Discovery of Widespread GTP-Binding Motifs in Genomic DNA and RNA

Edward A. Curtis1,2 and David R. Liu1,2,*
1Department of Chemistry and Chemical Biology
2Howard Hughes Medical Institute
Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA
*Correspondence: drliu@fas.harvard.edu
http://dx.doi.org/10.1016/j.chembiol.2013.02.015

SUMMARY

Biological RNAs that bind small molecules have been implicated in a variety of regulatory and catalytic processes. Inspired by these examples, we used in vitro selection to search a pool of genome-encoded RNA fragments for naturally occurring GTP aptamers. Several aptamer classes were identified, including one (the "G motif") with a G-quadruplex structure. Further analysis revealed that most RNA and DNA G-quadruplexes bind GTP. The G motif is abundant in eukaryotes, and the human genome contains ∼75,000 examples with dissociation constants comparable to the GTP concentration of a eukaryotic cell (∼300 μM). G-quadruplexes play roles in diverse cellular processes, and our findings raise the possibility that GTP may play a role in the function of these elements. Consistent with this possibility, the sequence requirements of several classes of regulatory G-quadruplexes parallel those of GTP binding.

INTRODUCTION

Once thought to function primarily as a passive carrier of genetic information, RNA is now known to play an active role in diverse cellular processes (Tucker and Breaker, 2005; Hüttenhofer et al., 2005; Bartel, 2009; Zhang et al., 2010). Although RNAs perform their biological functions using a variety of mechanisms, in a growing number of known cases, binding of cellular small molecules plays a critical role. Examples include riboswitches, which modulate gene expression in response to metabolites (Tucker and Breaker, 2005), and some ribozymes, which can require cofactors such as GTP and glucosamine-6-phosphate to catalyze reactions (Zhang et al., 2010). In most described cases, the role of RNA-bound small molecules is to modulate RNA folding in a manner that regulates gene expression (Tucker and Breaker, 2005). For example, in the presence of thiamine, the 5’ UTR of the Escherichia coli thiC messenger RNA (mRNA) adopts a secondary structure in which its Shine-Dalgarno sequence is inaccessible to the ribosome, resulting in decreased expression of downstream genes (Winkler et al., 2002). RNA-bound cofactors can also play catalytic roles in reactions catalyzed by ribozymes. The 3’ hydroxyl group of an RNA-bound GTP molecule, for example, acts as a nucleophile in the first step of the self-splicing reaction catalyzed by the Group I intron (Cech, 1990).

Most examples of naturally occurring small-molecule-binding RNAs have been identified using methods that search genomic databases for phylogenetically conserved RNA secondary structures (Barrick et al., 2004). This approach is especially suited to the discovery of riboswitches because they are often physically linked to the genes they regulate, greatly facilitating ligand identification (Winkler et al., 2002). Although they are a powerful way to identify riboswitches, bioinformatic methods are less applicable to aptamers whose ligands cannot be deduced from their genomic context, aptamers that lack canonical secondary structures, or aptamers whose degree of secondary structure conservation falls below a search threshold. A more general limitation of such methods is that they cannot be used to search for new motifs with biochemical functions specified by the experimenter.

Given these considerations, we speculated that additional examples of naturally occurring functional RNAs could be identified by directly selecting for motifs with specific biochemical activities. As an initial application of this approach, we used in vitro selection to search a pool of phylogenetically diverse genome-derived RNA fragments for new examples of naturally occurring GTP aptamers. We chose GTP as our initial target because it is an essential molecule in all known organisms, it is widely used as a substrate by protein enzymes (Dever and Merrick, 1989; Alberts et al., 2007), and two known naturally occurring ribozymes use GTP as a cofactor (Cech, 1990; Teixeira et al., 2004). We identified several classes of genome-encoded GTP aptamers using this approach, the most abundant of which (the "G motif") adopts a G-quadruplex structure. Further analysis revealed that virtually all G-quadruplexes of both RNA and DNA bind GTP. We estimate that ∼75,000 of the G-quadruplexes in the human genome bind GTP with a dissociation constant comparable to the GTP concentration of a typical eukaryotic cell, including members of several different classes of G-quadruplex regulatory elements. The sequence requirements of these elements parallel those of the G motif, suggesting that GTP binding may be involved in the function of some classes of regulatory G-quadruplexes.

RESULTS AND DISCUSSION

Construction of Pools of Genome-Derived RNA Fragments

To begin our search for new examples of naturally occurring small-molecule-binding RNAs, we constructed pools of
genome-derived RNA fragments from which such aptamers could be isolated (Figure 1A). Genomic DNA from several phylogenetically diverse eubacteria (Escherichia coli, Bacillus subtilis, and Bacteroides fragilis), archaeabacteria (Haloarcula marismortui, Aeropyrum pernix, and Methanococcus jannaschii), and eukaryotes (Homo sapiens and Gallus gallus) was randomly fragmented using DNase I, and fragments between 100 and 600 bp were purified by gel electrophoresis. We chose genomic DNA as the starting point of these pools in light of studies suggesting that a larger fraction of the genome is transcribed than was previously thought (Nielsen, 2011), and also because both low-abundance and tissue-specific transcripts would be better represented in such a library than in one generated from cellular RNA.

These genomic DNA fragments were ligated into a vector, amplified by PCR, and transcribed using T7 RNA polymerase to generate pools of genome-derived RNA fragments flanked by defined primer binding sites suitable for in vitro selection experiments (for details, see Supplemental Experimental Procedures available online). In contrast to previous methods to select genomic RNA sequences with specific biochemical properties (Gold et al., 1997; Salehi-Ashtiani et al., 2006; Zimmermann et al., 2010), our pool contained genome-derived RNA fragments from multiple eukaryotic, eubacterial, and archaeabacterial species. Pools from different species can be mixed without sacrificing representation because the number of molecules present at the beginning of a typical in vitro selection experiment (~10^15) is far greater than the number needed to encode even a large genome.

Isolation of Naturally Occurring GTP Aptamers by In Vitro Selection

The RNA pools (containing ~10^12 molecules from each species) were combined and allowed to fold in a buffer containing 20 mM MgCl2 and 200 mM KCl, similar to conditions in which naturally occurring aptamers and ribozymes are typically active. After the combined pools were incubated with GTP immobilized through covalent attachment of its γ-phosphate to agarose beads (Sassanfar and Szostak, 1993; Connell and Yarus, 1994), bound molecules were eluted with EDTA, amplified by RT-PCR, and transcribed to generate RNA for the next round of selection (Figure 1B). After only two rounds of selection (and three GTP-agarose purifications; see Supplemental Experimental Procedures for details), the GTP-binding activity of the enriched pool could be detected (Figure 1C), and after two additional rounds of selection, the surviving molecules (the round 4 pool) were cloned and sequenced.

We evaluated the GTP-binding activities of nine randomly selected clones from the round 4 pool by comparing the fraction of each purified RNA that bound to GTP-agarose with the fraction that bound to control resin lacking GTP. Five of these nine clones bound at least 10-fold more efficiently to GTP agarose than to agarose beads lacking GTP, and four RNAs also bound at least 10-fold more efficiently to GTP agarose than did an unselected random N48 RNA pool (Table S1). Based on the number of GTP-agarose purifications needed to generate a pool with detectable GTP-binding activity, as well as the GTP-binding activities of individual aptamers isolated in the selection, we estimate that one in 10^3-10^5 genome-derived RNA fragments in the starting pool possess sufficient GTP-binding activity to survive the selection (see Supplemental Experimental Procedures for details).

Analysis of 73 unique sequences from the round 4 pool revealed that virtually all were derived from the genomes of either H. sapiens (human, 42 sequences) or G. gallus (chicken, 25 sequences). Approximately half of these sequences mapped to
intergenic regions, but others occurred in exons, introns, antisense to exons, or antisense to introns. To obtain more insight into the potential roles of these GTP aptamers, we characterized their sequence requirements in greater detail.

**Sequence Requirements and Biochemical Characterization of the G Motif Aptamer**

Initial examination of the sequences isolated in the selection revealed that virtually all were guanosine rich (see Supplementary Information for sequences), and that clusters of three or more consecutive guanosines occurred ~4-fold more frequently in both human and chicken sequences that survived the selection than they did in the genomes of these species. To better understand the relationship between G clusters and GTP-binding activity, we characterized the sequence requirements of one of these aptamers, clone 4-56, in greater detail. This aptamer was chosen for initial characterization efforts because (1) it is known to be expressed (as part of a 486 nt transcript of unknown function; Oh et al., 2005); (2) pilot experiments revealed that its GTP-binding activity is conserved in primates (Figure S1), and that clusters of three or more guanosine clusters are required for efficient binding to GTP, and that in the context of adenosine spacers, as the spacer length increases, the binding efficiency decreases (Figure 2C; also see Table S1). Taken together, these sequence requirements are consistent with those of a G-quadruplex structure (Davis, 2004; Figure 2D). Consistent with this hypothesis, the circular dichroism (CD) spectrum of one of the most efficient GTP binders assayed is similar to that of previously described parallel G-quadruplex structures, with a positive peak at ~210 nm, a negative peak at ~240 nm, and a positive peak at ~260 nm (Kypr et al., 2009; Figure 2E). In contrast, CD spectra of RNA A-form helices contain a negative peak at ~210 nm, no peak at ~240 nm, and a positive peak at ~260 nm (Kypr et al., 2009).

Functional RNAs, such as aptamers and ribozymes, typically require metal ions for activity (Pyle, 2002). To characterize the metal ions requirements of the G motif, we first determined whether both Mg$^{2+}$ and K$^+$, the metal ions present in the selection buffer, were required for GTP-binding activity using a construct that binds GTP efficiently in both GTP-agarose pull-down (Table S1) and gel filtration (Table S2) assays. Titration experiments revealed that in the absence of K$^+$, the G motif binds GTP optimally at ~3 mM Mg$^{2+}$ (Figure S2A). In the absence of Mg$^{2+}$, the aptamer can also bind GTP, with a sigmoidal dependence on K$^+$ concentration and a plateau at ~1 M (Figure S2B). The binding efficiencies in Mg$^{2+}$ alone and K$^+$ alone were similar.
to GTP. Sequences compared were those in Figures 2B and 2C (RNA variants) and Figure S2C and S2D (DNA variants). The blue line shows the expected relationship if the GTP-binding activities of DNA and RNA variants of each sequence tested were equal.

For (A)–(C), experiments were performed using the sequence GGGGGAGGGGUGGG. Reported values indicate the average of three independent experiments, and error bars indicate 1 SD. See the Supplemental Experimental Procedures for additional experimental details. See also Figure S2.

The affinity of the G motif for GTP was determined by measuring the amount of GTP bound by the aptamer as a function of GTP concentration (Figure 3B), revealing a dissociation constant ($K_d$) of 270 μM. The affinity of the G motif for GTP is comparable to that of the Group I intron (Moran et al., 1993), a naturally occurring ribozyme that uses GTP as a cofactor, and is in a range that could be physiologically relevant in a eukaryotic cell (for example, the GTP concentration is 130 μM in HeLa cells (Finch et al., 1993), 300 μM in rat C6 glioma cells (Franklin and Twose, 1977), 340 μM in yeast (Koç et al., 2004), and 400 μM in Xenopus eggs (Woodland and Pestell, 1972).

To characterize the binding specificity of the G motif aptamer for GTP, we determined the ability of 16 GTP analogs to inhibit binding of radiolabeled GTP to the G motif aptamer using a gel filtration assay. GTPγS, GDP, GMP, guanosine, dGTP, ddGTP, and cGMP compete with GTP for binding to the G motif, whereas 7-deaza-dGTP, 7-methyl-GTP, 6-thio-GTP, 6-methylthio-GTP, ITTP, XMP, ATP, UTP, and CTP do not (Figure 3C). These results suggest that the G motif makes contacts with both the Hoogsteen and Watson-Crick faces of GTP, but not with the hydroxyl or phosphate groups (Figures 3C and 3D), and are consistent with a model in which the G motif binds GTP by incorporating it as one of the guanosines in the tetrad of a G-quadruplex formed by the aptamer itself (Figure 2D).

Both RNA and DNA sequences are known to form G-quadruplex structures (Davis, 2004), and we hypothesized that DNA versions of the G motif might also bind GTP. To test this hypothesis, we assayed 30 DNA variants of the G motif (identical in sequence to those characterized as RNA sequences in Figures 2B and 2C) for their ability to bind GTP. The results revealed that the RNA variants of the G motif that bind GTP are also typically active as DNA sequences, but the GTP-binding activities of the DNA variants of the G motif tend to be somewhat lower (3.7-fold lower on average in our assay) than their RNA counterparts (Figures 3E, S2C, and S2D).

The G Motif Aptamer Is Widespread in Eukaryotic Genomes

Both site-directed mutagenesis and CD experiments suggest that the G motif adopts a G-quadruplex structure. Based on this analysis, we hypothesized that the "quadruplex folding..."
Consistent with this finding, the $K_d$ of the construct with a $K_d$ of 270
control, and the GTP-binding activities of four of the sequences
sequences bound GTP more efficiently than a random sequence
randomly chosen examples from the human genome as DNA
motifs present in a data set of interest (such as a sequenced genome). To test this hypothesis, we first used the 
algorithm (Huppert and Balasubramanian, 2005), which can identify various types of G-quadruplexes in sequence databases, to search each of the eight genomes from which our RNA pool was derived for the G-quadruplex consensus sequence $G_{3-4}N_{1-7}G_{3-4}N_{1-7}G_{3-4}$. This analysis revealed that the number of potential quadruplex-forming sequences in these genomes ranged from six in $B. subtilis$ to $\sim 140,000$ in $G. gallus$ (Table S3). Previous work has shown that $\sim 376,000$ G-quadruplexes occur in the $H. sapiens$ genome (Todd et al., 2005; Huppert and Balasubramanian, 2005). When normalized for genome size, potential G-quadruplex-forming sequences occurred most frequently in the $A. pernix$, $G. gallus$, and $H. sapiens$ genomes (approximately once every $10^4$ nt), and least frequently in the genome of $B. subtilis$ (approximately once every $10^6$ nt; Table S3).

To estimate the fraction of these potential G-quadruplex-forming sequences that contain the G motif, we synthesized 20 randomly chosen examples from the human genome as DNA and tested them for their ability to bind GTP. All 20 of these sequences bound GTP more efficiently than a random sequence control, and the GTP-binding activities of four of the sequences matched or exceeded that of our reference G-motif (characterized in Figure 3) with a $K_d$ of 270 $\mu M$ (Figure 4A; Table S2). Consistent with this finding, the $K_d$ of the construct with the highest GTP-binding activity (construct 20) was 100 $\mu M$ (Figure 4B). To confirm that the tested sequences form G-quadruplexes, we characterized their structures by CD. Consistent with a previous test of the quadruplex folding rule (Huppert and Balasubramanian, 2007), all 20 sequences exhibited CD spectra consistent with G-quadruplex structures under the conditions tested (Figures 4C and S3).

Further analysis revealed two additional aspects of the relationship between G-quadruplexes and the G motif. First, the structures of the highest-affinity aptamers identified in this study are parallel-strand G-quadruplexes, although not all parallel-strand structures bind GTP efficiently (Figures 4A and S3). Second, the sequences of the highest-affinity aptamers typically contain short spacers connecting G clusters, consistent with the trends shown in Figures 2C and S2D. Indeed, sorting all tested sequences by the maximum allowed spacer length using quadparser revealed that sequences encoding G-quadruplexes with spacers no longer than 1 nt bound GTP $>10$-fold more efficiently than sequences encoding G-quadruplexes with longer spacers, and the average GTP-binding activity continued to decrease as the maximum allowed spacer length increased (Figure 4D; Table S4).

To more systematically characterize the phylogenetic distribution of the G motif aptamer, we used quadparser to search $\sim 80$ sequenced archaeabacterial, eubacterial, and eukaryotic genomes for G-quadruplexes with spacers no longer than 1 nt. This search revealed that the G motif is widespread in sequenced genomes, although its density (the number of examples of the G motif per nucleotide) varies by $>1,000$-fold (Figure 4E; Table S5). G motif density tends to increase with genome size and is $\sim 20$-fold higher in eukaryotic genomes than in those of eubacteria or archaeabacteria (Figure 4E; Table S5).

These results establish that the quadruplex folding rule can be used to identify new examples of the G motif in genomic sequence databases, and that the sequences that satisfy this rule, those with short spacers tend to bind GTP the most efficiently. In addition, they indicate that the density of the G motif is considerably higher in the genomes of eukaryotes than in those of archaea or eubacteria.

GTP-Binding Activity of G-Quadruplex Regulatory Elements

G-quadruplexes have been shown to play roles in the regulation of diverse cellular processes (Kostadinov et al., 2006; Kendrick and Hurley, 2010; Bugaut and Balasubramanian, 2012). To investigate the possibility that GTP binding might play a role in the cellular function of G-quadruplex regulatory elements, we tested five well-studied examples—a transcriptional repressor in the human c-MYC promoter (Siddiqui-Jain et al., 2002; Seenisamy et al., 2004), an enhancer of 3' end formation in the SV40 late transcript (Bagga et al., 1995), an internal ribosomal entry site (IRES) in the human vascular endothelial growth factor (VEGF) gene (Morris et al., 2010), a translational repressor in the 5' UTR of the human NRAS gene (Kumari et al., 2008), and a G-quadruplex derived from the vertebrate telomere sequence (Blackburn, 2001)—for their ability to bind GTP. The GTP-binding activity of each of these regulatory elements was confirmed, although the telomere-derived sequence could only bind GTP in a buffer containing $\mathrm{Sr}^{2+}$, which is known to promote the formation of parallel-strand telemeric G-quadruplexes (Pedroso et al., 2007; Figures 5A and S4). Mutations that are known to reduce the cellular activities of these elements also significantly reduced their abilities to bind GTP (Figure 5A). The dissociation constants of these G-quadruplex regulatory elements for GTP ranged between 60 $\mu M$ (for the c-MYC transcriptional repressor) and 600 $\mu M$ (for the SV40 RNA processing enhancer; Figures 5B–5E).

Although these results are consistent with the possibility that GTP binding plays a role in the cellular mechanism of regulatory G-quadruplexes, it is also possible that the proper functioning of these elements simply requires that they adopt a G-quadruplex structure. Since not all G-quadruplexes bind GTP efficiently (Figures 4A and S3), one way to distinguish between these possibilities is to compare the known sequence requirements of these regulatory elements with those of GTP binding. We performed this comparison using three different approaches. First, we investigated the extent to which the GTP-binding activity of these regulatory elements, rather than simply their ability to form G-quadruplex structures, has been conserved in evolution. This analysis indicated that, despite changes in the primary sequence, the GTP-binding activity of the c-MYC transcriptional repressor and the VEGF IRES has been conserved in primates, and that of the NRAS translational terminator has been conserved in placental mammals (Figure 6).

Second, we investigated the correlation between the ability of previously characterized variants of G-quadruplex regulatory elements to perform their cellular function and to bind GTP. The ability of variants of both the c-MYC transcriptional
repressor and the SV40 RNA processing enhancer to perform their cellular function was strongly correlated with their ability to bind GTP (Figure S5; Table S1). In contrast, we observed only a weak correlation between the ability of variants of the VEGF IRES to promote cap-independent translation and to bind GTP (Figure S5).

Figure 4. Identification of the G Motif in Sequenced Genomes
(A) GTP-binding activity of 20 randomly chosen G-quadruplexes from the human genome. G-quadruplexes were identified using the quadparser algorithm.
(B) Dissociation constant of one of these G-quadruplexes for GTP. Normalized GTP bound = amount of GTP bound at the indicated GTP concentration as measured by gel filtration / (amount of GTP bound at the lowest GTP concentration used in the assay).
(C) CD spectra of three of the human G-quadruplexes tested for the ability to bind GTP in (A). See Figure S3 for CD spectra of all 20 sequences.
(D) Average GTP-binding activity of intramolecular DNA G-quadruplexes as a function of spacer length. The height of each bar indicates the average GTP-binding activity of all G-quadruplexes characterized in this study with the indicated maximum allowed spacer length.
(E) Density of the G motif in phylogenetically diverse eubacterial, archaebacterial, and eukaryotic species as a function of genome size. Examples of the G motif were identified by searching for G-quadruplexes with spacers of no more than 1 nt using the quadparser algorithm.

For (A), (B), and (D), reported values indicate the average of three independent experiments, and error bars indicate 1 SD. See the Supplemental Experimental Procedures and Tables S2, S4, and S5 for additional experimental details. See also Figure S3.
Third, based on our observation that G-quadruplexes with short spacers bind GTP more efficiently than those with longer spacers (Figures 2C, 4D, and S2D), we used bioinformatic methods to determine whether, for any of the genomic contexts in which G-quadruplex regulatory elements are known to be enriched, G-quadruplexes with short spacers show higher enrichment values than those with longer spacers. Enrichment values (defined as density in a particular genomic context / density in the genome being examined) were determined for G-quadruplexes 100 bp upstream of human transcription start sites (Huppert and Balasubramanian, 2007), 100 nt downstream of human polyadenylation sites (Kostadinov et al., 2006), and in 5' UTRs of human mRNAs (Bugaut and Balasubramanian, 2012). G-quadruplexes with short spacers are enriched in each context.
Figure 6. Evolutionary Conservation of the GTP-Binding Activity of G-Quadruplex Regulatory Elements

(A) Conservation of the GTP-binding activity of the c-MYC transcriptional repressor. Top: Sequence alignment of variants of the c-MYC transcriptional repressor from primates, with G clusters containing two or more guanosines indicated in red. Because multiple overlapping G-quadruplexes can form in this region, 2–11

B

C

Human
Chimp
Gorilla
Orangutan
Rhesus monkey
Baboon
Marmoset
Mouse lemur

GGAGAGGGGGAGGAGGGA
GGAGAGGGGGAGGAGGGA
GGAGAGGGGGAGGAGGGA
GGAGAGGGGGAGGAGGGA
GGAGAGGGGGAGGAGGGA
GGAGAGGGGGAGGAGGGA
GGAGAGGGGGAGGAGGGA
GGAGAGGGGGAGGAGGGA

GGACCGGGGAGGAGGGC
GGACCGGGGAGGAGGGC
GGACCGGGGAGGAGGGC
GGACCGGGGAGGAGGGC
GGACCGGGGAGGAGGGC
GGACCGGGGAGGAGGGC
GGACCGGGGAGGAGGGC
GGACCGGGGAGGAGGGC

GGTTTA
GGTTTA
GGTTTA
GGTTTA
GGTTTA
GGTTTA
GGTTTA
GGTTTA

10000
1000
100
10
1

GTP-binding activity

10000
1000
100
10
1

GTP-binding activity

10000
1000
100
10
1

GTP-binding activity

Human
Orangutan
Baboon
Marmoset
Lemur
Bush baby
0%
8%
9%
11%
15%
25%

0%
Lemur
11%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%

0%
of these three contexts, and enrichment values show the expected decrease with increasing spacer length for G-quadruplexes upstream of transcription start sites (Table S6; see also Huppert and Balasubramanian, 2007). We also note that although the enrichment values of G-quadruplexes in human 5′ UTRs do not appear to increase with decreasing spacer length, an example of reduced spacer length leading to increased translational inhibition has been reported (Halder et al., 2009).

Taken together, these results show that representative members of several different classes of regulatory G-quadruplexes bind GTP, with $K_d$ values ranging between 60 and 600 µM. They also indicate that the sequence requirements of these elements, especially those involved in regulation of transcription, parallel the sequence requirements of the G motif.

Identification of a Cellular Protein that Interacts with G Motif-GTP Complexes

More than 30 cellular proteins that interact with various types of G-quadruplex structures have been identified (Fry, 2007). These include nucleic acid binding proteins that promote the folding of G-quadruplexes, helicases that unwind G-quadruplex structures, and nucleases that specifically cleave phosphodiester bonds in G-quadruplexes (Fry, 2007). The existence of such factors led us to hypothesize that proteins that bind G motif-GTP complexes might also exist. Identification of such proteins would demonstrate that G-motif-GTP complexes can interact with cellular components, and might provide clues about the potential biochemical roles played by these complexes in cells. To search for such proteins, we developed a gel filtration system that can separate biotinylated G-motif-GTP-streptavidin ternary complexes from both biotinylated G motif-GTP binary complexes and free GTP molecules (Figure 7A). By performing this assay using unlabeled aptamer, unlabeled protein, and radiolabeled GTP, one can readily distinguish G motif-GTP-protein complexes from G motif-protein complexes of a similar molecular weight. We screened previously identified G-quadruplex binding proteins using this approach and observed that the zinc-finger protein CNBP (Calcetara et al., 2010) forms a stable ternary complex with a DNA variant of the G motif and GTP (Figure 7B). Control experiments demonstrated that CNBP does not bind GTP by itself (Figure 7B), that G motif-GTP complexes do not nonspecifically interact with BSA (Figure 7B), and that the observed effect does not simply reflect CNBP increasing the amount of GTP bound by the G motif under these conditions (Figure 7C). No radiolabeled product was observed when these complexes were analyzed by either PAGE or SDS-PAGE, suggesting that neither GTP nor its gamma phosphate is covalently linked to either CNBP or the G motif.

CNBP is a 170 amino acid CCHC-type zinc finger protein that is highly conserved in vertebrates (Calcetara et al., 2010). Originally discovered in a screen for proteins that bind the sterol regulatory element (Rajavashisth et al., 1989), CNBP also plays roles in forebrain development and cell proliferation, and has been linked to the human diseases myotonic dystrophy and sporadic inclusion body myositis (Calcetara et al., 2010). Of particular relevance to our studies, CNBP regulates transcription of the c-MYC gene by binding a G-rich region in its promoter, which contains a phylogenetically conserved example of the G motif (Figures 5A, 5B, and 6A; Michelotti et al., 1995). CNBP has also been shown to promote formation of parallel-strand G-quadruplex structures in vitro, suggesting that under certain conditions it might act as a chaperone for the G motif (Borgognone et al., 2010). Our finding that CNBP binds G motif-GTP complexes indicates that of the many cellular proteins that interact with G-quadruplexes, at least one recognizes the same structural confirmation that binds GTP.

SIGNIFICANCE

RNAs play roles in diverse cellular processes, and in an increasing number of known cases the ability to bind small molecules is an important aspect of their function. Inspired by these examples, we used in vitro selection to search a pool of genome-derived RNA fragments for naturally occurring GTP aptamers. The results revealed that DNA and RNA G-quadruplexes possess intrinsic GTP-binding activity. The highest-affinity GTP-binding quadruplexes form structures with parallel strands. One way a G-quadruplex might bind GTP would be to incorporate the nucleotide into one of the tetrads in the structure. This binding mode is consistent with the nucleotide specificity of the G motif, as well as with previous observations that guanosine derivatives can assemble into G-quadruplex structures when incubated at millimolar concentrations (Gellert et al., 1962).

Our study also indicates that GTP aptamers are abundant in eukaryotic genomes. We estimate that ~75,000 of the G-quadruplexes in the human genome bind GTP with dissociation constants comparable to the GTP concentration of a eukaryotic cell, including motifs previously shown to regulate transcription, RNA processing, and translation. This observation raises the possibility that GTP may play a role in the cellular function of these elements. Consistent with
this hypothesis, the sequence requirements of several types of regulatory G-quadruplexes parallel those of GTP binding. While this manuscript was in review, a paper was published describing the discovery of an ATP aptamer in several bacterial and eukaryotic genomes (Vu et al., 2012). This motif was originally isolated from a random sequence pool (Sassanfar and Szostak, 1993) and had previously been identified in viruses (Shu and Guo, 2003) and archaebacteria (Laserson et al., 2005). Another recent study showed that the ydaO riboswitch binds ATP, and modulates gene expression in an ATP-dependent manner (Watson and Fedor, 2012).

EXPERIMENTAL PROCEDURES

Pools were generated by fragmentation of genomic DNA using DNase I followed by gel purification of ~100–600 bp fragments on agarose gels. Fragments with 3’ adenosine overhangs were generated by incubation first with DNA polymerase I and then with dATP and Taq DNA polymerase. These fragments were ligated into pGEM-T vectors and amplified by PCR using primers flanking the insertion site, one of which contained a T7 promoter at its 5’ end. Templates were transcribed using T7 RNA polymerase to generate starting pools for in vitro selection experiments. GTP aptamers were isolated by incubating pool RNA with GTP agarose, washing away unbound molecules with selection buffer, and eluting bound RNAs with EDTA. Eluted molecules were subjected to RT-PCR and transcribed to generate RNA for the next round of selection. After four rounds of selection, the pool was cloned using the TOPO TA kit (Invitrogen) and sequenced.

Bioinformatic analysis of the G motif was performed using the quadparser algorithm. Analyzed data sets were obtained from the NCBI website, the DOE Joint Genome Institute website, or the UCSC Genome Browser. CD experiments were performed using a JASCO J-715 spectropolarimeter.

See Supplemental Experimental Procedures for more details regarding the materials and methods used in this work.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, six tables, Supplemental Sequences, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2013.02.015.
ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute and the NIH/NIGMS (R01GM065865). E.A.C. and D.R.L. designed the research, E.A.C. performed the research and contributed new reagents/analytic tools, and E.A.C. and D.R.L. analyzed the data and wrote the paper.

Received: October 12, 2012
Revised: January 22, 2013
Accepted: February 25, 2013
Published: April 18, 2013

REFERENCES


