Author's Accepted Manuscript

www.elsevier.com/locate/talanta

A Genome-Inspired, Reverse Selection Approach to Aptamer Discovery

Christina M. Albanese, Suttipong Suttapitugsakul, Shruthi Perati, Linda B. McGown

PII: S0039-9140(17)30921-9
DOI: http://dx.doi.org/10.1016/j.talanta.2017.08.093
Reference: TAL17889

To appear in: *Talanta*

Received date: 12 June 2017
Revised date: 8 August 2017
Accepted date: 29 August 2017

Cite this article as: Christina M. Albanese, Suttipong Suttapitugsakul, Shruthi Perati and Linda B. McGown, A Genome-Inspired, Reverse Selection Approach to Aptamer Discovery, *Talanta,* http://dx.doi.org/10.1016/j.talanta.2017.08.093

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

A Genome-Inspired, Reverse Selection Approach to Aptamer Discovery

Christina M. Albanese, Suttipong Suttapitugsakul, Shruthi Perati, Linda B. McGown*

Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, 12180, USA

*Corresponding author: mcgowl@rpi.edu

Abstract

Limitations of Systematic Evolution of Ligands by Exponential Enrichment (SELEX) and related methods that depend upon combinatorial oligonucleotide libraries have hindered progress in this area. Our laboratory has introduced a new approach to aptamer discovery that uses oligonucleotides with sequences drawn from the human genome to capture proteins from biological samples. Specifically, we have focused on capture of proteins in nuclear extracts from human cell lines using G-quadruplex (G4) forming genomic sequences. Previous studies identified capture of several proteins both in vitro and in live cells by the Pu28-mer sequence from the *ERBB2* promoter region. Here we provide a more comprehensive study of protein capture from BT474 and MCF7 human breast cancer cells using G4-forming sequences from the *CMYC*, *RB*, *VEGF* and *ERBB2* human oncogene promoter regions. Mass spectrometric analysis and Western blot analysis of protein capture at oligonucleotide-modified surfaces revealed capture of nucleolin by all three of the oligonucleotides in BT474 and MCF7 cells, and also of ribosomal protein L19 (RPL19) in BT474 cells. Chromatin immunoprecipitation (ChIP) analysis confirmed the interaction of nucleolin with all three promoter sequences in MCF7 cells and with *RB* in BT474 cells. ChIP also revealed interactions of RPL19 with *CMYC* in BT474 cells and of both RPL19 and ribosomal protein L14 (RPL14) with *ERBB2* in BT474 cells. These results

Accepted manuscript

ACCEPTED MANUSCRIPT

offer the basis for development of new aptamers based on the G4 sequences from the *CMYC*, *RB*, *VEGF*, and *ERBB2* promoters toward proteins including nucleolin, RPL19 and RPL14. These interactions also may have biological and therapeutic significance.

Keywords: Aptamer, genomic DNA, G-quadruplex, cancer, oncogene promoter

Introduction

Since their introduction in 1990 [1,2], aptamers have become an increasingly mainstream alternative to antibodies as affinity reagents for targets ranging from small molecules to large proteins and even whole cells. They offer important, well described advantages over antibodies that make them attractive candidates for wide-ranging analytical and medical applications [3-7]. Typically, the process of aptamer discovery to a particular target begins with construction of a randomly generated library of 10^{14}-10^{15} oligonucleotides. The library is then reduced through an iterative selection process to a handful of sequences that exhibit high affinity to the target through a process known as Systematic Evolution of Ligands by Exponential Enrichment (SELEX) [1].

There is no question that aptamers, once discovered, can live up to their promise as powerful alternatives to antibodies for highly selective, affinity binding reagents. The limiting factor in achieving their full potential, however, is the relatively small number of targets to which aptamers have been selected despite intensive effort and substantial investments of time and money [8]. This limited success has been attributed to a number of factors. Most often cited is the SELEX process itself, which is laborious, non-standardized and difficult to fully automate [8]. Yet even if these problems are solved, there will remain fundamental drawbacks inherent in combinatorial approaches related to sequence bias of combinatorial libraries and PCR

Accepted manuscript

ACCEPTED MANUSCRIPT

amplification, leading to under-representation of certain structural motifs in the selection process [9,10].

Here we describe progress toward a new, genome-inspired selection process for aptamer discovery that reverses the selection process in order to overcome practical and fundamental limitations of combinatorial selection approaches. Rather than selecting an aptamer from a combinatorial library of oligonucleotides for high affinity binding to a specific target protein, the genome-inspired approach uses specific DNA sequences from the human genome to capture proteins from natural pools such as nuclear protein extracts. By using naturally occurring DNA sequences to capture target proteins from natural protein pools, the new approach takes advantage of eons of biological evolution of DNA sequences that interact selectively with particular proteins to perform biological functions. Linking aptamer discovery to nature increases the chances of uncovering protein-DNA affinity binding interactions that have biological significance as well as analytical utility.

The specific focus of the present work is on genomic G-rich sequences related to cancer that can form G-quadruplex (G4) structures. Although G4 is a highly successful motif for aptamers, it is underrepresented in combinatorial libraries; those that are present are generally limited to two-tier G4 structures [9,10]. Additionally, G4-forming sequences are prone to inefficient PCR amplification, further biasing SELEX away from these structures. In contrast to combinatorial libraries, nature provides a large diversity of sequences capable of forming multi-tiered G4 structures throughout the human genome that are promising candidates for aptamers [11-17]. In particular, possible G4-forming sequences are prevalent in gene promoter regions [15,16], particularly in oncogene promoters [16]. Since gene promoters are characterized by protein binding interactions, these regions represent a rich source of potential aptamers.

Accepted manuscript

ACCEPTED MANUSCRIPT

Our earlier discovery of high affinity binding between a G4-forming sequence from the insulin-linked polymorphic region of the insulin gene promoter and the proteins insulin and insulin-like growth factor 2 (IGF2) [18] resulted in identification of aptamers to these proteins [19,20]. This discovery led us to pursue a more general, reverse selection, genome-inspired pathway to aptamer discovery [19,21], in which a genomic sequence is used to capture potential targets from natural protein pools such as cell lysates or nuclear extracts, and to apply it to G4-forming sequences from human oncogene promoter regions [21]. Studies of binding between individual, G4-forming sequences from oncogene promoter regions and the oncogene protein products with the goal of aptamer discovery have since been reported [22].

Libraries of genomic sequences previously have been used in place of randomly generated oligonucleotide libraries in SELEX [11,12]. In contrast, the present approach eliminates the need not only for oligonucleotide libraries, but also for PCR amplification, and purified, immobilized target protein. Additionally, unlike the SELEX process, which doesn't reveal success or failure until the conclusion of a generally lengthy process, the genome-inspired approach provides rapid feedback about potential affinity binding interactions. Finally, by reversing the selection process, the present approach offers the ability to discover previously unknown DNA-protein binding interactions.

The G4-forming sequences used in the present work are from the promoter regions of the *CMYC*, retinoblastoma (*RB)*, vascular endothelial growth factor (*VEGF*) and ERBB2 Receptor Tyrosine Kinase 2 *(ERBB2)* human oncogenes. These genes are either overexpressed in cancer or express proteins that exhibit a loss of activity in cancer [23-29]. The sequences are shown in Figure 1, along with the proposed structures of *CMYC*, *RB* and *VEGF* [30-39]. The structure of *ERRB2* is not yet available, but is thought to form tetrad:heptad:heptad:tetrad (T:H:H:T) similar

Accepted manuscript

ACCEPTED MANUSCRIPT

to that reported for an analogous sequence from CMYB [38]. The oligonucleotides were attached to magnetic beads and incubated with nuclear protein extracts from cultured BT474 and MCF7 breast cancer cells. The captured proteins were analyzed using MALDI MS to screen for differential capture by the G4 oligonucleotides compared to a non-G4-forming control, and then separated using gel electrophoresis. Bands appearing only, or to a much greater extent, for the G4 oligonucleotides compared to a non-G4-forming control were extracted from the gels and analyzed using LC-MS/MS. Protein candidates were then selected for Western blot interrogation. Proteins confirmed by Western blot were then used in chromatin precipitation (ChIP) analysis to determine if the proteins bind to the G4-forming sequences in the live cells.

Accepted manuscript

Materials and methods

Materials

Custom sequence oligonucleotides were obtained from Integrated DNA Technologies (IDT) with a 5'-thiol modification for attachment to fused silica surfaces or a 5'-biotin modification for attachment to streptavidinated magnetic beads. Sequences of the G4 oligonucleotides are shown in Figure 1. The control oligonucleotide is G-rich sequence 5'-GGT GGT GGT TGT GGT-3' that does not form an intramolecular G4 structure. Reagent-grade solid chemicals were obtained from Sigma unless otherwise noted. BT474 and MCF7 adherent breast cancer cell lines were obtained from the American Type Culture Collection (ATCC) as vials of frozen cells.

Circular Dichroism Spectroscopy

Prior to analysis, stock DNA oligonucleotide solutions were heated at 95°C for 5 min and cooled slowly in the presence of potassium ions. All oligonucleotides were diluted in 100 mM

ACCEPTED MANUSCRIPT

KCl to a final concentration of 1.25 μM unless otherwise noted. Samples were measured in a 1-cm pathlength quartz cuvette on a Jasco 715 spectropolarimeter equipped with a Peltier temperature controller. A blank spectrum was taken for the 100 mM KCl solution and subtracted from all sample spectra. CD scans were run in triplicate at 25°C over a range of 200-350 nm and reported as the average. Instrument parameters were set to a 1 nm step size, 4 s response time, and 100 nm/min scan speed.

Cell Culture and Nuclear Extract Collection

BT474 and MCF7 adherent breast cancer cell lines were obtained from the American Type Culture Collection (ATCC) as vials of frozen cells. BT474 cells were grown in ATCC's Hybri-Care medium dissolved in cell culture grade water containing 1.5 g/L sodium bicarbonate (Sigma), 10% fetal bovine serum (ATCC), and 5% penicillin/streptomycin (Fisher). MCF7 cells were grown in ATCC-formulated Eagle's Minimum Essential Medium (EMEM) containing 0.01 mg/mL human recombinant insulin (Life Technologies), 10% fetal bovine serum, and 5% penicillin/streptomycin. Both cell types were grown in 75-cm^2 vented cap tissue culture treated flasks (CELLTREAT) and maintained at 37°C with 5% CO_2.

After reaching 70-80% confluence, cells were rinsed with ice-cold Dulbecco's phosphate buffered saline (ATCC), lifted from the surface using cell scrapers with a 20-mm lifter blade (CELLTREAT), and collected in microcentrifuge tubes (Eppendorf). All succeeding protein extraction steps were performed in a cold room at 4°C, with all materials and equipment having been precooled at least overnight. Each cell pellet was centrifuged at 3000 RPM for 2 min and the supernatant discarded. Hypotonic solution A (10 mM HEPES (pH 7.9), 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM EGTA, 1 mM EDTA, 0.1% NP-40, 1 mM DTT, 1% protease inhibitor cocktail)

Accepted manuscript

ACCEPTED MANUSCRIPT

was added to each pellet at a volume equal to the packed cell volume (1:1 ratio). Samples were centrifuged at 3000 RPM for 2 min and the supernatant was removed. Solution A was added to each pellet a second time at a 1.5:1 ratio. After a brief vortex mixing, the cell suspensions were swelled on ice for 10 min and centrifuged at 6000 RPM for 30 s. The supernatant, containing the cytoplasmic extract, was collected and stored at -20°C. Solution B (20 mM HEPES (pH 7.9), 1.5 mM $MgCl_2$, 20 mM KCl, 1 mM EGTA, 1 mM EDTA, 0.1% NP-40, 1 mM DTT, 1% protease inhibitor cocktail) was added to each remaining cell pellet at a 1:1 ratio and gently mixed. Samples were centrifuged at 12000 RPM for 10 min and the supernatant was discarded. The addition of solution B, centrifugation, and removal of supernatant was repeated a second time. Solution B was then added to each pellet a third time using a 0.5:1 ratio. Samples were transferred to a Mikura Ltd. Orbis microplate shaker and an equal volume of solution C (50 mM HEPES (pH 7.9), 1.5 mM $MgCl_2$, 1 mM EGTA, 1 mM EDTA, 0.1% NP-40, 1 mM DTT, 1% protease inhibitor cocktail) was added dropwise, slowly, while shaking. Following 30 min of mixing, samples were centrifuged at 12000 RPM for 30 min. The supernatant, containing the nuclear extract, was then collected and stored at -80°C until use.

Affinity Protein Capture on Streptavidinated Magnetic Beads

Dynabeads MyOne C1 streptavidinated magnetic beads (Life Technologies) were vortexed in their original container for approximately 30 s until homogeneously resuspended in solution. Following transfer of an appropriate volume to a microcentrifuge tube, beads were washed three times with 1X binding and washing (B&W) buffer (5 mM Tris-HCl, pH 7.5; 0.5 mM EDTA; 1 M KCl) to remove preservatives. Beads were then resuspended in 2X B&W buffer (10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 2 M KCl) at a volume equal to twice the initial

Accepted manuscript

ACCEPTED MANUSCRIPT

volume of beads used. An equal volume of 1X B&W buffer containing 200 pmol of the appropriate DNA oligonucleotide was then added to each tube and incubated for 10 min at room temperature with shaking. Custom sequence oligonucleotides were obtained from Integrated DNA Technologies with a 5'-biotin attachment. After oligonucleotide attachment, the beads were washed three times with 1X B&W buffer to remove unbound DNA.

For protein capture, the appropriate nuclear protein extract was added to each tube and incubated overnight at ambient temperature with shaking. The following morning, beads were washed three times with PBS (150 mM KCl, 5 mM KH_2PO_4, 8 mM K_2HPO_4, pH 7.3). Captured proteins were eluted into either deionized water at 60°C for 15 min or 5% formic acid at room temperature with shaking for 10 min. Beads were then separated from solution using a bar magnet. The supernatant, containing the captured proteins released from the beads, was collected. Each magnetic bead capture was performed in triplicate, and the three eluted protein mixtures were pooled together. These pooled samples were dried and concentrated using a Labconco CentriVap Complete Vacuum Concentrator at the Center for Functional Genomics at the State University of New York at Albany. Samples were then reconstituted in deionized water and aliquoted. Aliquots of captured protein were either used immediately or stored at -20°C for later use.

Accepted manuscript

MALDI MS Analysis

Captured proteins from the magnetic beads experiments were mixed 1:1 with sinapinic acid MALDI matrix (10 mg sinapinic acid/mL 50% ACN + 0.3% TFA) and spotted directly onto the MALDI target. Bruker protein calibration standard II was used to externally calibrate the mass spectrometer. All MALDI-TOF MS data was collected in linear positive mode using a

ACCEPTED MANUSCRIPT

Bruker Autoflex Speed MALDI MS at the Stewart's Advanced Instrumentation and Technology (SAInT) Center at Siena College. Laser frequency was set to 1000 Hz with a pulsed ion extraction of 6000 ns and suppression up to 4000 Da. Ten spectra were summed together for each measurement, for a total of 20,000 laser shots.

SDS-PAGE Separation

Captured proteins were mixed with 5X Laemmli buffer (10% SDS, 0.25 M DTT, 50% glycerol, 0.01% bromophenol blue in 0.3125 M Tris) in a 1:5 ratio and denatured at 95°C for 5 min. Following protein denaturation, 5 µL of sample was loaded into each well of a Bio-Rad precast 4-15% TGX polyacrylamide protein gel. SDS-PAGE separation was performed at 120 V, 30 mA for 45 min using a Bio-Rad Mini-PROTEAN Tetra Cell electrophoresis system containing 1X SDS-PAGE running buffer (25 mM Tris, pH 8.3; 190 mM glycine; 0.1% sodium dodecyl sulfate). A pre-stained protein ladder was also loaded on the gel to monitor electrophoretic separation and serve as a marker for approximate protein masses. Gels were rinsed three times with deionized water for 5 min each and stained with Bio-Safe G-250 Coomassie stain (Bio-Rad) for 1 h with shaking. De-staining was carried out in deionized water for at least 30 min, with shaking, until protein bands were visible and the blue background was minimized. Gels images were obtained using a Bio-Rad ChemiDoc XRS+ imaging system.

Trypsin Digest and LC-MS/MS Analysis

Gel bands of interest were excised using sterile surgical blades and transferred to individual microcentrifuge tubes. In-gel trypsin digestion and LC-MS/MS analysis were then performed at the Center for Functional Genomics at the State University of New York at Albany.

Accepted manuscript

ACCEPTED MANUSCRIPT

Gel bands were first washed with deionized water, smashed into smaller pieces using a homogenizer, and de-stained with 100 mM ammonium bicarbonate. The samples were centrifuged to collect the gel pieces at the bottom of the tube. The supernatant was removed and the gel pieces were shrunk with acetonitrile. Following centrifugation and discarding of the supernatant, the gel pieces were incubated with 20 mM tris(2-carboxyethyl)phosphine (TCEP) at 37°C for 1 h to reduce disulfide bonds. The pieces were dehydrated once again with acetonitrile and then alkylated with 40 mM iminodiacetic acid (IDA) for 45 min at 37°C to prevent reforming of disulfide bonds. Gel pieces were then dehydrated, dried on a speed vac, and incubated with trypsin at 37°C overnight. The next day, peptides were extracted from the gel pieces using acetonitrile, trifluoroacetic acid, and formic acid. LC-MS/MS data was obtained using a CapLC HPLC (Waters Co.) coupled with an ABSCIEX Applied Biosystems QSTAR XL mass spectrometer. The HPLC was equipped with a Jupiter C_{18} column (3 μm, 100 μm ID x 150 mm) from Phenomenex (Torrance, CA). A linear gradient flow was used at a flow rate of 250 nL/min. MASCOT 2.5 software from Matrix Science was used to search the NCBInr human sub database with the following paramters: a peptide mass tolerance of 0.3 Da, up to one missed cleavage, and possible variable modifications of methionine oxidation, cysteine carboxyamidomethylation, and deamination. Potential protein matches were selected for further analysis based on a score above the threshold, approximate mass consistent with that seen for the corresponding gel band, known G4 interactions, and potential role in cancer.

Western Blotting

Captured protein samples were first separated on a precast 4-15% gradient SDS-PAGE gel, as described previously, along with MagicMark XP Western protein standard (Thermo

Accepted manuscript

ACCEPTED MANUSCRIPT

Fisher Scientific). After three deionized water rinses of the gel, the separated proteins were transferred from the gel to a 0.2-μM pore polyvinylidene difluoride (PVDF) membrane that had been pre-wetted with methanol. Following a 15-min equilibration of the gel, membrane, filter paper, and foam pads in ice-cold transfer buffer (25 mM Tris, 190 mM glycine, 20% methanol); the transfer stack was assembled, placed into the electrode assembly in a Bio-Rad Mini Trans-Blot cell, and covered with ice-cold transfer buffer. Transfer was allowed to proceed at 70 V, 350 mA for 70 min with an ice pack and constant stirring of a magnetic stir bar to aid in heat dissipation. Successful transfer to the membrane was monitored using a pre-stained protein ladder.

Following protein transfer, the membrane was incubated for 1½ h with blocking buffer (5% Carnation dry milk in PBS containing 0.2% Tween-20 (PBS-T)) with shaking at room temperature. After washing the membrane three times with transfer buffer for 10 min each with shaking, the appropriate primary antibody, diluted according to manufacturer specifications in blocking buffer, was added to the membrane and incubated overnight at 4°C with shaking. Rabbit anti-nucleolin was purchased from Thermo Fisher and rabbit anti-RPL14 and mouse anti-RPL19 were from Novus Biologicals.

The following morning, the membrane was washed with blocking buffer three times for 10 min each with shaking. The membrane was then incubated at room temperature for 1 h with shaking with an appropriate secondary antibody, diluted in PBS-T according to manufacturer guidelines. The secondary antibodies used were goat anti-rabbit IgG (H+L) HRP-conjugated antibody (Life Technologies) and donkey anti-mouse IgG HRP-conjugated antibody (R&D Systems). After secondary antibody incubation, the membrane was rinsed three times with PBS-T for 10 min each on the shaker. SuperSignal West Femto Maximum Sensitivity Substrate (Life

Accepted manuscript

ACCEPTED MANUSCRIPT

Technologies), comprising equal parts peroxide solution and luminol enhancer, was added to the membrane and incubated for 5 min with shaking. After draining excess liquid, the membrane was imaged on the ChemiDoc XRS+ imaging system according to the settings for chemiluminescent detection.

Chromatin Immunoprecipitation and PCR

A ChIP-IT Express kit (Active Motif) was used to collect chromatin from each cell line and to perform chromatin immunoprecipitation (ChIP) experiments. Confluent cells were fixed with 1% formaldehyde, scraped into ice-cold PBS supplemented with 0.005% 100 mM phenylmethylsulfonyl fluoride (PMSF), and collected. Fixed cells were lysed with a dounce homogenizer and the released chromatin was sheared using a Vibra-Cell VCX500 ultrasonic processor (Sonics). Ten 1-min pulses were performed for each sample with 10 s of sonication at 30% power and 50 s of rest, with samples being kept on ice for the duration of the 10-min process. Following sonication, samples were centrifuged at 13000 RPM for 10 min at 4°C and the supernatant, containing sheared chromatin, was collected.

The DNA component of chromatin was isolated through DNA-protein cross-link reversal, RNA and protein degradation, and ammonium acetate-ethanol precipitation. Following removal of residual salt, the concentration of DNA was measured using a Nanodrop ND-1000 spectrophotometer. Approximate concentrations were on an average scale of 1200 ng/μL for BT474 and 400 ng/μL for MCF7. These values were used to calculate the volume of sheared chromatin necessary for ChIP experiments. ChIP reactions were set up in siliconized microcentrifuge tubes according to the ChIP-IT Express Kit protocol. Rabbit anti-nucleolin verified for use in ChIP was purchased from Novus Biologicals. ChIP-verified antibodies were

Accepted manuscript

ACCEPTED MANUSCRIPT

not commercially available for RPL14 and RPL19, so the antibodies from Western blotting were used. Reaction tubes were incubated at 4°C for 4 h on an end-to-end rotator and centrifuged briefly to collect liquid from the caps. After using a bar magnet to separate the beads from the liquid, the supernatant was discarded. The beads were then washed once with ChIP Buffer 1 and twice with ChIP Buffer 2. Beads were resuspended in Elution Buffer AM2, incubated for 15 min at room temperature with shaking, and briefly centrifuged. DNA-protein crosslinks were then reversed with the addition of Reverse Cross-linking Buffer and heating of the collected chromatin at 95°C for 15 min. Proteinase K (2 µL of 0.5 µg/µL) was added to each sample and mixed well followed by a 30-min incubation at 37°C. Samples were returned to room temperature and 2 µL of Proteinase K Stop Solution was added to each. The collected DNA was then purified using Active Motif's Chromatin Immunoprecipitation DNA Purification Kit. DNA was either used immediately or stored at -20°C until further use.

PCR primers were designed based on the human genomic sequence of each promoter and tested for primer efficiency using Oligo 6 software (Molecular Biology Insights, Inc., DBA Oligo, Inc., Colorado Springs, CO). Sequences of the anticipated PCR product for each promoter sequence including forward and reverse primers are shown in Supplementary Material Figure S1. The forward and reverse primers were custom ordered from IDT. AccuPrime GC-rich DNA polymerase was purchased from Life Technologies. PCR reactions were set up in 0.2-mL thin-wall PCR tubes and mixed thoroughly. Filtered pipette tips from Fisher were used for all mixture components. PCR was performed using a Bio-Rad T100 Thermal Cycler according to the following protocol: 2-min denaturing step at 94°C; 48 cycles of 94°C for 20 s, 66°C for 20 s, 68°C for 1 min; 2-min final extension at 68°C; hold step at 4°C. PCR products were purified using Active Motif's Chromatin IP DNA Purification Kit.

Accepted manuscript

ACCEPTED MANUSCRIPT

PCR products were mixed with 6X DNA loading buffer (30% glycerol (v/v), 0.25% bromophenol blue (w/v) in deionized water) and separated on a 2% agarose gel at 120V, 100 mA for 120 min using an Owl B2 horizontal electrophoresis system (Thermo Fisher). Gels were immediately stained with 1X SYBR gold in 1X TBE for 15 min with shaking. Gels were imaged on a Bio-Rad Chemi-Doc XRS+ imaging system using the appropriate settings for visualizing SYBR gold. After confirming the presence of a single band for each sample, the PCR products and appropriate primers were then sent overnight to Eurofins MWG Operon LLC (Louisville, KY) for DNA sequencing.

Results

Circular Dichroism Spectra of the G4 Oligonucleotides

The CD spectra of the G4-forming oligonucleotides and the non-G4-forming control are shown in Figure 2. The spectra of *CMYC* and *VEGF* indicate parallel, intramolecular G4 structures, which is consistent with previous reports for these oligonucleotides [30-33,36,37]. The spectrum of *RB*, which has previously been reported to have an antiparallel G4 structure [34,35], also indicates a parallel structure. This inconsistency may be explained by the presence of potassium ion, which can promote transitions between antiparallel and parallel structures [40]. The CD spectrum of *ERRB2* (not shown) was consistent with reports for the homologous sequence from *CMYB* that is thought to form a parallel, tetrad:heptad:heptad:tetrad (T:H:H:T) G4 structure [38].

MALDI MS Analysis of Proteins Captured on Oligonucleotide-Modified Beads

Protein capture from nuclear protein extracts of BT474 cultured cells at the surfaces of the oligonucleotide-modified beads was performed, followed by elution of the captured proteins

Accepted manuscript

ACCEPTED MANUSCRIPT

for analysis by MALDI MS. The use of MALDI MS provides rapid screening of protein capture that allows us to determine not only that capture has occurred, but also about the diversity of the captured proteins and the differences between capture at the various G4 and control surfaces. The peaks may correspond to individual proteins or protein fragments, as well as to different charge states of the same protein or fragment, and therefore was used only to compare capture profiles rather than to determine the number and identities of captured proteins. Furthermore, detectability in MALDI MS decreases with increasing m/z, biasing the spectra toward singly charged, lower mass proteins and multiply charged, higher mass proteins.

The MALDI mass spectral results are shown in Supplementary Materials Figures S2a and S2b. Comparison of the total nuclear extract with proteins captured by each of the oligonucleotides (Supplementary Materials Figures S2a) shows the ability to selectively capture high affinity binders in the presence of higher abundance, non-binding proteins. More detailed comparison of protein capture by the non-G4 control oligonucleotide and each of the G4-forming oligonucleotides (Supplementary Materials Figures S2b) shows that the G4 oligonucleotides capture proteins that are not captured by the control, and that the capture profiles vary among the different G4 oligonucleotides, indicating significant differences in selectivity.

The protein capture results are summarized in Table 1, which shows the approximate m/z values of peaks that were detected for the G4-forming oligonucleotides but not detected, or detected to a much lesser extent, for the control oligonucleotide. *VEGF* exhibited the greatest number of unique peaks, and *CMYC* the least. Each oligonucleotide has at least two unique peaks not observed for the others. None of the peaks were observed for all three oligonucleotides.

SDS-PAGE Separation and LC-MS/MS Analysis

Accepted manuscript

ACCEPTED MANUSCRIPT

SDS-PAGE was performed on proteins captured from BT474 and MCF7 cell nuclear extracts by the G4 oligonucleotides and the non-G4 control. Results for BT474 are shown in Supplementary Figure S3. Bands that appeared only for the G4 oligonucleotide or that had a much greater intensity relative to the control oligonucleotide were selected for analysis by LC-MS/MS. These included bands at approximately 110 kDa and 25 kDa, which are indicated in Supplementary Figure S3. Similar results were obtained for MCF7 (not shown).

Following in-gel trypsin digestion of the protein bands of interest, LC-MS/MS analysis with MASCOT scoring yielded lists of potential protein candidates. These lists were scrutinized to select protein candidates for further experimentation based on several criteria. First, proteins below the MASCOT score threshold of 40 were eliminated. Protein mass was then considered to select matches best corresponding with the sample protein bands extracted from the gels. These proteins were then researched for known interactions with G4 structures and/or a significant role in cancer. Finally, we searched for commercially available antibodies to the selected proteins, as these would be necessary for the subsequent steps of Western blotting and ChIP. After taking all of these factors into account, we chose three promising protein candidates for further analysis: nucleolin (NCL), which had MASCOT scores > 100 and has a predicted mass of 77 kDa and apparent mass of 100-110kDa [41]; ribosomal protein L19 (RPL19), which had a score of 63 and has a predicted mass of 23.4 kDa; ribosomal protein L14 (RPL14) which had a score of 57 and has a predicted mass of 23.8 kDa.

Western Blot Interrogation of Potential Protein Candidates

Western blotting with primary antibodies specific for the protein candidates selected from the LC-MS/MS results was used to determine if these proteins could be detected among the proteins captured from the nuclear extracts by the G4 oligonucleotides. The results, shown in

Accepted manuscript

ACCEPTED MANUSCRIPT

Supplementary Figure S4 and summarized in Table 2, For all three oligonucleotides, NCL and RPL19 were detected in the proteins captured from BT474 cells but only NCL was detected in the proteins captured from MCF7 cells. RPL-14 capture was not detected for any of the oligonucleotides in either cell line (not shown).

Chromatin Immunoprecipitation (ChIP) and DNA Sequencing

Following Western blotting, ChIP experiments were performed to determine whether the proteins identified as binders to the G4-forming oligonucleotides *in vitro* bind to the same G-rich sequences in the chromatin of live BT474 cells. The sequencing results for the ChIP experiments are shown in Supplementary Figure S5 and the ChIP results are summarized in Table 3. Results for BT474 cells indicate binding of *CMYC* to RPL-19, of *RB* to NCL, and of *ERBB2* to all three proteins; binding was not detected for *VEGF* to any of the proteins. For MCF7 cells, binding was detected for NCL with all four sequences but no binding was observed for RPL19 or RPL14 to any of the sequences. The two cell lines used in these studies differ in their expression of the protein ERBB2 – BT474 is known to overexpress ERBB2 while MCF7 cells do not [42-44]. It is also know that high RPL19 expression levels are coincident with ERBB2 overexpression [45]. Our results confirming capture of RPL19 from BT474 cells but not from MCF7 therefore, are consistent with reported gene expression levels.

Accepted manuscript

Discussion

These studies demonstrate the potential of our genome-inspired, reverse-selection method for aptamer discovery. The ability to selectively capture a few protein candidates from the complex protein mixtures in cell nuclear extracts is illustrated by the MALDI MS results and by gel electrophoresis. LC-MS/MS analysis of specific bands in the gels, followed by Western blots,

ACCEPTED MANUSCRIPT

identified NCL, RPL19, and RPL14 as protein targets for potential aptamers based on G4-forming sequences from *CMYC*, *RB*, *VEGF*, and *ERBB2* gene promoter regions.

Previous work on *ERBB2* indicated *in vitro* capture of NCL, Ku70, Ku80, PURA, and hnRNPK from BT474 nuclear extracts [21]. Of those proteins, only NCL was also detected in the present studies of *in vitro* capture from BT474 nuclear extracts, while capture of RPL19 and RPL14 was detected in the present work but not in the previous studies. Possible explanations are differences in the cycles of the cultured cells, or indeterminate differences in growth conditions or experimental conditions in the capture experiments.

The next step toward aptamer discovery will be refinement of the sequences to tune their selectivity and affinity. All of the sequences are candidates for development of aptamers to NCL and RPL19, while *ERBB2* appears to be the best starting sequence for an aptamer to RPL14. Association of NCL with multiple oligonucleotides including those in the present work is not surprising, given that NCL is an RNA binding protein that associates with a number of RNA and DNA sequences, including RNA G4 oligonucleotides [46]. Deselection toward NCL may therefore be necessary for development of G4-forming aptamers to other proteins.

In addition to providing the basis for development of aptamers to NCL, RPL19, and RPL14, this work suggests the participation of these proteins in regulation of gene expression in breast cancer through their association with G4-forming sequences in gene promoter regions. The ChIP experiments indicate such interactions in live breast cancer cells between NCL and *CMYC*, *RB*, *VEGF*, and *ERBB2*, between RPL19 and *CMYC* and *ERBB2*, and between RPL14 and *ERBB2*. While NCL has previously been reported to bind to the *CMYC* promoter in HeLa cervical cancer cells [47] and to the *VEGF* promoter in A498 kidney carcinoma cells [37], there do not appear to be previous reports of this interaction in breast cancer cells. Interaction of NCL

Accepted manuscript

ACCEPTED MANUSCRIPT

with the *RB* promoter region has not to our knowledge been previously reported. Interactions of RPL19 and RPL14 with gene promoters do not yet appear to have been studied, although RPL19 has been found to interact with certain RNA G-quadruplexes in HEK embryonic kidney cells and HeLa cells [46]. Further studies of the protein interactions with G4-forming regions of oncogene promoters reported here could lead to new insights into the role of these interactions in breast cancer and other diseases, and more generally of genomic G4 structures in gene regulation. Thus, the genomic, reverse selection approach offers the dual advantages of aptamer discovery and fundamental, biological discovery.

Acknowledgements

We are grateful to Dr. Qishan Lin and Jinghua Zhu at the Center for Functional Genomics at the State University of New York at Albany for trypsin digestion and LC-MS/MS analysis and Dr. Kristopher Kolonko at the Stewart's Advanced Instrumentation and Technology (SAInT) Center at Siena College for his guidance in MALDI MS analysis.

Funding

This work was supported by National Institutes of Health Grant 1R01GM112850.

Accepted manuscript

References

[1] C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase., Science. 249 (1990) 505–510. doi:10.1126/science.2200121.

[2] A.D. Ellington, J.W. Szostak, In vitro selection of RNA molecules that bind specific ligands, Nature. 346 (1990) 818–22. doi:10.1038/346818a0.

[3] L.B. McGown, M.J. Joseph, J.B. Pitner, G.P. Vonk, C.P. Linn, The nucleic acid ligand: a new tool for molecular recognition, Anal. Chem. 67 (1995) 663A–668A.

ACCEPTED MANUSCRIPT

doi:10.1021/ac00117a726.

[4] S.L. Clark, V.T. Remcho, Aptamers as analytical reagents, Electrophoresis. 23 (2002) 1335–1340. doi:10.1002/1522-2683(200205)23:9<1335::AID-ELPS1335>3.0.CO;2-E.

[5] S. Tombelli, M. Minunni, M. Mascini, Analytical applications of aptamers, Biosens. Bioelectron. 20 (2004) 2424–2434. doi:10.1016/j.bios.2004.11.006.

[6] S.D. Jayasena, Aptamers: an emerging class of molecules that rival antibodies in diagnostics, Clin. Chem. 45 (1999) 1628–1650.

[7] R. Mukhopadhyay, Aptamers are ready for the spotlight, Anal. Chem. 77 (2005) 115A–118A.

[8] G.S. Baird, Where are all the aptamers?, Am. J. Clin. Pathol. 134 (2010) 529–531. doi:10.1309/AJCPFU4CG2WGJJKS.

[9] S.A. McManus, Y. Li, Assessing the amount of quadruplex structures present within G2-tract synthetic random-sequence DNA libraries, PLoS One. 8 (2013) e64131. doi:10.1371/journal.pone.0064131.

[10] J. Gevertz, H.H. Gan, T. Schlick, In vitro RNA random pools are not structurally diverse: a computational analysis, RNA. 11 (2005) 853–863. doi:10.1261/rna.7271405.

[11] B.S. Singer, T. Shtatland, D. Brown, L. Gold, Libraries for genomic SELEX, Nucleic Acids Res. 25 (1997) 781–786. doi:10.1093/nar/25.4.781.

[12] B. Zimmerman, I. Bilusic, C. Lorenz, R. Schroeder, Genomic SELEX: a discovery tool for genomic aptamers, Methods. 52 (2010) 125–132. doi:10.1016/j.ymeth.2010.06.004.

[13] B. Gatto, M. Palumbo, C. Sissi, Nucleic acid aptamers based on the G-quadruplex structure: therapeutic and diagnostic potential, Curr. Med. Chem. 16 (2009) 1248–1265. doi:10.2174/092986709787846640.

[14] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic Acids Res. 33 (2005) 2908–2916. doi:10.1093/nar/gki609.

[15] J.L. Huppert, S. Balasubramanian, G-quadruplexes in promoters throughout the human genome, Nucleic Acids Res. 35 (2007) 406–413. doi:10.1093/nar/gkl1057.

[16] H. Han, L.H. Hurley, G-quadruplex DNA: a potential target for anti-cancer drug design, TiPS. 21 (2000) 136–142.

[17] A.K. Todd, M. Johnston, S. Neidle, Highly prevalent putative quadruplex sequence motifs in human DNA, Nucleic Acids Res. 33 (2005) 2901–2907. doi:10.1093/nar/gki553.

Accepted manuscript

ACCEPTED MANUSCRIPT

[18] A.C. Connor, K.A. Frederick, E.J. Morgan, L.B. McGown, Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide, J. Am. Chem. Soc. 128 (2006) 4986–4991. doi:10.1021/ja056097c.

[19] J. Xiao, J.A. Carter, K.A. Frederick, L.B. McGown, A genome-inspired DNA ligand for the affinity capture of insulin and insulin-like growth factor-2, J. Sep. Sci. 32 (2009) 1654–1664. doi:10.1002/jssc.200900060.

[20] W. Yoshida, E. Mochizuki, M. Takase, H. Hasegawa, Y. Morita, H. Yamazaki, K. Sode, K. Ikebukuro, Selection of DNA aptamers against insulin and construction of an aptameric enzyme subunit for insulin sensing, Biosens. Bioelectron. 24 (2009) 1116–1120. doi:10.1016/j.bios.2008.06.016.

[21] T. Zhang, H. Zhang, Y. Wang, L.B. McGown, Capture and identification of proteins that bind to a GGA-rich sequence from the ERBB2 gene promoter region, Anal. Bioanal. Chem. 404 (2012) 1867–1876. doi:10.1007/s00216-012-6322-y.

[22] W. Yoshida, T. Saito, T. Yokoyama, S. Ferri, K. Ikebukuro, Aptamer selection based on G4-forming promoter region, PLoS One. 8 (2013). doi:10.1371/journal.pone.0065497.

[23] D.J. Slamon, J.B. DeKernion, I.M. Verma, M.J. Cline, Expression of cellular oncogenes in human malignancies, Science. 224 (1984) 256–262. doi:10.1126/science.6538699.

[24] G. Gasparini, Prognostic value of vascular endothelial growth factor in breast cancer, Oncologist. 5 (2000) 37–44. doi:10.1634/theoncologist.5-suppl_1-37.

[25] J. Adams, P.J. Carder, S. Downey, M.A. Forbes, K. Maclennan, V. Allgar, S. Kaufman, S. Hallam, R. Bicknell, J.J. Walker, F. Cairnduff, P.J. Selby, T.J. Perren, M. Lansdown, R.E. Banks, Vascular endothelial growth factor (VEGF) in breast cancer: comparison of plasma, serum, and tissue VEGF and microvessel density and effects of tamoxifen, Cancer Res. 60 (2000) 2898–2905.

[26] Y. Liu, R.M. Tamimi, L.C. Collins, S.J. Schnitt, H.L. Gilmore, J.L. Connolly, G.A. Colditz, The association between vascular endothelial growth factor expression in invasive breast cancer and survival varies with intrinsic subtypes and use of adjuvant systemic therapy: results from the Nurses' Health Study, Breast Cancer Res. Treat. 129 (2011) 175–184. doi:10.1007/s10549-011-1432-3.

[27] E.E. Bosco, Y. Wang, H. Xu, J.T. Zilfou, K.E. Knudsen, B.J. Aronow, S.W. Lowe, E.S. Knudsen, The retinoblastoma tumor suppressor modifies the therapeutic response of breast cancer, J. Clin. Invest. 117 (2007) 218–228. doi:10.1172/JCI28803.218.

[28] J.S. Ross, J.A. Fletcher, HER-2/neu (c-erb-B2) gene and protein in breast cancer, Am. J. Clin. Pathol. 112 (1999) S53–S67.

[29] M.A. Liebert, L. Delacroix, D. Begon, G. Chatel, P. Jackers, R. Winkler, Distal ERBB2

Accepted manuscript

ACCEPTED MANUSCRIPT

promoter fragment displays specific transcriptional and nuclear binding activities in ERBB2 overexpressing breast cancer cells, DNA Cell Biol. 24 (2005) 582–594. doi:10.1089/dna.2005.24.582.

[30] T. Simonsson, P. Pecinka, M. Kubista, DNA tetraplex formation in the control region of c-myc, Nucleic Acids Res. 26 (1998) 1167–1172. doi:10.1093/nar/26.5.1167.

[31] A. Ambrus, D. Chen, J. Dai, R.A. Jones, D. Yang, Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization, Biochemistry. 44 (2005) 2048–2058. doi:10.1021/bi048242p.

[32] D. Yang, L.H. Hurley, Structure of the biologically relevant G-quadruplex in the c-MYC promoter., Nucleos. Nucleot. Nucl. 25 (2006) 951–968. doi:10.1080/15257770600809913.

[33] 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. 99 (2002) 11593–11598. doi:10.1073/pnas.182256799.

[34] A.I.H. Murchie, D.M.J. Lilley, Retinoblastoma susceptibility genes contain 5' sequences with a high propensity to form guanine-tetrad structures, Nucleic Acids Res. 20 (1992) 49–53. doi:10.1093/nar/20.1.49.

[35] Y. Xu, H. Sugiyama, Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb), Nucleic Acids Res. 34 (2006) 949–954. doi:10.1093/nar/gkj485.

[36] D. Sun, K. Guo, J.J. Rusche, L.H. Hurley, Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents, Nucleic Acids Res. 33 (2005) 6070–6080. doi:10.1093/nar/gki917.

[37] D. Sun, K. Guo, Y. Shin, Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene, Nucleic Acids Res. 39 (2011) 1256–1265. doi:10.1093/nar/gkq926.

[38] S.L. Palumbo, R.M. Memmott, D.J. Uribe, Y. Krotova-Khan, L.H. Hurley, S.W. Ebbinghaus, A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity, Nucleic Acids Res. 36 (2008) 1755–1769. doi:10.1093/nar/gkm1069.

[39] J. Choi, T. Majima, Conformational changes of non-B DNA, Chem. Soc. Rev. 40 (2011) 5893–5909. doi:10.1039/C1CS15153C.

[40] C. Rehm, I.T. Holder, A. Groβ, F. Wojciechowski, M. Urban, M. Sinn, J.S. Hartig, A bacterial DNA quadruplex with exceptional K+ selectivity and unique structural

Accepted manuscript

ACCEPTED MANUSCRIPT

polymorphism, Chem. Sci. 5 (2014) 2809–2818. doi:10.1039/C4SC00440J.

[41] M.M. Tajrishi, R. Tuteja, N. Tuteja, Nucleolin: the most abundant multifunctional phosphoprotein of nucleolus, Commun. Integr. Biol. 4 (2011). doi:10.4161/cib.4.3.14884.

[42] American Type Culture Collection (ATCC) www.atcc.org (accessed May 24, 2016).

[43] A.C. Society, How is breast cancer classified? http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-classifying (accessed May 24, 2016).

[44] Z. Inic, M. Zegarac, M. Inic, I. Markovic, Z. Kozomara, I. Djurisic, I. Inic, G. Pupic, S. Jancic, Difference between Luminal A and Luminal B subtypes according to Ki-67, tumor size, and progesterone receptor negativity providing prognostic information, Clin. Med. Insights Oncol. 8 (2014) 107–111. doi: 10.4137/CMO.S18006.

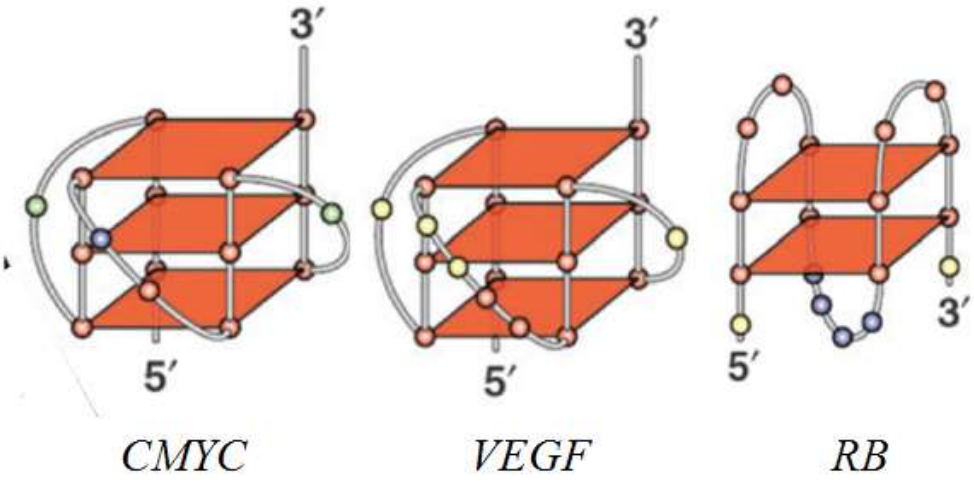
[45] J.L. Henry, D.L. Coggin, C.R. King, High-level expression of the ribosomal protein L19 in human breast tumors that overexpress erbB-2, Cancer Res. 53 (1993) 1403–1408.

[46] A. Hacht, O. Seifert, M. Menger, T. Schütze, A. Arora, P. Neubauer, A. Wagner, C. Weise, J. Kurreck, Identification and characterization of RNA guanine-quadruplex binding proteins, Nucleic Acids Res. 42 (2014) 6630–6644. doi: 10.1093/nar/gku290.

[47] V. González, 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. doi:10.1074/jbc.M109.018028.

Accepted manuscript

ACCEPTED MANUSCRIPT



CMYC	5'-TGG GGA GGG TGG GGA GGG TGG GGA AGG-3'
VEGF	5'-GGG CGG GCC GGG GGC GGG-3'
RB	5'-CGG GGG GTT TTG GGC GGC-3'
ERBB2	5′-AGGAGAAGGAGGAGGTGGAGGAGGAGGG-3′
CMYC	5'-GGG T GGG GA GGG T GGG -3'
VEGF	5'-GGG C GGG CCGG GGG C GGG-3'
RB	5'-G G GG G G TTTT G G GC G G-3'

Figure 1. Proposed structures and sequences of G4-forming oligonucleotides (adapted from [39]). Sequences on bottom show nucleotides that form the guanine tetraplexes grouped together.

ACCEPTED MANUSCRIPT

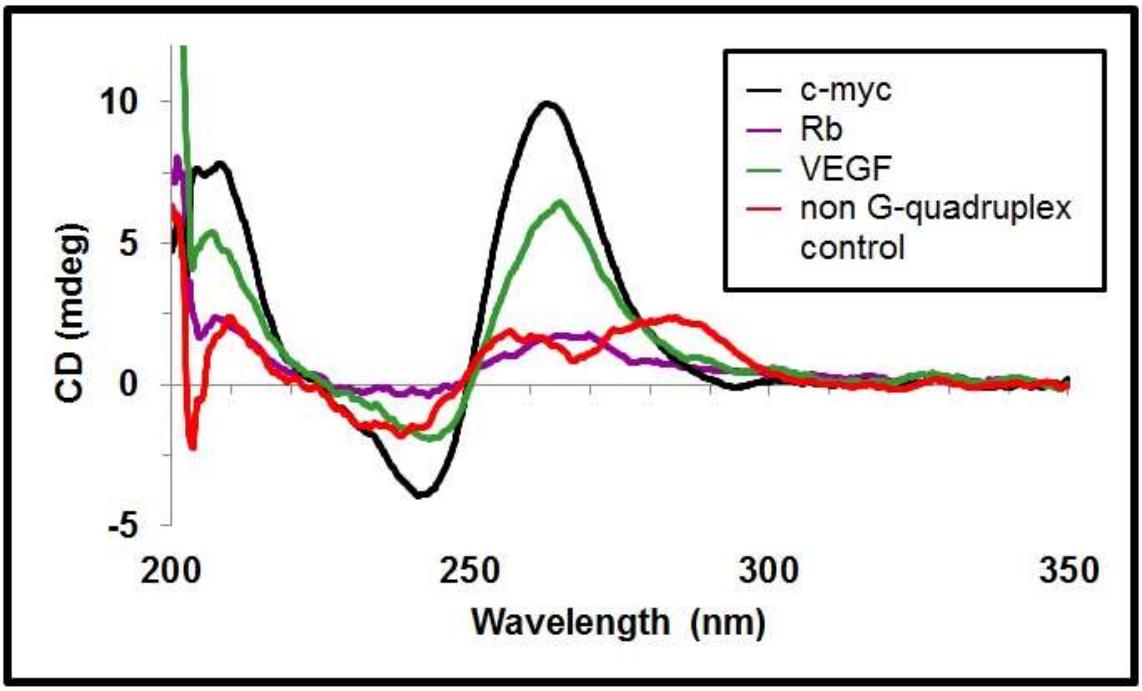


Figure 2. CD spectra of G4-forming oligonucleotides and control oligonucleotide.

Table 1. Summary of MALDI MS peaks of proteins captured from BT474 cultured cells on G4-forming oligonucleotide-modified beads but not on the control oligonucleotide-modified beads.

m(kDa)/z	*CMYC*	*VEGF*	*RB*
8.3	✓		
10.1	✓	✓	
11.6		✓	✓
13.4-13.5	✓		
14.2		✓	✓
14.4			✓
15.2			✓
16.8		✓	✓
19.7		✓	✓
27.2	✓		
36		✓	✓
43.9		✓	
47.8		✓	
51.1		✓	
53.5		✓	

Table 2. Summary of Western blot results. "✓" indicates the protein was detected.

	CMYC		*VEGF*		*RB*		*ERBB2*	
Protein	BT474	MCF7	BT474	MCF7	BT474	MCF7	BT474	MCF7

ACCEPTED MANUSCRIPT

NCL	✓	✓	✓	✓	✓	✓	✓[a]	✓
RPL19	✓		✓		✓		-[b]	-[b]
RPL14							-[b]	-[b]

[a] From reference 21.

[b] Experiment not performed.

Table 3. Summary of ChIP results. "✓" indicates that association of the protein with the region of the gene promoter containing the G4-forming sequence was detected.

	CMYC		***VEGF***		***RB***		***ERBB2***	
Protein	BT474	MCF7	BT474	MCF7	BT474	MCF7	BT474	MCF7
NCL		✓		✓	✓	✓	✓[a]	✓
RPL19	✓						✓	
RPL14							✓	

[a] From reference 21.

Accepted manuscript

Highlights

- A new, genome-inspired, reverse selection approach to aptamer discovery is described.
- G-quadruplex forming sequences from oncogene promoter regions will provide a basis for aptamers to proteins associated with breast cancer.
- In some cases, the proteins were also found to bind to the G4-forming promoter regions in live cells.
- The results have both analytical and biological significance.