides with two constant lateral (1, 3 or 6 nts) and variable central loops (1, 3 or 6 nts). (B) EMSA analysis of NCL binding to each G4 with different loop-length combination. Oligonucleotides identity is shown above each lane. (C) Quantification of NCL/G4 complex bands obtained in the EMSA in panel (B). (D) Native gel electrophoresis of G4-folded oligonucleotides. Oligonucleotides identity is shown above each lane. Lanes 1–4 are marker oligonucleotides of known length and unable to fold into secondary structures. The symbol \* indicates oligonucleotides showing two bands with different migration rate.



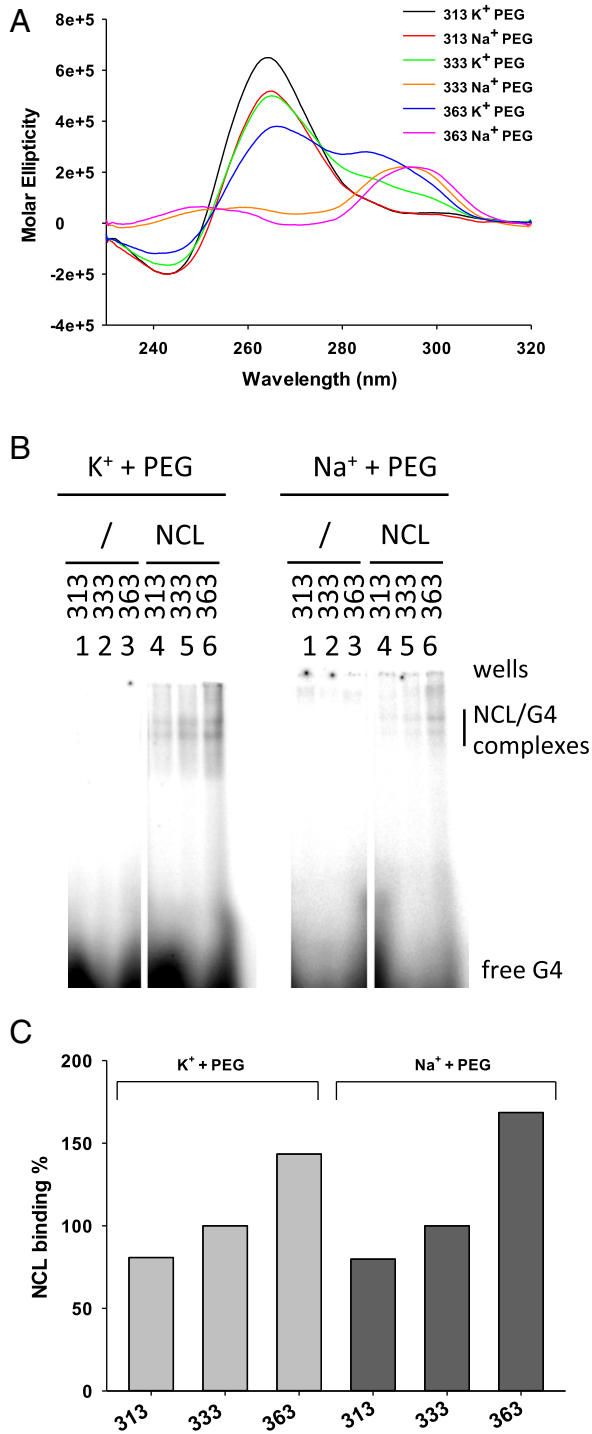
**Fig. 4.** Analysis of monomeric and dimeric G4 folding of relevant oligonucleotides. (A) Native gel electrophoresis at increasing oligonucleotide concentration (20, 45, 120, 415, 1600, 6200 nM). Samples identity is indicated above the gel image. Marker oligonucleotides with known length and unable to fold into secondary structures are loaded. (B) CD spectra of G4s showing dimeric folding and their corresponding monomolecular form when one T base was added at each of the two ends. (C) EMSA analysis of NCL binding. Oligonucleotides identity is shown above each lane. (D) Quantification of NCL/G4 complex bands obtained in the EMSA in panel (C).

to the parallel conformation [66,67]. In general, G4s with shorter loops were more stable than those with longer loops (Table 3).

EMSA binding analysis showed a general increase in NCL binding upon loop extension. In particular, considering the series where the lateral loops were kept constant, an increase in binding was observed with the increasing length of the central loop; in the series where the central loop was kept constant, an increased binding was observed upon the increasing length of the lateral loops (Fig. 3B and C). Two notable exceptions were found: when all loops were 1 nt- or 6-nt long, NCL binding was higher or lower, respectively, than expected by the above-described trend (oligonucleotides 111 and 666 in Fig. 3B and C).

The observed behavior was apparently independent of the G4 conformation: oligonucleotides adopted both mixed and parallel-like G4 conformations (Fig. 3A), while no straightforward correlation with the intensity of NCL binding could be found (Fig. 3B and C). However,

when testing G4 folding in EMSA, we observed one main band for all oligonucleotides, except oligonucleotides 111 and 131, which displayed two bands running at different migration rates (Fig. 3D). To test the molecularity of the upper running band, each labeled oligonucleotide was added with increasing amounts of the same unlabeled nucleic acid. As shown in Fig. 4A, the amount of the upper running band increased with increasing amounts of the unlabeled oligonucleotide, indicating the dimeric nature of the upper band. It has been reported that the presence of one T base at the end of a G4 forming oligonucleotide induces a preferential monomeric G4 [68,69]. Indeed, when one T base was added at each end of the 111, 131 and 161 oligonucleotides (T111T, T131T, T161T, Table 1), the oligonucleotides maintained the G4 conformation (Fig. 4B) and only the monomeric faster running band was obtained (Fig. 4C). These T-added oligonucleotides bound NCL to a lower extent than those without T, probably due to steric



**Fig. 5.** Analysis of NCL binding towards oligonucleotides with variable G4s conformation. (A) CD spectra of 313, 333, 363 oligonucleotides folded in the presence of 20% (w/v) PEG200 and K<sup>+</sup> or Na<sup>+</sup>. (B) EMSA analysis of NCL binding to 313, 333, 363 oligonucleotides folded in the presence of 20% (w/v) PEG200 and K<sup>+</sup> or Na<sup>+</sup> as indicated above the gel image. (C) Quantification of NCL/G4 complex bands obtained in the EMSA in panel (B). For the K<sup>+</sup> or Na<sup>+</sup> experiment series, binding was normalized to the intensity of the 333 NCL/G4 complex band in the corresponding salt condition.

hindrance, however, the binding trend was restored: the increase in the amount of NCL/DNA complex paralleled the lengthening of the loops (Fig. 4D).

The above data indicate that in G4-forming sequences formed by GGG tracts and T loops, NCL binding is in general favored by the presence of one or two long loops (i.e. 6 nt), whereas three long loops are

less favored. The presence of one or two 1-nt long loops in the absence of at least one 6-nt loops disfavors NCL binding.

#### 3.4. NCL binding to G4 oligonucleotides is independent of the G4 conformation

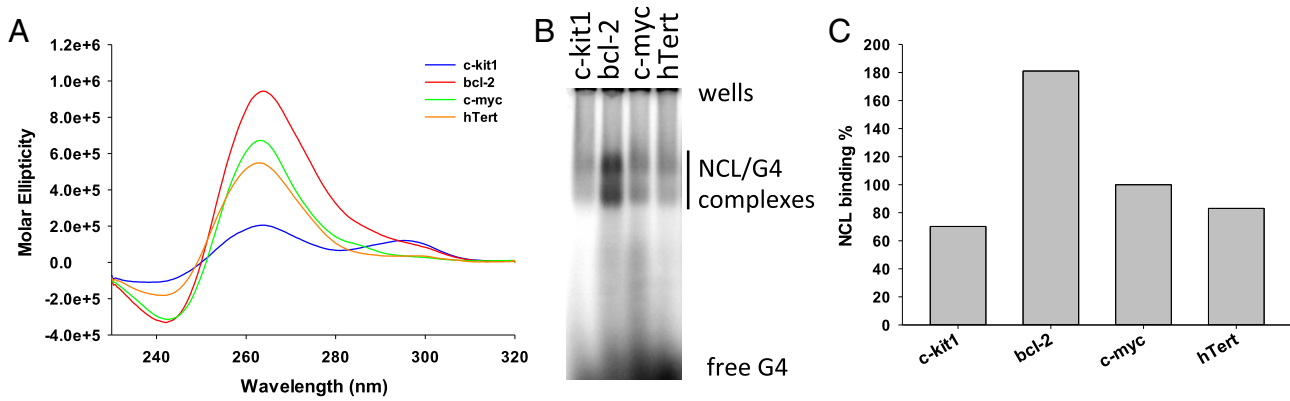
It has been reported that oligonucleotide G4 folding depends on the presence of the monovalent cation in solution. For instance, the human telomeric sequence adopts a hybrid conformation in K<sup>+</sup> solution [70], while it is antiparallel in a Na<sup>+</sup> solution [71]. Moreover, the presence in solution of cosolvents, such as polyethylene glycol (PEG), during G4 folding is commonly used to perturb G4 topology [72]. The human telomeric G4 DNA in the presence of PEG is reported to adopt a parallel G4 conformation, which is not the predominant form in dilute solutions [73–76]. To investigate the effect of G4 conformation on NCL binding, we folded oligonucleotides previously tested in KCl alone (Fig. 3B), in KCl and NaCl in the presence of PEG 20% (w/v). In these conditions we observed in most cases a clear-cut conformational change (Fig. 5A): for instance, oligonucleotide 333 was mixed with a prevalent antiparallel conformation in K<sup>+</sup>, while it shifted to a prevalent parallel topology in K<sup>+</sup> + PEG (compare Figs. 3A and 5A). The binding of NCL was assessed on these oligonucleotides folded in PEG (Fig. 5B). We observed the maintenance of the increased NCL binding to longer loop sequences trend (Fig. 5C). These data indicate that binding of NCL to telomeric G4 oligonucleotides is in general independent of the conformation and that loop length is the major determinant of NCL binding efficiency to telomeric G4s.

#### 3.5. Sequence dependence: NCL binding to oncogenes promoter G4s

To verify if NCL binding determinants observed in telomeric (GGGTTT)<sub>n</sub> G4s were maintained in G4s with more complex sequences, we tested NCL binding towards four oncogene promoters i.e. c-myc, c-kit1, bcl-2 and hTert 1–4 [77–80]. These were shown to form stable G4s whose conformations have been characterized by deep structural analysis. All these oncogene promoter G4s display loops with different length, composition and conformation (Table 1 and Fig. 6A). NCL bound preferentially to bcl-2, followed by c-myc; hTert 1–4 and c-kit1 displayed similar and lower NCL binding (Fig. 6B and C). Among these G4s, bcl-2 is the only structure exhibiting a long loop (7 nts) and two shorter loops (3 and 1 nts), whereas c-myc, c-kit1 and hTert 1–4 they all have loops ≤4 nts (Table 1).

It has very recently been shown that the hTert full-length sequence is composed of 12 G-tracts that fold in two adjacent G4s: our tested hTert 1–4 comprises the first four G-tracts; the second G4, spanning G-tracts 5–12, displays a peculiar structure where a 26 nt-long loop forms an extended stem-hairpin [81]. We wished to test NCL binding to the two separate hTert G4s (hTert 1–4 and hTert 5–12) and to the full-length sequence (hTert 1–12). All three hTert G4s have similar CD spectra with a major peak at 260 nm and a shoulder at 295 nm, with molar ellipticity values at 260 nm increasing with the number of G-tracts in the G4-forming oligonucleotides (Fig. 7A). NCL binding to hTert 1–4 and hTert 5–12 was similar, indicating that the stem-hairpin suppresses the enhancing binding effect of the long loop (Fig. 7B and C). Binding of NCL to hTert 1–12 was slightly enhanced (Fig. 7B and C), retracing the behavior observed in the (GGGTTA)<sub>n</sub> and (GGGTTT)<sub>n</sub> oligonucleotides, where NCL preferentially bound to oligonucleotides with longer loops when multiple G-tracts were available for folding (Figs. 1 and 2). It is thus possible that also in the case of hTert 1–12 NCL induces folding of a less stable but favorite long-looped G4. In alternative, the 7 nt spacer between the two G4s (hTert 1–4 and hTert 5–12) is recognized as a long loop and therefore stimulates NCL binding. We next tested bcl-2 and c-myc in the presence of PEG 20% (w/v) to rule out an effect of PEG in NCL binding. Both oligonucleotides in K<sup>+</sup> are folded in a parallel G4 conformation, which is not modified in the presence of PEG (Fig. 7D). NCL maintained a similar trend and degree of





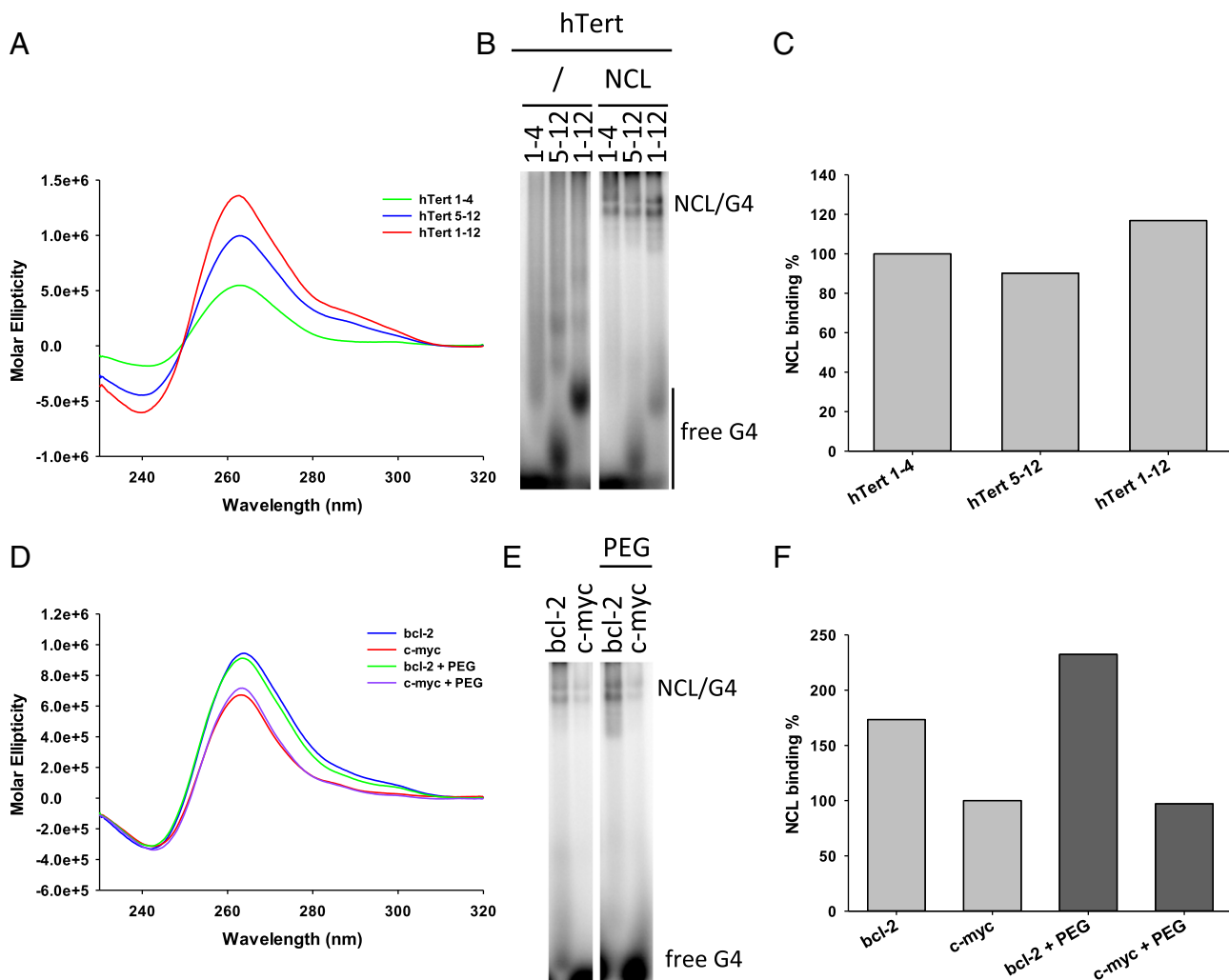
**Fig. 6.** NCL binding to human oncogene promoter G4s. (A) CD spectra of G4s formed in the c-kit1, bcl-2, c-myc and hTert 1–4 promoters. (B) EMSA analysis of NCL binding to oncogene promoter G4s. (C) Quantification of NCL/G4 complexes obtained in the EMSA in panel (B).

binding to these G4s in the absence/presence of PEG, confirming that PEG did not significantly alter NCL binding (Fig. 7E).

These data are in perfect agreement with the NCL binding trend observed in telomeric (GGGTTT)<sub>n</sub> G4s. In addition, they also fit with NCL binding to HIV-1 LTR promoter G4s, previously reported and obtained by both EMSA and FRET analysis [45]. In that case, as well, the best

bound G4 LTR-III displayed 1, 11, 3 nt-long loops, followed by LTR-II with 8, 11, 1 nt-long loops; in contrast LTR-IV, which presented 1, 4, 1 nt-long loops, was bound to a negligible extent [21,45].

These data indicate that the results collected on (GGGTTT)<sub>n</sub> G4s apply in general to G4s with random sequences and that, again, the loop length is the major determinant in driving NCL binding to G4s.



**Fig. 7.** NCL binding to G4s in special conditions. (A) CD spectra of G4s formed by hTert 1–4, hTert 5–12 and hTert 1–12 oligonucleotides. Numbers indicate G4-tracts. (B) EMSA analysis of NCL binding to hTert G4s. (C) Quantification of NCL/G4 complexes obtained in the EMSA in panel (B). (D) CD spectra of G4s formed by c-myc and bcl-2 in the presence/absence of PEG 20% (w/v). (E) EMSA analysis of NCL binding to c-myc and bcl-2 G4s in the presence/absence of PEG 20% (w/v). (F) Quantification of NCL/G4 complexes obtained in the EMSA in panel (E).

## 4. Discussion

Despite the fact that NCL binding to G4 nucleic acids has been proposed to be responsible for human neurodegenerative diseases [16], pathological molecular mechanisms in cells [41–44] and viruses [48–56], a deep characterization of the G4 features that induce NCL binding have been attempted only by Gonzalez et al. [41]. This group, who performed pioneer studies on NCL binding to the oncogene promoter c-myc G4, reported a preferential binding of NCL to parallel G4s with short loops. In contrast, our present work indicates that long loops in the G4 structure are preferred and this behavior is independent of the G4 conformation or loop/oligonucleotide sequence. These apparent differences may be explained by several reasons: 1) the use of a recombinant NCL purified in bacteria [41] vs the full-length human NCL purified from human cells (this work): the recombinant protein included all RBDs and C-terminal domain of the natural human protein but lacked the N-terminal domain which in turn was fused to the *Escherichia coli* maltose-binding protein; even if the nucleic acid binding activity was maintained in these conditions, the specific recognition of G4s, and in particular of their steric features, may be severely distorted by the introduction of a different N-terminal domain. Indeed, the N-terminal domain of NCL is a long unstructured (300 amino acids) domain that, besides its reported interaction with other proteins, may support the binding specificity of the RBD and RGG domains that directly make contact with the nucleic acid. 2) It has been demonstrated that short loops tend to constrain G4 structures into a parallel conformation, regardless of base composition [66,67], therefore the suggested preferential binding to short-looped G4s in the parallel conformation may derive from the analysis of two dependent variables. 3) In our present work we performed a systematic analysis of loop sequence and oligonucleotide conformation working on a simplified (GGGTTT)<sub>n</sub> sequence. We unambiguously showed that NCL prefers G4s with at least one long loop (≥3 nts), while three short loops (1 nt) or three long loops (6 nts) are disfavored. The analysis on the simplified sequences allowed us to modify the conformation of the oligonucleotides working in different salt (K<sup>+</sup> or Na<sup>+</sup>) and crowding conditions (PEG 20% w/v) [68,72,75] to show that NCL binding is in general independent of the oligonucleotide conformation and that loop lengths is the driving force that enhances NCL binding to G4s. By extending the results from the telomeric sequences to cellular oncogene and HIV-1 promoter G4s, we were able to show that loop length is the major determinant in the enhancement of NCL binding to G4s even in more variable and complex sequences. This observation is in line with the reported recognition of G4s by the cellular protein EWS: EWS shares with NCL the RGG domain which mediates recognition of G4s [82] and binds to G4s independently of base composition in long loops [64].

Finally, our results may shed some light on NCL recognition of repeats of G4 forming regions, such as the hexanucleotide repeat involved in the neurodegenerative diseases amyotrophic lateral sclerosis and frontotemporal dementia [16] and telomeric repeats [83]. We (Fig. 2A present work) and others [65] have shown that regions containing multiple G-tracts can fold into several conformations involving the alternative use of different G-tracts. We here indicate that NCL binding stimulates folding of these regions to G4s containing long loops (i.e. some G-tracts may be included to the loop moiety) therefore modifying the overall 3D structure and steric hindrance of these nucleic acid repeated sequences. Interestingly, indeed, NCL preferentially stabilized G4s that were less intrinsically stable, likely because of the possibility in this condition to fold them in the best protein-fitting conformation, i.e. G4s with longer and loose loops: this conformational modification may represent a marker to molecular signaling and disease pathogenesis. Our data perfectly fit with the very recent demonstration that protein binding to G4 and iMotif structures modify complex stability and the overall nucleic acid conformation to provide a mechanosensor-regulated cellular mechanism [84].

## 5. Conclusions

Our work has unambiguously shown that loop length of G4 structures is a fundamental feature recognized by NCL to promote its binding to G4-folded nucleic acids. Understanding the determinants that regulate NCL/G4 binding is the first step towards the development of therapies that specifically alter this molecular switch that triggers cellular signaling, such as initiation of transcription in the HIV-1 promoter and hence viral production, virus infection, tumor formation, maintenance and metastasis, and neurodegeneration.

## Transparency Document

The Transparency document associated with this article can be found, in online version.

## Acknowledgments

Funding: This work was supported by the Bill & Melinda Gates Foundation, Seattle, WA [GCE grant numbers OPP1035881, OPP1097238]; the European Research Council [ERC Consolidator grant 615879]; and the Italian Ministry of University and Research [FIRB-Ideas RBID082ATK].

## References

- [1] H.J. Lipps, D. Rhodes, G-quadruplex structures: in vivo evidence and function, *Trends Cell Biol.* 19 (2009) 414–422.
- [2] D. Sen, W. Gilbert, A sodium-potassium switch in the formation of four-stranded G4-DNA, *Nature* 344 (1990) 410–414.
- [3] A. Verma, K. Halder, R. Halder, V.K. Yadav, P. Rawal, R.K. Thakur, F. Mohd, A. Sharma, S. Chowdhury, Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species, *J. Med. Chem.* 51 (2008) 5641–5649.
- [4] S.G. Hershman, Q. Chen, J.Y. Lee, M.L. Kozak, P. Yue, L.S. Wang, F.B. Johnson, Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*, *Nucleic Acids Res.* 36 (2008) 144–156.
- [5] N. Beäume, R. Pathak, V.K. Yadav, S. Kota, H.S. Misra, H.K. Gautam, S. Chowdhury, Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of *D. radiodurans* involves G4 DNA-mediated regulation, *Nucleic Acids Res.* 41 (2013) 76–89.
- [6] P. Rawal, V.B. Kumarasetti, J. Ravindran, N. Kumar, K. Halder, R. Sharma, M. Mukerji, S.K. Das, S. Chowdhury, Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation, *Genome Res.* 16 (2006) 644–655.
- [7] R.S. Thakur, A. Desingu, S. Basavaraju, S. Subramanya, D.N. Rao, G. Nagaraju, *Mycobacterium tuberculosis* DinG is a structure-specific helicase that unwinds G4 DNA: implications for targeting G4 DNA as a novel therapeutic approach, *J. Biol. Chem.* 289 (2014) 25112–25136.
- [8] M. Wieland, J.S. Hartig, Investigation of mRNA quadruplex formation in *Escherichia coli*, *Nat. Protoc.* 4 (2009) 1632–1640.
- [9] L.M. Harris, C.J. Merrick, G-quadruplexes in pathogens: a common route to virulence control? *PLoS Pathog.* 11 (2015), e1004562.
- [10] M. Metifiot, S. Amrane, S. Litvak, M.L. Andreola, G-quadruplexes in viruses: function and potential therapeutic applications, *Nucleic Acids Res.* 42 (2014) 12352–12366.
- [11] N. Maizels, G4-associated human diseases, *EMBO Rep.* 16 (2015) 910–922.
- [12] D. Rhodes, H.J. Lipps, G-quadruplexes and their regulatory roles in biology, *Nucleic Acids Res.* 43 (2015) 8627–8637.
- [13] B. Zhou, C. Liu, Y. Geng, G. Zhu, Topology of a G-quadruplex DNA formed by C9orf72 hexanucleotide repeats associated with ALS and FTD, *Sci. Rep.* 5 (2015) 16673.
- [14] P. Mani, V.K. Yadav, S.K. Das, S. Chowdhury, Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination, *PLoS One* 4 (2009), e4399.
- [15] S. Nakken, T. Rognes, E. Hovig, The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts, *Nucleic Acids Res.* 37 (2009) 5749–5756.
- [16] A.R. Haeuser, C.J. Donnelly, G. Periz, E.A. Simko, P.G. Shaw, M.S. Kim, N.J. Maragakis, J.C. Troncoso, A. Pandey, R. Sattler, J.D. Rothstein, J. Wang, C9orf72 nucleotide repeat structures initiate molecular cascades of disease, *Nature* 507 (2014) 195–200.
- [17] J.F. Fiset, D.R. Montagna, M.R. Mihalescu, M.S. Wolfe, A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing, *J. Neurochem.* 121 (2012) 763–773.
- [18] M. Fry, L.A. Loeb, The fragile X syndrome d(CGG)<sub>n</sub> nucleotide repeats form a stable tetrahelical structure, *Proc. Natl. Acad. Sci. U. S. A.* 91 (1994) 4950–4954.
- [19] R. Perrone, M. Nadai, J.A. Poe, I. Frasson, M. Palumbo, G. Palu, T.E. Smithgall, S.N. Richter, Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity, *PLoS One* 8 (2013), e73121.

- [20] S. Amrane, A. Kerkour, A. Bedrat, B. Viallet, M.L. Andreola, J.L. Mergny, Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development, *J. Am. Chem. Soc.* 136 (2014) 5249–5252.
- [21] R. Perrone, M. Nadai, I. Frasson, J.A. Poe, E. Butovskaya, T.E. Smithgall, M. Palumbo, G. Palu, S.N. Richter, A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter, *J. Med. Chem.* 56 (2013) 6521–6530.
- [22] D. Piekna-Przybylska, M.A. Sullivan, G. Sharma, R.A. Bambara, U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence, *Biochemistry* 53 (2014) 2581–2593.
- [23] R. Perrone, E. Butovskaya, D. Daelemans, G. Palu, C. Pannecouque, S.N. Richter, Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19, *J. Antimicrob. Chemother.* 69 (2014) 3248–3258.
- [24] R. Perrone, F. Doria, E. Butovskaya, I. Frasson, S. Botti, M. Scalabrin, S. Lago, V. Grande, M. Nadai, M. Freccero, S.N. Richter, Synthesis, binding and antiviral properties of potent core-extended naphthalene diimides targeting the HIV-1 long terminal repeat promoter G-quadruplexes, *J. Med. Chem.* 58 (2015) 9639–9652.
- [25] S. Artusi, M. Nadai, R. Perrone, M.A. Biasolo, G. Palu, L. Flamand, A. Calistri, S.N. Richter, The herpes simplex virus-1 genome contains multiple clusters of repeated G-quadruplex: implications for the antiviral activity of a G-quadruplex ligand, *Antivir. Res.* 118 (2015) 123–131.
- [26] P. Murat, J. Zhong, L. Lekiéff, N.P. Cowieson, J.L. Clancy, T. Preiss, S. Balasubramanian, R. Khanna, J. Tellam, G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation, *Nat. Chem. Biol.* 10 (2014) 358–364.
- [27] J. Nourse, F.B. Johnson, P.M. Lieberman, Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment, *J. Virol.* 83 (2009) 10336–10346.
- [28] J. Tan, C. Vonrhein, O.S. Smart, G. Bricogne, M. Bollati, Y. Kusov, G. Hansen, J.R. Mesters, C.L. Schmidt, R. Hilgenfeld, The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes, *PLoS Pathog.* 5 (2009), e1000428.
- [29] K. Tluckova, M. Marusic, P. Tothova, L. Bauer, P. Sket, J. Plavec, V. Viglasky, Human papillomavirus G-quadruplexes, *Biochemistry* 52 (2013) 7207–7216.
- [30] S.R. Wang, Y.Q. Min, J.Q. Wang, C.X. Liu, B.S. Fu, F. Wu, L.Y. Wu, Z.X. Qiao, Y.Y. Song, G.H. Xu, Z.G. Wu, G. Huang, N.F. Peng, R. Huang, W.X. Mao, S. Peng, Y.Q. Chen, Y. Zhu, T. Tian, X.L. Zhang, X. Zhou, A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target, *Sci. Adv.* 2 (2016), e1501535.
- [31] A.M. Fleming, Y. Ding, A. Alenko, C.J. Burrows, Zika virus genomic RNA possesses conserved G-quadruplexes characteristic of the Flaviviridae family, *ACS Infect. Dis.* 2 (2016) 674–681.
- [32] S.R. Wang, Q.Y. Zhang, J.Q. Wang, X.Y. Ge, Y.Y. Song, Y.F. Wang, X.D. Li, B.S. Fu, G.H. Xu, B. Shu, P. Gong, B. Zhang, T. Tian, X. Zhou, Chemical targeting of a G-quadruplex RNA in the ebola virus L gene, *Cell Chem. Biol.* 23 (2016) 1113–1122.
- [33] C. Sissi, B. Gatto, M. Palumbo, The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective, *Biochimie* 93 (2011) 1219–1230.
- [34] W. Palm, T. de Lange, How shelterin protects mammalian telomeres, *Annu. Rev. Genet.* 42 (2008) 301–334.
- [35] S. Cogo, M. Paramasivam, A. Membrino, K.K. Yokoyama, L.E. Xodo, 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 (2010) 22003–22016.
- [36] S. Cogo, A.E. Shchekotikhin, L.E. Xodo, HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property, *Nucleic Acids Res.* 42 (2014) 8379–8388.
- [37] P. Mohaghegh, J.K. Karow, R.M. Brosh Jr., V.A. Bohr, I.D. Hickson, The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases, *Nucleic Acids Res.* 29 (2001) 2843–2849.
- [38] S. Seal, D. Thompson, A. Renwick, A. Elliott, P. Kelly, R. Barfoot, T. Chagtai, H. Jayatilake, M. Ahmed, K. Spanova, B. North, L. McGuffog, D.G. Evans, D. Eccles, C. Breast Cancer Susceptibility, D.F. Easton, M.R. Stratton, N. Rahman, Truncating mutations in the Fanconi anemia J gene BRP1 are low-penetrance breast cancer susceptibility alleles, *Nat. Genet.* 38 (2006) 1239–1241.
- [39] M. Paramasivam, A. Membrino, S. Cogo, 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.
- [40] A.C. Girvan, Y. Teng, L.K. Casson, S.D. Thomas, S. Juliger, M.W. Ball, J.B. Klein, W.M. Pierce Jr., S.S. Barve, P.J. Bates, AGRO100 inhibits activation of nuclear factor-kappaB (NF-kappaB) by forming a complex with NF-kappaB essential modulator (NEMO) and nucleolin, *Mol. Cancer Ther.* 5 (2006) 1790–1799.
- [41] V. Gonzalez, K. Guo, L. Hurley, D. Sun, Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein, *J. Biol. Chem.* 284 (2009) 23622–23635.
- [42] V. Gonzalez, L.H. Hurley, The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity, *Biochemistry* 49 (2010) 9706–9714.
- [43] D.J. Uribe, K. Guo, Y.J. Shin, D. Sun, Heterogeneous nuclear ribonucleoprotein K and nucleolin as transcriptional activators of the vascular endothelial growth factor promoter through interaction with secondary DNA structures, *Biochemistry* 50 (2011) 3796–3806.
- [44] P.C. Wei, Z.F. Wang, W.T. Lo, M.I. Su, J.Y. Shew, T.C. Chang, W.H. Lee, A cis-element with mixed G-quadruplex structure of NPGPx promoter is essential for nucleolin-mediated transactivation on non-targeting siRNA stress, *Nucleic Acids Res.* 41 (2013) 1533–1543.
- [45] E. Tosoni, I. Frasson, M. Scalabrin, R. Perrone, E. Butovskaya, M. Nadai, D. Fabris, S.N. Richter, Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription, *Nucleic Acids Res.* 43 (2015) 8884–8897.
- [46] Y. Daniely, D.D. Dimitrova, J.A. Borowiec, Stress-dependent nucleolin mobilization mediated by p53-nucleolin complex formation, *Mol. Cell. Biol.* 22 (2002) 6014–6022.
- [47] S. Storck, M. Shukla, S. Dimitrov, P. Bouvet, Functions of the histone chaperone nucleolin in diseases, *Subcell. Biochem.* 41 (2007) 125–144.
- [48] C. Callebaut, J. Blanco, N. Benkirane, B. Krust, E. Jacotot, G. Guichard, N. Seddiki, J. Svab, E. Dam, S. Muller, J.P. Briand, A.G. Hovanessian, Identification of V3 loop-binding proteins as potential receptors implicated in the binding of HIV particles to CD4(+) cells, *J. Biol. Chem.* 273 (1998) 21988–21997.
- [49] W. Gao, M. Li, J. Zhang, Tandem immunoprecipitation approach to identify HIV-1 Gag associated host factors, *J. Virol. Methods* 203 (2014) 116–119.
- [50] T. Ueno, K. Tokunaga, H. Sawa, M. Maeda, J. Chiba, A. Kojima, H. Hasegawa, Y. Shoya, T. Sata, T. Kurata, H. Takahashi, Nucleolin and the packaging signal, psi, promote the budding of human immunodeficiency virus type-1 (HIV-1), *Microbiol. Immunol.* 48 (2004) 111–118.
- [51] C.A. Balinsky, H. Schmeisser, S. Ganesan, K. Singh, T.C. Pierson, K.C. Zoon, Nucleolin interacts with the dengue virus capsid protein and plays a role in formation of infectious virus particles, *J. Virol.* 87 (2013) 13094–13106.
- [52] B.J. Bender, D.M. Coen, B.L. Strang, Dynamic and nucleolin-dependent localization of human cytomegalovirus UL84 to the periphery of viral replication compartments and nucleoli, *J. Virol.* 88 (2014) 11738–11747.
- [53] Y.L. Chen, C.D. Liu, C.P. Cheng, B. Zhao, H.J. Hsu, C.L. Shen, S.J. Chiu, E. Kieff, C.W. Peng, Nucleolin is important for Epstein-Barr virus nuclear antigen 1-mediated episome binding, maintenance, and transcription, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 243–248.
- [54] B.L. Strang, S. Boulant, T. Kirchhausen, D.M. Coen, Host cell nucleolin is required to maintain the architecture of human cytomegalovirus replication compartments, *MBio* 3 (2012).
- [55] P.Y. Su, Y.F. Wang, S.W. Huang, Y.C. Lo, Y.H. Wang, S.R. Wu, D.B. Shieh, S.H. Chen, J.R. Wang, M. Lai, C.F. Chang, Cell surface nucleolin facilitates enterovirus 71 binding and infection, *J. Virol.* 89 (2015) 4527–4538.
- [56] F. Tayyari, D. Marchant, T.J. Moraes, W. Duan, P. Mastrangelo, R.G. Hegele, Identification of nucleolin as a cellular receptor for human respiratory syncytial virus, *Nat. Med.* 17 (2011) 1132–1135.
- [57] A.G. Hovanessian, Midkine, a cytokine that inhibits HIV infection by binding to the cell surface expressed nucleolin, *Cell Res.* 16 (2006) 174–181.
- [58] S. Nisole, E.A. Said, C. Mische, M.C. Prevost, B. Krust, P. Bouvet, A. Bianco, J.P. Briand, A.G. Hovanessian, The anti-HIV pentameric pseudopeptide HB-19 binds the C-terminal end of nucleolin and prevents anchorage of virus particles in the plasma membrane of target cells, *J. Biol. Chem.* 277 (2002) 20877–20886.
- [59] E.A. Said, J. Courty, J. Svab, J. Delbe, B. Krust, A.G. Hovanessian, Pleiotrophin inhibits HIV infection by binding the cell surface-expressed nucleolin, *FEBS J.* 272 (2005) 4646–4659.
- [60] M. Metifiot, S. Amrane, J.L. Mergny, M.L. Andreola, Anticancer molecule AS1411 exhibits low nanomolar antiviral activity against HIV-1, *Biochimie* 118 (2015) 173–175.
- [61] R. Perrone, E. Butovskaya, S. Lago, A. Garzino-Demo, C. Pannecouque, G. Palu, S.N. Richter, The G-quadruplex-forming aptamer AS1411 potentially inhibits HIV-1 attachment to the host cell, *Int. J. Antimicrob. Agents* 47 (2016) 311–316.
- [62] M. McCormick-Graham, D.P. Romero, A single telomerase RNA is sufficient for the synthesis of variable telomeric DNA repeats in ciliates of the genus *Paramecium*, *Mol. Cell. Biol.* 16 (1996) 1871–1879.
- [63] K.N. Luu, A.T. Phan, V. Kuryavly, L. Lacroix, D.J. Patel, Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold, *J. Am. Chem. Soc.* 128 (2006) 9963–9970.
- [64] K. Takahama, C. Sugimoto, S. Arai, R. Kurokawa, T. Oyoshi, Loop lengths of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's sarcoma, *Biochemistry* 50 (2011) 5369–5378.
- [65] T.S. Dexheimer, D. Sun, L.H. Hurley, Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter, *J. Am. Chem. Soc.* 128 (2006) 5404–5415.
- [66] A. Bugaut, S. Balasubramanian, A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes, *Biochemistry* 47 (2008) 689–697.
- [67] R. Tippana, W. Xiao, S. Myong, G-quadruplex conformation and dynamics are determined by loop length and sequence, *Nucleic Acids Res.* 42 (2014) 8106–8114.
- [68] E. Lary, A. Marchand, S. Amrane, V. Gabelica, J.L. Mergny, Quadruplex turncoats: cation-dependent folding and stability of quadruplex-DNA double switches, *J. Am. Chem. Soc.* 138 (2016) 2780–2792.
- [69] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E.S. Baker, M.T. Bowers, E. De Pauw, V. Gabelica, G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation, *J. Am. Chem. Soc.* 130 (2008) 10208–10216.
- [70] A. Ambrus, D. Chen, J. Dai, T. Bialis, R.A. Jones, D. Yang, Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution, *Nucleic Acids Res.* 34 (2006) 2723–2735.
- [71] Y. Wang, D.J. Patel, Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex, *Structure* 1 (1993) 263–282.
- [72] R. Buscaglia, M.C. Miller, W.L. Dean, R.D. Gray, A.N. Lane, J.O. Trent, J.B. Chaires, Polyethylene glycol binding alters human telomere G-quadruplex structure by conformational selection, *Nucleic Acids Res.* 41 (2013) 7934–7946.
- [73] B. Heddi, A.T. Phan, Structure of human telomeric DNA in crowded solution, *J. Am. Chem. Soc.* 133 (2011) 9824–9833.
- [74] Y. Xue, Z.Y. Kan, Q. Wang, Y. Yao, J. Liu, Y.H. Hao, Z. Tan, Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K+ solution under molecular crowding condition, *J. Am. Chem. Soc.* 129 (2007) 11185–11191.
- [75] M. Boncina, G. Vesnaver, J.B. Chaires, J. Lah, Unraveling the thermodynamics of the folding and interconversion of human telomere G-quadruplexes, *Angew. Chem.* 55 (2016) 10340–10344.

- [76] D. Miyoshi, A. Nakao, N. Sugimoto, Molecular crowding regulates the structural switch of the DNA G-quadruplex, *Biochemistry* 41 (2002) 15017–15024.
- [77] J. Dai, D. Chen, R.A. Jones, L.H. Hurley, D. Yang, NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region, *Nucleic Acids Res.* 34 (2006) 5133–5144.
- [78] K.W. Lim, L. Lacroix, D.J. Yue, J.K. Lim, J.M. Lim, A.T. Phan, Coexistence of two distinct G-quadruplex conformations in the hTERT promoter, *J. Am. Chem. Soc.* 132 (2010) 12331–12342.
- [79] S. Rankin, A.P. Reszka, J. Huppert, M. Zloh, G.N. Parkinson, A.K. Todd, S. Ladame, S. Balasubramanian, S. Neidle, Putative DNA quadruplex formation within the human c-kit oncogene, *J. Am. Chem. Soc.* 127 (2005) 10584–10589.
- [80] A. Siddiqui-Jain, C.L. Grand, D.J. Bearss, L.H. Hurley, Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 11593–11598.
- [81] H.J. Kang, Y. Cui, H. Yin, A. Scheid, W.P. Hendricks, J. Schmidt, A. Sekulic, D. Kong, J.M. Trent, V. Gokhale, H. Mao, L.H. Hurley, A pharmacological chaperone molecule induces cancer cell death by restoring tertiary DNA structures in mutant hTERT promoters, *J. Am. Chem. Soc.* (2016), <http://dx.doi.org/10.1021/jacs.6b07598>.
- [82] P. Thandapani, T.R. O'Connor, T.L. Bailey, S. Richard, Defining the RGG/RG motif, *Mol. Cell* 50 (2013) 613–623.
- [83] F. Ishikawa, M.J. Matunis, G. Dreyfuss, T.R. Cech, Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)<sub>n</sub>, *Mol. Cell. Biol.* 13 (1993) 4301–4310.
- [84] C. Sutherland, Y. Cui, H. Mao, L.H. Hurley, A mechanosensor mechanism controls the G-quadruplex/i-motif molecular switch in the MYC promoter NHE III1, *J. Am. Chem. Soc.* (2016), <http://dx.doi.org/10.1021/jacs.6b09196>.