*7290–7301* *Nucleic Acids Research, 2009, Vol. 37, No. 21* *Published online 26 September 2009*
*doi:10.1093/nar/gkp793*

# The HSV-1 ICP27 RGG box specifically binds flexible, GC-rich sequences but not G-quartet structures

**Kara A. Corbin-Lickfett[1], I-Hsiung Brandon Chen[1], Melanie J. Cocco[2,*] and Rozanne M. Sandri-Goldin[1,*]**

[1]Department of Microbiology and Molecular Genetics and [2]Department of Molecular Biology and Biochemistry, University of California Irvine, Irvine, CA 92697, USA

Received June 22, 2009; Revised September 8, 2009; Accepted September 9, 2009

## ABSTRACT

**Herpes simplex virus 1 (HSV-1) protein ICP27, an important regulator for viral gene expression, directly recognizes and exports viral RNA through an N-terminal RGG box RNA binding motif, which is necessary and sufficient for RNA binding. An ICP27 N-terminal peptide, including the RGG box RNA binding motif, was expressed and its binding specificity was analyzed using EMSA and SELEX. DNA oligonucleotides corresponding to HSV-1 glycoprotein C (gC) mRNA, identified in a yeast three-hybrid analysis, were screened for binding to the ICP27 N-terminal peptide in EMSA experiments. The ICP27 N-terminus was able to bind most gC substrates. Notably, the ICP27 RGG box was unable to bind G-quartet structures recognized by the RGG domains of other proteins. SELEX analysis identified GC-rich RNA sequences as a common feature of recognition. NMR analysis of SELEX and gC sequences revealed that sequences able to bind to ICP27 did not form secondary structures and conversely, sequences that were not able to bind to ICP27 gave spectra consistent with base-pairing. Therefore, the ICP27 RGG box is unique in its recognition of nucleic acid sequences compared to other RGG box proteins; it prefers flexible, GC-rich substrates that do not form stable secondary structures.**

## INTRODUCTION

The Herpes Simplex Virus type I (HSV-1) infected cell protein 27 (ICP27) is an essential, multifunctional protein that is expressed immediately after infection. ICP27 contributes to the shut off of host cell gene expression and interacts with a myriad of viral and cellular proteins and viral mRNA to promote the expression of viral gene products at the transcriptional, posttranscriptional and translational levels. Early during infection, ICP27 inhibits host cell splicing by mediating the stalling of spliceosomal complexes (1) and recruits cellular RNA polymerase II to sites of viral transcription (2). Later in infection, ICP27 functions as a shuttling protein and directly promotes the export of viral intronless mRNA by binding viral RNA in the nucleus (3,4) and facilitating its export to the cytoplasm via interactions with the nuclear export factor Aly/REF (5–7) and the nuclear export receptor TAP/NFX1(5). ICP27 has also been shown to promote translation of a subset of transcripts in the cytoplasm, possibly through its interaction with translation initiation factors (8,9). The regions of ICP27 required for important protein–protein and protein–RNA interactions have been mapped to multiple functional domains within the N- and C-termini of ICP27 by mutational analysis (Figure 1). The N-terminus of ICP27 contains nuclear import and export signals and an RGG box RNA binding motif, which is required for binding viral RNA (3,10) and the RNA export factor Aly/REF. The C-terminus of ICP27 contains three predicted hnRNP K-homology (KH) RNA binding domains (11); however, these domains have not been shown to be required for RNA binding. ICP27 binds zinc and there is a zinc finger-like domain in the C-terminus (12), a region involved in many protein–protein interactions.

RGG box nucleic acid binding motifs are found in a number of cellular DNA and RNA binding proteins and are defined as arginine- and glycine-rich sequences that usually contain closely spaced RGG repeats interspersed with aromatic residues (13). A well-studied RNA binding RGG box containing protein is the Fragile X Mental Retardation Protein (FMRP). In addition to its RGG box RNA binding motif, FMRP has two hnRNP K-homology (KH) RNA binding domains and a novel N-terminal domain of FMRP (NDF) RNA binding motif in the N-terminus (14,15). Even though FMRP contains additional RNA binding motifs (NDF and KH domains) all mRNA binding so far has been mapped to

*To whom correspondence should be addressed. Tel: +1 949 856 7570; Fax: +1 949 824 8598; Email: rmsandri@uci.edu
Correspondence may also be addressed to Melanie J. Cocco. Fax: +1 949 824 8551; Email: mcocco@uci.edu

the RGG box (16), which is also true for ICP27 (3,10). RGG boxes were originally thought to mediate only nonspecific interactions with nucleic acids; however, it has been shown that structure of the FMRP RGG box peptide is flexible and unstructured but able to specifically recognize G-quartet RNA structures (16,17). G-quartets are complex tertiary structures that can readily form in DNA or RNAs containing three or more consecutive guanines. Four guanosine bases can interact via Hoogsteen base–base interactions to create a planar ring, which can stack upon each other vertically with additional G-quartet planes to form a higher order structure. The G-quartet structure is very stable *in vitro* and can coordinate monovalent ions, such as potassium and sodium, between and within the plane of the guanine tetrads (18–21). A G-quartet RNA structure was found in the FMRP RNA targets analyzed (17) and also appears to be recognized by proteins in G-rich sequences such as telomeres (22,23) promoter regions of the oncogenes Rb (24) and c-myc (25) and immunoglobulin switch regions (19).

The RNA binding specificity of the ICP27 RGG box *in vivo* to viral RNA has not been well defined. *In vitro*, ICP27 prefers poly (G) to poly (U) RNA homopolymers and the RGG box peptide alone can mediate RNA binding (10). ICP27 was shown to specifically bind seven HSV-1 intronless RNAs, but not an intron-containing RNA, during HSV-1 infection in UV-cross-linking experiments (3). The cross-linked RNAs belonged to all three kinetic classes of HSV-1 genes; immediate early (ICP4 and ICP27), early (ICP8 and TK) and late genes (glycoproteins C and D.) Additionally, the export of these RNAs to the cytoplasm was reduced in the absence of ICP27 and required the ICP27 RGG box (3). A yeast three-hybrid analysis identified 31 HSV-1 sense RNAs 35–225 nucleotides in length and mapped to 28 open reading frames, which interacted specifically with ICP27 (26). *In situ* hybridization experiments with a probe specific for the glycoprotein C (gC) RNA in cells infected with an ICP27 RGG box deletion mutant virus (d4-5) found that the glycoprotein C (gC) RNA is not efficiently exported from the nucleus, even when the ICP27 RGG box mutant protein is efficiently exported (27). Microarray analysis of nuclear and cytoplasmic fractions from cells infected with d4-5 confirmed that this effect was not specific to gC and that a majority of viral transcripts from all kinetic classes require an intact RGG box for export, further supporting the importance of the RGG box in facilitating the export of viral RNA (27). However, it is not clear how ICP27 specifically binds and exports HSV-1 intronless transcripts and whether there is a recognition motif on the substrate RNAs that provides a mechanism for preferential association with and export by ICP27. Using HSV-1 sequences identified in a yeast three-hybrid analysis and an *in vitro* selection of RNA substrates (SELEX), we found that an ICP27 N-terminal peptide, including the RGG box, does not recognize a particular nucleotide sequence or G-quartet structure, but prefers G/C-rich substrates that are flexible and do not form secondary or tertiary structures.

## MATERIALS AND METHODS

### Plasmids

The HSV-1 ICP27 N-terminal peptide was PCR amplified from a plasmid containing the full length KOS ICP27 gene with a 5′ primer that introduced an NdeI site for cloning (5′-dGCGCATATGGCGACTGACATTGATATG-3′) and a 3′ primer that introduced a BamHI site and a stop codon (5′-dGCGGGATCCTTACGACAAACCATCGG CGGC-3′). The 480-nt PCR product consisted of the N-terminal 160 amino acids of ICP27 and was cloned into the pET15b plasmid (Novagen) and verified by sequencing. This pET15b-ICP27 N-terminus construct has a N-terminal 6X His tag. The codon optimized ICP27 gene was obtained from Verdezyne, Inc. (formerly CODA genomics). PCR was used to amplify the sequence corresponding to the ICP27 N-terminal 160 amino acids with the 5′ primer that introduced an NdeI site (5′-dGGCCATATGGCGACCGACATCGACATG CTG-3′) and a 3′ primer that introduced an XhoI site and a C-terminal 6× His tag (5′-dGGTGCTCGAGTTA TTAGTGATGGTGATGGTGGTGAGACAGACCGT CGTCAGCAGCACC-3′) The PCR product was cloned into the pET21b plasmid (Novagen) and verified by sequencing.

### Protein expression and purification

pET plasmids expressing the His tagged ICP27 N-terminal 160 amino acids (wild-type or codon optimized) were transformed into BL21 Rosetta strain *E. coli* (Novagen). Cultures were grown in Luria Bertani (LB) broth containing 50 μg/ml Ampicillin to an OD 600 of 0.8–1.0 and induced with 100 mM IPTG (Sigma) for 3 h at 37°C. The His-tagged ICP27 N-terminal proteins were purified using Ni-NTA agarose (Qiagen) according to the manufacturer's recommendation for native protein purification. ICP27 N-terminus was dialyzed into 50 mM Tris buffer pH 8 and concentrated with a Centriprep centrifugal filter device (Millipore). The Coomassie Plus Bradford assay kit (Pierce) was used to determine protein concentration.

### Yeast three-hybrid analysis

The analysis was done as described in the user manual of the RNA–Protein Hybrid Hunter Kit Version A (Invitrogen). Specifically, the yeast strain provided, L40*ura*MS2, produces the protein hybrid of the LexA DNA binding domain fused to the bacteriophage MS2 coat protein. Full-length ICP27 was cloned in frame into the plasmid pYESTrp2 to generate a second hybrid protein consisting of ICP27 fused to the B42 activation domain (pYESTrp2/ICP27). The RNA library was generated by digesting the SalI fragment of the plasmid pGC Sal/Hind L (provided by Dr Edward Wagner), containing the HSV-1 open reading frames for the UL43 and UL44 (gC) genes (nucleotides 95 155–98 129), with XmaI. The XmaI digest yielded DNA fragments ranging from 50 to 500 bp in size that were inserted into the XmaI site of the plasmid pRH5′. This placed the UL43 and 44 genes downstream from the MS2 RNA sequence to generate

MS2-UL43/UL44 hybrid RNA molecules (pRH5′/UL43/UL44). The plasmids pYESTrp2/ICP27 and pRH5′/UL43/UL44 were co-transformed into L40*ura*MS2 and allowed to grow for 6 days at 30°C. A positive RNA–protein interaction was determined by growth on YC-UWH medium (Invitrogen) lacking uracil, tryptophan and histidine and supplemented with 5 mM 3-aminotriazole to reduce false-positive clones, and filter lift assays on colonies to detect β-galactosidase expression from positive RNA–protein interactions.

### Elecrophoretic mobility shift assay

Five pmoles of gC 30-mer DNA oligonucleotides, G-quartet DNA (dTAGGGGTT) or SELEX 20-mer DNA oligonucleotides (Operon, Inc.) were radiolabeled with γ-$^{32}$P ATP with Optikinase (USB) and purified using the Qiaquick nucleotide removal kit (Qiagen). For the RNA electrophoretic mobility shift assay (EMSA), RNA oligonucleotides gC 1–30 (5′-rCGCCGACCCUC CGUUGUAUUCUGUCACCGG-3′) and 31–60 (5′- rG CCGCUGCCGACCCAGCGGCUGAUUAUCGG-3′) (Integrated DNA Technologies) were radiolabeled and purified in the same manner as the DNA oligonucleotides Twenty fmoles of each radiolabeled G-quartet DNA, gC DNA or RNA oligonuleotide was incubated with 0.5–62.5 μM of purified His-tagged ICP27 N-terminus in 1× binding buffer (20 mM Tris pH 8, 150 mM KCl, 1 mM EDTA pH 8 and 1 mM DTT) with 10% glycerol and 300 μg/ml BSA for 30 min at 37°C. Samples were loaded onto a prerun 5% acrylamide/bisacrylamide gel with 2.5% w/v glycerol and 1× Tris Acetate buffer and subjected to electrophoresis for 2 h at 35 mAmp. Gels were vacuum-dried onto filter paper and exposed to film.

### SELEX analysis

A 2.5 pmol of the SELEX oligonucleotide (5′-dGGAAGC TTAGAATAAACGCTCAANNNNNNNNNNNNN NNNNNNTTCGACATGAGGCCCTCdTAGACG G-3′) or three SELEX control oligonucleotides, A control (5′-GGAAGCTTAGAATAAACGCTCAAAAAAAAA AAAAAAAAAAAAATTCGACATGAGGCCCTCTA GACGG-3′), T control (5′-dGGAAGCTTAGAATAAA CGCTCAATTTTTTTTTTTTTTTTTTTTCGACAT GAGGCCCTCTAGACGG-3′) or AT control (5′-dGGA AGCTTAGAATAAACGCTCAAATATATATATATAT ATATATTTCGACATGAGGCCCTCTAGACGG-3′) were PCR amplified with 500 pmol of the 5′ primer (5′-dT AATACGACTCACTATAGGGAAGCTTAGAATAAA C-3′) to attach a T7 promoter to the 5′-end and 500 pmol of the 3′ primer (5′-dGCCGTCTAGAGGGCCTCATGT CGAA-3′) with Taq DNA polymerase (Invitrogen). This PCR product was then used as a template to make the SELEX RNA pool using T7 RNA polymerase-plus (Ambion). SELEX RNA was subjected to DNAse digestion with RQ1 DNAse I (Promega) to remove the SELEX DNA PCR template and purified with a NucAway Spin Column (Ambion, Inc.) The ICP27 N-terminal peptide was purified with Ni-NTA agarose (Qiagen) under native purification conditions from a 50 ml culture of Rosetta E. coli grown and induced for expression in LB. ICP27 N-terminal peptide bound to Ni-NTA was incubated with SELEX RNA for 1 h at room temperature in 1× binding buffer (20 mM Tris pH 8, 150 mM NaCl). Protein:SELEX RNA complexes were washed and eluted with His elution buffer (250 mM Imidizole, 300 mM NaCl and 50 mM $NaH_2PO_4$ pH 8) and protein was removed by phenol:chloroform extraction. Eluted SELEX RNA was reverse transcribed using M-MLV Reverse Transcriptase (Ambion) and a primer specific to the 3′ sequence flanking the 20 nucleotide random region (5′-dGCCGTCTAGAG GGCCTCATGTCGAA-3′). The cDNA was then PCR amplified using Taq DNA polymerase (Invitrogen) with primers specific to the 5′- and 3′-flanking sequences with the 5′ primer incorporating a T7 promoter sequence. This was repeated for 10 rounds of selection and the final PCR products were cloned into pUC18 and sequenced.

### NMR analysis

All DNA oligonucleotides (Operon) were resuspended in 50 mM Tris pH 8, 150 mM KCl and 1 mM EDTA pH 8 at a final concentration of 10 mM. Chelex 100 Resin (Bio-Rad) was added to each oligonucleotide solution and incubated for 1 h at room temperature with rotation to remove contaminating metals. The Chelex resin was removed by centrifugation. The G-quartet oligonucleotide $[dTTGGGGTT]_4$ was heated to 95°C for 5 min and allowed to cool to room temperature prior to NMR analysis. Homogeneity of the G-quartet sample was confirmed by the appearance of only one set of downfield NMR peaks. All oligonucleotides were diluted to 0.5 mM in resuspension buffer with 10% D2O for NMR analysis. One-dimensional spectra were collected on a Varian 800 MHz NMR Spectrometer at 25°C using WATERGATE solvent elimination (28). The 16 K data points were collected to define a spectral width of 15 000 Hz with a recycle delay of 3 s in between scans. Data were processed with NMR pipe (29) using 5 Hz line broadening.

## RESULTS

### Expression of the ICP27 N-terminus

To investigate the interactions of the ICP27 N-terminus with HSV-1 sequences, the N-terminal 160 amino acids of ICP27, which includes the RGG box RNA binding motif, were expressed in *Escherichia coli* (Figure 1A.) The ICP27 coding sequence was cloned into an IPTG-inducible expression vector with a N-terminal His tag. When this wild-type ICP27 N-terminal peptide was expressed in Rosetta *E. coli*, a BL21 bacterial strain expressing additional tRNAs for optimal eukaryotic gene expression, two proteins species were purified (data not shown) and MALDI mass spectrometry showed that the two proteins were 19 kDa and 15 kDa in size (data not shown). The 19 kDa protein species was the expected size of the full-length ICP27 peptide with an N-terminal His affinity tag. To improve the amount of full-length protein and to eliminate the truncation products, the ICP27 gene was codon optimized and synthesized for expression in *E. coli* using CODA genomics technology. Codon

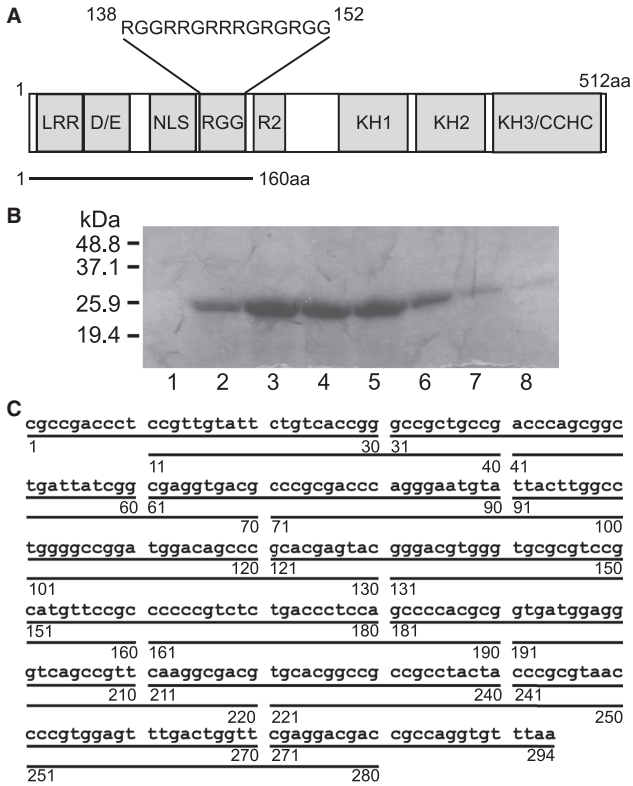


**Figure 1.** Schematic of the ICP27 functional domains and bacterial expression of the ICP27 N-terminal 160 amino acids. (**A**) ICP27 functional domains include an N-terminal leucine-rich region (LRR), acidic region (D/E), a nuclear localization signal (NLS), arginine-/glycine-rich RGG box (RGG) and a second arginine-rich region (R2.) The C-terminus contains three predicted hnRNP K homology (KH) domains and a zinc finger-like motif (CCHC). The region of the ICP27 N-terminus expressed in *E. coli* is underlined. (**B**) Codon optimized ICP27 N-terminal peptide was expressed in Rosetta *E. coli* with a C-terminal His tag. Expressed protein was purified using a Nickel column under native conditions. Elution fractions were collected from the Nickel column and fractions 1 through 8 were run on a 10–20% SDS–PAGE gradient gel and stained with Coomassie Blue. (**C**) Glycoprotein C (gC) gene sequence identified in the yeast three hybrid screen with HSV-1 UL 43 and gC RNA and full-length ICP27. DNA sequence of the 294-nt region of the gC gene identified in a yeast three-hybrid screen interacting with ICP27. This DNA sequence corresponds to nucleotides 96 946–97 239 of the KOS HSV-1 gC mRNA sequence that were identified interacting with ICP27. Sequences underlined and numbered denote the overlapping 30-mer gC DNA oligonucleotides used in EMSA with the ICP27 N-terminal peptide.

optimization changes the codon sequence without affecting the amino acid composition of the protein to prevent ribosomal stalling and enhance expression and solubility. The codon-optimized gene was cloned into an expression vector with a C-terminal His tag and expressed in Rosetta BL21 *E. coli*. Expression and purification of the codon optimized ICP27 N-terminal 160 amino acids yielded a single protein species migrating at 26 kDa on a Tris–Tricene gradient gel (Figure 1B.) The purified protein had a molecular weight of 17.8 kDa as determined by MALDI-Mass spectrometry (data not shown) which correlates with its correct mass. The full length, codon optimized ICP27 N-terminus protein with a C-terminal His tag is smaller than its N-terminal His tagged counterpart due to the use of a different cloning scheme, which eliminated vector sequences from being incorporated into the translated ICP27 N-terminal peptide. Therefore, the codon-optimized ICP27 gene was expressed and used for the rest of the study.

### EMSA binding studies with gC sequences to investigate ICP27 binding specificity

To investigate the ICP27 RNA binding specificity, HSV-1 RNA sequences that interact with ICP27 were first identified. A yeast three hybrid analysis identified 31 HSV-1 sense RNAs, mapping to 28 different open reading frames, which interacted specifically with ICP27 *in vivo* (26). ICP27 was also found to specifically associate with seven HSV-1 RNAs in *in vivo* UV-crosslinking studies. The late RNA for the glycoprotein C (gC) gene was one of the RNAs that UV-cross-linked to ICP27 in infected cells (3). To further refine the pools of possible RNA binding sequences, a yeast three-hybrid screen was performed with an RNA library generated from nucleotides 95 155–98 129 of the KOS HSV-1 genome, which encompasses two late genes, UL43 and UL44 (gC). Sequences corresponding to both the mRNA coding and noncoding strands from UL43 and gC were identified in the yeast three hybrid screen as interacting with ICP27 (data not shown). No obvious consensus sequences were identified in these sequences; consequently EMSA was used to screen these sequences for those that bound to ICP27 with high affinity. A 294-nt region of gC, corresponding to nucleotides 96 946–97 239 of the coding strand of the KOS HSV-1 gC mRNA, was chosen as an EMSA binding substrate. This region of the gC mRNA was chosen over the other sequences because it was within the coding region of the gC mRNA and was small enough to screen. Nineteen overlapping 30-mer DNA oligonucleotides corresponding to coding strand of the gC RNA were used in EMSA experiments (Figure 1C). RGG box motifs in other proteins recognize both RNA and DNA quadruplex structures (30,31), so gC DNA was used instead of gC RNA for this initial screening to provide a more stable substrate for analysis. The ICP27 N-terminus was incubated with individual radiolabeled gC DNA 30mers at 37°C and resolved by nondenaturing acrylamide gel electrophoresis. Figure 2A shows a gC DNA sequence that the ICP27 N-terminus was able to shift its migration to a more slowly migrating band, gC 11–40, and a second sequence, gC 31–60, whose migration was not affected by addition of protein. Some sequences' migration were moderately affected (Figure 2B; 121–150), whereas most gC sequences were shifted well by the ICP27 N-terminus (Figure 2B; gC 1–30, 91–120 and Figure 2C; 61–90.) Of the gC DNA oligonucleotides tested, fourteen were shifted, two were shifted moderately and three showed no discernable shift (Table 1).

The ICP27 N-terminal RGG box motif binds directly to HSV-1 RNA. The RGG boxes of FMRP and other proteins have been shown to specifically recognize G-quartet RNA structures (16,17,25,31). Therefore, the ability of the ICP27 N-terminus to bind G-quartets was

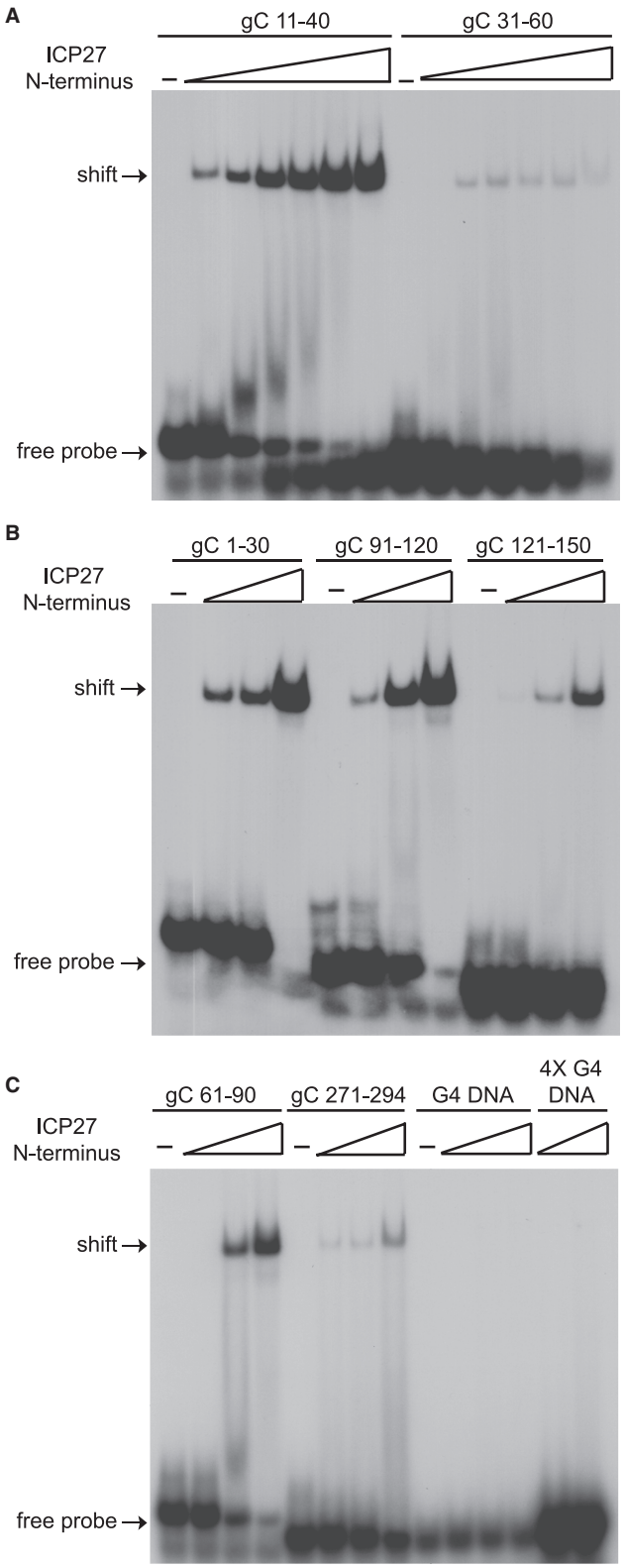


**Figure 2.** The ICP27 N-terminus binds to a majority of gC DNA sequences but not a G-quartet DNA sequence. (**A**) A 20 fmol of radiolabeled gC DNA oligonucleotides gC 11–40 and gC 31–60 (see figure for sequence) were either incubated with no protein or with 0.5 to 20 μM of the ICP27 N-terminal peptide. Samples were electrophoresed on a prerun acrylamide gel and dried gels were exposed to film. Arrows indicate the migration of free probe and the shift due to ICP27 N-terminal peptide binding. (**B**) A 20 fmol of radiolabeled gC DNA oligonucleotides gC 1–30, gC 91–120 or gC 121–150 were incubated with no protein or 0.5 to 12.5 μM of the ICP27 N-terminal peptide. Samples were electrophoresed and dried gels were exposed to film. (**C**) A 20 fmol of radiolabeled gC DNA oligonucleotides gC 61–90, gC 271–294 and G4 DNA $[TAGGGGTT]_4$ were incubated with 0.5 to 12.5 μM of the ICP27 N-terminal peptide. In the lanes denoted 4X G4 DNA (far right lanes), 80 fmol of G4 DNA were used.

tested by EMSA. A parallel intermolecular G-quartet structure can form when four individual oligonucleotides align $[dTAGGGGTT]_4$ in the presence of potassium cations (21). This G-quartet DNA (G4 DNA) was also shown previously to bind to the nucleolin RGG box peptide (32). The G4 DNA was incubated with increasing amounts of the ICP27 N-terminal peptide. The ICP27 N-terminus did not slow the migration G-quartet oligonucleotide, even when it was added at four times the molar concentration (Figure 2C; G4 DNA). Another oligonucleotide that forms a intramolecular G-quartet of the 'chair' classification [dGGTTGGTGTGGTTGG] when runs of G's align in a plane and T's form loops of varying lengths (33) was also not shifted by the ICP27 N-terminus (data not shown). These data suggest that the region of the ICP27 N-terminus that binds gC sequences does not bind G-quartet structures *in vitro*. Since ICP27 is an RNA binding protein, the ability of the ICP27 N-terminus to bind to gC sequences as RNA molecules was determined using EMSA. Radiolabeled DNA and RNA oligonucleotides for gC 1–30, a sequence the ICP27 N-terminus bound, and gC 31–60, a sequence not bound, were analyzed in parallel EMSAs for their interaction with the ICP27 N-terminus (Figure 3A). The ICP27 N-terminus was able recognize and shift more radiolabeled probe for the gC 1–30 RNA oligonucleotide compared with the corresponding gC1–30 DNA sequence when equivalent amounts of oligonucleotide and ICP27 N-terminus protein was added, suggesting that the ICP27 N-terminus has a stronger affinity for RNA sequences *in vitro*. The ICP27 N-terminus did not bind to RNAs indiscriminately, since the gC 31–60 RNA was not shifted (Figure 3A) and this sequence is also not shifted as a DNA oligonucleotide (Figures 2A and 3A)

HSV-1 gC DNA sequences were used in competition EMSA experiments to determine if some sequences were preferential substrates for the ICP27 N-terminus. Two of the competitors used, gC 31–60 and gC 271–294, do not appear to bind ICP27 strongly (Figure 2A and C), and, consequently, at high concentrations these are useful in probing competitive binding established through nonspecific interactions. A third competitor's migration, gC 11–40, was strongly affected by the addition of the ICP27 N-terminus in EMSA experiments (Figure 2A and C), most likely through specific recognition. Radiolabeled sequence gC 71–100 (Figure 1C), a sequence shifted well by ICP27, was incubated with the ICP27 N-terminus alone or with nonradioactive gC competitors at 5, 10 and 50 times the molar concentration of the radiolabeled sequence. Sequences that were not shifted well by the ICP27 N-terminus (gC 31–60; Figure 2A and gC 271–294; Figure 2C.) were not able to compete with gC 71–100 at molar concentrations five

**Table 1.** Summary of EMSA results and mfold analysis

| Number[a] | Sequence | Binding in EMSA[b] | No. of mfold structures | Average mfold $\Delta G$ (kcal/mol) |
|---|---|---|---|---|
| 1–30 | cgccgaccctccgttgtattctgtcaccgg | + | 2 | −0.57 |
| 11–40 | ccgttgtattctgtcaccgggccgctgccg | + | 2 | −1.36 |
| 31–60 | gccgctgccgacccagcggctgattatcgg | − | 1 | −8.34 |
| 41–70 | acccagcggctgattatcggcgaggtgacg | + | 3 | −1.50 |
| 61–90 | cgaggtgacgcccgcgacccagggaatgta | + | 4 | −1.63 |
| 71–100 | cccgcgacccagggaatgtattacttggcc | + | 2 | −1.66 |
| 91–120 | ttacttggcctggggccggatggacagccc | + | 2 | −3.09 |
| 101–130 | tggggccggatggacagcccgcacgagtac | + | 1 | −2.99 |
| 121–150 | gcacgagtacgggacgtgggtgcgcgtccg | ± | 1 | −5.95 |
| 131–160 | gggacgtgggtgcgcgtccgcatgttccgc | + | 1 | −5.88 |
| 151–180 | catgttccgcccccgtctctgaccctcca | + | 2 | −0.11 |
| 161–190 | cccccgtctctgaccctccagccccacgcg | + | 4 | −0.68 |
| 181–210 | gccccacgcggtgatggagggtcagccgtt | + | 1 | −2.40 |
| 191–220 | gtgatggagggtcagccgttcaaggcgacg | ± | 1 | −2.56 |
| 211–240 | caaggcgacgtgcacggccgccgcctacta | − | 2 | −5.57 |
| 221–250 | tgcacggccgccgcctactacccgcgtaac | + | 7 | −1.46 |
| 241–270 | cccgcgtaaccccgtggagtttgactggtt | + | 1 | −3.11 |
| 251–280 | cccgtggagtttgactggttcgaggacgac | + | 2 | −0.92 |
| 271–294 | cgaggacgaccgccaggtgtttaa | − | 2 | −1.37 |

[a]Number indicates the nucleotides of the gC sequence (Figure 2).
[b]+ indicates good binding in EMSA, ± indicates detectable binding in EMSA but not as strong as +, − indicates poor binding in EMSA.

and ten times that of gC 71–100 (Figure 3B). When nonradiolabeled gC 31–60 was added at the highest molar concentration (50×), about half of the radiolabeled gC 71–100 was displaced. However, another strongly shifted sequence, gC 11–40, was able to efficiently compete with gC 71–100 for binding to the ICP27 N-terminus at all concentrations of competitor tested. This suggests that both sequences shifted well by the ICP27 N-terminus, gC 11–40 and gC 71–100, are competing for the same region of the protein, possibly the RGG box, and that weak binding sequences cannot compete with strong binding sequences, even when added in molar excess. Therefore, the ICP27 N-terminal peptide containing the RGG box has a preference for binding certain gC sequences and there may be a higher order structure that determines binding specificity.

### Mfold and NMR analysis of gC DNA sequences

To investigate the possibility of a conserved binding motif or secondary structure within the gC DNA sequences that were shifted by the ICP27 N-terminus in the EMSAs, a motif-based sequence analysis tool (MEME) (34) and mfold analysis using the Zuker algorithm (35) were performed on the nineteen gC DNA sequences. Representative predicted mfold secondary structures for some gC DNA sequences are shown in Figure 4. MEME analysis failed to find a motif present in all gC sequences that were shifted by the ICP27 N-terminus (data not shown). Figure 4 shows representative mfold structures of gC DNAs that were shifted (gC 1–30, gC 11–40, gC 91–120 and gC151–180), not shifted (gC 31–60 and gC 211–240) or intermediately shifted (gC 121–150 and gC 191–220) by the ICP27 N-terminus. Mfold analysis revealed that sequences moderately shifted or not shifted by ICP27 had more stable secondary structures (relatively low $\Delta G$ values $< -5.0$ kcal/mol) predicted by mfold compared to those that were shifted well by ICP27 (Table 1). Specifically, the gC 31–60, gC 121–150, gC 211–240, which were either weakly or not shifted by the ICP27 N-terminus, had the lowest $\Delta G$ values. In contrast, the majority of the gC sequences that were shifted by the ICP27 N-terminus are predicted to have less stable secondary structures (high $\Delta G$ values between $-0.11$ and $-3.1$ kcal/mol). This suggested that sequences not shifted by the ICP27 N-terminus by EMSA had a higher propensity to form secondary structures and that may be interfering with ICP27 binding. To test this hypothesis, representative individual gC sequences were analyzed by NMR for secondary structure. Amino, imino and ring hydrogen atoms attached to nitrogen involved in DNA secondary structures are resistant to solvent exchange and have unique downfield chemical shifts between 10 and 14 p.p.m. Guanosine imino resonances diagnostic for G-quartets appear in the region between 10 and 12 p.p.m. in the spectra and Watson–Crick pairing causes iminos and aminos to resonate in the region between 12 and 14 p.p.m. (36) One-dimensional proton spectra were collected on three sequences that were shifted well, gC 1–30, gC 11–40 and gC 91–120, one intermediately shifted sequence, gC 191–220, and two sequences that were not shifted by the ICP27 N-terminus, gC 31–60 and G-quartet DNA (G4) (Figure 5). Four downfield peaks were observed between 10.5 and 12.5 p.p.m. for the G4 DNA $[dTTGGGGTT]_4$ (Figure 5), which have been shown to represent hydrogen-bonded imino resonances for each guanine quartet (21,37). Observation of these peaks is good evidence for quadruplex formation and confirms homogenous preparation of the sample. These G4 DNA peaks were distinct from downfield peaks observed in the gC 31–60 and gC 191–220 spectra, which resonate between 12.5 and 13.5 p.p.m., indicative of Watson–Crick base-pairing. In two of the gC sequences that were not shifted by the ICP27 N-terminus, two and three clear peaks were observed in

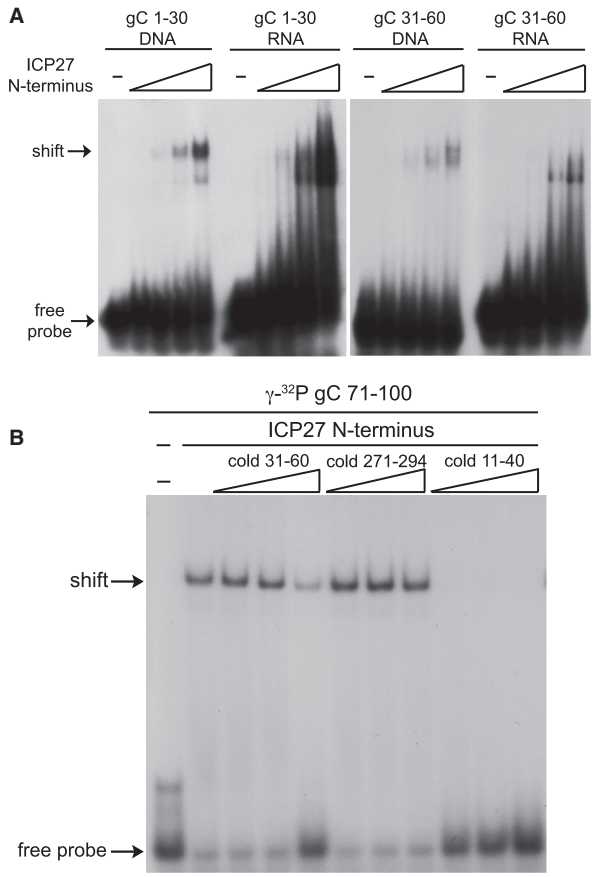


**Figure 3.** The ICP27 N-terminus binds to gC RNA sequences and the interaction between ICP27 N-terminus and individual gC DNA sequences can be competed by other high affinity binding gC DNA sequences but not by weak binding DNA sequences. (**A**) A 20 fmol of radiolabeled DNA or RNA oligonucleotides corresponding to the gC 1–30 or gC 31–60 sequence (see Figure 1C for sequence) were incubated with no protein (−) or 2.5, 7.5, 22.5 and 67.5 μM of the ICP27 N-terminal peptide. Samples were electrophorsesed and dried gels were exposed to film. Arrows indicate the position of free and shifted probes. (**B**) A 20 fmol of radiolabeled gC oligonucleotide gC 71–100 was incubated with no protein (lane 1), 5 μM of the ICP27 N-terminal peptide (lane 2) or 5 μM of the ICP27 N-terminal peptide with 5× (100 fmol), 10× (200 fmol) or 50× (1000 fmol) of nonradioactive competitor gC DNA oligonucleotides (gC 31–60, gC 271–294 or gC 11–40.) Samples were electrophoresed and dried gels were exposed to film. Arrows indicate the position of the free and shifted probes.

the gC 191–220 and gC 31–60 spectra, respectively. These two sequences have a secondary structure distinct from that of a G-quartet and are predicted to form a hairpin structure stabilized by Watson–Crick pairing. Spectra collected on three gC sequences that were shifted well by the ICP27 N-terminus (gC 1–30, gC 11–40 and gC 91–120) showed weak or no downfield peaks, indicating that no stable secondary structures were formed for these sequences. Notably, the sequences that showed downfield chemical shifts, indicative of a secondary structure, showed no shift by the ICP27 N-terminus in EMSA (Figure 2), suggesting that formation of secondary structures may preclude binding by the ICP27 N-terminus.

## Investigation of ICP27 RNA binding specificity using SELEX

To further investigate the binding specificity of the ICP27 N-terminus, a modified SELEX analysis was performed with recombinant ICP27 N-terminal protein and a random RNA library (38). An RNA library was generated from a pool of DNA oligonucleotide templates with a twenty-nucleotide random region flanked by unique sequences for PCR amplification and cloning. PCR was used to add a T7 promoter to the double-stranded DNA oligonucleotide, and RNA was synthesized using T7 RNA polymerase. SELEX RNA was incubated with the ICP27 N-terminal peptide bound to Ni-NTA agarose. The RNA that bound to the ICP27 N-terminal peptide was eluted, purified and reverse transcribed into cDNA. Ten rounds of selection were done and the pool of cDNA from the 10th round was cloned and sequenced. SELEX control RNAs were also used that contained twenty adenosines, thymidines or an equal mix of both in place of the twenty-nucleotide random region (see 'Materials and Methods' section). No SELEX PCR product was detected when these control RNAs were used (data not shown), indicating there was no detectable binding to these SELEX control RNAs and that the common flanking sequence was not contributing to the selection of the SELEX RNAs. The 20-nt random sequence for 19 clones is shown in Table 2. The HSV-1 genome has a G/C content of 68% and the average G/C content for the 19 clones was 70%, suggesting that ICP27 may have a preference for G/C RNA sequences.

## EMSA analysis of SELEX sequences

To determine if there were sequences within the SELEX pool that had high affinity for ICP27, EMSA were performed with the 19 20-nt SELEX sequences. DNA oligonucleotides corresponding to the SELEX RNA sequences were radiolabeled and incubated with increasing amounts of ICP27 N-terminal peptide. The gC DNA sequence gC 11–40 (Figures 1C and 2A) was used as a positive control to compare binding of ICP27 N-terminus with the SELEX sequences. Figure 6 shows a representative EMSA performed with selected SELEX clones. There were some sequences whose migration was not significantly affected by the addition of the ICP27 N-terminus, such as SELEX 2 (Figure 6), but for the majority of sequences, the ICP27 N-terminus was able to shift the migration of the SELEX DNA sequences. Only three out of the nineteen SELEX DNA sequences were not shifted to a slower migrating band (Table 2). It is possible that the G/C percentage of a particular sequence might influence the binding by the ICP27 N-terminus. Two of the sequences that were not shifted, SELEX 2 and 9, had a lower relative G/C content of 65 and 55%, respectively, compared to the rest of the pool (Table 2). However, other sequences that were shifted by the ICP27 N-terminus, such as SELEX 12, was only 55% G/C, suggesting that the G/C content, or primary sequence, is not the predominant factor influencing recognition and binding by the ICP27 N-terminus.

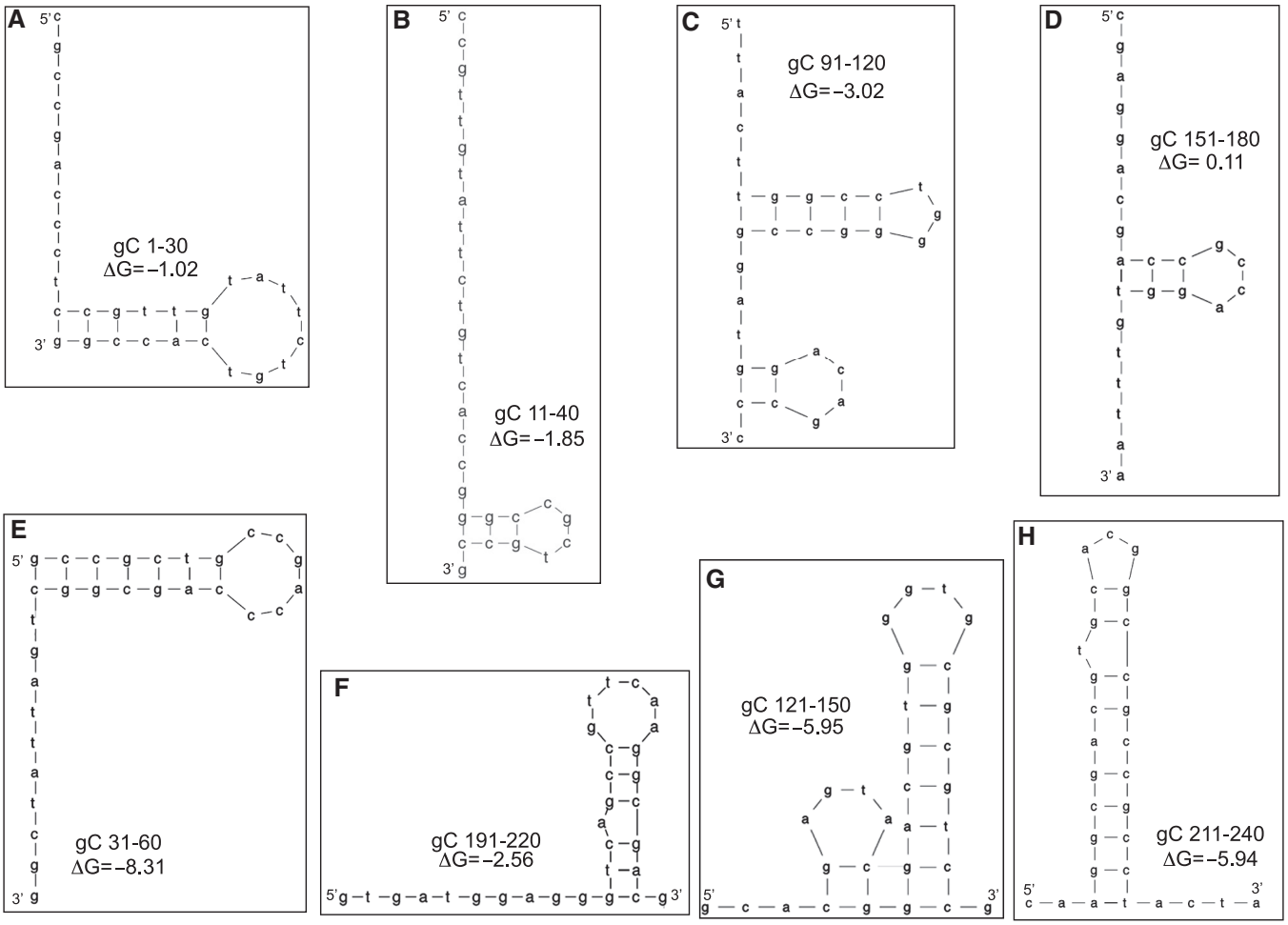


**Figure 4.** mfold analysis and representative predicted secondary structures of selected gC DNA sequences. All gC 30-mer DNA oligonucleotides were analyzed by mfold using the DNA mfold server version 3.2 (35). Default constraints were used which included a folding temperature of 37°C and an upper bound of 10 on the number of computed foldings. (**A** through **H**) The predicted mfold structure with the lowest $\Delta G$ in kcal/mol for eight out of the nineteen gC DNA sequences (Table 1) (A) gC 1–30, (B) gC 11–40, (C) gC 91–120, (D) gC 151–180, (E) gC 31–60, (F) gC 191–210, (G) gC 121–150 and (H) gC 211–240.

### Mfold and NMR analysis of SELEX sequences

The Zuker mfold algorithm was used to analyze the SELEX RNA sequences for possible secondary structure formation (35,39). Figure 7 shows representative mfold structures of SELEX RNAs that were shifted (SELEX 1,4, 6 and 13), not shifted (SELEX 2 and 5) or intermediately shifted (SELEX 3) by the ICP27 N-terminus. Varying lengths of hairpin secondary structure were predicted for the SELEX RNAs with an average $\Delta G$ of $-1.2$ to $-7.1$ kcal/mol. 1D proton NMR analysis was performed on DNA oligonucleotides corresponding to SELEX RNAs 2, 4 and 13 to determine if secondary structures were forming. Only the SELEX 2 sequence, which is not shifted by the ICP27 N-terminus, showed some broad downfield peaks between 10.5 and 12.5 p.p.m. (Figure 8). The two SELEX sequences that were shifted by the ICP27 N-terminus had either a single downfield peak distinct from G-quartets, possibly one Watson–Crick pair (SELEX 4), or no downfield field peak (SELEX 13). These data support the EMSA and NMR results obtained from HSV-1 gC sequences, which suggests that the ability of the ICP27 N-terminus to bind sequences is not dependent on the recognition of a binding motif or structure. Thus, it appears that the N-terminus of ICP27 recognizes unstructured RNA substrates with relatively high GC content.

## DISCUSSION

In this study, we identified gC sequences that bind specifically to the N-terminus of ICP27 *in vitro* with high affinity. Competition EMSAs between high affinity binding sequences suggest a single binding site within the ICP27 N-terminus, possibly the RGG box. Current studies with RGG box arginine substitution mutants revealed that indeed the RGG box is responsible for the binding of gC sequences by the ICP27 N-terminal peptide (manuscript in preparation.) An RNA SELEX experiment identified a pool of GC-rich RNAs that were able to selectively bind ICP27. We also observed that the ICP27

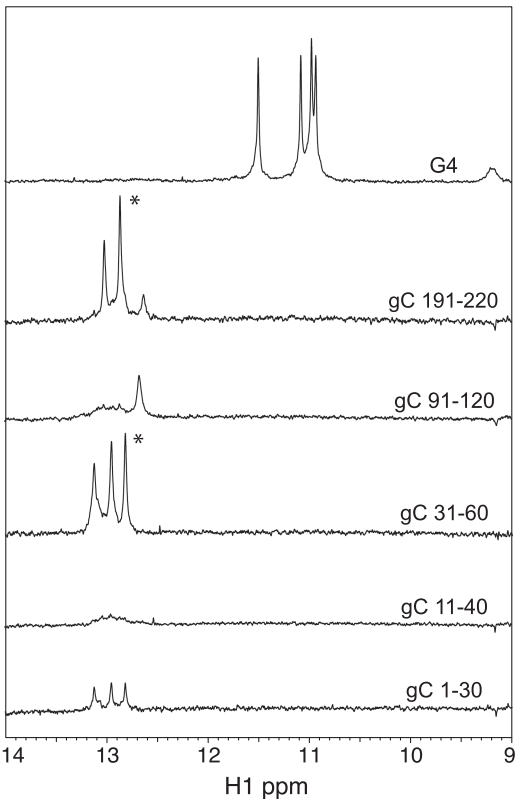


**Figure 5.** 1D $^{1}$H spectra of G-quartet DNA and selected gC 30-mer DNA oligonucleotides. A portion of the 1D $^{1}$H spectra (9–14 p.p.m.) collected on 0.5 mM G4 DNA $[TTGGGGTT]_4$, gC 191–220, gC 91–120, gC 31–60, gC 11–40 and gC 1–30 at 25°C in 50 mM Tris pH 8, 150 mM KCl and 10% $D_2O$. Spectra were scaled using a resolved H1' ribose signal. Asterisks (*) indicate prominent peaks detected in gC 191–220 and gC 31–60 sequences, two sequences that were not shifted by the ICP27 in EMSA experiments (Figure 2).

N-terminus can interact with both DNA (gC) and RNA (gC or SELEX) sequences with similar affinities but does not bind to DNAs that form a G-quartet structure. We did not identify a consensus sequence or structure from either the gC or SELEX sequences that the ICP27 N-terminus specifically recognizes, but instead found by NMR analysis of the binding sequences, that folding of the substrate into a secondary structure correlates with ICP27's failure to recognize a particular substrate. Therefore, the ICP27 RGG box prefers DNA or RNA substrates with no stable secondary structure and we hypothesize that the formation of secondary structures may interfere with recognition and binding. It is notable that there was a discrepancy between the secondary structures predicted by the mfold algorithm and the actual structures detected by NMR. For some mfold structures, Watson–Crick base pairing to form the secondary structure was accurately predicted (gC 31–60 and gC 191–220), but for others, complex hairpins in which many base pairs were predicted to exist (gC 91–120 and SELEX 13) only one or no base pairs, respectively, were observed in the NMR spectra. Thus, there are limitations in the accuracy of the structure prediction algorithms and for the sequences used in this study it was important to determine whether the predicted secondary structures were actually formed.

We report in these studies that the ICP27 N-terminal RGG box RNA binding motif is unique in its recognition and binding of DNA and RNA and, in contrast to the FMRP RGG box, did not recognize G-quartet RNAs or structured RNAs, but actually preferred substrates without secondary structure. This was surprising since it had been shown in a previous report that the ICP27 RGG box peptide (amino acids 135–157) could specifically bind to Sc1 RNA, which is a G-quartet-forming RNA also recognized by FMRP, in EMSA and fluorescence

**Table 2.** Summary of SELEX EMSA results and mfold analysis

| Clone number | Sequence | Binding in EMSA[a] | GC content (%) | No. of mfold structures | Average mfold ΔG (kcal/mol) |
|---|---|---|---|---|---|
| 1 | ggccgggucaccucgcaccg | + | 80 | 1 | −6.85 |
| 2 | uccacggcagauacgccacg | − | 65 | 1 | −7.09 |
| 3 | ggguguгcaggggggauaugu | ± | 60 | 2 | −1.31 |
| 4 | gagacguagaggacggggcc | + | 70 | 1 | −4.56 |
| 5 | cccuggccaagcguccgccc | − | 80 | 1 | −4.25 |
| 6 | cgcagcgcgcccgucccgcg | + | 90 | 6 | −3.87 |
| 7 | ugacgucacagaccccgccg | ± | 70 | 3 | −1.96 |
| 8 | gguccgugguccgcugccg | ± | 80 | 4 | −4.75 |
| 9 | ccggcauagucacacguccg | − | 65 | 5 | −2.52 |
| 10 | cgcccugcacacguuucugg | + | 65 | 5 | −1.96 |
| 11 | gguguacuggugcgggcugc | + | 70 | 7 | −2.60 |
| 12 | ggcacugcaccuguuuguga | + | 55 | 1 | −5.57 |
| 13 | cgguacguggacgaugcugg | + | 65 | 1 | −5.50 |
| 14 | acgcccgucgccugcacuug | + | 65 | 3 | −1.79 |
| 15 | cagcgcagacaccguccgcc | + | 75 | 2 | −5.44 |
| 16 | caacgcucgggugauuacgg | + | 55 | 6 | −1.20 |
| 17 | ggugagguaggggucgucg | ± | 65 | 1 | −2.45 |
| 18 | cggcacgccucggcccccca | + | 85 | 1 | −6.07 |
| 19 | ucuacggggcgggggcuggg | + | 80 | 2 | −4.24 |

[a] + indicates good binding in EMSA, ± indicates detectable binding in EMSA but not as strong as +, − indicates poor binding in EMSA.

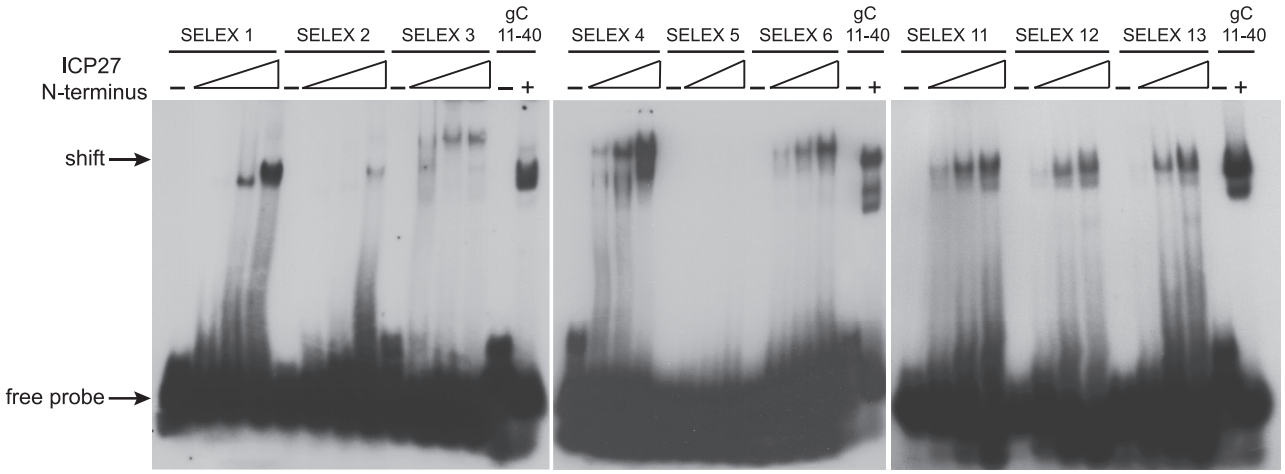


**Figure 6.** EMSA of selected SELEX DNA oligonucleotides. A 20 fmol of radiolabeled SELEX 20-mer DNA oligonucleotides (see Table 2) or the gC 11–40 DNA oligonucleotide were either incubated no protein or with 2.5–62.5 μM of the ICP27 N-terminal peptide. A 62.5 μM of the ICP27 N-terminal peptide was used in the (+) gC 11–40 lanes. Samples were electrophoresed on a prerun acrylamide gel with Tris Acetate Buffer and dried gels were exposed to film. Arrows indicate the migration of free probe and the shift due to ICP27 N-terminal peptide binding.

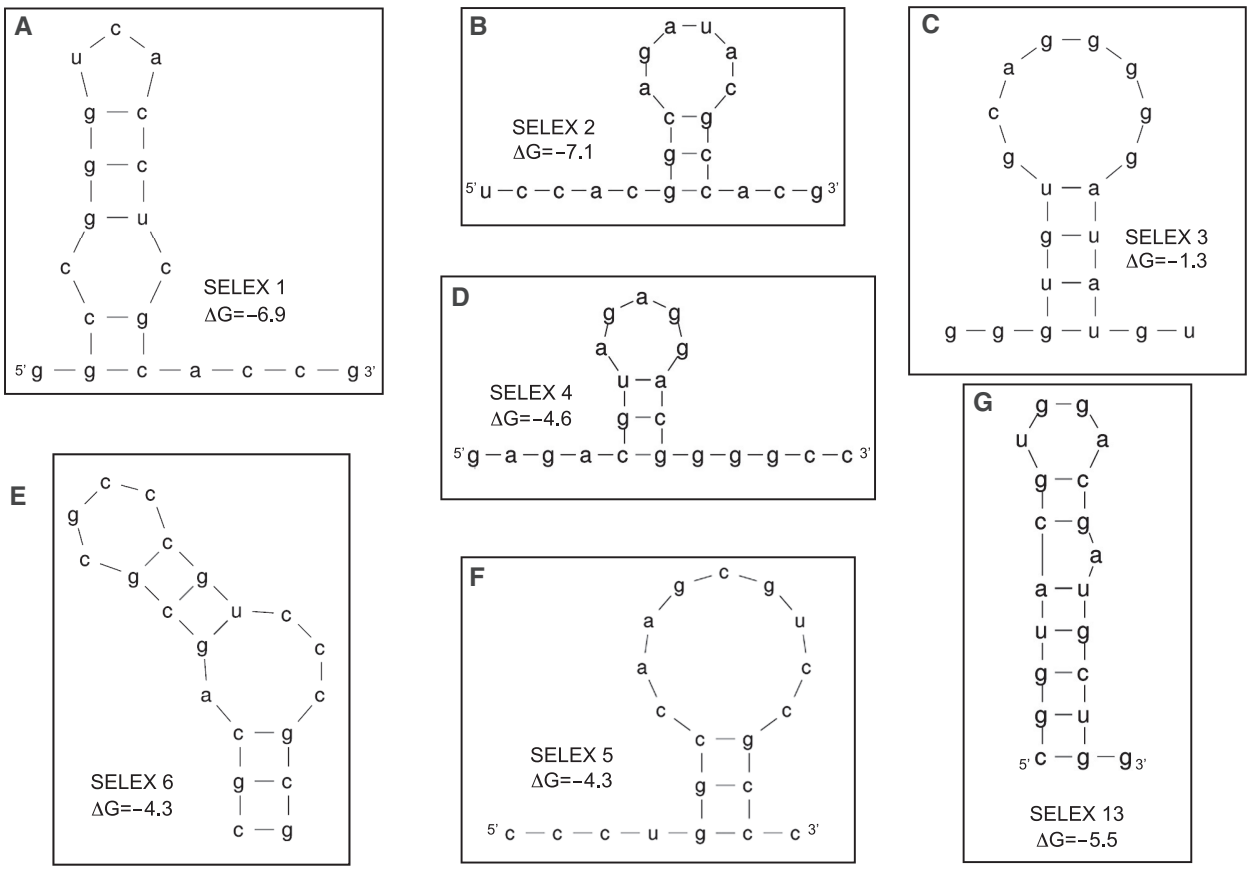


**Figure 7.** mfold analysis and representative predicted secondary structures of selected SELEX RNA sequences. All SELEX RNA sequences were submitted for mfold analysis using the RNA mfold version 3.2 (35,39). Default constraint settings were used including a folding temperature of 37°C. (**A** through **G**) The predicted mfold structure with the lowest ΔG in kcal/mol for seven out of the nineteen SELEX sequences identified. (A) SELEX 1, (B) SELEX 2, (C) SELEX 3, (D) SELEX 4, (E) SELEX 6, (F) SELEX 5 and (G) SELEX 13.

spectroscopy binding studies (40). In contrast to FMRP, the ICP27 RGG box did not stabilize the Sc1 G quartet (38). It is possible that we did not see the same interaction with either G-quartet substrate we tested due to a different molar ratio of protein to substrate and the use of much lower concentration of binding substrates in our EMSA studies compared to the Sc1 EMSA (20–80 fmol G4 DNA compared to 40 μM Sc1 RNA). Although binding

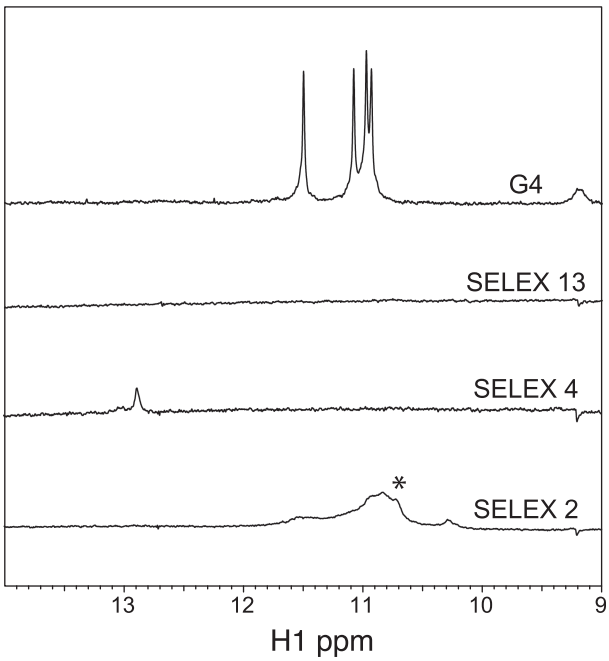


**Figure 8.** 1D $^{1}$H spectra of G-quartet DNA and selected SELEX DNA sequences. A portion of the 1D $^{1}$H spectra (9–14 p.p.m.) collected on 0.5 mM G4 DNA $[dTTGGGGTT]_4$, SELEX 13, SELEX 4 and SELEX 2 at 25°C in 50 mM Tris pH 8, 150 mM KCl and 10% $D_2O$. Spectra were scaled using a resolved H1' ribose signal. The asterisk (*) indicates broad peaks specifically seen in SELEX 2, a sequence that was not shifted by the ICP27 in EMSA experiments (Figure 6).

substrate concentrations were lower in the fluorescence spectroscopy experiments, it is possible that our construct including the entire 160 amino acid portion of the ICP27 N-terminus influenced the lack of G-quartet recognition compared to the RGG box peptide fragment that appears to bind. The ICP27 N-terminal domain used in this study specifically preferred sequences that did not form stable secondary structures and this suggests that not all RGG box binding motifs may recognize the same sequence or structure and that the spacing of the arginine and glycine residues, as well as other protein domains, may determine the specificity in RNA and DNA binding.

The ICP27 N-terminus is not dramatically different from other proteins with RGG box RNA binding motifs in preferring to bind GC-rich sequences; however it differs in the ability to bind a stable structure. The RGG box DNA or RNA binding motifs in the cellular hnRNP U protein (30), nucleolin (31) and Epstein–Barr Virus (EBV) EBNA1 protein (41) also prefer G-rich sequences but these proteins also recognize structured substrates. The nucleolin RGG box can bind to G4 DNA, predicted to form in rDNA regions of the genome, in which nucleolin plays an important role in ribosome biogenesis (31). The EBV EBNA1 protein is involved in recruiting origin recognition complex (ORC) proteins to the EBV episomal mini-chromosome origin of replication and binds specifically G-rich EBV RNAs and a G-quartet RNA sequence. It was found the binding by EBNA1 to the ORC proteins was RNA dependent and that cooperative binding of ORC1 and EBNA1 to G-rich RNA may be a mechanism for ORC recruitment (41). It is possible that the nucleolin and EBNA1 proteins with very specific roles in ribosome biogenesis and EBV genome maintenance, respectively, may require a specific DNA or RNA 'anchor' to recruit and maintain the stable association of the proteins with the specific target sequence. This is in contrast to ICP27, which binds nuclear viral RNA and mediates its export to the cytoplasm in cooperation with cellular export factors. Based on this study, ICP27 does not prefer nucleic acids that form structures and may simply prefer GC-rich, flexible nucleic acid substrates. The gC substrates bound by the ICP27 N-terminus had a very high G/C content, which is typical for HSV-1 genes, and the SELEX sequences identified ranged from 55% to 90% G/C content. Therefore, it is possible that specificity of binding by the ICP27 N-terminal RGG box may be determined first by the primary sequence of the substrate (GC-rich) and second by the substrate's lack of secondary structure, and this may be the mechanism by which ICP27 specifically binds and exports viral RNA. Since the HSV-1 genome is 68% G/C content, compared to 41%G/C content for the human genome, and late during infection there is an overall abundance of viral transcripts in the cell, ICP27's ability to specifically bind GC-rich viral RNAs may be the mechanism by which viral RNA is preferentially exported. In addition, other functional domains within the full length ICP27 protein, such as the KH domains, may confer more binding specificity for viral RNAs. It is also possible that ICP27's preference for GC-rich, unstructured single-stranded viral RNAs may be a mechanism to avoid binding to cellular GC-rich DNA and RNA and those that form secondary structures, such as rDNA, telomeres, tRNAs and regions involved in splicing, during infection. To investigate the full spectrum of RNAs specifically binding to ICP27, cross-linking immunoprecipitation (CLIP) experiments are being performed with wild-type ICP27 and several RGG box mutants in infected cells. This high throughput technique will increase the pool of RNA sequences that interact with ICP27 in vivo and determine the importance of the ICP27 RGG box for binding to specific RNAs. This type of analysis will allow further delineation of the specificity of RNA binding by ICP27 and the importance of the RGG box.

## ACKNOWLEDGEMENTS

We would like to thank Teresa Lehmann and Evgeny Fadeev for NMR technical assistance, and Stuart Souki for helpful review of the article.

## FUNDING

The National Institutes of Health; National Institute of Allergy and Infectious Diseases (RO1 AI61397 to R.M.S-G.; F32 AI062033 to K.C.L.). Funding for open access charge: National Institutes of Health (AI 61397 and AI 21515)

*Conflict of interest statement*. None declared.

## REFERENCES


1. Hardy,W.R. and Sandri-Goldin,R.M. (1994) Herpes simplex virus inhibits host cell splicing, and regulatory protein ICP27 is required for this effect. *J. Virol.*, **68**, 7790–7799.
2. Dai-Ju,J.Q., Li,L., Johnson,L.A. and Sandri-Goldin,R.M. (2006) ICP27 interacts with the C-terminal domain of RNA polymerase II and facilitates its recruitment to herpes simplex virus 1 transcription sites, where it undergoes proteasomal degradation during infection. *J. Virol.*, **80**, 3567–3581.
3. Sandri-Goldin,R.M. (1998) ICP27 mediates HSV RNA export by shuttling through a leucine-rich nuclear export signal and binding viral intronless RNAs through an RGG motif. *Genes Dev.*, **12**, 868–879.
4. Ingram,A., Phelan,A., Dunlop,J. and Clements,J.B. (1996) Immediate early protein IE63 of herpes simplex virus type 1 binds RNA directly. *J. Gen. Virol.*, **77 (Pt 8)**, 1847–1851.
5. Chen,I.H., Li,L., Silva,L. and Sandri-Goldin,R.M. (2005) ICP27 recruits Aly/REF but not TAP/NXF1 to herpes simplex virus type 1 transcription sites although TAP/NXF1 is required for ICP27 export. *J. Virol.*, **79**, 3949–3961.
6. Chen,I.H., Sciabica,K.S. and Sandri-Goldin,R.M. (2002) ICP27 interacts with the RNA export factor Aly/REF to direct herpes simplex virus type 1 intronless mRNAs to the TAP export pathway. *J. Virol.*, **76**, 12877–12889.
7. Koffa,M.D., Clements,J.B., Izaurralde,E., Wadd,S., Wilson,S.A., Mattaj,I.W. and Kuersten,S. (2001) Herpes simplex virus ICP27 protein provides viral mRNAs with access to the cellular mRNA export pathway. *EMBO J.*, **20**, 5769–5778.
8. Larralde,O., Smith,R.W., Wilkie,G.S., Malik,P., Gray,N.K. and Clements,J.B. (2006) Direct stimulation of translation by the multifunctional herpesvirus ICP27 protein. *J. Virol.*, **80**, 1588–1591.
9. Fontaine-Rodriguez,E.C. and Knipe,D.M. (2008) Herpes simplex virus ICP27 increases translation of a subset of viral late mRNAs. *J. Virol.*, **82**, 3538–3545.
10. Mears,W.E. and Rice,S.A. (1996) The RGG box motif of the herpes simplex virus ICP27 protein mediates an RNA-binding activity and determines in vivo methylation. *J. Virol.*, **70**, 7445–7453.
11. Soliman,T.M. and Silverstein,S.J. (2000) Identification of an export control sequence and a requirement for the KH domains in ICP27 from herpes simplex virus type 1. *J. Virol.*, **74**, 7600–7609.
12. Vaughan,P.J., Thibault,K.J., Hardwicke,M.A. and Sandri-Goldin,R.M. (1992) The herpes simplex virus immediate early protein ICP27 encodes a potential metal binding domain and binds zinc in vitro. *Virology*, **189**, 377–384.
13. Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.
14. Adinolfi,S., Bagni,C., Musco,G., Gibson,T., Mazzarella,L. and Pastore,A. (1999) Dissecting FMR1, the protein responsible for fragile X syndrome, in its structural and functional domains. *RNA*, **5**, 1248–1258.
15. Adinolfi,S., Ramos,A., Martin,S.R., Dal Piaz,F., Pucci,P., Bardoni,B., Mandel,J.L. and Pastore,A. (2003) The N-terminus of the fragile X mental retardation protein contains a novel domain involved in dimerization and RNA binding. *Biochemistry*, **42**, 10437–10444.
16. Darnell,J.C., Jensen,K.B., Jin,P., Brown,V., Warren,S.T. and Darnell,R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489–499.
17. Ramos,A., Hollingworth,D. and Pastore,A. (2003) G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA*, **9**, 1198–1207.
18. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871–880.
19. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
20. Sen,D. and Gilbert,W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*, **344**, 410–414.
21. Wang,Y. and Patel,D.J. (1992) Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry*, **31**, 8112–8119.
22. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,I. Jr. and Blackburn,E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **51**, 899–908.
23. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
24. Murchie,A.I. and Lilley,D.M. (1992) Retinoblastoma susceptibility genes contain 5′ sequences with a high propensity to form guanine-tetrad structures. *Nucleic Acids Res.*, **20**, 49–53.
25. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
26. Sokolowski,M., Scott,J.E., Heaney,R.P., Patel,A.H. and Clements,J.B. (2003) Identification of herpes simplex virus RNAs that interact specifically with regulatory protein ICP27 in vivo. *J. Biol. Chem.*, **278**, 33540–33549.
27. Johnson,L.A. and Sandri-Goldin,R.M. (2009) Efficient nuclear export of herpes simplex virus 1 transcripts requires both RNA binding by ICP27 and ICP27 interaction with TAP/NXF1. *J. Virol.*, **83**, 1184–1192.
28. Piotto,M., Saudek,V. and Sklenar,V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
29. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
30. Kiledjian,M. and Dreyfuss,G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.*, **11**, 2655–2664.
31. Hanakahi,L.A., Sun,H. and Maizels,N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.*, **274**, 15908–15912.
32. Cocco,M.J., Hanakahi,L.A., Huber,M.D. and Maizels,N. (2003) Specific interactions of distamycin with G-quadruplex DNA. *Nucleic Acids Res.*, **31**, 2944–2951.
33. Marathias,V.M. and Bolton,P.H. (1999) Determinants of DNA quadruplex structural type: sequence and potassium binding. *Biochemistry*, **38**, 4355–4364.
34. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
35. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
36. Feigon,J., Koshlap,K.M. and Smith,F.W. (1995) 1H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.*, **261**, 225–255.
37. Smith,F.W. and Feigon,J. (1992) Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature*, **356**, 164–168.
38. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
39. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
40. Zanotti,K.J., Lackey,P.E., Evans,G.L. and Mihailescu,M.R. (2006) Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA. *Biochemistry*, **45**, 8319–8330.
41. Norseen,J., Thomae,A., Sridharan,V., Aiyar,A., Schepers,A. and Lieberman,P.M. (2008) RNA-dependent recruitment of the origin recognition complex. *EMBO J.*, **27**, 3024–3035.