BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2009,** *48,* 11487–11495 11487
DOI: 10.1021/bi901420k

Repression of Translation of Human Estrogen Receptor α by G-Quadruplex Formation[†]

Graham D. Balkwill,[‡] Kamila Derecka,[§] Thomas P. Garner,[‡] Charlie Hodgman,[§] Anthony P. F. Flint,[§] and Mark S. Searle*,[‡]

[‡]*Centre for Biomolecular Sciences, School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, U.K., and* [§]*School of Biosciences, University of Nottingham, Sutton Bonington, Leicestershire LE12 5RD, U.K.*

Received August 14, 2009; Revised Manuscript Received October 21, 2009

ABSTRACT: Tissue-specific expression of the human estrogen receptor α gene (*ESR1*) is achieved through multiple promoter sequences resulting in various mRNA transcripts encoding a common protein but differing in their 5′-untranslated region (5′-UTR). Many cancers are estrogen-sensitive with neoplastic growth stimulated through the estrogen receptor, a transcription factor that regulates developmental genes. We demonstrate that the human *ESR1* gene is rich in potential quadruplex-forming sequences with 3 of 20 identified within exonic regions. In particular, we show using CD, UV, and NMR spectroscopy that a stable DNA G-quadruplex motif is formed within the exon C gene sequence. This motif, which PCR shows is transcribed in normal and neoplastic endometrium and in MCF-7 cells, forms a stable RNA quadruplex demonstrable by CD and UV analysis. Cloning the exon C G-quadruplex sequence upstream of a luciferase reporter gene caused a 6-fold reduction of enzymatic activity compared to a mutant sequence. We conclude that the exon C G-quadruplex motif is present in the 5′-UTR of the mRNA transcript, where it modulates the efficiency of translation.

DNA sequences containing runs of guanine bases can spontaneously assemble into four-stranded quadruplex structures stabilized by G-quartets. These structures occur throughout the genome with one motif on average every 10000 bases (*1*, *2*). As many as 376000 potential quadruplexes exist within the human genome (*2*), with telomeres, centromeres, immunoglobulin switch regions, oncogenic promoters, and repeat elements implicated in triplet repeat expansion diseases all found to be G-rich sequences (*3*, *4*). As a consequence, there is considerable interest in the possible existence and biological roles of G-quadruplexes. Putative quadruplex-forming sequences are abundant in human gene promoter sequences with >40% containing one or more quadruplex motif (*2*). This led to the suggestion that promoter quadruplex structures may downregulate gene transcription by acting as molecular switches or as steric blocks, either preventing transcription machinery attaching or by binding a protein that blocks transcription. In addition, where G-quadruplexes are present in transcribed regions of DNA, and appear in the 5′-untranslated region (5′-UTR)[1] of RNA transcripts, they may modulate translation rates (*5*). As a result, quadruplexes formed within oncogenes could represent important drug targets with much current interest in cancer-related genes such as c-myc (*6*), c-kit (*7*), k-ras (*8*), VEGF (*9*), and bcl-2 (*10*).

The steroidal estrogen hormones regulate growth, differentiation, and function in a wide range of target tissues and affect numerous aspects of human physiology as well as being implicated in cancer, cardiovascular disease, Alzheimer's disease, osteoporosis, obesity, arthritis, and social behavior (*11*). The biological effects of estrogens are mediated by their intracellular receptors. These molecules, like other nuclear receptors, are ligand-inducible transcription factors capable of regulating gene expression either by binding to specific estrogen response elements (EREs) located in the promoter regions of target genes or through protein−protein interactions which modulate the function of other transcription factors (*12*). Currently, two isoforms of the estrogen receptor have been described, ERα and ER*β*, each encoded by a separate gene (*ESR1* and *ESR2*) located on different chromosomes. These two receptors, which are comprised of 595 and 530 amino acids, respectively, show sequence homology and share structural and functional domains. They both bind estrogens with high affinity and interact with EREs in a similar manner. However, they exhibit different tissue distributions, ligand affinities, and transcriptional activities. This has led to the general observation that the α isoform is ubiquitously expressed in different tissues, whereas there is a complete lack of *β* isoform expression in several tissues and organs (*13*). The hormone-activated ERs form both homodimers and heterodimers, and both isoforms are coexpressed in many cell types.

The estrogen receptors, along with other nuclear hormone receptors, activate transcription within genes associated with cell proliferation and differentiation, and the level of such activity is dependent on the target tissue. ERα gene expression is upregulated in estrogen receptor-positive breast cancers and promotes the growth of several other types of malignancy (*11*). As a result, the ERα protein has become an important drug target for the development of antagonists or through the inhibition of gene expression.

Tissue-specific expression of the human ERα gene (*ESR1*) is achieved through multiple promoter regions which respond to distinct combinations of factors present in different tissues that regulate alternative splicing and multiple promoter usage. Various mRNA transcripts are formed which encode a common

[†]We thank the EPSRC of the U.K. and the School of Chemistry of the University of Nottingham for funding G.D.B. and T.P.G. and the School of Biosciences for financial support to K.D. and A.P.F.F.

*Corresponding author. E-mail: Mark.Searle@nottingham.ac.uk. Tel: (44) 115 951 3567. Fax: (44) 115 846 6059.

[1]Abbreviations: QS, quadruplex sequence; RQ, RNA quadruplex; UTR, untranslated region; *ESR1*, estrogen receptor α gene; ERα, estrogen receptor α; CD, circular dichroism; NMR, nuclear magnetic resonance.

Published on Web 10/27/2009

protein but differ in their 5′-UTR. Certain pathological conditions employ alternative mRNA transcripts or express levels of these transcripts not seen in the normal physiological state. A possible therapeutic approach would be to alter the efficiency of expression of a particular mRNA transcript by identifying and targeting quadruplex formation either within the gene, within the gene promoter sequence, or within a specific mRNA transcript. We demonstrate here that the human *ESR1* gene is rich in potential DNA quadruplex-forming sequences including a stable RNA G-quadruplex motif within the 5′-UTR of mRNA transcribed from exon C. We have used a range of biophysical tools to characterize this quadruplex motif in both DNA and RNA and have used a reporter assay to show that the RNA quadruplex modulates the efficiency of translation *in vitro*.

MATERIALS AND METHODS

Bioinformatics. The *Homo sapiens ESR1* gene, located on chromosome 6 at 6q25.1, was studied for putative G-quadruplex formation. The nucleotide sequence of this gene encompasses 8 coding exons and 295721 nucleotides (Ensembl Human Gene ID, ENSG00000091831, or NCBI EntrezGene ID, 2099). This investigation additionally required the promoter and 3′-UTR nucleotide sequences (Ensembl genome browser, http://www.ensembl.org). This longer genomic sequence, comprising 514622 nucleotides, is between nucleotide positions 151977826 and 152492447 (Vega Locus ID, OTTHUMG00000016103). The sequence was obtained in the FASTA format prior to further analysis.

The *quadparser* program was used to search the nucleotide sequence for potential quadruplex-forming sequences (PQS) (*2*). The program internally parsed the sequence data searching for at least four G-tracts, comprising runs of three or more guanines separated by loop sequences of one to seven bases in length. The program returned 20 hits that matched the search criteria. An identical search was performed for C-rich sequences within the gene that could correspond to putative quadruplex-forming sequences on the complementary strand. The program returned 19 hits that matched the search criteria. The sequence and location of the 8 coding exons was identified using the data provided in the Ensembl exon report for the *ESR1* gene. The sequence and location of the 9 noncoding 5′-UTR exons were identified using the work of Koš et al. (*14*) and references cited therein. The single nucleotide polymorphisms for the *H. sapiens ESR1* gene were obtained from the Entrez SNP database on the NCBI Web site (http://www.ncbi.nlm.nih.gov/). The 1642 deposited SNPs (as of January 2008) were compared against the PQS for potentially destabilizing mutations.

Oligonucleotide Samples. DNA and RNA oligonucleotides synthesized on a 50 nmol scale were supplied lyophilized (Invitrogen, Paisley, U.K.). Individual stock solutions (30−50 nM) were prepared with Milli-Q water. DNA oligonucleotides for NMR studies were synthesized on a 200 nmol scale, HPLC purified, and supplied lyophilized (Invitrogen, Paisley, U.K.).

CD and UV/Vis Spectroscopy. The CD and UV experiments were recorded on a PiStar-180 spectrophotometer interfaced with an Acorn Archimedes computer with inbuilt software (Applied Photophysics Ltd., Surrey, U.K.). The spectrophotometer's optical system was configured with a 75 W Xe lamp, circular light polarizer, and end-mounted photomultiplier. Temperature was regulated using a RTE-300 circulating programmable water bath (Neslab Inc., Portsmouth, NH) and a thermoelectric temperature controller (Melcor, Trenton, NJ). The machine was calibrated with *d*-camphorsulfonic acid before the start of the project. Typically, samples were prepared in the concentration range 3−6 μM in 100 mM potassium chloride containing 10 mM potassium hydrogen phosphate (pH 7). All samples were then heated to 90 °C and allowed to cool slowly to 10 °C over 4 h. The concentration dependence of the spectra was assessed over a range of 3−30 μM DNA.

UV melting curves were recorded by measuring the absorbance at 295 nm of the sample in a tightly stoppered 1 cm path length cuvette. The preannealed samples were equilibrated at 5 °C for 20 min before being heated to 95 °C and cooled to 5 °C at 0.5 °C/min, to obtain both the melting and annealing curves. Melting and annealing temperatures were obtained from a van't Hoff analysis of the melting curves, assuming a two-state equilibrium. CD spectra were recorded at 5 °C over the range 325−215 nm using a 1 cm path length cuvette. Data points were recorded every 1 nm. The number of counts was set to 100000, adaptive sampling was enabled and set to 500000, and entrance/exit slits of 2 nm were employed. Each trace reported is the average of three to five scans. A sample containing buffer alone was subjected to the same protocol and its CD spectrum recorded using the same cuvette. Following subtraction of the buffer baseline, the CD data were normalized to zero at 325 nm and converted to molar ellipticity [θ] units (deg cm^2 dmol^{-1}).

NMR Spectroscopy. Spectra were acquired using a Bruker Avance 600 spectrometer with either a TXI triple resonance inverse probe with *z*-axis gradient or a broad-band inverse probe with *z*-axis gradient. One-dimensional spectra were obtained at 5 °C using the WATERGATE and 1,1 jump-return pulse sequences to achieve water suppression. All one-dimensional NMR experiments were recorded over 32768 data points, with a spectral width of 20.028 ppm and a relaxation delay time of 2 s. The high scan experiments typically employed 20000−30000 scans to achieve an acceptable signal-to-noise ratio. Samples for NMR consisted of 70−100 μM DNA dissolved in 50−100 mM KCl containing 10 mM potassium hydrogen phosphate (pH 7). Samples were prepared in 600 μL of 10% D_2O/90% H_2O buffer and were then heated to 90 °C for 15 min, allowed to slowly cool to room temperature, and stored at 4 °C overnight prior to data collection.

Expression of Exon C. Total RNA from MCF-7 cells and total RNA from normal and neoplastic endometrial tissues (Ambion, catalogue 7232_N and T) were treated with DNase I (Ambion AM1906). Subsequently, 400 ng of each RNA was used in RT-PCR with a Beckman GeXP kit (Beckman Coulter UK Ltd., High Wycombe) and resolved on capillary electrophoresis followed by sequencing. The positions of the 3 PCR amplicons (numbered as seq X62462.1) for exon C expression were as follows: a fragment located 5′ to the quadruplex, nucleotides 272−372; a fragment covering the quadruplex, nucleotides 514−652; and an exon/intron spanning fragment to control for DNA contamination, nucleotides 687−1030. In each case the products included 37 nucleotides corresponding to the sequence of the universal primers used in the Beckman GeXP system, giving products of 137, 175, and 380 nucleotides, respectively. Under some conditions the product ran 29 bp longer than expected, as a result of formation of quadruplex secondary structures which affected mobility.

Plasmid Construction. A fragment of approximately 500 bp representing positions 1726−2235 of exon C in the human *ESR1* promoter region (AF082876) was amplified by PCR, *Nco*I digested, and cloned into pGL3-Basic vector at the *Hin*dIII/*Nco*I

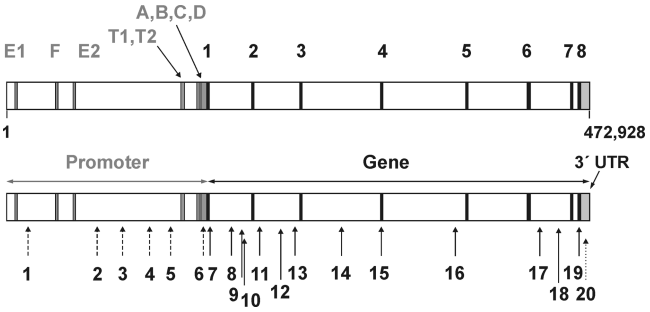

FIGURE 1: (a) The human *ESR1* gene, including promoter and 3′-UTR. The gene spans 472928 nucleotides encompassing eight coding exons (numbered 1−8) (black lines) and nine noncoding 5′-untranslated exons (gray lines) corresponding to alternative promoter sequences labeled A−F, T1, and T2 (*14*). The lines representing the exons are not to scale, with respect to the nucleotide sequence, but merely indicate the relative positions. (b) The relative locations of the 20 putative quadruplex-forming sequences are shown with arrows. Three quadruplex sequences (QS) fall within exons (C in the promoter region and 4 and 8). Dotted arrows represent possible motifs within the promoter region, and solid arrows show motifs arising within the gene sequence. Finally, QS20 is located within the 3′-UTR. The numbers correspond to sequences shown in Figure 2.

sites. The forward primer included the *Hin*dIII site followed by the T7 promoter sequence. The quadruplex-forming potential of the sequence gggta**GGGG**caaaggggctgggg (positions 1881−1904) was compromised by mutating one of the G-runs into As as follows, gggta**AAAA**caaaggggctgggg, using a QuickChange mutagenesis kit (Stratagene).

In Vitro Transcription and Translation. Plasmid DNA with either functional or mutated quadruplex sequences was linearized (at a *Sal*I site), and 1 μg of each construct was used for *in vitro* transcription using the mMessage mMachine T7 ultra kit (Ambion) according to the manufacturer's instructions and as described (*5*). After confirming RNA integrity and size on 1% agarose gels, the capped and poly(A)-tailed RNA was used in an *in vitro* translation procedure with rabbit reticulocyte lysate (Promega) according to producer's protocol. Luciferase was measured using a MicroLumenotPlus LB 96 V luminometer (Berthold Technologies) and firefly luciferase assay kit (Biotium Inc., Hayward, CA).

RESULTS

Human ESR1: Bioinformatics Analysis. The human *ESR1* gene spans ~473 kilobases. In addition to the eight coding exons, the promoter region includes nine noncoding exons with a series of alternative promoters (Figure 1) (*14*). These noncoding exons do not contribute toward the final protein sequence but are present in various combinations in the 5′-UTR of the mRNA (*15*). The transcription of these exons and the use of their promoters are tissue-specific and reflect the highly regulated expression of *ESR1* (*14*). A bioinformatics search for G_{3-5}-N_{1-7}-G_{3-5}-N_{1-7}-G_{3-5}-N_{1-7}-G_{3-5} sequences using the *quadparser* program (where N = any base, with connecting loops of one to seven bases) (*2*) identified the presence of 20 possible quadruplex-forming motifs distributed throughout the *ESR1* gene and promoter sequence (Figure 2). This corresponded to ~1 quadruplex motif every 24000 bases. The majority of these sequences (17 out of 20) are located within introns, but three were identified within exons. Two occurred in exon 4 (QS15) and exon 8 (QS19), but of particular interest in this paper was QS6, located within the 5′-untranslated region of exon C, which we have confirmed is transcribed into RNA (see below).

DNA Quadruplex Formation in Sequences Derived from Exon C of the Human ESR1 Gene. QS6 has an abundance of guanines with three G_4 tracts and one G_3 tract separated by potential loop sequences of different lengths. The formation of a DNA quadruplex was assessed initially using CD and UV spectroscopy (Figure 3). A 27-mer, 5′-d(CAAGGGTAGGGGCAAA GGGGCTCCCCT), which included capping nucleotides on the 5′- and 3′-ends to discourage intermolecular association, was analyzed. UV spectra recorded at 5 and 95 °C showed a hyperchromic shift around 295 nm, indicating quadruplex formation (*7*). The far-UV CD spectrum of the 27-mer showed maxima at 260 and 295 nm in the ratio of approximately 3:2. This could be interpreted as either the presence of a mixture of different parallel and antiparallel quadruplex structures in equilibrium (giving the 260 and 290 nm bands, respectively) or a single species with mixed strand alignment. We can readily distinguish between these two possibilities from the melting curves shown in Figure 3b. The two curves recorded at different wavelengths (260 and 290 nm) give T_m values differing by ~10 °C. If the CD spectrum was that of a single mixed structure, then we would expect to see a common T_m when monitored at the two wavelengths, reflecting the same cooperative unfolding process. This is not the case, and the data are consistent with a mixture of structures with different stabilities. The temperature dependence of the spectra showed that the peak at 295 nm initially decreased while the 260 nm band increased, suggesting either the conversion of some of the lower stability antiparallel structure (T_m ~ 65 °C) into the parallel form (T_m ~ 74 °C) or a temperature-sensitive change in the structure of the parallel form that results in a small CD spectral shift (*16*, *17*). Thus, QS6 within the exon C promoter sequence shows evidence for formation of stable DNA quadruplex structures.

NMR Confirms Quadruplex Formation in QS6 Derived from the Exon C Promoter. One-dimensional ^{1}H NMR experiments on QS6 at 600 MHz showed the presence of NH resonances in the 10.5−12.5 ppm region of the spectrum characteristic of guanine imino proton signals from hydrogen-bonded G-tetrads. However, there is evidence for both a single discrete monomeric species, which gives sharp resonances, and other more dynamic interconverting conformations that give rise to a background of broader, less well resolved signals (Figure 3c). Some redundancy of guanines within the sequence may explain this structural heterogeneity.

We synthesized several QS6 analogues containing inosine substitutions to reduce the number of available G-tetrads. Inosine lacks a 2-amino group and cannot form the same stable hydrogen-bonded structure as guanine. Under these conditions quadruplexes with longer loops tend to adopt antiparallel arrangements (*18*). We explored this possibility through two inosine substitutions (singly at position 9 and a double substitution at positions 9 and 20) to favor loops of three nucleotides:

d(CAA**GGG**TA**G**^{9}**GGG**CAAA**GGG**^{20}**G**CT**GGGG**T)

The NMR spectra revealed that neither G9I nor G9I/G20I stabilized a single species (data not shown), although changes in the ratio of parallel and antiparallel conformers were evident from CD studies such that G9I/G20I with loop 1 = TAI and loop 3 = ICT now appeared to be predominantly the antiparallel structure (ca. 60%) as judged by the intensity of the band at 295 nm.

Number	Position of quadruplex in gene sequence	Located in Exon?	Length	Sequence
Promoter quadruplex sequences				
QS1	12,577 - 12,607	No	31	**GGGGGTAGTGTTGGGGGAGTGGGGAGAGGGG**
QS2	66,532 - 66,560	No	29	**GGGGTGGGAGGGGAAGAGGGAGGTGAGGG**
QS3	86,548 - 86,572	No	25	**GGGCATGTAGGGCTTGGGCATGGGG**
QS4	107,266 - 107,286	No	21	**GGGGGCTGGGAAGGGGAAGGG**
QS5	123,506 - 123,535	No	30	**GGGCCTGTCGGGGGGTTGGGGGTCAAGGGG**
QS6	148,770 - 148,792	Exon C	23	**GGGTAGGGGCAAAGGGGCTGGGG**
Gene quadruplex sequences				
QS7	151723 - 151748	No	26	**GGGAGGGAGGGAGGGAGGGAGAAGGG**
QS8	169375 - 169400	No	26	**GGGATGCGGGGAGGGGGGTGGCGGGG**
QS9	177847 - 177876	No	30	**GGGGTCTATCAGGGGGTTGGGGGCTAAGGG**
QS10	178182 - 178219	No	38	**GGGCAGGCAAGGGTGAGGGTGCAGGGGGTCATGCTGGG**
QS11	191638 - 191659	No	22	**GGGGTCGGGGGAGGGGGGAGGG**
QS12	208248 - 208269	No	22	**GGGGTGGGGGGAGGGGGGAGGG**
QS13	219441-219459	No	19	**GGGAGGTGGGGAGGGAGGG**
QS14	256217-256250	No	34	**GGGGCCTGTCGGGAGTTGGGGGACTAGGGGAGGG**
QS15	287541-287564	Exon 4	24	**GGGGAGGGCAGGGGTGAAGTGGGG**
QS16	346271-346291	No	21	**GGGGATGAGGGTGGGGATGGG**
QS17	412884-412908	No	25	**GGGGTGGTGTAGGGTGGGGTGGGGG**
QS18	427570-427603	No	34	**GGGGCCTGTCAGGGGTTGGGGGGCTGGGGGAGGG**
QS19	443582-443603	Exon 8	22	**GGGGCACGGGAGAAGGGTGGGG**
Quadruplex sequence in 3′-UTR				
QS20	461,923 - 461,944	No	22	**GGGGTGGGGGGCAAGGGGAGGG**

FIGURE 2: The putative quadruplexes identified in the promoter sequence, gene sequence, and 3′-UTR (untranslated region) of the human *ESR1* gene. Sequences with four or more G-tracts of three or more guanine bases were identified using the *quadparser* program (*2*). Loop lengths of between one and seven bases were selected. The number assigned to the quadruplex sequence corresponds to the numbered arrow shown in Figure 1. The position of the quadruplex within the genomic 472928 nucleotide sequence is given, and whether this is an exonic region of the sequence is indicated.

Drug-Induced Changes in Quadruplex Conformational Equilibria. We extended the CD analysis to examine effects of small molecule ligands on conformational heterogeneity by inducing quadruplex formation. Titration of QS6 with RHPS4, a quadruplex-binding fluorinated polycyclic acridinium cation (*19*−*21*), produced a marked increase in the proportion of antiparallel structures upon melting and reannealing. In the absence of salt, QS6 gives a CD spectrum consistent with a largely unfolded state with a maximum at 255 nm. By analogy with telomestatin (*22*), addition of the RHPS4 acridinium cation induced quadruplex structure in QS6 (Figure 4a). Successive 0.5 mol equiv of the drug added to QS6 resulted in a decrease in single-stranded DNA and an increase in the G-quadruplex DNA band at 292 nm showing exclusively the antiparallel conformation. A point of saturation at a drug/DNA ratio of 2 suggests the drug binds with a 2:1 stoichiometry (Figure 4b), with a single ligand molecule stacked on each end of the quadruplex (*18*). The sequential formation of first a 1:1 and then a 2:1 complex by RHPS4 is evident from ESI-MS analysis (*17*) and supports such a binding model rather than the stacking of a ligand dimer on one or the other end of the structure (results not shown). The CD spectrum, with a strong positive band at 292 nm, suggests a significant population of the (2 + 2) antiparallel basket form observed for the human telomeric repeat sequence (*h*-Tel) in Na^+ solution (Figure 4c). Specific interactions of RHPS4 with diagonal loops appear to favor formation of the (2 + 2) antiparallel arrangement of strands, where the ligand is held in place by a combination of stacking interactions with the G-tetrads and specific interactions with bases in the loop (*23*, *24*). A similar ligand-induced preference of RHPS4 for the antiparallel conformation of *h*-Tel has also been observed (*17*).

Quadruplex Sequence Is Present in Transcripts of ESR1 Containing Exon C. The three sets of PCR primers were used to amplify a sequence located 5′ to the quadruplex, nucleotides 272−372, the fragment covering the quadruplex (nucleotides 514−645), and an exon/intron spanning fragment (nucleotides 867−1030). The primers were each hybridized to three different RNA preparations from human uterus, human endometrial cancer, and human breast cancer (MCF-7) cells. When the products were run on capillary electrophoresis (Figure 5a−d) there was no amplification of the 3′ exon C/intron boundary or the sequence located 5′ to the quadruplex. However, in each tissue a product corresponding to the quadruplex-containing sequence in exon C was isolated. Reamplification and sequencing confirmed the presence of the quadruplex in the product (Figure 5e). This product did not reflect DNA contamination of the amplified RNA as the RNA samples were treated with DNase to eliminate this possibility, and this was confirmed by the absence of products on amplification of the 3′ exon C/intron boundary or the 5′ sequence. We conclude that the quadruplex sequence is present in transcripts of the estrogen receptor gene containing exon C, confirming earlier data (*41*, *42*).

Quadruplex Formation in the 5′-UTR of the Exon C mRNA Transcript. Quadruplex formation within the 5′-UTR

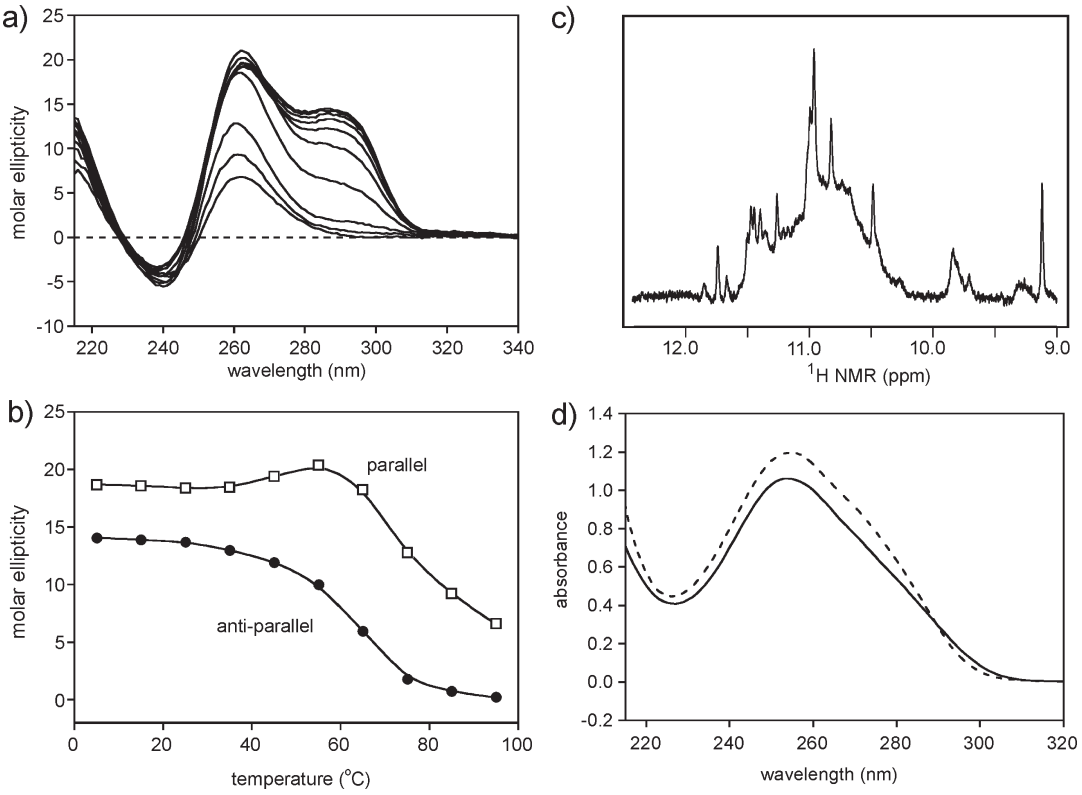

FIGURE 3: Spectroscopic data on the human exon C quadruplex sequence QS6 (see Figure 2). (a) Far-UV CD spectra in 110 mM K^+, pH 7.0, at temperatures between 5 and 95 °C. Bands at 260 and 295 nm show the presence of parallel and antiparallel species, respectively. (b) Changes in ellipticity at 260 and 295 nm as a function of temperature showing different melting profiles at the two wavelengths support the presence of a mixture of conformers which appear to interconvert as the temperature is increased (initial small increase in intensity at 260 nm as the less stable antiparallel structure begins to melt). We estimate T_m values for the two forms of ~65 °C (antiparallel) and ~75 °C (parallel). (c) NMR spectrum showing the resonances in the 10.5−12.5 ppm region characteristic of quadruplex formation, recorded at 278 K, 119 μM DNA, 100 mM KCl, and $KHPO_4$, pH 7.0. (d) UV scans at high (dotted line) and low temperatures, demonstrating the hyperchromic shift characteristic of quadruplex-forming sequences. The samples for CD and UV analysis consisted of 6 μM DNA in 100 mM KCl and 10 mM $KHPO_4$, pH 7.0.

of mRNAs has already been shown to play a critical role in the regulation of translation (*5*, *25*). The 5′-UTR of the mature exon C mRNA transcript was studied by CD and UV analysis using a synthetic 27-mer RNA (RQ6). In solutions containing 100 mM K^+, we observed a CD spectrum of RQ6 indicative of an exclusively parallel quadruplex structure with a strong band at 265 nm (Figure 6a). In contrast, its DNA counterpart QS6 showed a mixture of parallel/antiparallel structures with a significant contribution to the CD spectrum from both *syn* and *anti* glycosidic conformations with a strong band also evident at 295 nm. However, comparison of the relative stabilities of the two structures from CD melting experiments showed that the mRNA sequence was significantly more stable in 100 mM K^+ and at pH 7.0 with a T_m > 85 °C (Figure 6b). The ribose 2′-OH group evidently affects the conformational equilibrium between different structural topologies.

Effects of the Exon C mRNA (RQ6) on the Efficiency of Translation. The potential regulatory role of quadruplex formation by RQ6 on RNA translation was investigated by cloning ~500 base pairs of the exon C sequence into the expression vector pGL3 between the T7 promoter and the firefly luciferase reporter gene. The quadruplex-forming sequence of exon C was located 154 nucleotides downstream of the T7 promoter sequence. Two plasmids were prepared, one carrying the native exon C gene sequence, GGGTAGGGGCAAAGGGGCTGGGG (+Qx), and the other a mutated sequence, GGGTAAAAACAAAGG-GGCTGGGG (*mut*Q), in which one G_4 tract of QS6 had been substituted with A_4 to eliminate the possibility of quadruplex formation, while maintaining the length of the insert. One microgram of RNA was subsequently used in *in vitro* translation with rabbit reticulocyte lysate, and the efficiency of translation was measured from the luminescent output resulting from the luciferase catalytic activity (Figure 7). The results showed that the adenine mutations resulted in a 6-fold increase in the rate of translation, showing that the formation of a stable quadruplex within the mRNA transcript directly affects levels of translation when situated relatively close to the 5′-cap of the transcript (*25*).

DISCUSSION

G-rich repeats tend to be underrepresented in exonic regions of gene sequences, with C-rich, A-rich, and T-rich repeats occurring more frequently (*2*). This may be the result of an evolutionary selective pressure to reduce the number of potential mRNA quadruplex-forming sequences to minimize the undesirable downregulation of mRNA translation, including interference with posttranscriptional processing such as gene splicing events. When located close to polyadenylation and splice sites, sequences capable of forming G-quadruplexes act as regulators by interacting with hnRNP F and H proteins (*26*, *27*). Our search highlighted two sequences (QS7 and QS15 in Figure 1) which are both within 60 nucleotides of a 3′-splice site. There also appears to be a higher incidence of potential quadruplex motifs within the first two introns where a larger number of regulatory sequences are likely to be found (*28*). Thus, although the four quadruplex motifs identified within the first intron (Figure 1, QS7

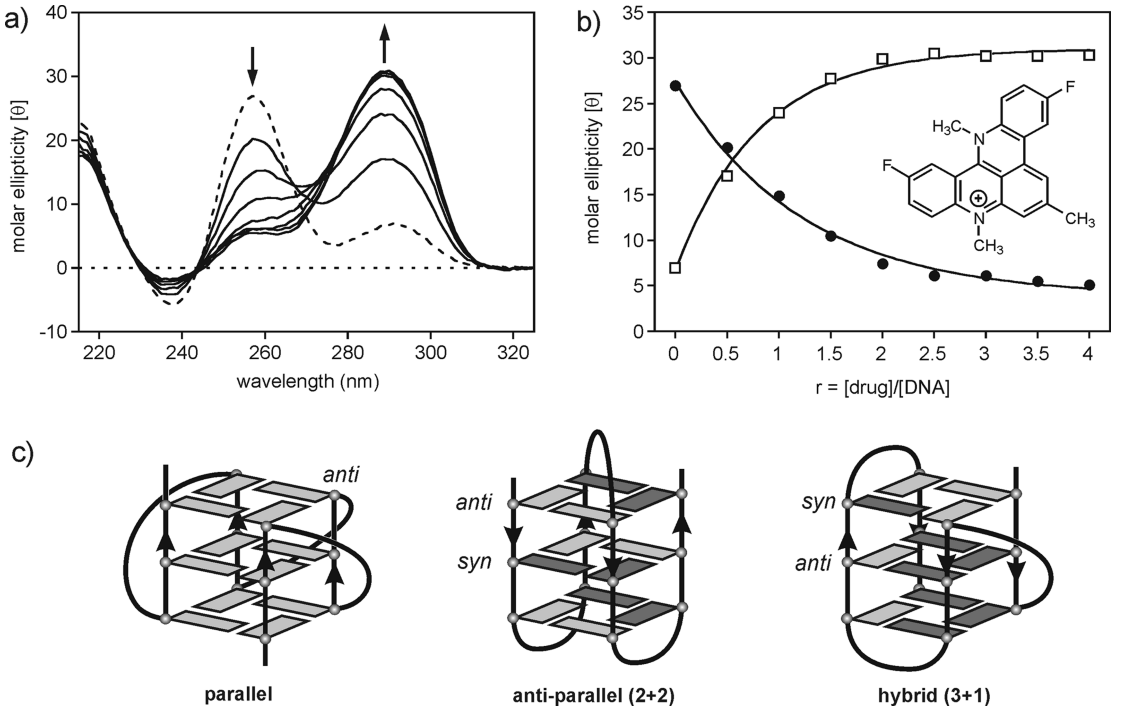

FIGURE 4: (a) A stacked plot representing the CD titration of the salt-free exon C DNA sequence with increasing half-equivalents of the telomerase inhibitor RHPS4 (see inset in (b)). The initial signal at 257 nm corresponds to the unfolded single-stranded sequence (CD spectrum shown as a dashed line). Addition of drug induces antiparallel structure in the exon C sequence as evident from the strong band at 292 nm. (b) Two titration curves representing the changes in ellipticity with the addition of each RHPS4 half-equivalent as monitored at 257 and 292 nm. A saturation point is observed for the two curves at close to $r = 2$ (where r is the drug/DNA fraction). (c) Schematic representation of the possible parallel, (2 + 2) antiparallel, and (3 + 1) hybrid quadruplex folds. The (2 + 2) and (3 + 1) nomenclature refers to the number of parallel strands. In the case of (2 + 2) structure, the basket arrangement is shown in which there is an equal distribution of strands running in opposite directions, resulting in diagonal or lateral loops at either end of the structure. In the (3 + 1) hybrid, the distribution of strands is asymmetric with three strands having the same alignment; as a consequence one of the external loops adopts a propeller orientation on the outside of the structure. These arrangements are achieved by different combinations of *syn* and *anti* glycosidic torsion angles, as illustrated. In the parallel structures, all strands are orientated in the same direction, resulting in propeller loops; all glycosidic torsion angles are *anti*.

to QS10) are untranslated, they may affect the splicing machinery or the binding of other regulatory proteins that determine rates of gene expression. Genome-wide studies suggested that quadruplex sequences immediately downstream of transcription start sites (TSS) are associated with increased rates of gene expression (*29*). QS7, which is within 900 bases of the TSS of *ESR1*, has the potential to adopt an extremely stable quadruplex structure with three single nucleotide loops that could prevent rehybridization with the template strand. In addition, an analysis of the *ESR1* coding strand for C-rich sequences indicates that no putative quadruplex structures can form in the same region of the complementary template strand. An analysis of the whole 472928 nucleotide sequence, however, identified 19 C-rich motifs capable of forming quadruplex structures in the template strand. Thus, the number of QSs in the coding and template strands is essentially identical (20 versus 19, respectively).

Promoter sequences tend to be enriched with quadruplex sequences, and the probability of finding such a sequence is directly related to its proximity to the transcription start site (TSS) (*30*). While there are six such sequences within the *ESR1* promoter, the more general distribution of motifs may be a consequence of the complicated array of noncoding exons and the use of alternative promoters. A number of the identified sequences are found with four discrete G-tracts, but these are seldom of the same length. This suggests that a number of these motifs may form multiple structures utilizing different guanines. In addition, a number of these sequences are comprised of large G-rich regions interspersed with single A or T bases, analogous to the sequences identified for c-myc (*31*, *32*). Interestingly, the exon sequences tend to be comprised of four distinct G-tracts, although longer G-rich regions are found in the intron sequences.

Genome-wide studies on quadruplex motifs and the distribution of loop sequences (*1*) and loop lengths (*2*) show significant deviation from a random distribution. The most frequently occurring loops were those containing single A and single T bases, with the quadruplex motif comprising three single nucleotide loops occurring most commonly. Over half the sequences identified in the analysis of the *ESR1* gene potentially incorporate a single nucleotide loop, while QS7 could readily accommodate three single nucleotide loops, resulting in both higher stability *in vitro* (*33*) and reduced competition for duplex formation due to a shorter sequence. The presence of longer loops in a number of the identified motifs might reduce the chance of them forming in a biological context or allow tighter regulation through specific loop recognition and induced folding.

The NMR spectra suggested QS6 adopted a number of interconverting structures, and there is evidence for this from other genes. In the promoter sequence of *c-myc*, the sequence d(A**GGGT**-**GGGGAGGGTGGGG**) forms an ensemble of four structures each utilizing a different combination of guanines to generate three G-tetrads (*7*). A quadruplex derived from the human *bcl-2* promoter region with a GGGGG tract, d(**GGG**CGC**GGG**AGG-AA**GGGGGCGGG**), also exhibited a variety of imino resonances in the 1D NMR spectrum under similar conditions (*34*). However, when thymine substitutions reduced the longer tract to only three guanines, d(**GGG**CGC**GGG**AGGAA<u>TT</u>**GGGCGGG**),

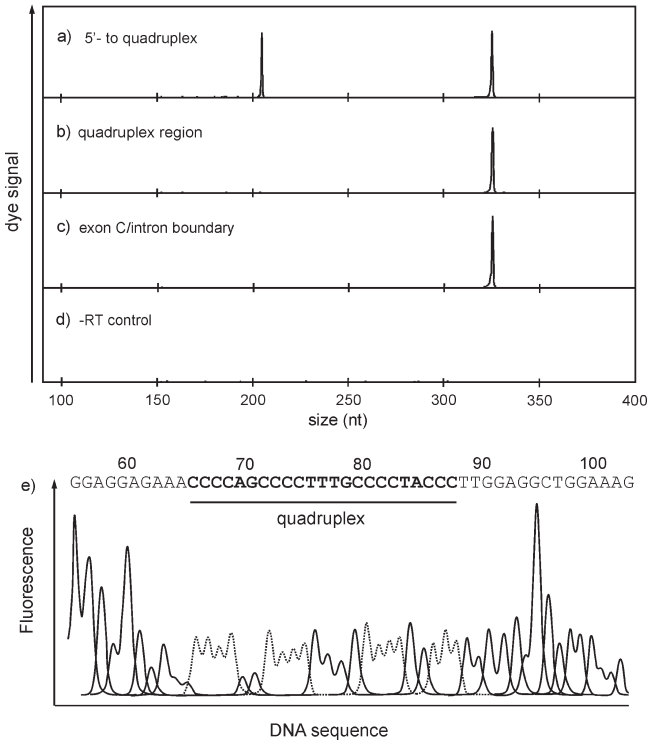

FIGURE 5: Expression of exon C in normal human uterus identified by duplex PCR reactions. Three pairs of primers specific for (a) the region upstream (5′) of exon C, aplicon size of 137 nt, (b) the quadruplex region of exon C, size 175, or (c) the exon C/intron boundary (expected peak size of 380 nt) were used together with primers to amplify kanamycin RNA (added to each duplex reaction mixture as an internal control, peak size of 325 nt) (d). (e) Output of sequencing of the 204 nucleotide product with a reverse primer (the C-rich motif is highlighted with the C nucleotides shown as dotted lines in the sequencing data). Note the apparent size difference between the amplicons in (b), 204 versus 175 nucleotides, is due to the composition of the PCR buffer used in the Beckman GeXP method, as the same product runs at the expected position when amplified in other buffers (data not shown).

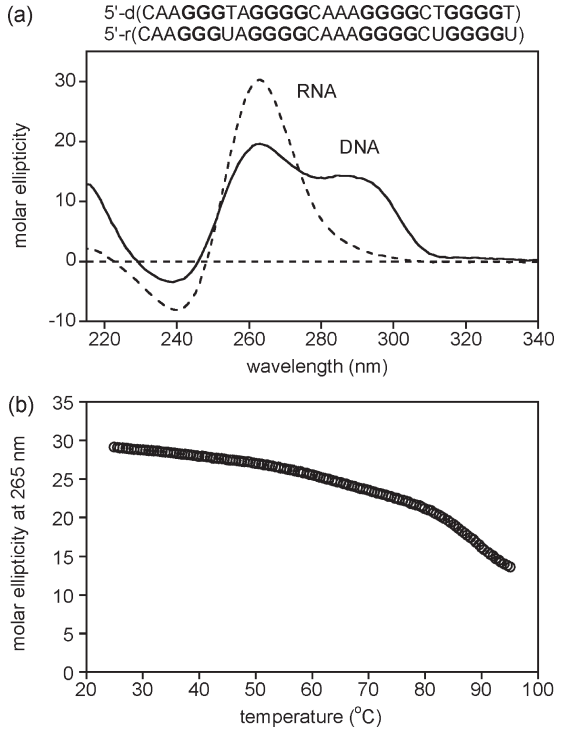

FIGURE 6: (a) CD spectra of the exon C DNA quadruplex sequence (QS6) and that of the equivalent RNA transcript RQ6 (both sequences shown at the top of the figure). Spectra were recorded under identical conditions at 298 K in 100 mM K^+ solution at pH 7.0. The ellipticity profiles show different conformational preferences. The RNA sequence has a single strong positive band at 262 nm, indicative of exclusively parallel-folded structure, whereas the DNA sequence has a double positive with a strong band also evident at 290 nm indicative of a significant contribution from an antiparallel conformer. (b) Temperature dependence of the ellipicity at 265 nm of the RNA quadruplex RQ6 showing a $T_m > 85$ °C.

well-resolved imino resonances became apparent for a single major conformer.

CD analysis of the RNA quadruplex (RQ6) showed the adoption of a stable, exclusively parallel structure with a strong maximum ellipticity at 262 nm. In contrast, the DNA sequence (QS6) gives a CD signature consistent with a mixture of both parallel and antiparallel structures with positive bands at 260 and 295 nm. Similar differences in the conformational preferences are evident in the CD spectra of human DNA and RNA telomeres, where a more detailed structural analysis has revealed a mixture of (2 + 2) antiparallel and (3 + 1) hybrid DNA structures (see Figure 4c), with the DNA conformational equilibrium sensitive to the nature of the monovalent cation and the flanking nucleotide sequence (*35*−*38*). The conformational preference of the RNA ribose sugar for the C3′-endo pucker is known to favor the A-form over the B-form of duplex DNA; however, the origin of the effects on the structural topology of parallel versus antiparallel quadruplexes is less clear.

The molecular link between exon C and endometrial cancer is not currently understood. One theory is that CpG island methylation renders the exon incapable of acting as a template for receptor production (*39*). However, it is also possible that the formation of a quadruplex structure prevents the expression of exon C mRNA. Thus, *in vivo* generated quadruplex structures may regulate *ESR1* expression in tissues susceptible to cancer. This sequence was shown by CD, UV, and NMR spectroscopy to be predisposed toward quadruplex formation *in vitro*. However, *in vivo* protein-induced quadruplex folding may favor a specific conformation. Many alternatively spliced mRNA transcripts with a deletion of one or more coding exons or truncation in the coding region have been observed, especially in breast cancer (*40*). In particular, a variant mRNA that lacks exon 4 has been identified in human breast cancer cell lines (*40*), and one possible hypothesis is that quadruplex formation in the 5′-UTR of the mRNA downregulates *ESR1* expression in normal human cell lines and must be inactivated in rapidly proliferating cancer cells.

We have used the exon nomenclature for the *ESR1* gene proposed by Koš et al. (*14*), with the extended 5′ boundary of exon C 3067 bp upstream of exon A. This locates the quadruplex sequence (QS6) within exon C. However, there has been some uncertainty over the 5′ boundary of exon C, with the structure proposed by Koš et al. (*14*) reflecting this. In particular, it was unclear whether the proposed 5′ extension to exon C (*41*, *42*) in *ESR1* transcripts expressed in MCF-7 cells was in fact transcribed. Transcripts with short 5′ untranslated regions have been identified (*43*, *44*), but these authors did not specifically search for longer variants. However, on the basis of genomic sequence it has been suggested that the 5′ exon C boundary is located 524 bp upstream of the original position (*42*, *44*), which includes QS6 in the transcribed region. Our data confirm that QS6 is present in

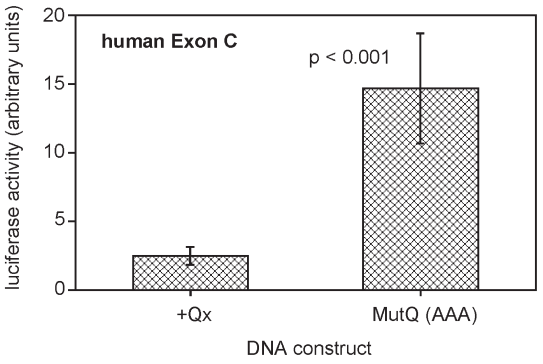

FIGURE 7: Translational efficiency for different DNA constructs cloned upstream of the luciferase reporter gene. The wild-type exon C sequence (+Qx) was compared with a mutated sequence (MutQ) containing a GGGG to AAAA substitution which destabilizes the exon C quadruplex. The latter shows a 6-fold enhancement of enzymatic activity, consistent with the exon C motif modulating the efficiency of translation. Luciferase activity, arbitrary units (mean ± SE for three separate experiments).

ESR1 transcripts from normal and cancerous endometrium and in MCF-7 cells. As a result, this sequence may lead to reduced estrogen receptor translation rates in tissues expressing transcripts containing exon C, relative to those not expressing this sequence (Figure 7).

ESR1 gene expression occurs from up to 10 transcription start sites and results in multiple transcript structures formed through alternative exon splicing in a tissue-specific manner (*15*, *42*, *45*). Considerable attention has been paid to control of exon usage through epigenetic mechanisms such as CpG methylation (*46*, *47*), high levels of methylation leading to gene silencing, and differential methylation patterns are known to contribute to differential levels of expression of the *ESR1* gene in neoplasia (*47*, *48*). However the present results suggest that G-quadruplex formation in RNA transcripts may also affect levels of receptor protein expression through differential effects on translation rates. Thus transcription of a quadruplex-containing exon would be expected to reduce levels of the translated product. The situation may arise, therefore, that in a tissue where epigenetic signals lead to transcription of a quadruplex-forming exon, translation of the resulting transcript is low relative to those transcripts without the quadruplex.

In the present experiments there was no indication that differential exon usage is associated with neoplasia, as the exon C transcript was present in both normal and cancerous endometrium. Other reports suggest there is differential expression of exon C in normal and neoplastic tissue (*49*). An explanation for this difference may be that the GeXP PCR system used here is highly sensitive and may have resulted in the detection of rare transcripts. Therefore, there may be significant quantitative, if not qualitative, differences between exon C expression in the normal and neoplastic tissue which was not observed in this work. If, as has been suggested, differential exon usage is characteristic of neoplastic transformation (*14*, *42*, *49*), the possibility arises that G-quadruplexes may play a significant role in determining estrogen receptor levels in neoplasia. Opportunities to disrupt or stabilize these sequences may therefore offer new avenues toward clinical management of cancer. Indeed, the repression of translation by a stable G-quadruplex within the 5′-UTR of the MT3 matrix metalloproteinase mRNA suggests a similar therapeutic approach to regulation of cancer cell invasion and metastasis (*50*).

ACKNOWLEDGMENT

We thank Emma Gregson for providing MCF-7 cells, Prof. Malcolm Stevens for providing a sample of the drug RHPS4, and Prof. Bob Webb from the School of Biosciences, University of Nottingham, for encouragement and support.

REFERENCES

1. Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res. 33*, 2901–2907.
2. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acid Res. 33*, 2908–2916.
3. Shafer, R. H., and Smirnov, I. (2001) Biological aspects of DNA/RNA quadruplexes. *Biopolymers 56*, 209–227.
4. Sinden, R. P. (1999) Biological implications of the DNA structures associated with disease-causing triplet repeats. *Am. J. Hum. Genet. 64*, 346–353.
5. Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5′-UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol. 3*, 218–221.
6. Ambrus, A., Chen, D., Dai, J., Jones, R. A., and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the c-myc promoter. Implications for G-quadruplex stabilisation. *Biochemistry 44*, 2048–2058.
7. Fernando, H., Reszka, A. P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A. R., Neidle, S., and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry 45*, 7854–7860.
8. Cogoi, S., and Xodo, L. E. (2006) G-quadruplex formation within the promoter of the *KRAS* proto-oncogene and its effect on transcription. *Nucleic Acids Res. 34*, 2536–2549.
9. De Armond, R., Wood, S., Sun, D., Hurley, L. H., and Ebbinghaus, S. W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1α promoter. *Biochemistry 44*, 16341–16350.
10. Dexheimer, T. S., Fry, M., and Hurley L. H. (2006) in DNA quadruplexes and gene regulation. Quadruplex nucleic acids (Neidle, S., and Balasubramanian, S., Eds.) pp 180−207, Royal Society of Chemistry, Cambridge.
11. Deroo, B. J., and Korach, K. S. (2006) Estrogen receptors and clinical disease. *J. Clin. Invest. 116*, 561–570.
12. O'Lone, R., Frith, M. C., Karlsson, E. K., and Hansen, U. (2004) Genomic targets of nuclear estrogen receptors. *Mol. Endocrinol. 18*, 1859–1875.
13. Pfaffl, M. W., Lange, I. G., Daxenberger, A., and Meyer, H. H. D. (2001) Tissue-specific expression pattern of estrogen receptors (ER): Quantification of ER alpha and ER beta mRNA with real-time RT-PCR. *Acta Pathol., Microbiol. Immunol. Scand. 109*, 345–355.
14. Koš, M., Reid, G., Denger, S., and Gannon, F. (2001) Genomic organisation of the human ER alpha promoter region. *Mol. Endocrinol. 15*, 2057–2063.
15. Okuda, Y., Hirata, S., Watanabe, N., Shoda, T, Kato, J., and Hoshi, K. (2003) Novel splicing events of untranslated first exons in human estrogen receptor a (ERα) gene. *Endocr. J. 50*, 97–104.
16. Balkwill, G. D., Garner, T. P., and Searle, M. S. (2009) Folding of single-stranded DNA quadruplexes containing an autonomously stable mini-hairpin loop. *Mol. Biosyst. 5*, 542–547.
17. Garner, T. P., Williams, H. E. L., Katarzyna, G. I., Roe, S., Oldham, N. J., Stevens, M. F. G., Moses, J. E., and Searle, M. S. (2009) Selectivity of small molecule ligands for parallel and anti-parallel DNA G-quadruplex structures. *Org. Biomol. Chem. 7*, 4194–4200.
18. Hazel, P., Huppert, J. L., Balasubramanian, S., and Neidle, S (2004) Loop-length dependent folding of G-quadruplexes. *J. Am. Chem. Soc. 126*, 16405–16415.
19. Gowan, S. M., Heald, R., Stevens, M. F. G., and Kelland, L. R. (2001) Potent inhibition of telomerase by small molecule pentacyclic acridines capable of interacting with G-quadruplexes. *Mol. Pharmacol. 60*, 981–988.
20. Gavathiotis, E., Heald, R. A., Stevens, M. F. G., and Searle, M. S. (2003) Drug recognition and stabilisation of the parallel-stranded DNA quadruplex $d(TTAGGGT)_4$ containing the human telomeric repeat. *J. Mol. Biol. 334*, 25–36.
21. Heald, R., Modi, C., Cookson, J. C., Hutchinson, I., Laughton, C. A., Gowan, S. M., Kell, L. R., and Stevens, M. F. G. (2002) Anti-tumor polycyclic acridines. 8. Synthesis and telomerase-inhibitory activity of methylated pentacyclic acridinium salts. *J. Med. Chem. 45*, 590–597.
22. Rezler, E. M., Gokhale, V., Seenisamy, J., Bashyam, S., Kim, M. -Y., White, E., Wilson, W. D., and Hurley, L. H. (2005) Telomestatin and

diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J. Am. Chem. Soc. 127*, 9439–9447.
23. Gavathiotis, E., Heald, R. A., Stevens, M. F. G., and Searle, M. S. (2001) Recognition and stabilisation of quadruplex DNA by a potent new telomerase inhibitor: NMR studies of the 2:1 complex of a pentacyclic methylacridinium cation with $d(TTAGGGT)_4$. *Angew. Chem., Int. Ed. 40*, 4749–4751.
24. Haider, S. M., Parkinson, G. N., and Neidle, S. (2003) Structure of a G-quadruplex-ligand complex. *J. Mol. Biol. 326*, 117–152.
25. Kumari, S., Bugaut, A., and Balasubramanian, S. (2008) Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within The 5′-UTR of The NRAS proto-oncogene. *Biochemistry 47*, 12664–12669.
26. Bagga, P. S., Arhin, G. K., and Wilusz, J. (1998) DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. *Nucleic Acids Res. 26*, 5343–5350.
27. Arhin, G. K., Boots, M., Bagga, P. S., Milcarek, C., and Wilusz, J. (2002) Downstream sequence elements with different affinities for the hnRNP H/H′ protein influence the processing efficiency of mammalian polyadenylation signals. *Nucleic Acids Res. 30*, 1842–1850.
28. Majewski, J., and Ott, J. (2002) Distribution and characterisation of regulatory elements in the human genome. *Genome Res. 12*, 1827–1836.
29. Du, Z., Zhao, Y. Q., and Li, N. (2008) Genome-wide analysis reveals regulatory role of G4 DNA in gene expression. *Genome Res. 18*, 233–241.
30. Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res. 35*, 406–413.
31. Simonsson, T., Pecinka, P., and Kubista, M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res. 26*, 1167–1172.
32. Seenisamy, J., Rezler, E. M., Powell, T. J., Tye, D., Gokhale, V., Joshi, C. S., Siddiqui-Jain, A., and Hurley, L. H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc. 126*, 8702–8709.
33. Bugaut, A., and Balasubramanian, S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry 47*, 689–697.
34. Dai, J., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc. 128*, 1096–1098.
35. 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/anti-parallel strands in potassium solution. *Nucleic Acids Res. 34*, 2723–2735.
36. Phan, A. T., Kuryavyi, V., Luu, K. N., and Patel, D. J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences. *Nucleic Acids Res. 35*, 6517–6525.
37. Xue, Y., Kan, Z.-Y., Yao, Y., Liu, J., Hao, Y.-H., and Tan, Z. (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K^+ solution under molecular crowding condition. *J. Am. Chem. Soc. 129*, 11191–11185.
38. Phan, A. T., and Patel, D. J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and anti-parallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties and folding/unfolding kinetics. *J. Am. Chem. Soc. 125*, 15021–15027.
39. Sasaki, M., Kotcherguina, L., Dharia, A., Fujimoto, S., and Dahiya, R. (2001) Cytosine-phosphoguanine methylation of estrogen receptors in endometrial cancer. *Cancer Res. 61*, 3262–3266.
40. Pfeffer, U., Fecarotta, E., Castagnetta, L., and Vidali, G (1993) Estrogen-receptor variant messenger-RNA lacking exon-4 in estrogen-responsive human breast-cancer cell-lines. *Cancer Res. 53*, 741–743.
41. Piva, R., Bianchi, N., Aguiari, G. L., Gambari, R., and del Senno, L. (1993) Sequencing of an RNA transcript of the human estrogen receptor gene: evidence for a new transcriptional event. *J. Steroid Biochem. Mol. Biol. 46*, 531–538.
42. Flouriot, G., Griffin, C., Kenealy, M., Sonntag-Buck, V., and Gannon, F. (1998) Differentially expressed messenger RNA isoforms of the human estrogen receptor-alpha gene are generated by alternative splicing and promoter usage. *Mol. Endocrinol. 12*, 1939–1954.
43. Amaral, S., Schroth, W., Kugler, S., Fritz, P., Simon, W., and Brauch, H. (2008) The promoter C specific ERα isoform is associated with tamoxifen outcome in breast cancer. *Breast Cancer Res. Treat*. (DOI 10.1007/s10549-008-0241-9).
44. Koš, M., Denger, S., Reid, G., and Gannon, F. (2002) Upstream open reading frames regulate the translation of the multiple mRNA variants of the estrogen receptor. *J. Biol. Chem. 277*, 37131–37138.
45. Szreder, T., and Zwierzchowski, L. (2004) Polymorphism within the bovine estrogen receptor-α gene 5′-region. *J. Appl. Genet. 45*, 225–236.
46. Penolazzi, L., Lambertini, E., Giordano, S., Sollazzo, V., Traina, G., del Senno, L., and Piva, R. (2004) Methylation análisis of the promoter F of estrógeno receptor α gene: Effects on the level of transcription on human osteoblastic cells. *J. Steroid Biochem. Mol. Biol. 91*, 1–9.
47. Berger, J., and Daxenbichler, G. (2002) DNA methylation of nuclear receptor genes—Possible role in malignancy. *J. Steroid Biochem. Mol. Biol. 80*, 1–11.
48. Asada, H., Yamagata, Y., Taketani, T., Matsuoka, A., Tamura, H., Hattori, N., Ohgane, J., Hattori, N., Shiota, K., and Sugino, N. (2008) Potential link between estrogen receptor-α gene hypomethylation and uterine fibroid formation. *Mol. Hum. Reprod. 14*, 539–545.
49. Weigel, R. J., Crooks, D. L., Iglehart, J. D., and deConinck, E. C. (1995) Quantitative analysis of the transcriptional start sites of estrogen receptor in breast cancer. *Cell Growth Differ. 6*, 707–711.
50. Morris, M. J., and Basu, S. (2009) An unusually stable G-quadruplex within the 5′-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry 48*, 5313–53.