BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 7625–7633 7625
DOI: 10.1021/bi100804f

Occurrence of a Quadruplex Motif in a Unique Insert within Exon C of the Bovine Estrogen Receptor α Gene (ESR1)†

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

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

Received May 20, 2010; Revised Manuscript Received July 21, 2010

ABSTRACT: The 5′ end of exon C of the bovine estrogen receptor α gene (*bov*-ESR1) includes a unique G-rich insert, not found in other closely related mammalian genes, which lies close to both a double E-box transcription factor binding site and the site of a single nucleotide (G/A) polymorphism. Biophysical studies, using CD and UV absorbance measurements, show that this 22 base insert leads to the formation of a family of stable G-quadruplex structures which are unaffected by the G/A polymorphism. Multiplex PCR shows that the region including the G-quadruplex is transcribed into RNA, and studies with a synthetic RNA transcript sequence demonstrated formation of a highly stable parallel-folded quadruplex structure. Luciferase reporter constructs demonstrate that the G-rich sequence reduces rates of translation when present in the 5′-UTR of mRNA transcripts. Mutations (GGG to AAA) that destabilize the quadruplex lead to a 15-fold enhancement of translational efficiency, suggesting that a possible biological role of the insert in exon C of the *bov*-ESR1 is to regulate translation of this exon.

Estrogens are steroid hormones that regulate tissue-specific cell growth and differentiation not only in reproductive physiology but in numerous aspects of human health and disease including cancer, cardiovascular disease, Alzheimer's disease, osteoporosis, obesity, arthritis, and social behavior (*1*). The biological effects of estrogens are mediated by nuclear estrogen receptors (ER), which act as ligand-inducible transcription factors to regulate gene expression through dimerization and binding to specific estrogen response elements (EREs) located in the promoter regions of target genes. However, it is now recognized that estrogen receptor signaling pathways can deviate from the classical model. For example, a third of human genes that are regulated by ERs do not contain ERE sequences. It is thought that activated ERs can regulate gene expression without direct DNA binding, employing protein−protein interactions to modulate the function of other transcription factors (*2*).

Currently, two isoforms of the estrogen receptor have been described, ERα[1] and ER*β*, each of which is coded for by a separate gene (ESR1 and ESR2, respectively) located on different chromosomes. These two receptors, ERα and ER*β* (in humans, comprised of 595 and 530 amino acids, respectively; in cattle, 596 and 527), show significant overall sequence homology, sharing common structural and functional domains. They both bind estrogens with high affinity and interact with EREs in an analogous manner. However, they have different tissue distributions, ligand affinities, and transcriptional activities. The ERα isoform is widely expressed throughout different tissues, whereas expression of the *β* isoform has a more limited distribution (*3*). The hormone-activated ERs can form both homodimers and heterodimers, as both forms, α and *β*, are coexpressed in many cell types.

Nuclear hormone receptors, including the estrogen receptors, activate transcription within genes associated with cell proliferation and differentiation, with the level of such activity dependent on the target tissue. ERα gene expression is known to be upregulated in estrogen receptor-positive breast cancers and can promote the growth of several other types of malignancy (*1*). Estrogen is essential in the initiation and development of breast and endometrial cancer and may also influence the development of prostrate cancer. As a result of this, the ERα protein has become an important drug target for the development of antagonists.

Initially, muscle-specific transcription factors were thought to only bind specifically to the major groove of the double-stranded CANNTG sequence (known as the E-box motif). However, it is now recognized that MyoD, the major member of the family of myogenic transcription factors, can also bind quadruplex structures. Yafe et al. identified a number of G-rich sequences in regulatory regions of human mitochondrial creatine kinase (sMtCK), mouse MCK, and mouse α7 integrin (*4*). The authors demonstrated, using nondenaturing gel electrophoresis, dimethyl sulfate (DMS) footprinting, and CD spectroscopy, that these sequences adopt bimolecular quadruplex structures. For instance, the two G-rich sequences located 85 nucleotides apart within the mouse α7 integrin promoter adopt a heterodimeric bimolecular quadruplex (*4*). Thus although a homodimeric quadruplex could not form *in vivo* as there is only a single copy of each sequence on each chromosome, a heterodimeric sequence formed from two separate G-rich sequences might. It is possible therefore that a

†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 DEFRA 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; UTR, untranslated region; *ESR1*, estrogen receptor α gene; ERα, estrogen receptor α; CD, circular dichroism; NMR, nuclear magnetic resonance.

Published on Web 07/26/2010

bimolecular quadruplex forms *in vivo* when the DNA is transiently single stranded, for example, during a transcription event.

A role has been proposed for quadruplexes in modulating the transcriptional activity of MyoD (*5*, *6*). By virtue of its helix−loop−helix DNA recognition motif, MyoD forms homodimers as well as heterodimers with other ubiquitously expressed helix−loop−helix proteins, such as E12 and E47 (*7*). Interestingly, the heterodimers bind E-box DNA more tightly than the homodimers and act as strong activators of transcription. In contrast, the homodimers bind bimolecular quadruplex structures more tightly than heterodimers. Based on these studies, it was proposed that switching between different DNA targets (quadruplex versus E-box) may provide a mechanism for regulating the timed activation of myogenic genes. Alternatively, it has been suggested that G-quadruplexes may bind hormones themselves in a specific manner, leading to modulation of gene expression (*8*).

The functional role that estrogens play in reproduction makes the estrogen receptors and their genes excellent biomarkers for animal fertility. Single nucleotide polymorphisms (SNPs) that accumulate in reproduction-related genes over several generations have led to unexpected traits in animal fertility. One such SNP was identified in exon C of the bovine ESR1 gene (*9*). We show that a G/A polymorphism, within exon C, is located close to a double E-box motif and only seven nucleotides downstream of a unique G-rich insert not found in other closely related mammalian genes (*9*). It has been shown that many potential quadruplex-forming sequences exist within the human genome (*10*), playing an important role in regulating gene expression at both the transcriptional and translational level (*11*−*14*). This has led to oncogenic promoter sequences with putative quadruplex-forming regions becoming important anticancer drug targets (*15*−*19*). We have probed quadruplex formation *in vitro* within the DNA insert of the bovine ESR1 gene and within the corresponding mRNA transcript using circular dichroism experiments and demonstrate significant influence on the efficiency of mRNA translation.

MATERIALS AND METHODS

Bioinformatics. The bovine (*Bos taurus*) ESR1 gene (*bov*-ESR1), located on chromosome 9, was studied for putative G-quadruplex formation. The 137928 nucleotide sequence was extracted from the Ensembl database (ENSBTAG00000007159). This sequence was extended in the 5′ direction by 5000 nucleotides to form a promoter region encompassing the three known 5′ UTR exons. The sequences of exons A, B, and C were identified using the accession numbers AY332655, AY641987, and AY641988, respectively. The quadparser program was used to search through the nucleotide sequence and identify possible quadruplex-forming sequences (*10*). 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 five 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 sequence and location of the coding exons were identified using the data provided in the Ensembl exon report for the ESR1 gene. The sequence and location of the 5′ UTR exons were identified using the work of Kos et al. and the references contained therein (*20*). The single nucleotide polymorphisms 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 putative quadruplex sequences for potentially destabilizing mutations.

The ESR1 sequence alignment was performed with the Basic Local Alignment Search Tool (BLAST) hosted on the NCBI webpages (http://www.ncbi.nlm.nih.gov/blast). Alignments were performed on the gene and an arbitrarily defined promoter region of 5000 nucleotides. The following species were aligned: *Bos taurus*, *Homo sapiens*, *Canis familiariz*, *Pan troglodytes*, *Oryctolagus cuniculus*, and *Equus caballus*.

Oligonucleotide Samples. Oligonucleotide sequences were derived from the sequence located in exon C in the *B. taurus* ESR1 gene. DNA oligonucleotides, purchased from Invitrogen (Paisley, U.K.), were synthesized on a 50 μmol scale and supplied lyophilized. Individual stock solutions (30−50 μM) were prepared with Milli-Q water.

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. The 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 prior to the start of the project. Typically, samples were prepared in the concentration range 3−6 μM in a buffer comprising 100 mM potassium chloride and 10 mM potassium hydrogen phosphate (pH 7). All samples were then heated to 90 °C and allowed to cool to 10 °C over a period of 4 h.

UV melting curves were recorded by measuring the absorbance at 295 nm of the sample in a tightly stopped 1 cm path length cuvette. The preannealed samples were equilibrated at 5 °C for 20 min before being heated to 95 °C and then cooled to 5 °C at a rate of 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 at every 1 nm step. 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 the 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}). Unless otherwise stated, the spectra were found to be concentration independent across the range 3−30 μM DNA. This was interpreted as evidence of intramolecular quadruplex formation.

Expression of the Exon C Quadruplex-Forming Sequence. Several bovine fetal and adult uterine tissues obtained from local slaughterhouses were frozen in liquid nitrogen. RNA was extracted using a Macherey-Nagel nucleic acid purification kit. DNA contamination of RNA was eliminated by digestion with DNase included in the kit. Additionally, 100 ng of RNA before reaction was subjected to DNA digestion (2 units, 10 min at 37 °C) with Turbo DNase (Ambion).

Table 1: PCR Primer Sequences

primer name	product size	left sequence	right sequence
Qx, quadruplex-forming region	137	cggcaacagtgtgtctgttc	aagccttctgggatccactt
H2A	155	tccagtgttggtgattccag	gcagaaatttggttggttgg
GNRHR_prom_DNA cont	167	cagtcgtgtccggacttctt	atgcatgcttacggagaagg
exC	182	cagacagcaagcctctcctt	atgtgtttgcatgtgggatg
ESR1	225	cccaactcctcctcatcctc	ctggctctgattcacgtcct
IGF1	265	tcttctatctggccctgtgc	acatctccagcctcctcaga
Kan(r)	325	atcatcaggattgcattcgattcctgtttg	attccgactcgtccaacatc
exon/intron boundary 2	235	tccagcagggtaggactgtt	ttgcttagaattcgcgaggt
exon/intron boundary 3	237	caagcccatggaacatttct	ctgctttcaggaccgtaagg

A sensitive, multiplex RT-PCR (GeXP, A21019, A25295; Beckman) was used to investigate whether the quadruplex in the *estrogen receptor* exon C is expressed in bovine tissues *in vivo*. The GeXP system utilized a conversion from a reaction initiated with specific primers to PCR based on universal primers. The universal sequence flanked the 5′ end of each of the primers and added an additional 37 nucleotides to the length of each amplicon. The sequence of specific primers for all genes and size of amplicons given are in Table 1. To exclude the possibility that amplicons arise due to DNA contamination of RNA, additional duplex reactions were performed on RNA used in the study. As DNA contamination controls, pairs of primers were used for the bovine gonadotrophin releasing hormone receptor (GnRHR) promoter (167 nucleotides), as well as two pairs for exon C/intron boundaries (235 and 237 nucleotides). Each of these primer pairs was used in a separate reaction spiked with kanamycin resistance gene mRNA. The sequences of exon/intron boundary primers are given in Table 1.

The primers included in the multiplex PCR were designed to cover the potential quadruplex-forming region of exon C (Qx), exon C fragment 320 nt downstream of Qx (ExC), histone 2A (H2A) as a house keeping gene, ESR1 and IGF1 as RNA condition controls, and the GnRHR promoter as a DNA contamination control. Kanamycin resistance gene mRNA added to the reaction acted as an internal positive control for the system. DNase-treated RNA (60 ng) was used in every reaction according to the manufacturer's protocols. Products were run on capillary electrophoresis and analyzed using Beckman SEQ and Profiler software. The amplicons of the *ESR1* promoter and exon C were reamplified in PCR and sequenced.

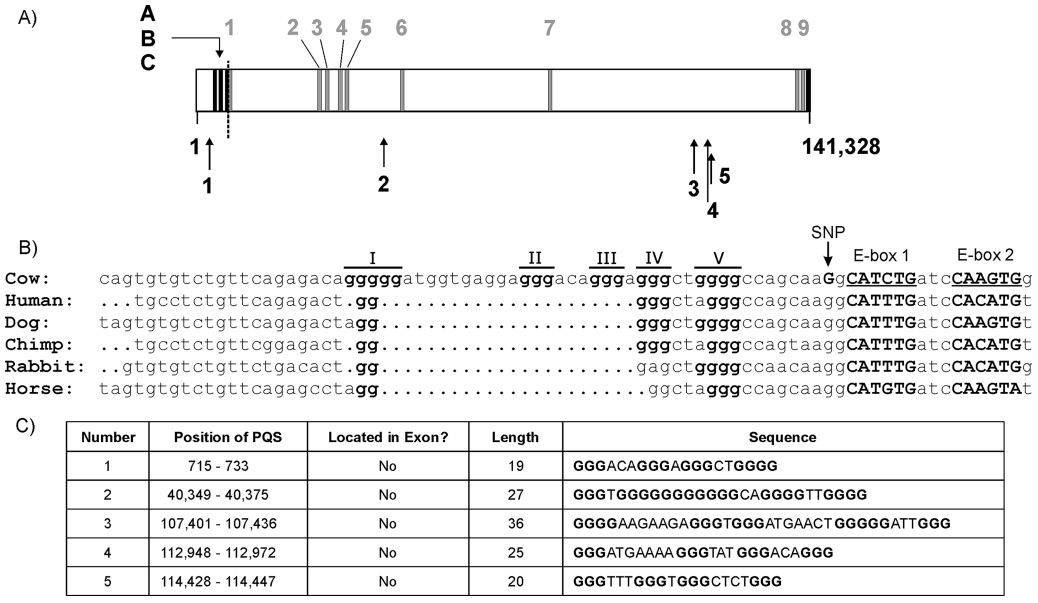
DNA Constructs for Studies of Rates of Translation. Complementary DNA primers containing the T7 promoter site, with and without the quadruplex sequence (±Q), G or A single nucleotide polymorphism, and double E-box motif were supplied by Sigma. After annealing, the double-stranded DNA fragments were cloned into pGL3 vectors Basic and Control, at *Hin*dIII/*Nco*I sites, upstream of the luciferase gene. Success of cloning was confirmed by sequencing. The last three guanine residues in the sequence of the T7 site (taatacgactcactataggg) were incorporated into the transcript to become part of the first five Gs present in the bovine ERα insert. The sequence of primers was as follows: *Hin*dII site, T7 promoter, ±Qx, SNP (shown in bold capital letters), the double E-box (sequences are underlined), and *Nco*I site; in *Mut*Q the ggg to aaa mutations within the quadruplex-forming region are underlined and italicized. +Q sequence: AAAaagcttTAATACGACTCACTATAGGGgatggtgaggagggacaggagggctggggccagcaa**G**gcatctgatccaagtgggatcccagaaggctCCATGGaa. −Q sequence: AAAaagcttTAATACGACTCACTATAGGctggggccagcaa**G**gcatctgatccaagtgggatcccagaaggctCCATGGaa. *Mut*Q: AAAaagcttTAATACGACTCACTATAGGGgatggtgaggagggacaaaaagggctggggccagcaa**G**gcatctgatccaagtgggatcccagaaggctCCATGGaa.

In Vitro Transcription and Translation in a Cell-Free System. Plasmid DNA with T7 promoter with or without the quadruplex or mutation/SNP/E-box was linearized (at the *Sal*I site), and 1 μg of each construct was used for *in vitro* transcription using the mMessage mMachime T7 ultra kit (Ambion) according to the manufacturer's instructions and as described in Kumari et al. (*21*) The RNA produced using the T7 promoter was capped and poly(A) tailed and used in the *in vitro* translation procedure after confirming RNA integrity and size on 1% agarose gels as suggested in Ambion's instructions. One microgram of the transcripts with incorporated regulatory sequences containing ±quadruplex or mutation/SNP/E-box upstream of the luciferase gene was denatured and used as a template in the *in vitro* translation reaction with a rabbit reticulocyte lysate extract (Promega), according to the manufacturer's protocol. The outcome of the translation reaction (2.5 μL of cell lysate) containing luciferase reporter protein was measured using a MicroLumenotPlus LB 96 V (Berthold Technologies) with a Biotium kit.

In Vivo Translation in BST Cells. *In vitro* transcription/translation was repeated with cells in culture. Bovine endometrial stromal (BST) cells were used because they express the estrogen receptor and have a complement of transcription factors which interact with the estrogen receptor promoter. Cells isolated on day 16 of the estrous cycle were maintained as described previously (*22*). Cells were harvested 68 h after electroporation with the appropriate plasmid construct (with or without the quadruplex or mutation/SNP/E-box upstream of the luciferase gene). The cells were lysed with lysis buffer (Biotium), and 20 μL of cell lysate was used for luciferase translation assessment as described above.

RESULTS AND DISCUSSION

Quadruplex-Forming Sequences in the bov-ESR1 Gene. A sequence alignment of the ESR1 gene and its promoter sequence across a number of species demonstrated the presence of a unique 23-nucleotide G-rich insert immediately upstream of the E-box transcription factor binding site of *bov*-ESR1 (Figure 1). The biological relevance of this extra motif within the bovine sequence is unknown; however, the presence of additional guanine runs leads to the unique possibility of formation of an intramolecular quadruplex structure. Whether other potential quadruplex-forming sequences are present within the *bov*-ESR1 gene was also addressed. A bioinformatics search, using the quadparser program (*10*), identified the presence of five possible quadruplex-forming motifs distributed throughout the ESR1 gene and promoter sequences (Figure 1A). This corresponded



Number	Position of PQS	Located in Exon?	Length	Sequence
1	715 - 733	No	19	GGGACAGGGAGGGCTGGGG
2	40,349 - 40,375	No	27	GGGTGGGGGGGGGGCAGGGGTTGGGG
3	107,401 - 107,436	No	36	GGGGAAGAAGAGGGTGGGATGAACTGGGGGATTGGG
4	112,948 - 112,972	No	25	GGGATGAAAAGGGTATGGGACAGGG
5	114,428 - 114,447	No	20	GGGTTTGGGTGGGCTCTGGG

FIGURE 1: (A) The bovine ESR1 gene. The gene spans 141328 nucleotides encompassing nine coding exons (1−9) and the three known noncoding 5′ UTR exons (A, B, and C). The positions of the five putative quadruplex-forming sequences are shown by the arrows. (B) A section of the ESR1 promoter sequence showing alignment across a number of different species; the G-rich tracts (I−V) are marked on the bovine sequence. The E-box motif (underlined) is conserved across these species, as is the G-rich section immediately prior to it (bold). The bovine gene includes a G-rich insert of 22 nucleotides not present in the other species listed, which has the potential to adopt an intramolecular quadruplex motif. The BLAST sequence alignment was performed using the ESR1 gene information for *Bos taurus* (cow), *Homo sapiens* (human), *Canis familiariz* (dog), *Pan troglodytes* (chimpanzee), *Oryctolagus cuniculus* (rabbit), and *Equus caballus* (horse). (C) Positions and sequences of the five putative quadruplex-forming sequences (PQS) in the bovine gene.

DNA sequences

I II III IV V

$QS1_{1-36}$ GGGGGATGGTGAGGAGGGACAGGGAGGGCTGGGGCC
$QS1_{6-36}$ ATGGTGAGGAGGGACAGGGAGGGCTGGGGCC
$QS1_{10-36}$ TGAGGAGGGACAGGGAGGGCTGGGGCC
$QS1_{15-36}$ AGGGACAGGGAGGGCTGGGGCC
$QS1_{1-30}$ GGGGGATGGTGAGGAGGGACAGGGAGGGCT
$QS1_{1-25}$ GGGGGATGGTGAGGAGGGACAGGGA
$S1_{human}$ GGGGGCTAGGGCCAGTAAGGCATTTGATCCACATGT

RNA sequences

QSR_{10-36} UGAGGAGGGACAGGGAGGGCUGGGGCC
QSR^{MutQ}_{10-36} UGAGGAGGGACAAAAAGGGCUGGGGCC
QSR^{MutQ}_{15-36} AGGGACAAAAAGGGCUGGGGCC

FIGURE 2: Family of DNA and RNA sequences derived from QS1 identified in the bovine ESR1 gene and used in biophysical studies of G-quadruplex folding and stability. The five G-tracts (I−V) are highlighted above the $QS1_{1-36}$ sequence.

to ~1 quadruplex motif every 28000 bases. The promoter region was arbitrarily defined to be 5000 nucleotides to incorporate the three known 5′-UTR exons. None of the quadruplex sequences that were identified were found in any of the previously identified exons (Figure 1C).

We subsequently used biophysical techniques to probe the folding of a family of synthetic oligonucleotides derived from the *bov*-ESR1 gene sequence (Figure 2). DNA quadruplex structures exhibit characteristic circular dichroism (CD) and ultraviolet (UV) spectra that are dependent on structural topology. Parallel-stranded folds exhibit a maximum positive ellipticity around 260 nm with a corresponding negative intensity around 240 nm. Antiparallel-stranded quadruplexes show a characteristic maximum around 295 nm with a negative signal around 260 nm. Upon thermal unfolding to form unstructured single strands, the CD signal weakens considerably while a significant hyperchromic absorbance shift is identified at 295 nm in the UV spectrum (*23*−*26*). We have used both of these spectroscopic signatures to identify and characterize quadruplex formation and stability.

We first examined a 36-mer ($QS1_{1-36}$) incorporating the entire bovine G-rich insert (Figure 2). This sequence has 23 guanine bases including five runs (I−V) of three or more consecutive Gs. The CD spectrum displays clear evidence for quadruplex formation with a maximum around 260 nm, a minimum at 240 nm, and also a smaller maximum arising at 295 nm (Figure 3a). The spectrum suggests a major population of parallel-stranded quadruplex species in equilibrium with a minor population of antiparallel or hybrid species (*26*, *27*). The presence of a quadruplex structure was also verified by temperature-dependent changes in the UV absorption spectrum. The high- and low-temperature UV spectra show a well-defined hyperchromic shift at 295 nm (Figure 3b), and a clear melting transition is observed with a midpoint at 60 °C (Figure 3c). Cooling the structure leads to melting and annealing curves with minimal hystersis.

The presence of five G-rich tracts within the *bov*-ESR1 sequence lends itself to a number of possible combinations of guanines that could lead to a mixture of quadruplex structures with a range of different loop lengths and folded topologies. The presence of a single nucleotide loop and other short loops is likely to favor parallel quadruplex structures, giving the intense band at 260 nm in the CD spectrum. However, not all reported quadruplex structures follow the rules of four runs of consecutive guanines separated by three loop regions; the human c-kit promoter sequence is a particular example (*28*).

5′- and 3′-Truncated Analogues Form Quadruplex Structures in Solution. In order to probe which guanines contribute to tetrad formation, a number of truncated sequences were designed to limit the number of possibilities (Figure 2). Initially, $QS1_{1-36}$ was truncated sequentially from the 5′ end by removing G-tract I

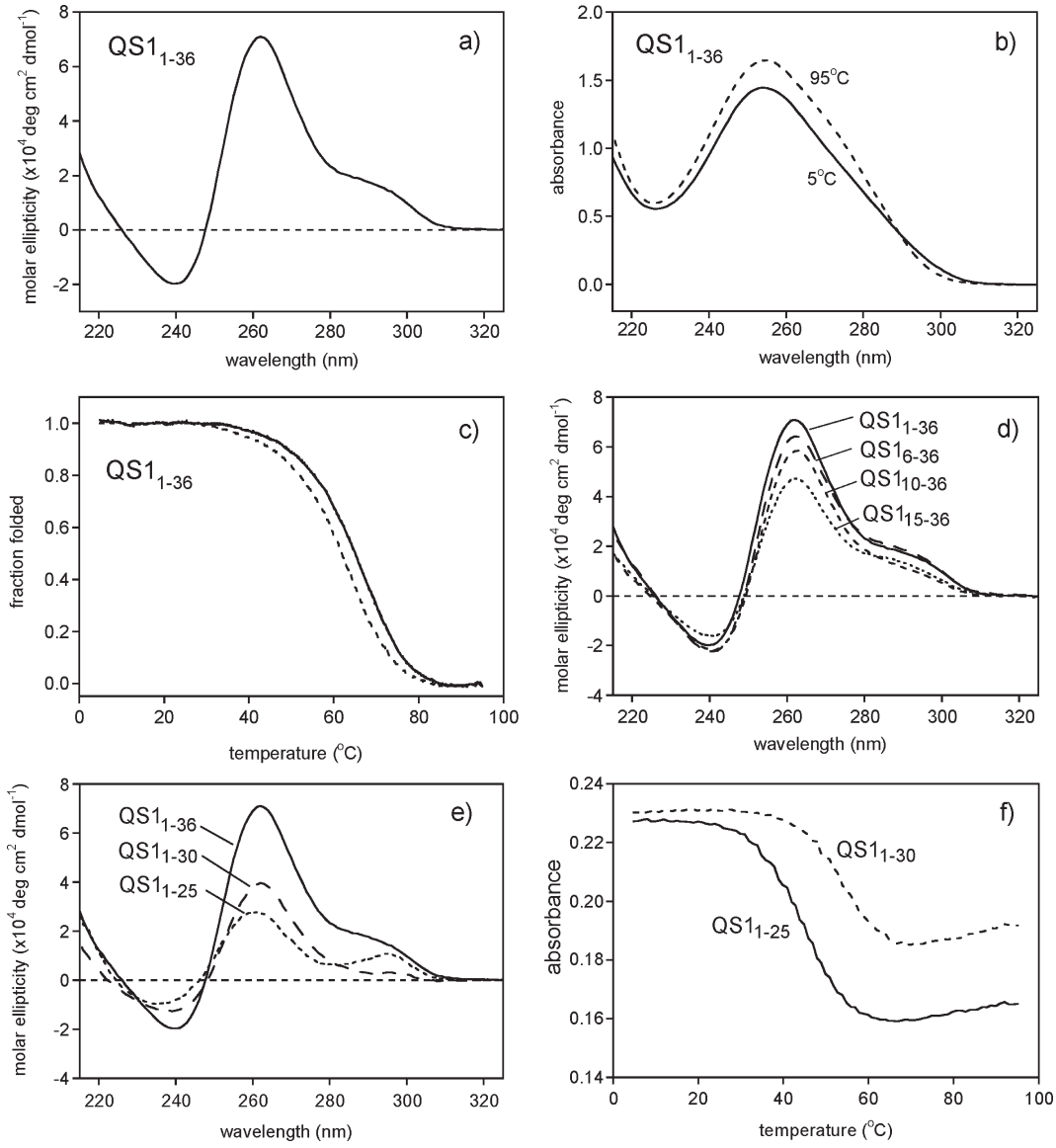


FIGURE 3: (a) The CD spectrum of $QS1_{1-36}$ demonstrating the presence of an intense band at 260 nm, characteristic of a parallel quadruplex structure, with the possibility of a minor antiparallel form as evident from the shoulder at 290 nm. Samples of 6 μM concentration were used containing 100 mM KCl and 10 mM potassium phosphate, pH 7. (b) UV spectra at high (95 °C) and low (25 °C) temperature highlighting the hyperchromic shift around 295 nm. (c) Normalized UV melting and annealing curves, as monitored at 295 nm, showing a small amount of hysteresis indicative of slow refolding kinetics. (d) Corresponding far-UV CD spectra for $QS1_{1-36}$ and the 5′-truncated sequences $QS1_{6-36}$, $QS1_{10-36}$, and $QS1_{15-36}$ (see Figure 2). CD spectra for sequences truncated from the 3′ end of $QS1_{1-36}$ to remove guanine bases in tracts V and IV ($QS1_{1-25}$ and $QS1_{1-30}$) are shown in (e) and UV melting curves (absorbance at 295 nm) in (f).

and the following loop sequence. These sequences all generated CD spectra with similar spectroscopic characteristics to the full-length sequence (Figure 3d). This suggested that the four guanine tracts on the 3′-terminal end of the sequence (II−V) contribute most significantly to quadruplex formation in the full-length sequence. In each case, quadruplex formation was confirmed by a similar hyperchromic shift in the UV spectrum around 295 nm, with only small differences in melting temperatures (data not shown).

An analogous set of sequences were systematically truncated from the 3′-side (removing the closely spaced tracts V and IV) (Figure 3e,f), and these sequences also demonstrated intramolecular quadruplex formation. However, the reduction in signal intensity in the CD spectrum, together with changes in the absorption envelope and lower stabilities, suggested that the 3′-truncations had a greater influence on the quadruplex structure. The shortest sequence ($QS1_{1-25}$; tracts I−III) contains only three tracts of three consecutive guanine bases but an extended loop of 10 nucleotides containing five additional guanine bases. Although the CD signal is weaker than the others at 260 nm, the UV melting curve shows a well-defined structural transition with a $T_m \sim 45$ °C (Figure 3f), indicating that the sequence is able to fold but must adopt a nonstandard topology involving noncontiguous guanines within the first loop sequences (Figure 2).

Thus, on the basis of the CD studies of the truncated sequences of *bov*-ESR1, we conclude that a number of possible quadruplex structures of $QS1_{1-36}$ can form from this single-stranded sequence. However, the four G-tracts (II−V) within the 3′-terminal part of the sequence lead to the most stable fold with loop lengths of three, one, and two nucleotides, respectively, resulting in a predominantly parallel structure. QS1, located immediately upstream of the E-box motif of the *bov*-ESR1 exon C sequence, is evidently predisposed to form quadruplex structures *in vitro*.

The alignment of the bovine and human gene sequences upstream of the E-box motifs (Figure 1) demonstrates that the

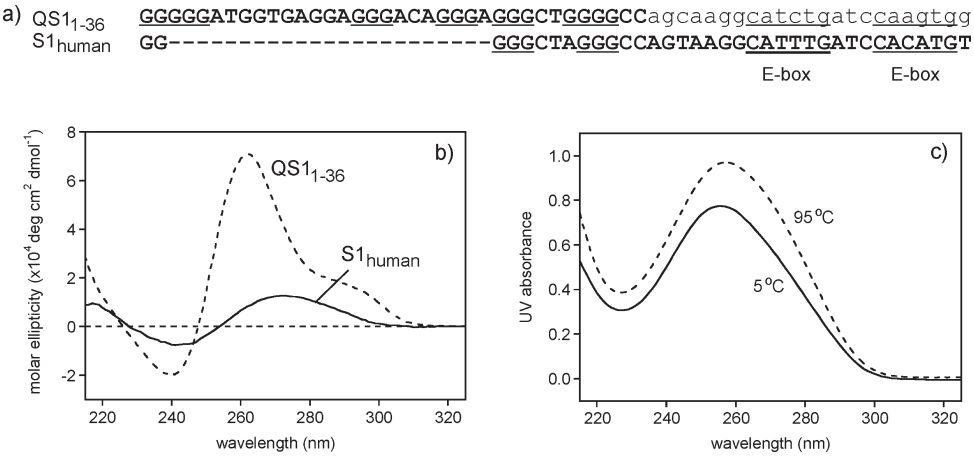


FIGURE 4: (a) Sequence of $QS1_{1-36}$ from bovine ESR1 compared with a portion of the human sequence ($S1_{human}$) from the same region of the gene but lacking the G-rich insert. Guanines have been highlighted in bold but do not conform to the required pattern for quadruplex formation. (b) The far-UV CD spectra at 25 °C comparing $QS1_{1-36}$ and $S1_{human}$ show that the latter does not give the characteristic band at 260 nm indicative of quadruplex structure. (c) UV absorbance spectra for S1-human at 5 and at 95 °C show no evidence for a hyperchromic shift at 295 nm, indicating that $S1_{human}$ is largely unstructured.

quadruplex-forming motif is unique to the *bov*-ESR1 gene. As a control, the equivalent human sequence ($S1_{human}$) was subjected to the same CD and UV analysis (Figure 4). $S1_{human}$ does not form a quadruplex structure as evident from the weak broad CD spectrum and lack of characteristic UV hyperchromic absorbance shift. Therefore, the G-rich insert found in *bov*-ESR1 uniquely bestows the ability to form a quadruplex structure in solution.

The Role of the SNP in Quadruplex Formation. The G/A SNP located between the E-box motif and the quadruplex-forming sequence was similarly investigated by CD and UV analysis. Its close proximity to both the E-box and the 3′ end of the quadruplex-forming sequence suggests that the SNP is more likely to play a role in E-box recognition rather than quadruplex assembly. CD experiments on sequences with the G/A polymorphism show only minor differences and confirm that the SNP does not perturb quadruplex formation and stability.

CD Studies of Quadruplex Formation in Synthetic mRNA Transcript. CD and UV studies were extended to the analysis of RNA folding by constructing a synthetic RNA sequence equivalent to that transcribed from $QS1_{10-36}$ (Figure 2). Biophysical studies of $QS1_{10-36}$ (DNA) and QSR_{10-36} (RNA) sequences revealed a similar far-UV CD spectrum dominated by a strong band at 260 nm indicative of analogous parallel-stranded structure (Figure 5a). CD melting studies reveal that the RNA quadruplex has a midpoint transition at 88 °C (Figure 5c), which is some 20 °C higher in stability than the corresponding DNA sequence (*21*), although the full RNA melting curve is not as well-defined because of the high thermal stability. The UV absorbance spectra recorded at 25 and 95 °C show the characteristic hyperchromic shift around 295 nm (Figure 5b) and a thermal difference spectrum (Figure 5d) with a double maximum with peaks at 250 and 280 nm, all of which are evidence for formation of a highly stable parallel-stranded RNA quadruplex structure within the *bov*-ESR1 mRNA transcript.

As controls, we also characterized two mutant sequences, $QSR_{10-36}{}^{MutQ}$ and $QSR_{15-36}{}^{MutQ}$, in which GGG in tract III was replaced by AAA to disrupt RNA quadruplex formation (Figure 2). CD analysis reveals that $QSR_{10-36}{}^{MutQ}$ is still able to form folded structure but with reduced stability ($T_m \sim 70$ °C) (Figure 5). We attribute this to the presence of a group of nonconsecutive Gs at the 5′ end of the sequence that is also implicated in the folding of the $QS1_{1-25}$ DNA sequence (Figure 3d). Further truncation of the RNA sequence ($QSR_{15-36}{}^{MutQ}$) completely eliminates any evidence for stable structure. The CD melting curves for the two mutant RNA sequences are included in Figure 5b.

Expression of Exon C and the Quadruplex-Forming Region in Bovine Tissues. We addressed the question whether exon C is present in ESR1 transcripts in bovine tissues, as it is in human (*20*, *29*) and rodent (*20*). A multiplex RT-PCR showed that exon C transcripts were present in several reproductive tissues. Thus bovine fetal ovaries (week 21, 26, 31) and testes (week 17, 19, 35), as well as adult ovarian follicles and uterine tissues (endometrium and myometrium), exhibited the presence of exon C transcripts. Furthermore, a fragment containing the quadruplex-forming sequence, which is present in the bovine genome, is also present in these transcripts (Figure 6a). Its identity was confirmed by subsequent sequencing of the 137 nt product. Bovine kidney used as a negative control tissue did not express either of the fragments (Figure 6b,c), despite expressing the estrogen receptor at a low level. Quadruplex/exon C expression was also tested against the exon/intron boundary and GnRHR promoter, neither of which resulted in amplification (Figure 6c).

Regulation of Translational Efficiency by RNA Quadruplex Formation in Vitro. The potential influence of the *bov*-ESR1 G-quadruplex in regulating the efficiency of translation was evaluated *in vitro* by cloning the 5′ UTR of *bov*-ESR1 into the pGL3 vector upstream of the firefly luciferase reporter gene and 12 nucleotides downstream of the T7 promoter sequence, using the methodology previously described for an RNA quadruplex in the 5′ UTR of the *NRAS* proto-oncogene (*21*). We generated several control DNA sequences in addition to the native sequence which contained the G-quadruplex sequence (+Qx) and the G or A SNP polymorphism (+QxG and +QxA) or a mixture of the two (+QxG/A). Three controls had the 32 base pairs of the quadruplex-forming sequence removed but with the same combination of SNP substitutions (−QxG, −QxA, and −QxG/A), and finally, a mutant (*Mut*Q) was generated with the GGG to AAA substitutions described above for $QSR_{10-36}{}^{MutQ}$ and $QSR_{15-36}{}^{MutQ}$. The length of this sequence was identical to that

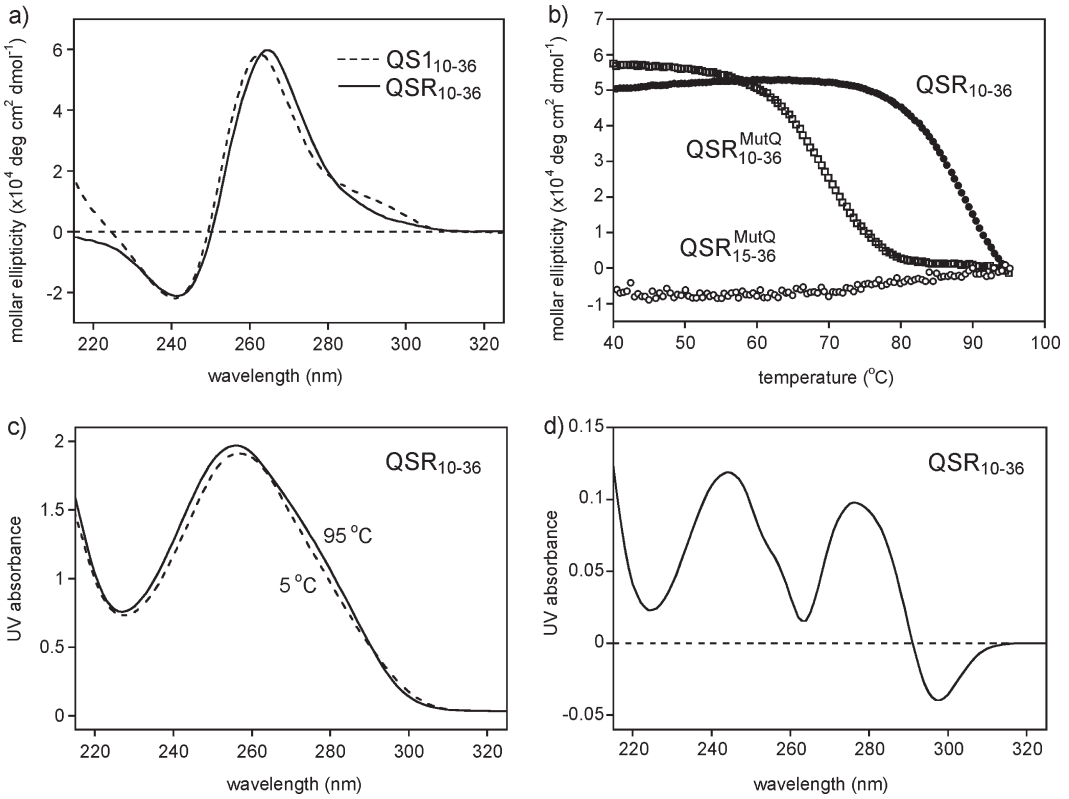


FIGURE 5: (a) Far-UV CD spectra of $QS1_{10-36}$ (DNA) and QSR_{10-36} (RNA) recorded at 25 °C (6 μM concentration containing 100 mM KCl and 10 mM potassium phosphate, pH 7) with both showing a characteristically strong band at 260 nm indicative of predominantly parallel quadruplex structures. (b) CD melting curves for QSR_{10-36} (T_m estimated to be 88 °C) and two mutant sequences (QSR_{10-36}^{MutQ} and QSR_{15-36}^{MutQ}) containing destabilizing GGG to AAA substitutions (see Figure 2). (c) UV spectra for QSR_{10-36} at high (95 °C) and low (25 °C) temperature highlighting the small hyperchromic shift around 295 nm. (d) UV difference spectrum illustrating the double maxima at 245 and 280 nm typical of a parallel fold and the hyperchromic shift that gives a negative difference around 295 nm.

of the natural 5′-UTR, with the mutations anticipated to destabilize the mRNA quadruplex. Using the luciferase reporter gene, the RNA transcript was generated by *in vitro* transcription using T7 RNA polymerase. The subsequent efficiency of RNA translation resulting from having the quadruplex motif so close to the T7 promoter sequence was evaluated on the basis of the catalytic activity of the luciferase enzyme estimated on the basis of its luminescent properties (*21*). The mutations described lead to significant changes in the relative translational efficiency (Figure 7a).

The G/A SNP alone resulted in no significant differences; however, complete removal of the quadruplex-forming insert, leading to a truncation of the native sequence, resulted in a 2-fold increase in translational efficiency, with a further 2-fold increase realized by introducing the G/A SNP simultaneously (+QxG versus −QxA). However, substitution of GGG for AAA in *Mut*Q led to a much larger effect, enhancing translational efficiency by 15-fold compared to the native sequence. The observations indicate that RNA quadruplex formation suppresses translational efficiency; however, the largest effects occur in the context of the full-length native sequence, rather than in truncated sequences in which the insert has been removed. These observations are consistent with those recently reported for the 5′ UTR of the *NRAS* proto-oncogene (*21*), where enhanced translational efficiency (<4-fold increase) was detected as a consequence of completely deleting the quadruplex-forming sequence and from introducing a similar GGG to AAA substitution (*Mut*Q) and maintaining sequence length. Subsequent studies suggested that the proximity of the quadruplex motif to the T7 promoter sequence also significantly influences translational efficiency, with the effects falling off with distance (*14*).

Modulation of Translation in Cells in Culture. To investigate whether the effect of the quadruplex-forming motif on translation of exon C transcripts occurred in whole cells, BST cells were transfected with DNA constructs comprising various sequences derived from exon C upstream of a luciferase reporter gene. On measuring reporter gene product in these cells, there was a >15-fold increase when the quadruplex-forming motif was removed and 7-fold increase with the construct containing the mutated quadruplex sequence (Figure 7b). These observations were consistent with the effect on transcription shown in the cell-free system.

Conclusions. The bovine estrogen receptor α gene (*bov*-ESR1) possesses a G-rich insert not found in other mammals, lying close to a double E-box transcription factor binding site for MyoD. We have shown that this 22 base insert bestows the ability of this region of the *bov*-ESR1 to form a G-quadruplex structure within the single-stranded gene sequence. *In vitro* biophysical studies on a series of 5′- and 3′-truncated sequences that accommodate different G-tracts from the five available have indicated that a number of quadruplex structures with different stabilities could be assembled *in vivo*. The equivalent human control sequence (lacking the G-rich insert) fails to achieve this. However, G-quadruplex formation was not affected by the G/A mutation of the polymorphism.

We have demonstrated by multiplex PCR that the G-quadruplex-forming region is transcribed. PCR primers were designed to amplify the exon C sequence, and the products were analyzed

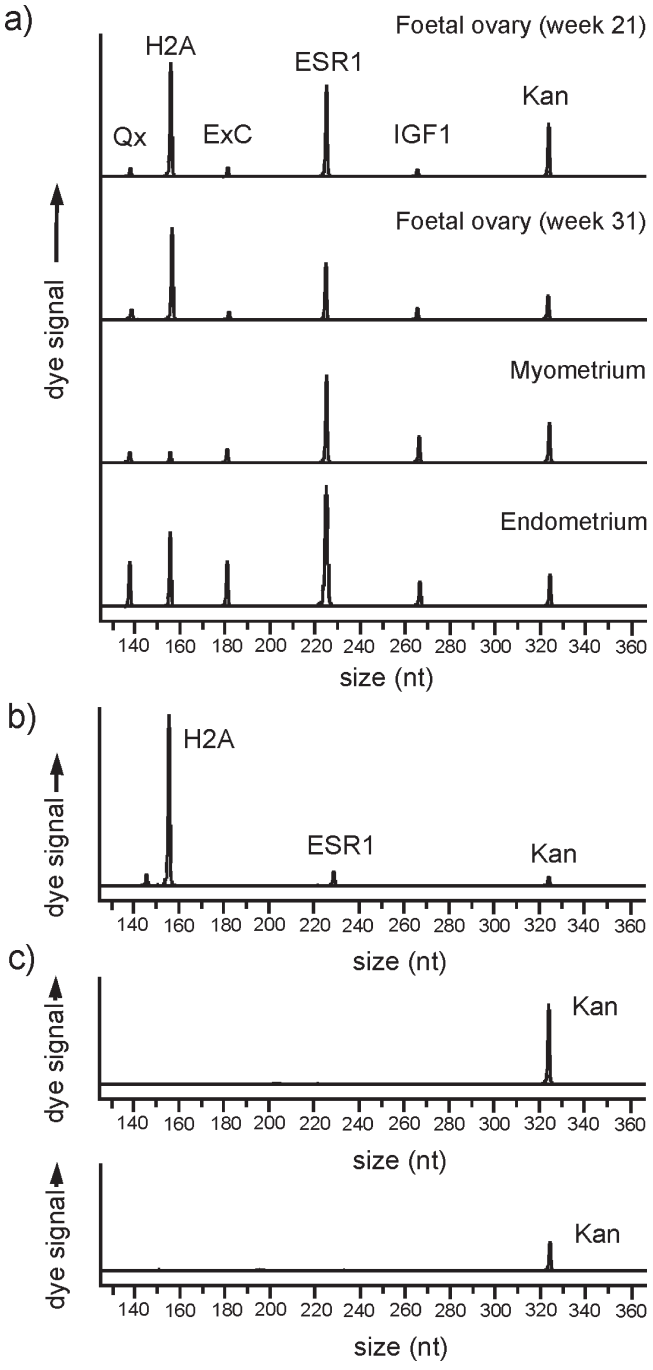


FIGURE 6: Expression of exon C in bovine tissues. Multiplex RT-PCR reactions investigating transcription of the quadruplex-forming region of the bovine ESR1 promoter: (a) expression of exon C and quadruplex-forming region in bovine fetal (21 and 31 weeks) and adult tissues (myometrium and endometrium). (b) The same multiplex reaction performed with bovine kidney RNA (the 144 nt peak is a nonspecific product). (c) To exclude the possibility that amplicons arise due to DNA contamination of RNA, additional duplex reactions were performed on RNA used in the study. As DNA contamination controls, pairs of primers were used for the bovine gonadotrophin releasing hormone receptor (GnRHR) promoter (167 nucleotides), as well as two pairs for exon C/intron boundaries (235 and 237 nucleotides). Each of these primer pairs was used in a separate reaction spiked with kanamycin resistance gene mRNA. The sequences of exon/intron boundary primers are given in Table 1.

by capillary electrophoresis. GeXP PCR experiments identified an amplified 137 base fragment which indicated that the quadruplex-forming sequence is incorporated, at low levels, into the transcript along with exon C.

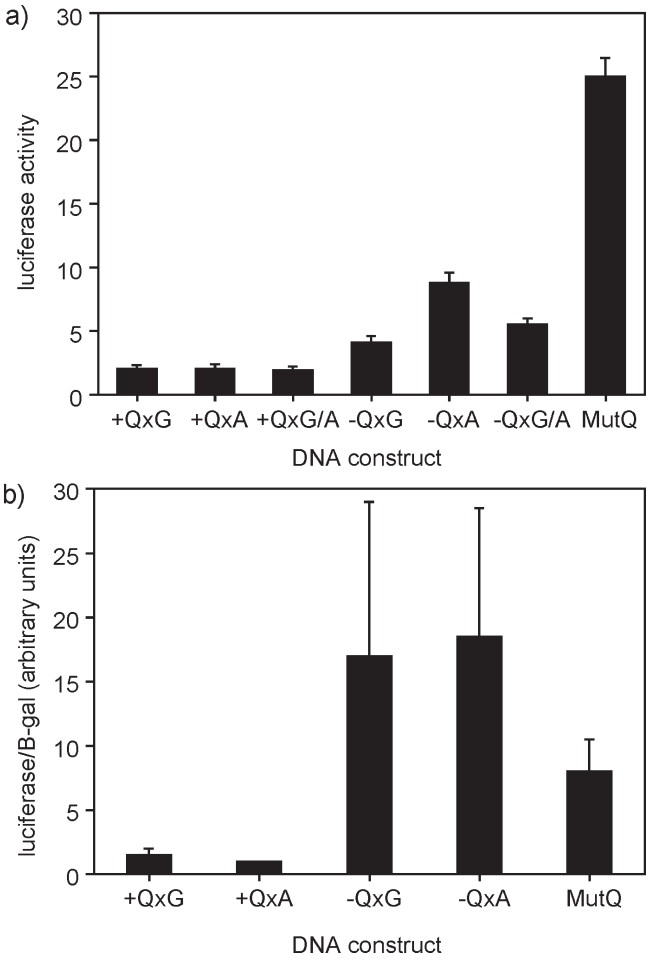


FIGURE 7: (a) Translational efficiency assayed *in vitro* investigating the role of the SNP and/or quadruplex formation on levels of RNA translation using a luciferase reporter gene containing *bov*-ESR1 inserts cloned downstream of a T7 promoter sequence for RNA polymerase. The putative quadruplex-forming sequence is 12 nucleotides from the T7 promoter site. +Qx and −Qx indicate inclusion or absence of the 25 nucleotide quadruplex sequence insert, respectively, and QxA or QxG is with either the A or G polymorphism (or G/A, representing a mixture of both). MutQ contains the GGG to AAA mutation blocking quadruplex formation. (b) Levels of protein reporter product (luciferase) measured in BST cells following transfection with luciferase reporter constructs in which the quadruplex-forming sequence was either present (+Qx), absent (−Qx), or mutated to prevent qudruplex formation (MutQ). In addition, the SNP was present as either allele (A or G). Luciferase levels are expressed in terms of the levels of β-galactosidase expressed from a cotransfected control plasmid.

In vitro translation assays and whole cell measurements of reporter protein showed that the quadruplex-forming sequence reduced translational efficiency. In BST cells, luciferase protein level, reflecting rates of both translation and transcription, was increased both by removal of the G-quadruplex and by its G/A mutation. An effect on translation rate was confirmed in studies utilizing *in vitro* translation, and these experiments also suggested that in the absence of the G-quadruplex sequence the G/A SNP mutation may also enhance translation. These differential effects on transcription and translation show that in cell culture both processes are affected by G-quadruplex formation. This is consistent with the observed formation of a highly stable RNA quadruplex ($T_m \sim 88$ °C) and with previous observations on other transcripts (*21*, *29*).

The biological role of the insert in *bov*-ESR1, which is absent at this site from other mammalian ESR genes, is presumed to be related to its effect on translation rate. The presence of similar quadruplex-forming sequences in the ESR in other species, albeit at other sites (*29*), supports the conclusion that these structures have functional significance. One possible role is to provide an additional level of estrogen receptor protein production through modulation of transcript stability/availability, as a result of interactions with chaperone molecules (*30*). This may extend the period for which transcripts are available for translation, although they are translated at a slower rate.

ACKNOWLEDGMENT

We acknowledge the help of L. Sheldrick with the BST cell culture work.

REFERENCES

1. Deroo, B. J., and Korach, K. S. (2006) Estrogen receptors and human disease. *J. Clin. Invest. 116*, 561–570.
2. O'Lone, R., Frith, M. C., Karlsson, E. K., and Hansen, U. (2004) Genomic targets of nuclear estrogen receptors. *Mol. Endocrinol. 18*, 1859–1875.
3. 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.
4. Yafe, A., Etzioni, S., Weisman-Shomer, P., and Fry, M. (2005) Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Res. 33*, 2887–2900.
5. Etzioni, S., Yafe, A., Khateb, S., Weisman-Shomer, P., Bengal, E., and Fry, M. (2005) Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes. *J. Biol. Chem. 280*, 26805–26812.
6. Shklover, J., Weisman-Shomer, P., Yafe, A., and Fry, M. (2010) Quadruplex structures of muscle gene promoter sequences enhance in vivo MyoD-dependent gene expression. *Nucleic Acids Res.* (Advance Access, Jan 6, 2010, pp 1−9).
7. Dexheimer, T. S., Sun, D., and Hurley, L. H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl- 2 P1 promoter. *J. Am. Chem. Soc. 128*, 5404–5415.
8. Connor, A. C., Frederick, K. A., Morgan, E. J., and McGown, L. B. (2006) Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide. *J. Am. Chem. Soc. 128*, 4986–4991.
9. Szreder, T, and Zwierzchowski, L. (2004) Polymorphism within the bovine estrogen receptor-alpha gene 5′-region. *J. Appl. Genet. 45*, 225–236.
10. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res. 33*, 2908–2916.
11. Fletcher, T., Sun, D., Salazar, M., and Hurley, L. H. (1998) Effect of DNA secondary structure on human telomerase activity. *Biochemistry 37*, 5536–5541.
12. Mergny, J.-L., Riou, J., Mailliet, P., Teulade-Fichou, M., and Gilson, E. (2002) Natural and pharmacological regulation of telomerase. *Nucleic Acids Res. 30*, 839–865.
13. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A. 99*, 11593–11598.
14. Kumari, S., Bugaut, A., and Balasubramanian, S. (2008) Position and stability are determining factors for translation repression by an RNA G-quadruplex sequence within the 5′ UTR of the *NRAS* proto-oncogene. *Biochemistry 47*, 12664–12669.
15. Ambrus, A., Chen, D., Dai, J., Jones, R. A., and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry 44*, 2048–2058.
16. Fernando, H., Reszka, A. P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A. R., Neidle, S., and Balasubramanian, S. (2006) A conserved quadruplex motif in a transcription activation site of the human c-kit oncogene. *Biochemistry 45*, 7854–7860.
17. 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.
18. 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 1alpha promoter. *Biochemistry 44*, 16341–16350.
19. Dexheimer, T. S., Fry, M., and Hurley, L. H. (2006) in Quadruplex Nucleic Acids (Neidle, S., and Balasubramanian, S., Eds.) pp 180−207, The Royal Society of Chemistry, Cambridge, U.K.
20. Kos, M., Reid, G., Denger, S., and Gannon, F. (2001) Minireview: genomic organization of the human ERalpha gene promoter region. *Mol. Endocrinol. 15*, 2057–2063.
21. 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.
22. Sheldrick, E. L. R., Derecka, K., Marshall, E., Chin, E. C., Hodges, L., Wathes, D. C., Abayasekara, D. R. E., and Flint, A. P. F. (2007) Peroxisome-proliferator-activated receptors and the control of levels of prostaglandin-endoperoxide synthase 2 by arachidonic acid in the bovine uterus. *Biochem. J. 406*, 175–183.
23. Mergny, J.-L., Phan, A. T., and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett. 435*, 74–78.
24. Ambrus, A., Chen, D., Dai, J., Bialis, B., Jones, R. A., and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res. 34*, 2723–2735.
25. Xue, Y., Kan, Z. Y., Wang, Q., 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*, 11185–11191.
26. 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. Biosystems 5*, 542–547.
27. Phan, A. T., Kuryavyi, V., Ngoc, K., and Patel, D. J. (2007) in Quadruplex Nucleic Acids (Neidle, S., and Balasubramanian, S., Eds.) pp 81−99, Royal Society of Chemistry Publishing, Cambridge, U.K.
28. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc. 129*, 4386–4392.
29. Balkwill, G. D., Derecka, K., Garner, T. P., Hodgman, C., Flint, A. P. F., and Searle, M. S. (2009) Repression of translation of human estrogen receptor a by G-quadruplex formation. *Biochemistry 48*, 11487–11495.
30. De Cian, A, and Mergny, J.-L. (2007) Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. *Nucleic Acids Res. 35*, 2483–2493.