9706 *Biochemistry* **2010,** *49,* 9706–9714
DOI: 10.1021/bi100509s

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

The C-Terminus of Nucleolin Promotes the Formation of the c-*MYC* G-Quadruplex and Inhibits c-*MYC* Promoter Activity[†]

Verónica González[‡] and Laurence H. Hurley*[,‡,§,∥]

[‡]*College of Pharmacy, Department of Pharmacology and Toxicology, University of Arizona, Tucson, Arizona 85721, United States,* [§]*BIO5 Institute, University of Arizona, Tucson, Arizona 85721, United States, and* [∥]*Arizona Cancer Center, University of Arizona, Tucson, Arizona 85724, United States*

Received April 5, 2010; Revised Manuscript Received October 4, 2010

ABSTRACT: Nucleolin, the most abundant nucleolar phosphoprotein of eukaryotic cells, is known primarily for its role in ribosome biogenesis and cell proliferation. It is, however, a multifunctional protein that, depending on the cellular context, can drive either cell proliferation or apoptosis. Our laboratory recently demonstrated that nucleolin can function as a repressor of c-*MYC* transcription by binding to and stabilizing the formation of a G-quadruplex structure in a region of the c-*MYC* promoter responsible for controlling 85−90% of c-*MYC*'s transcriptional activity. In this study, we investigate the structural elements of nucleolin that are required for c-*MYC* repression. The effect of nucleolin deletion mutants on the formation and stability of the c-*MYC* G-quadruplex, as well as c-*MYC* transcriptional activity, was assessed by circular dichroism spectropolarimetry, thermal stability, and in vitro transcription. Here we report that nucleolin's RNA binding domains 3 and 4, as well as the arginine-glycine-glycine (RGG) domain, are required to repress c-*MYC* transcription.

The mechanisms that regulate c-*MYC* transcription are complex and involve multiple promoters, start sites, and cis elements (NHEs)[1] (Figure 1A) (*1*). NHE III_1, which is located −142 to −115 base pairs upstream of the P1 promoter, has been shown to control 85−90% of c-*MYC* transcription. The template strand of this element consists of a G-rich sequence that can equilibrate between transcriptionally active forms (duplex and single-stranded DNA) and a silencer structure (G-quadruplex) (*2*).

It is well known that G-rich DNA has the ability to form G-quadruplex structures under physiological conditions (*3*). The fundamental structural unit of a G-quadruplex is known as a G-tetrad, which is composed of four guanines aligned in a planar ring configuration in which each guanine interacts with two adjacent guanines via Hoogsteen hydrogen bonding (Figure 1B, left and center) (*4*−*6*). Two or more G-tetrads can stack to form a G-quadruplex structure (Figure 1B, right). Formation of a G-quadruplex in the c-*MYC* promoter has been shown to be facilitated by the negative supercoiling generated by transcription (*7*, *8*). In addition, when the G-rich strand of NHE III_1 is assembled into a G-quadruplex, the binding sites of c-*MYC* transcriptional activators such as Sp1 and CNBP are masked, thus silencing c-*MYC* transcription (*9*).

Recent reports have demonstrated that putative G-quadruplex motifs are highly prevalent in human promoter regions (*10*−*12*). In addition, the presence of G-quadruplex motifs has been shown to correlate with gene function, with oncogenes having a disproportionately high incidence of G-quadruplex motifs, whereas the promoters of tumor suppressors exhibit an extremely low potential for G-quadruplex formation (*13*). Importantly, G-quadruplex topological diversity arises from variations in strand directionality, loop length and sequence, and the number of G-tetrad stacks, allowing for specific structural recognition by G-quadruplex-interactive proteins (*14*−*19*).

We have previously identified nucleolin as a selective c-*MYC* G-quadruplex-binding protein that has the ability to induce the stable formation of the parallel c-*MYC* G-quadruplex from single-stranded DNA (*20*). In addition, we have established that nucleolin interacts with the c-*MYC* promoter in vivo in HeLa cells. Most importantly, our laboratory has demonstrated that nucleolin can repress c-*MYC* transcription in a significant and dose-dependent manner (*20*).

Nucleolin, one of the most abundant nonribosomal proteins of eukaryotes, has been shown to play a role in ribosome biogenesis and cell proliferation. It is, however, a multifunctional protein whose function is dependent on cellular context. A number of stress stimuli have been shown to induce changes in the cellular localization of nucleolin (*21*−*28*), and there is strong evidence that nucleolin can function as a stress-sensitive tumor suppressor (*22*, *24*, *29*). In addition, stress- or cisplatin-activated p53 has been shown to translocate to the nucleolus, where it forms a complex with nucleolin, causing nucleolin to relocate from the nucleolus to the nucleoplasm (*30*, *31*). This binding is believed to be one of the mechanisms regulating the reactivation of the p53 protein in cisplatin-treated human cervix carcinoma cells (*29*, *31*). Interestingly, the interaction with p53 is transient, leading to the

[†]This work was supported, in whole or in part, by the National Institutes of Health (GM085585 and Under-Represented Minority Supplements to CA94166 and CA95060).

*To whom correspondence should be addressed: College of Pharmacy, University of Arizona, 1703 E. Mabel, Tucson, AZ 85721. Phone: (520) 626-5622. Fax: (520) 626-0035. E-mail: hurley@pharmacy.arizona.edu.

[1]Abbreviations: RBD, RNA binding domain; RBD1, RNA binding domain 1; RBD2, RNA binding domain 2; RBD3, RNA binding domain 3; RBD4, RNA binding domain 4; NHE, nuclease hypersensitive element; CD, circular dichroism; RGG, arginine-glycine-glycine; MBP, maltose-binding protein; Pu47ss, purine-rich 47-mer containing c-*MYC* NHE III_1; pDel-4, reporter plasmid containing the NHE III_1 region of the c-*MYC* promoter; EMSA, electromobility shift assay.

 Published on Web 10/11/2010

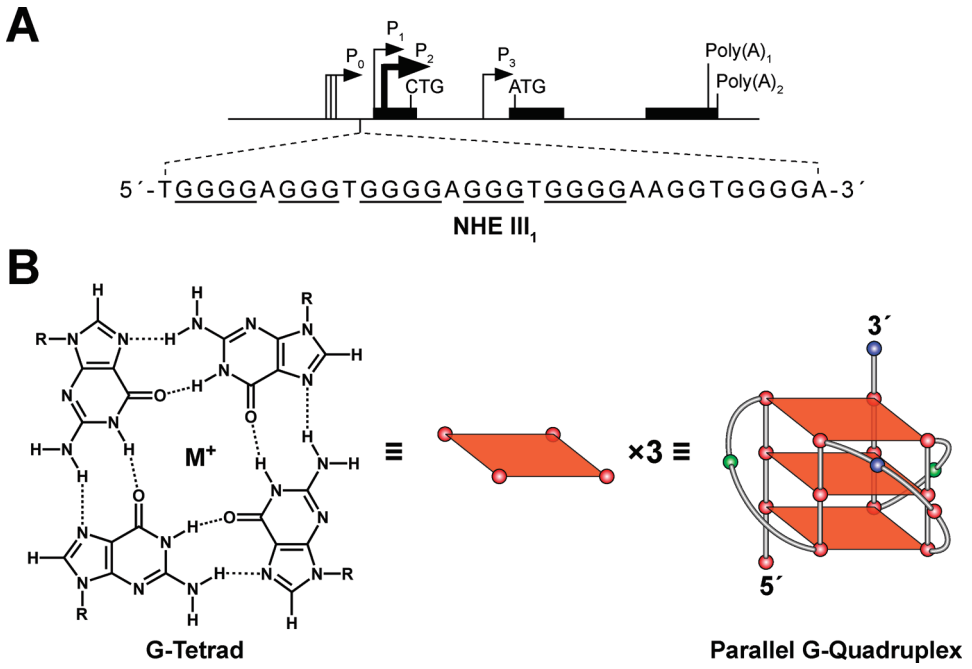

FIGURE 1: Promoter structure of the c-*MYC* gene and scheme of its G-quadruplex structure. (A) Location of the NHE III_1 region within the c-*MYC* promoter. Runs of guanines that can participate in G-quadruplex formation are underlined. (B) Scheme of a guanine tetrad and cartoon of the c-*MYC* G-quadruplex structure: left, hydrogen bonding pattern in a G-tetrad; center, schematic diagram of a G-tetrad; right, cartoon representing a G-quadruplex structure that is found in the c-*MYC* promoter region.

accumulation of nucleolin in the nucleoplasm. This could allow nucleolin to subsequently interact with additional proteins or DNA targets in the nucleoplasm, which would further stimulate the execution of the apoptotic program (*21*−*23*).

Nucleolin is a modular protein composed of an N-terminal domain rich in acidic residues, a central region containing four globular RNA binding domains (RBDs) separated by flexible linker loops, and a C-terminal domain rich in arginine and glycine residues (RGG domain) (Figure 2A) (*32*). This modular architecture allows for greater versatility of the protein, because by combining multiple domains, nucleolin can construct various interaction surfaces that can be assembled and disassembled as needed. Consequently, nucleolin can recognize a large number of targets. For example, by combining its first two RBDs, nucleolin interacts with the stem−loop RNA structure formed by the nucleolin recognition element (*33*, *34*), while all four RBDs are required for binding to a single-stranded RNA motif found in pre-rRNA (*35*−*37*). In this study, we investigate the structural elements of nucleolin that are required for repression of c-*MYC* transcription. The effects of both N-terminal and C-terminal deletion mutants of nucleolin on c-*MYC* G-quadruplex formation and transcriptional repression were assessed.

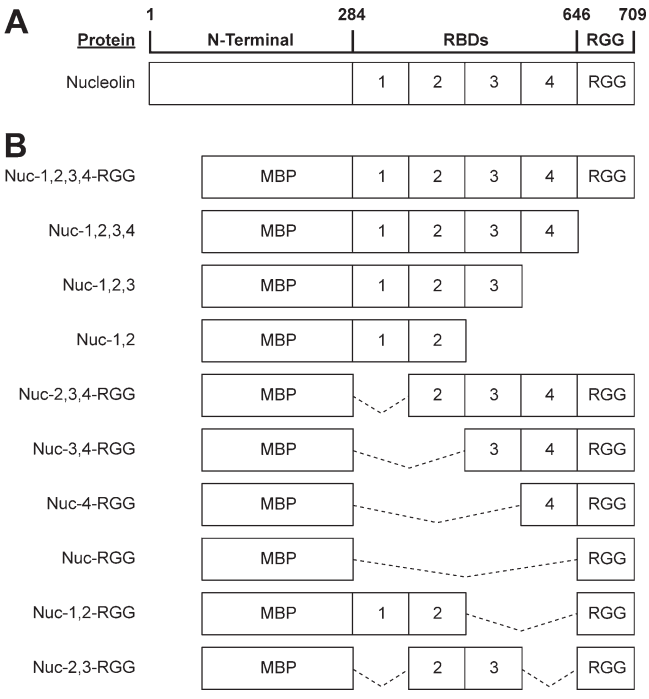

FIGURE 2: Nucleolin deletion mutants. (A) Diagram of nucleolin structure. (B) Diagram of the nucleolin deletion mutants used in this study. Dashed lines indicate regions of the nucleolin peptide that have been deleted from the Nuc-1,2,3,4-RGG construct. All proteins were overexpressed in *Escherichia coli* fused at the N-terminus to the maltose-binding protein (MBP).

MATERIALS AND METHODS

Plasmid Constructs. The reporter plasmid containing the NHE III_1 region of the human c-*MYC* promoter (pDel-4) linked to the firefly luciferase gene was kindly provided by B. Vogelstein (Johns Hopkins University, Baltimore, MD) (*38*). Plasmids pNuc-1,2,3,4-RGG, pNuc-3,4-RGG, pNuc-1,2-RGG, and pNuc-RGG for the expression of recombinant nucleolin deletion mutants were generously provided by L. A. Hanakahi (Johns Hopkins Bloomberg School of Public Health, Baltimore, MD). The pNuc-1,2,3,4-RGG plasmid was used as a template for the amplification of nucleolin's DNA coding sequence and construction of pNuc-2,3-RGG, pNuc-1,2,3, pNuc-2,3,4-RGG, and pNuc-4-RGG. Inverse polymerase chain reaction (PCR) amplification and linearization of the vector backbone were performed as previously described (*39*). Briefly, each PCR mixture contained 0.01 pmol of template DNA (pNuc-1,2,3,4-RGG) and 15 pmol of reverse primer REV (5′-AGCCGTAGTCGGTTCTGTGCCT-TCCACTTTCTGTTTCTTGGCTTCAGGAGCTGCT-3′) and

Table 1: List of Primer Pairs Used for the Construction of Nucleolin Deletion Mutants[a]

Nuc-1,2,3	
FWD, 5′-GAACCGACTACGGCTTTCAATCTCTTTGTTGGAAACCTAAACTTT-3′	
REV, 5′-GGATGGCTGGCTTCTGGCATTAGGTGATCCTTGTGGCTTGTGGTC-3′	
Nuc-1,2,3,4	
FWD, 5′-GAACCGACTACGGCTTTCAATCTCTTTGTTGGAAACCTAAACTTT-3′	
REV, 5′-CTTAGGTTTGGCCCAGTCCAAGGTAACTTTTTGTGGCTTGTGGTC-3′	
Nuc-4-RGG	
FWD, 5′-GAACCGACTACGGCTAAAACTCTGTTTGTCAAAGGCCTGTCTGAG-3′	
REV, 5′-CTTAGGTTTGGCCCAGTCCAAGGTAACTTTGCCACCTTCACCCTT-3′	
Nuc-2,3-RGG	
FWD, 5′-GAACCGACTACGGCTAAAGACAGTAAGAAAGAGCGAGATGCGAGA-3′	
REV, 5′-GGATGGCTGGCTTCTGGCATTAGGTGATCCGCCACCTTCACCCTT-3′	
Nuc-2,3,4-RGG	
FWD, 5′-GAACCGACTACGGCTAAAGACAGTAAGAAAGAGCGAGATGCGAGA-3′	
REV, 5′-CTTAGGTTTGGCCCAGTCCAAGGTAACTTTGCCACCTTCACCCTT-3′	

[a]Insert segments were generated by PCR using primers containing a fragment-specific sequence and 15 bp overhangs that overlap with the plasmid backbone (underlined bases) to allow for restriction-independent and ligase-free cloning.

forward primer FWD1 (5′-GACCACAAGCCACAAGGAAA-GAAGACGAAGTTTGAATAGGAATTCCTCGACCTGC-3′) to amplify the empty vector backbone for construction of pNuc-1,2,3 and pNuc-1,2,3,4, or forward primer FWD2 (5′-AAGGG-TGAAGGTGGCTTCGGGGGTCGTGGTGGAGGCAGA-GGCGGCTTTGGAGGAC-3′) to amplify the vector backbone containing the RGG domain for construction of pNuc-4-RGG, pNuc-2,3-RGG, and pNuc-2,3,4-RGG. 1× Pfu reaction buffer, dNTPs (200 μM each), and 2.5 units of Pfu DNA polymerase (Fermentas, Burlington, ON) were mixed in a final volume of 50 μL. The temperature cycles were as follows: 95 °C for 3 min, 18 cycles of 95 °C for 45 s, 62 °C for 1 min, and 68 °C for 2 min/kb (12 min), with a final extension at 68 °C for 2 min/kb (12 min). Linearized vectors were subsequently treated with 10 units of DpnI for 3 h at 37 °C to remove parental templates. Deletion mutants were constructed by using the In-Fusion cloning system (Clontech) according to the manufacturer's recommendations. Briefly, each sequence of interest was amplified using Advantage HD polymerase (Clontech) for 20 cycles by PCR using the primer pairs indicated in Table 1. The amplified DNA inserts were purified with the QIAquick gel extraction kit (Qiagen) and subcloned into the linear vector using the In-Fusion cloning kit. Each cloning reaction mixture contained 100 ng of linear vector, 200 ng of insert DNA, 1× In-Fusion Reaction Buffer, and 1 μL of In-Fusion enzyme in a final volume of 10 μL. The In-Fusion cloning reaction mixture was incubated for 15 min at 37 °C, followed by 15 min at 50 °C and 5 min on ice; 40 μL of TE buffer (pH 8) was added to each reaction mixture, and 2.5 μL of the In-Fusion mixture was then used to transform Origami B(DE3) competent cells. All clones generated by chain reaction amplification were sequenced throughout the amplified region.

Purification of Recombinant Nucleolin. All recombinant nucleolin deletion mutants were fused at the N-terminus to *Escherichia coli* maltose-binding protein (MBP). The MBP-fused proteins were purified on an amylose column (New England Biolabs) following the manufacturer's protocol. The proteins were then dialyzed and concentrated in assay buffer [20 mM Tris-HCl (pH 7.4), 5 mM NaCl, and 1 mM EDTA, in 50% glycerol]. All purified fusion proteins migrated as single species via SDS−PAGE. Protein concentrations were determined by the Bradford protein assay (Bio-Rad).

CD Spectroscopy. Oligonucleotide stocks were diluted to 5 μM in 50 mM Tris-HCl (pH 7.4). Pu47ss (single-stranded/unstructured c-*MYC* NHE III_1 oligo) samples were incubated with either assay buffer or 5 μM recombinant nucleolin protein at room temperature for 30 min to reach equilibrium prior to CD spectroscopy. To assemble the Pu47 oligo into the c-*MYC* G-quadruplex conformation, we heated the oligo at 95 °C for 10 min and left it to cool gradually to room temperature. The assembled G-quadruplex was then incubated with either assay buffer or 5 μM recombinant protein at room temperature for 30 min prior to CD spectroscopy. CD spectra were recorded on a Jasco (Easton, MD) model 810 spectropolarimeter at room temperature, using a quartz cell with a 1 mm optical path length and an instrument scanning speed of 100 nm/min, with a response time of 1 s and over a wavelength range of 225−325 nm. The reported spectrum of each sample represents the average of three scans. The spectral contributions from buffer and protein were subtracted as appropriate. To determine the stability of the G-quadruplex structures induced by the different nucleolin deletion mutants, the molar ellipticity versus temperature profiles (CD melting curves) of the G-quadruplexes were measured at 262 nm for the G-quadruplex using a temperature gradient of 1 °C/min from 20 to 95 °C.

Electrophoretic Mobility Shift Assay (*EMSA*). The Pu47 oligonucleotide containing the c-*MYC* NHE III_1 region was radiolabeled via incubation of the DNA oligo with [γ-^{32}P]dATP and T4 polynucleotide kinase (Fermentas). Radiolabeled Pu47 was preassembled into a G-quadruplex structure via incubation for 10 min at 95 °C in the presence of 100 mM KCl and allowed to gradually cool to room temperature. The radiolabeled oligo was then purified by electrophoresis on a 12% nondenaturing polyacrylamide gel. Binding of G-quadruplex DNA was conducted in 20 μL reaction mixtures containing 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1 mM DTT, 50 ng/μL poly(dI-dC), 4 μg/mL BSA, and 25 mM KCl. Glycerol (5%) was added to each EMSA reaction mixture immediately before it was loaded onto a 4% nondenaturing polyacrylamide gel containing 0.5× TBE. Protein complexes were resolved by running the gel at 10 mA for 1 h at room temperature. Dried EMSA gels were exposed on a phosphor screen for 24 h, after which bound and unbound DNA was visualized and quantified using a Storm 820 phosphorimager and ImageQuant (Molecular Dynamics).

In Vitro Transcription. In vitro transcription assays were performed using the HeLaScribe Nuclear Extract In Vitro Transcription System (Promega) following the manufacturer's protocol, with one exception: the DNA template used was negatively supercoiled plasmid DNA containing the c-*MYC* promoter region (a gift from B. Vogelstein). Briefly, 25 μL reaction mixtures

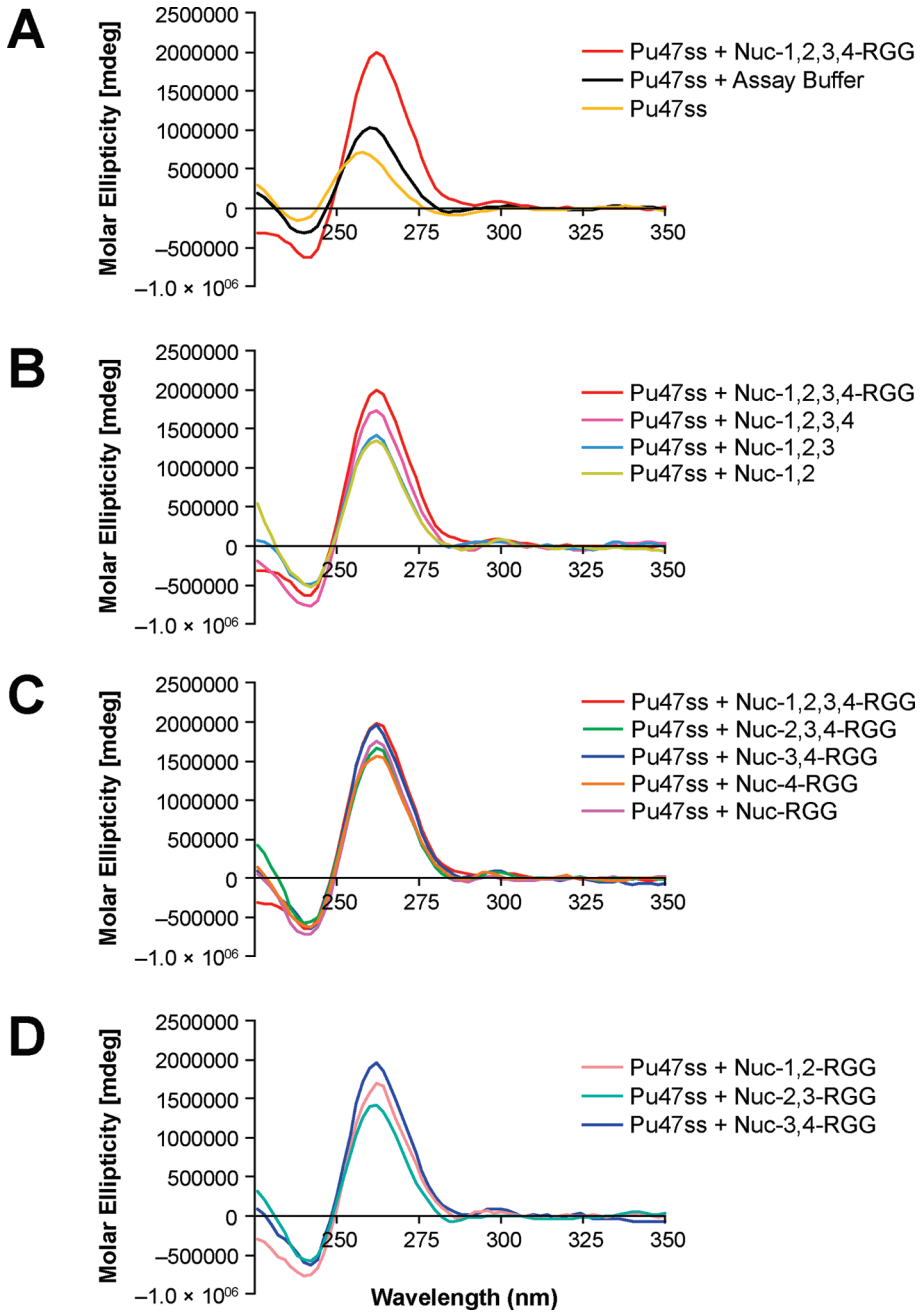

FIGURE 3: Effect of deletion mutagenesis on the ability of nucleolin to induce c-*MYC* G-quadruplex formation. (A) CD spectra of Pu47ss after incubation with Nuc-1,2,3,4-RGG, or assay buffer containing 20 mM Tris-HCl (pH 7.4), 5 mM NaCl, and 1 mM EDTA, in 50% glycerol. (B) CD spectra of Pu47ss after incubation with C-terminal nucleolin deletion mutants. (C) CD spectra of Pu47ss after incubation with nucleolin's N-terminal deletion mutants. (D) Comparison of the effect of RBD substitution on the ability of Nuc-3,4-RGG to induce G-quadruplex formation. Formation of a parallel G-quadruplex structure is reflected by the change in wavelength from 258 nm (single-stranded DNA) to 262 nm (G-quadruplex DNA) and increased molar ellipticity at 262 nm.

containing 1 mg of template DNA, 1× reaction buffer, 3 mM $MgCl_2$, 0.4 mM rATP, 0.016 mM rUTP, 0.4 mM rCTP, 0.4 mM rGTP, and 10 μCi of [α-^{32}P]rUTP (Amersham Biosciences) were assembled on ice. Transcription was initiated via addition of 8 units (50 μg) of HeLa nuclear extract, and the reaction mixture was allowed to incubate for 60 min at 30 °C. The reactions were then terminated with 175 μL of stop solution. After phenol/chloroform extraction, RNA transcripts were ethanol precipitated, dried, and redissolved in formamide loading buffer. RNA transcripts were resolved by denaturing gel electrophoresis (6% PAGE, 1× TBE). Radioactive transcription products were detected by autoradiography using a Storm 820 phosphorimager and ImageQuant (Molecular Dynamics).

RESULTS

Generation of Nucleolin Deletion Mutants. To gain an understanding, at the molecular level, of which of the nucleolin domains contain the c-*MYC* G-quadruplex binding activity and to establish the biological relevance of the various domains of nucleolin on c-*MYC* repression, we designed a number of nucleolin deletion mutants (Figure 2B). Full-length nucleolin cannot be expressed in *E. coli*, but deletion of the N-terminus has been shown to permit adequate expression of recombinant nucleolin (*40*). The Nuc-1,2,3,4-RGG expression vector, which carries the coding region for RBDs 1−4, as well as the RGG region, was generously provided by L. A. Hanakahi and was used to generate

the different nucleolin deletion mutants listed in Figure 2B. All recombinant proteins were purified and dialyzed in assay buffer as described above to concentrate the protein and remove excess salt.

C-Terminal Deletion Mutagenesis Strongly Impairs Nucleolin's Ability To Promote G-Quadruplex Formation. Conversions of the Pu47ss oligomer containing the c-*MYC* G-quadruplex-forming motif to the G-quadruplex conformation, induced by the different nucleolin deletion mutants, were monitored via CD. Each deletion mutant was incubated with Pu47ss for 1 h at room temperature before CD analysis. Consistent with our previous results (*20*), we found that Nuc-1,2,3,4-RGG can strongly induce the formation of a parallel c-*MYC* G-quadruplex structure, as observed by a shift of the maximum positive peak from 258 nm to the G-quadruplex signature peak at 262 nm (Figure 3A). Similar results were obtained for all the nucleolin deletion mutants; however, the intensity of the peaks varied among the different deletion mutants, with the C-terminal domain deletion mutants having a more dramatic weakening of their ability to induce G-quadruplex formation (Figure 3B) than the N-terminal deletion mutants (Figure 3C). Deletion of the RGG and RBD4 domains was most detrimental, suggesting that these domains play an important role in the induction of the c-*MYC* G-quadruplex structure. Deletion mutant Nuc-3,4-RGG induced G-quadruplex formation to the same extent as Nuc-1,2,3,4-RGG. In addition, substitution of RBDs 3 and 4 with either RBDs 1 and 2 or RBDs 2 and 3 in Nuc-3,4-RGG did not lead to equal G-quadruplex formation, suggesting that the RBDs are not equivalent and that RBDs 3 and 4 are critical for G-quadruplex stabilization (Figure 3D). Furthermore, removal of the RGG domain from Nuc-1,2-RGG, Nuc-2,3-RGG, or Nuc-3,4-RGG completely abolished the protein's ability to induce G-quadruplex formation (data not shown).

The C-Terminal Region of Nucleolin Containing RBDs 3 and 4 and the RGG Domain Induces Formation of a Stable c-MYC G-Quadruplex Structure. To assess the effect of each deletion mutant on the stability of the G-quadruplex structure, we determined the thermal stability by measuring the molar ellipticity of the G-quadruplex at 262 nm at increasing temperatures. Table 2 lists the effect of each deletion mutant on c-*MYC* G-quadruplex stability. Consistent with previous studies in our laboratory (*20*), we found that Nuc-1,2,3,4-RGG induces the formation of a stable c-*MYC* G-quadruplex structure, as observed by a shift to the right of the melting curve at 262 nm and a melting temperature of 59 °C (Figure 4A). None of the deletion mutants lacking the RGG domain were able to induce a G-quadruplex structure that was as stable as the one induced by Nuc-1,2,3,4-RGG (Figure 4B). Several of the N-terminal deletion mutants, however, were able to induce the formation of a stable G-quadruplex structure, particularly Nuc-3,4-RGG and Nuc-RGG, which induced the formation of G-quadruplexes with melting temperatures of 58 and 55 °C, respectively (Figure 4C). In addition, substitution of RBDs 3 and 4 with RBDs 1 and 2 or RBDs 2 and 3 in the presence of the RGG domain demonstrated that the different RBDs are not equivalent, as these proteins were not able to stabilize the c-*MYC* G-quadruplex to the same extent as the Nuc-3,4-RGG protein (Figure 4D). In summary, these results demonstrate that the C-terminal domain of nucleolin plays a critical role in the formation of the c-*MYC* G-quadruplex structure. In addition, we demonstrate that the RGG domain is essential for c-*MYC* G-quadruplex stabilization. Furthermore, our results confirm previous reports that show that the RGG domain plays a critical role in G-quadruplex binding (*41*−*43*).

Table 2: Effect of Nucleolin Deletion Mutants on c-*MYC* G-Quadruplex Thermal Stability

nucleolin deletion mutant	T_m (°C)	$-\Delta T_m$ in comparison to that of Nuc-1,2,3,4-RGG
Nuc-1,2,3,4-RGG	59	–
Nuc-1,2,3,4	43	16
Nuc-1,2,3	41	18
Nuc-1,2	43	16
Nuc-2,3,4-RGG	51	8
Nuc-3,4-RGG	58	1
Nuc-4-RGG	43	16
Nuc-RGG	55	4
Nuc-1,2-RGG	52	7
Nuc-2,3-RGG	43	16

The Minimal c-MYC G-Quadruplex Binding Domain of Nucleolin Consists of RBDs 3 and 4 and the RGG Domain. In a previous study, we showed that a recombinant protein containing nucleolin RBDs 1−4, as well as the RGG domain, binds preferentially and with high affinity to parallel c-*MYC*-like G-quadruplex structures (*20*). In an attempt to define the smaller nucleolin subdomains capable of interacting with the c-*MYC* G-quadruplex, we performed gel EMSAs on select nucleolin deletion mutants with a G-quadruplex formed from the Pu47 oligonucleotide. On the basis of our CD spectropolarimetric and thermal stability studies, we chose Nuc-1,2,3,4-RGG, Nuc-3,4-RGG, and Nuc-RGG for comparison of their c-*MYC* G-quadruplex binding affinities (Figure 5). The deletion proteins were expressed in *E. coli* as chimeric MBP fusion proteins and purified to homogeneity. Consistent with our previous results, Nuc-1,2,3,4-RGG bound with high affinity to the c-*MYC* G-quadruplex structure, as determined by a strong single shift in the mobility of the radiolabeled c-*MYC* G-quadruplex (Figure 5A). The single mobility shift observed at both low and high protein concentrations suggests that nucleolin is binding to each G-quadruplex as a monomer. In addition, we found that the minimal c-*MYC* G-quadruplex-binding domain of nucleolin consists of RBDs 3 and 4, as well as the RGG domain (Figure 5B). Interestingly, the interaction between Nuc-3,4-RGG and the c-*MYC* G-quadruplex structure was significantly weaker than that of the Nuc-1,2,3,4-RGG protein with the G-quadruplex (Figure 5B). Taken together with the results from our CD and thermal stability studies, the weaker binding of Nuc-3,4-RGG indicates that while RBDs 1 and 2 are not essential for the induction of formation of the c-*MYC* G-quadruplex, these domains are important for stabilizing the interaction between nucleolin and the c-*MYC* G-quadruplex. It appears that the C-terminus of nucleolin is critical for the initial recognition of the c-*MYC* NHE III_1 sequence, which in turn promotes G-quadruplex formation, while the rest of the protein may serve to further stabilize the interaction of the protein with the assembled G-quadruplex. The high affinity of nucleolin for the c-*MYC* G-quadruplex is most likely achieved via a combination of multiple weak interactions between the different protein domains and the c-*MYC* G-quadruplex. Accordingly, deletion of these domains would result in a lower binding affinity, as shown in our EMSA studies. Furthermore, while Nuc-RGG was able to effectively induce the formation of the c-*MYC* G-quadruplex in solution, no binding between the Nuc-RGG deletion mutant and the c-*MYC* G-quadruplex was detected by this method (Figure 5C). This suggests that while the RGG domain is essential for the induction of a stable c-*MYC* G-quadruplex structure, it is not sufficient for achieving a

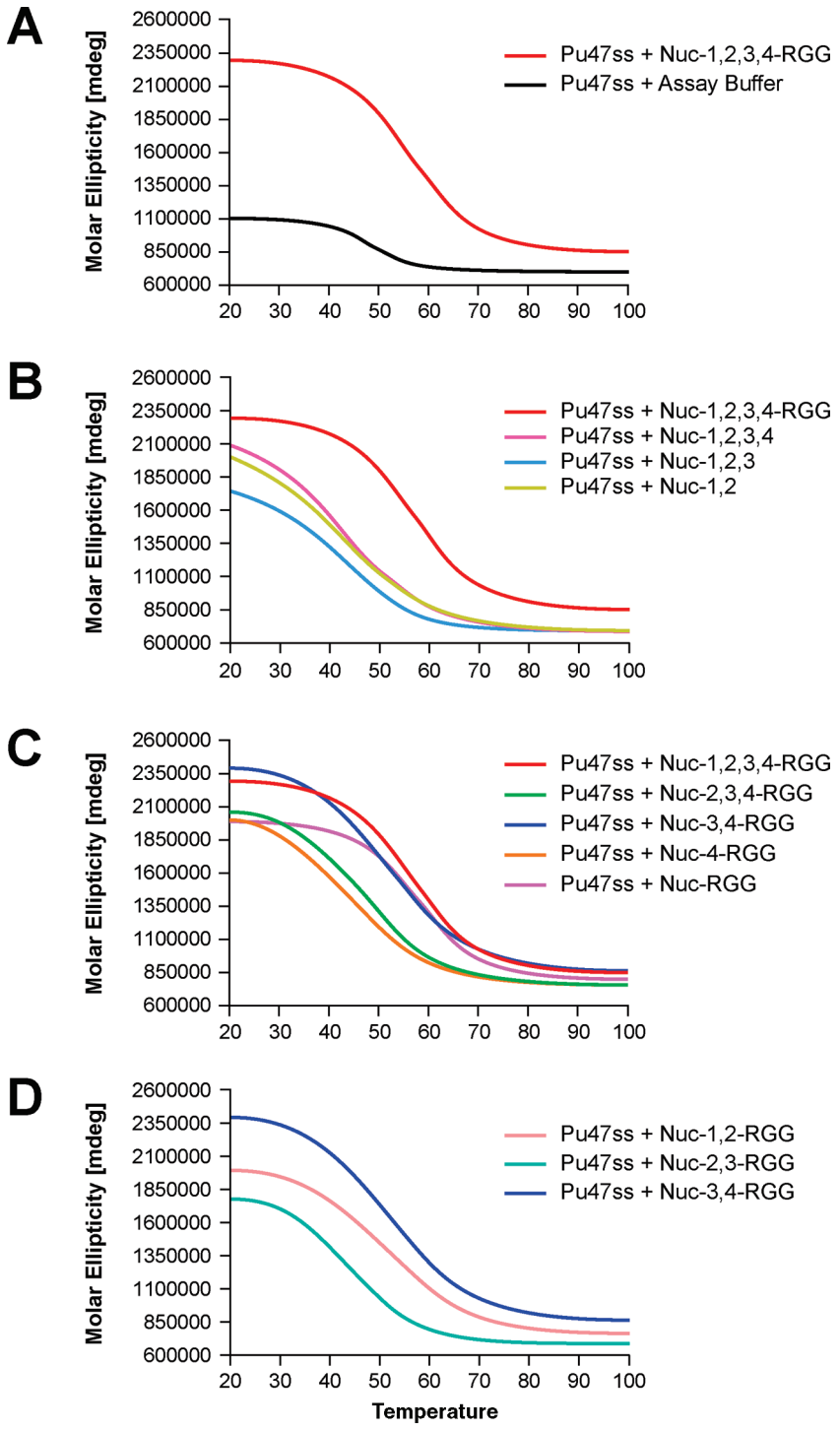

FIGURE 4: Effect of deletion mutagenesis on the ability of nucleolin to promote the formation of a thermally stable c-*MYC* G-quadruplex. (A) Melting curves obtained for Pu47 containing the c-*MYC* G-quadruplex motif after incubation with the various nucleolin deletion mutants, including Nuc-1,2,3,4-RGG. (B) Effect of nucleolin C-terminal deletions on c-*MYC* G-quadruplex stability. (C) Effect of nucleolin N-terminal deletions on c-*MYC* G-quadruplex stability. (D) Comparison of the effects that Nuc-1,2-RGG, Nuc-2,3-RGG, and Nuc-3,4-RGG have on c-*MYC* G-quadruplex stabilization.

stable interaction with the c-*MYC* G-quadruplex. It is possible that the RGG domain interacts very weakly with the c-*MYC* G-quadruplex structure and that these interactions are easily disrupted as a result of the stress that the sample undergoes during the electromobility separation (Figure 5C). Taken together, these results suggest that the C-terminal region of nucleolin containing RBDs 3 and 4 as well as the RGG domain is the minimal domain able to stably bind to the c-*MYC* G-quadruplex structure.

Nuc-3,4-RGG Inhibits c-MYC Promoter Activity in Vitro. Our laboratory has previously demonstrated by a luciferase assay that transient expression of full-length nucleolin in mammalian cells can exert a strong and dose-dependent inhibitory effect on the luciferase activity of a c-*MYC* promoter-driven construct (*20*). In this study, we used the same c-*MYC* reporter plasmid (Figure 6A) to investigate by in vitro transcription whether the Nuc-1,2,3,4-RGG recombinant protein retains the dose-dependent inhibitory activity on c-*MYC* promoter activity. The multiple bands observed in the gel correspond to the transcripts from the several transcription initiation sites found in the luciferase coding region (*44*). Here we demonstrate that Nuc-1,2,3,4-RGG strongly

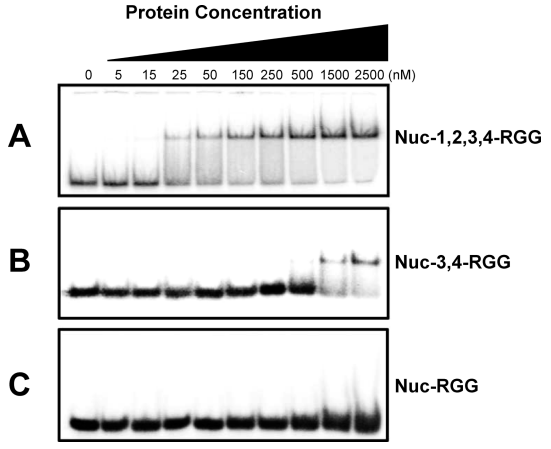

FIGURE 5: Comparison of the binding affinity of Nuc-1,2,3,4-RGG, Nuc-3,4-RGG, and Nuc-RGG for the c-*MYC* G-quadruplex. The radiolabeled oligomer (10000 cpm) was assembled into a G-quadruplex structure and incubated with Nuc-1,2,3,4-RGG (A), Nuc-3,4-RGG (B), or Nuc-1,2,3,4-RGG (C) at the indicated concentrations in a 20 μL reaction mixture containing 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1 mM DTT, 50 ng/μL poly(dI-dC), 4 μg/mL BSA, and 25 mM KCl for 1 h at room temperature.

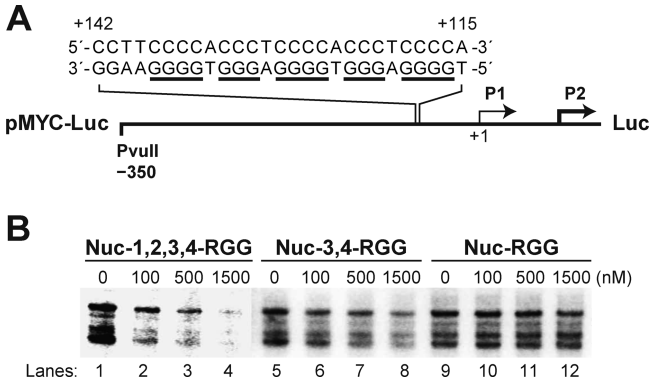

FIGURE 6: Effect of Nuc-1,2,3,4-RGG, Nuc-3,4-RGG, and Nuc-RGG on c-*MYC* promoter activity as determined by in vitro transcription. (A) Diagram of the p*MYC*-Luc reporter construct containing the c-*MYC* NHE III_1 region that can assemble into a G-quadruplex structure. (B) Effect of Nuc-1,2,3,4-RGG (lanes 1−4), Nuc-3,4-RGG (lanes 5−8), and Nuc-RGG (lanes 9−12) on c-*MYC* promoter activity as determined by in vitro transcription assay. Reaction mixtures contained 1 μg of p*MYC*-WT (Del-4) plasmid and the indicated concentrations of nucleolin. The plasmid and recombinant protein were incubated for 1 h at 4 °C prior to the addition of HeLa nuclear extract. In vitro transcription reactions were conducted at 42 °C for exactly 60 min, after which all reactions were stopped by the addition of stop buffer, as described in Materials and Methods. The multiple bands correspond to the transcripts from the several transcription initiation sites found in the luciferase coding region (*44*).

inhibits c-*MYC* promoter activity in a dose-dependent manner, as shown by the substantial decrease in the amount of radiolabeled luciferase transcript synthesized after incubation of the reporter plasmid with increasing concentrations of Nuc-1,2,3,4-RGG (Figure 6B, lanes 1−4). Similarly, we show that Nuc-3,4-RGG is able to inhibit c-*MYC* promoter activity in a dose-dependent manner, although this repression is significantly weaker than that exerted by Nuc-1,2,3,4-RGG (Figure 6B, lanes 5−8). Furthermore, we report that while the RGG domain is essential for the induction of formation of a stable c-*MYC* G-quadruplex, this domain is not sufficient to affect c-*MYC* promoter activity (Figure 6B, lanes 9−12). Consequently, it appears that the minimal c-*MYC* G-quadruplex binding region retaining the transcription repression activity of nucleolin consists of RBDs 3 and 4 and the RGG domain.

DISCUSSION

A growing number of G-quadruplex-interactive proteins are being identified in diverse organisms (*45*). The existence of proteins that interact with high affinity and selectivity for G-quadruplex structures to modulate or stabilize them provides strong arguments for the biological relevance of G-quadruplex structures in vivo. Our laboratory has previously identified nucleolin as a c-*MYC* G-quadruplex-binding protein that represses c-*MYC* promoter activity by inducing the formation of the c-*MYC* G-quadruplex structure (*20*). Here we report that the C-terminal region of nucleolin containing RBDs 3 and 4 and the RGG domain is essential for c-*MYC* G-quadruplex binding and stabilization. Furthermore, we show that this segment of the protein is the minimal region able to repress c-*MYC* transcription.

Nucleolin is a modular protein that is found in organisms ranging from yeast to mammals (*32*). The N-terminal domain of nucleolin comprises long acidic stretches interspersed with basic repeats, and its length is quite variable among the different species (*32*). This domain is highly phosphorylated, and it is believed to assume a nonglobular extended structure (*46*). The N-terminal domain of nucleolin has been shown to play a role in chromatin condensation, protein−protein interactions, and nucleolin functional regulation (*32*, *47*). On the other hand, the four RBDs found in the central region of nucleolin are highly conserved in human, rat, mouse, hamster, chicken, and *Xenopus laevis* (*32*). Interestingly, the RBDs found in these species are less conserved within the same protein than between RBDs from the different species (*32*). The RGG domain of nucleolin is defined by spaced Arg-Gly-Gly repeats interspersed with aromatic amino acids. CD and homology studies of this domain suggest that this region can adopt a highly flexible helical conformation consisting of repeated β-turns with arginine and phenylalanine side chains projecting outside the spiral structure (*48*). It was originally thought that this domain would create electrostatic and hydrophobic ridges prone to interacting nonspecifically with RNA and DNA (*48*); however, a number of recent studies have shown that this structure is able to confer substrate specificity (*42*, *49*, *50*). These findings are further supported by our previous EMSA and filter binding studies in which we showed that recombinant nucleolin containing the RGG domain, as well as all four RBDs, can discriminate between different G-quadruplex structures and bind preferentially to parallel c-*MYC*-like G-quadruplex structures (*20*).

While the results of the EMSA and the effect on transcription of the 3,4-RGG and 1,2,3,4-RGG regions are in reasonable accord, i.e., 3,4-RGG is much less effective as a transcriptional inhibitor than 1,2,3,4-RGG and EMSA showing a similar loss of binding of 3,4-RGG to the *MYC* G-quadruplex relative to the 1,2,3,4-RGG, there is little difference between the T_m values of the *MYC* G-quadruplex in the presence of the two proteins. There are a number of possible reasons for this discrepancy. While in the CD experiment the concentrations of the DNA and protein are equivalent, in the EMSA and transcription assays they are very different, with the protein greatly in excess by several orders of magnitude. However, there are other possibilities for this discrepancy, including limitation in the window of change that can occur in the ΔT_m after the G-quadruplex is formed. If the 3,4-RGG and 1,2,3,4-RGG regions are both

effective in facilitating the formation of the G-quadruplex, then the T_m values are likely to be approximately the same. However, if additional interactions due to RBDs 1 and 2 are required to form a stable complex that will survive the EMSA and competition with other proteins in cells, then the apparent discrepancy between the results can be explained. Another problem in correlating the cell-free data with the cellular data is that in cells the G-quadruplex is formed under negative superhelicity (*7*), whereas the cell-free system uses a single-stranded template. Thus, the dynamics are quite different in addition to competition in cells with other proteins, such as NM23-H2, Sp1, and CNBP, which also bind to this element. Thus, the more dynamic state of the EMSA may be a better measure of the effect on transcription than a T_m that measures just the ability of the protein to capture the folded species. The first identified G-quadruplex inducer protein is the β subunit of the *Oxytricha* telomere end-binding protein β (*51*, *52*). This protein has been reported to promote the formation of a G-quadruplex structure in telomeric DNA in a cell cycle−dependent manner in which the telomere end-binding protein β phosphorylation state is directly linked with G-quadruplex formation (*53*). Because nucleolin function is also regulated by phosphorylation, it would be interesting to investigate whether changes in nucleolin phosphorylation in these parameters would affect nucleolin−c-*MYC* promoter interaction as well as c-*MYC* G-quadruplex formation.

Furthermore, a number of stress stimuli have been shown to modulate nucleolin localization and function (*28*). For example, nucleoplasmic localization of nucleolin has been associated with the pro-apoptotic effects of a number of anticancer drugs (*30*, *54*, *55*). Specifically, the mechanisms of action of cisplatin and Quarfloxin are linked with the redistribution of nucleolin from the nucleoli to the nucleoplasm (*30*, *54*). Consequently, it has been suggested that nucleolin can function as a stress-sensitive tumor suppressor (*28*). If this is the case, it is possible that stress-induced nuclear localization would allow nucleolin to interact with the c-*MYC* promoter to induce G-quadruplex formation. Because nucleolin is a bona fide target of the c-*MYC* proto-oncogene (*56*), it is not unreasonable to suggest that nucleolin may form part of a negative feedback mechanism to prevent aberrant c-*MYC* expression.

In summary, there is a large body of evidence supporting the function of nucleolin as a tumor suppressor. Our findings that nucleolin represses c-*MYC* expression by inducing the formation of a G-quadruplex structure provide an explanation for the importance of nucleolin translocation from the nucleolus to the nucleoplasm after cellular stress or drug treatment.

ACKNOWLEDGMENT

We are grateful to Dr. Leslyn A. Hanakahi for generously providing plasmids pNuc-1,2,3,4-RGG, pNuc-3,4-RGG, pNuc-1,2-RGG, and pNuc-RGG and to Dr. Bert Vogelstein for kindly providing reporter plasmid pDel-4. We thank Dr. David Bishop for preparing, proofreading, and editing the final version of the manuscript and figures.

REFERENCES

1. Wierstra, I., and Alves, J. (2008) The *c-myc* promoter: Still *M*yster*Y* and *C*hallenge. *Adv. Cancer Res. 99*, 113–333.
2. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A. 99*, 11593–11598.
3. Davis, J. T. (2004) G-quartets 40 years later: From 5′-GMP to molecular biology and supramolecular chemistry. *Angew. Chem., Int. Ed. 43*, 668–698.
4. Arnott, S., Chandrasekaran, R., and Marttila, C. M. (1974) Structures for polyinosinic acid and polyguanylic acid. *Biochem. J. 141*, 537–543.
5. Ghosal, G., and Muniyappa, K. (2006) Hoogsteen base-pairing revisited: Resolving a role in normal biological processes and human diseases. *Biochem. Biophys. Res. Commun. 343*, 1–7.
6. Zimmerman, S. B., Cohen, G. H., and Davies, D. R. (1975) X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J. Mol. Biol. 92*, 181–192.
7. Brooks, T. A., and Hurley, L. H. (2009) The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer 9*, 849–861.
8. Sun, D., and Hurley, L. H. (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: Implications for drug targeting and control of gene expression. *J. Med. Chem. 52*, 2863–2874.
9. González, V., and Hurley, L. H. (2010) The c-MYC NHE III_1: Function and regulation. *Annu. Rev. Pharmacol. Toxicol. 50*, 111–129.
10. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res. 33*, 2908–2916.
11. Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res. 35*, 406–413.
12. Verma, A., Halder, K., Halder, R., Yadav, V. K., Rawal, P., Thakur, R. K., Mohd, F., Sharma, A., and Chowdhury, S. (2008) 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*, 5641–5649.
13. Eddy, J., and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res. 34*, 3887–3896.
14. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res. 34*, 2723–2735.
15. Dai, J., Chen, D., Jones, R. A., Hurley, L. H., and 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.
16. Matsugami, A., Okuizumi, T., Uesugi, S., and Katahira, M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem. 278*, 28147–28153.
17. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Am. Chem. Soc. 129*, 4386–4392.
18. Simonsson, T. (2001) G-quadruplex DNA structures: Variations on a theme. *Biol. Chem. 382*, 621–628.
19. Yang, D., and Hurley, L. H. (2006) Structure of the biologically relevant G-quadruplex in the c-MYC promoter. *Nucleosides, Nucleotides Nucleic Acids 25*, 951–968.
20. González, V., Guo, K., Hurley, L., and Sun, D. (2009) Identification and characterization of nucleolin as a c-*myc* G-quadruplex-binding protein. *J. Biol. Chem. 284*, 23622–23635.
21. Daniely, Y., and Borowiec, J. A. (2000) Formation of a complex between nucleolin and replication protein A after cell stress prevents initiation of DNA replication. *J. Cell Biol. 149*, 799–810.
22. Daniely, Y., Dimitrova, D. D., and Borowiec, J. A. (2002) Stress-dependent nucleolin mobilization mediated by p53-nucleolin complex formation. *Mol. Cell. Biol. 22*, 6014–6022.
23. Kim, K., Dimitrova, D. D., Carta, K. M., Saxena, A., Daras, M., and Borowiec, J. A. (2005) Novel checkpoint response to genotoxic stress mediated by nucleolin-replication protein A complex formation. *Mol. Cell. Biol. 25*, 2463–2474.
24. Klibanov, S. A., O'Hagan, H. M., and Ljungman, M. (2001) Accumulation of soluble and nucleolar-associated p53 proteins following cellular stress. *J. Cell Sci. 114*, 1867–1873.
25. Martelli, A. M., Robuffo, I., Bortul, R., Ochs, R. L., Luchetti, F., Cocco, L., Zweyer, M., Bareggi, R., and Falcieri, E. (2000) Behavior of nucleolar proteins during the course of apoptosis in camptothecin-treated HL60 cells. *J. Cell. Biochem. 78*, 264–277.
26. Mi, Y., Thomas, S. D., Xu, X., Casson, L. K., Miller, D. M., and Bates, P. J. (2003) Apoptosis in leukemia cells is accompanied by alterations in the levels and localization of nucleolin. *J. Biol. Chem. 278*, 8572–8579.

27. Mongelard, F., and Bouvet, P. (2007) Nucleolin: A multiFACeTed protein. *Trends Cell Biol. 17*, 80–86.
28. Storck, S., Shukla, M., Dimitrov, S., and Bouvet, P. (2007) Functions of the histone chaperone nucleolin in diseases. *Subcell. Biochem. 41*, 125–144.
29. Saxena, A., Rorie, C. J., Dimitrova, D., Daniely, Y., and Borowiec, J. A. (2006) Nucleolin inhibits Hdm2 by multiple pathways leading to p53 stabilization. *Oncogene 25*, 7274–7288.
30. Kito, S., Morimoto, Y., Tanaka, T., Haneji, T., and Ohba, T. (2005) Cleavage of nucleolin and AgNOR proteins during apoptosis induced by anticancer drugs in human salivary gland cells. *J. Oral Pathol. Med. 34*, 478–485.
31. Wesierska-Gadek, J., Schloffer, D., Kotala, V., and Horky, M. (2002) Escape of p53 protein from E6-mediated degradation in HeLa cells after cisplatin therapy. *Int. J. Cancer 101*, 128–136.
32. Ginisty, H., Sicard, H., Roger, B., and Bouvet, P. (1999) Structure and functions of nucleolin. *J. Cell Sci. 112*, 761–772.
33. Serin, G., Joseph, G., Faucher, C., Ghisolfi, L., Bouche, G., Amalric, F., and Bouvet, P. (1996) Localization of nucleolin binding sites on human and mouse pre-ribosomal RNA. *Biochimie 78*, 530–538.
34. Serin, G., Joseph, G., Ghisolfi, L., Bauzan, M., Erard, M., Amalric, F., and Bouvet, P. (1997) Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *J. Biol. Chem. 272*, 13109–13116.
35. Ginisty, H., Amalric, F., and Bouvet, P. (1998) Nucleolin functions in the first step of ribosomal RNA processing. *EMBO J. 17*, 1476–1486.
36. Ginisty, H., Amalric, F., and Bouvet, P. (2001) Two different combinations of RNA-binding domains determine the RNA binding specificity of nucleolin. *J. Biol. Chem. 276*, 14338–14343.
37. Ginisty, H., Serin, G., Ghisolfi-Nieto, L., Roger, B., Libante, V., Amalric, F., and Bouvet, P. (2000) Interaction of nucleolin with an evolutionarily conserved pre-ribosomal RNA sequence is required for the assembly of the primary processing complex. *J. Biol. Chem. 275*, 18845–18850.
38. He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B., and Kinzler, K. W. (1998) Identification of c-MYC as a target of the APC pathway. *Science 281*, 1509–1512.
39. Williams, M., Louw, A. I., and Birkholtz, L. M. (2007) Deletion mutagenesis of large areas in *Plasmodium falciparum* genes: A comparative study. *Malar. J. 6*, 64.
40. Hanakahi, L. A., Dempsey, L. A., Li, M.-J., and Maizels, N. (1997) Nucleolin is one component of the B cell-specific transcription factor and switch region binding protein, LR1. *Proc. Natl. Acad. Sci. U.S.A. 94*, 3605–3610.
41. Hanakahi, L. A., Sun, H., and Maizels, N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem. 274*, 15908–15912.
42. Ramos, A., Hollingworth, D., and Pastore, A. (2003) G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA 9*, 1198–1207.
43. Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J. 20*, 4803–4813.
44. Vopálenský, V., Mašek, T., Horváth, O., Vicenová, B., Mokrejš, M., and Pospíšek, M. (2008) Firefly luciferase gene contains a cryptic promoter. *RNA 14*, 1720–1729.
45. Fry, M. (2007) Tetraplex DNA and its interacting proteins. *Front. Biosci. 12*, 4336–4351.
46. Sapp, M., Richter, A., Weisshart, K., Caizergues-Ferrer, M., Amalric, F., Wallace, M. O., Kirstein, M. N., and Olson, M. O. (1989) Characterization of a 48-kDa nucleic-acid-binding fragment of nucleolin. *Eur. J. Biochem. 179*, 541–548.
47. Erard, M., Lakhdar-Ghazal, F., and Amalric, F. (1990) Repeat peptide motifs which contain β-turns and modulate DNA condensation in chromatin. *Eur. J. Biochem. 191*, 19–26.
48. Ghisolfi, L., Joseph, G., Amalric, F., and Erard, M. (1992) The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. *J. Biol. Chem. 267*, 2955–2959.
49. Corbin-Lickfett, K. A., Chen, I. H., Cocco, M. J., and Sandri-Goldin, R. M. (2009) The HSV-1 ICP27 RGG box specifically binds flexible, GC-rich sequences but not G-quartet structures. *Nucleic Acids Res. 37*, 7290–7301.
50. Darnell, J. C., Jensen, K. B., Jin, P., Brown, V., Warren, S. T., and Darnell, R. B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell 107*, 489–499.
51. Fang, G., and Cech, T. R. (1993) The β subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell 74*, 875–885.
52. Fang, G., and Cech, T. R. (1993) Characterization of a G-quartet formation reaction promoted by the β-subunit of the *Oxytricha* telomere-binding protein. *Biochemistry 32*, 11646–11657.
53. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and 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.
54. 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., and Rice, W. G. (2009) Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis. *Cancer Res. 69*, 7653–7661.
55. Teng, Y., Girvan, A. C., Casson, L. K., Pierce, W. M., Jr., Qian, M., Thomas, S. D., and Bates, P. J. (2007) AS1411 alters the localization of a complex containing protein arginine methyltransferase 5 and nucleolin. *Cancer Res. 67*, 10491–10500.
56. Greasley, P. J., Bonnard, C., and Amati, B. (2000) Myc induces the nucleolin and BN51 genes: Possible implications in ribosome biogenesis. *Nucleic Acids Res. 28*, 446–453.