Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA: a structural scaffold for nuclear, cytoplasmic and viral proteins

Joanna Sztuba-Solinska[1],*, Jason W. Rausch[1], Rodman Smith[2], Jennifer T. Miller[1], Denise Whitby[3] and Stuart F.J. Le Grice[1],*

[1]Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA, [2]AIDS and Cancer Virus Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA and [3]Viral Oncology Section, AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

ABSTRACT

Kaposi's sarcoma-associated herpes virus (KSHV) polyadenylated nuclear (PAN) RNA facilitates lytic infection, modulating the cellular immune response by interacting with viral and cellular proteins and DNA. Although a number nucleoprotein interactions involving PAN have been implicated, our understanding of binding partners and PAN RNA binding motifs remains incomplete. Herein, we used SHAPE-mutational profiling (SHAPE-MaP) to probe PAN in its nuclear, cytoplasmic or viral environments or following cell/virion lysis and removal of proteins. We thus characterized and put into context discrete RNA structural elements, including the *cis*-acting Mta responsive element and expression and nuclear retention element (1,2). By comparing mutational profiles in different biological contexts, we identified sites on PAN either protected from chemical modification by protein binding or characterized by a loss of structure. While some protein binding sites were selectively localized, others were occupied in all three biological contexts. Individual binding sites of select KSHV gene products on PAN RNA were also identified in *in vitro* experiments. This work constitutes the most extensive structural characterization of a viral lncRNA and interactions with its protein partners in discrete biological contexts, providing a broad framework for understanding the roles of PAN RNA in KSHV infection.

INTRODUCTION

Global transcriptional analyses of the human genome indicates that the number of non-coding RNAs far exceeds that of protein-coding mRNAs, (3,4). Among non-coding RNAs, long non-coding or lncRNAs are especially abundant. They are classified as having a length >200 nucleotides (nts), are mostly produced by RNA polymerase II (4), contain a 5′ terminal cap and are often spliced and polyadenylated (5,6). A major feature of lncRNAs is their propensity to assume higher order structures that mediate diverse functions ranging from protein recognition to 'sensing' viral infection and modulating cell immunity (7–9). In contrast, herpesviruses counteract host immunity by encoding lncRNAs. Such RNAs are less immunogenic than proteins and can thus escape immune surveillance (10).

The 1.1 kB polyadenylated nuclear (PAN) lncRNA of Kaposi's sarcoma-associated herpes virus (KSHV) (11,12) is produced during lytic infection and suppresses expression of host genes involved in the antiviral response. Transcription factor IRF4 (13), histone H3K27 demethylases UTX and JMJD3, and the lysine methyltransferase MLL2 (14) have all been shown to bind PAN RNA and, analogous to cellular lncRNAs, PAN RNA interacts with transcriptional regulators and chromatin modifiers. Mapping studies have found numerous PAN RNA interaction sites on the KSHV episome as well as the cellular genome, indicating that this viral lncRNA actively modulates expression of host and viral genes (15). Once processed by cellular machinery, PAN RNA resembles an mRNA in that it has a 5′ cap and a 3′ poly(A) tail, however, it contains no introns and is therefore not spliced. Despite its mainly nuclear localization (11), recent reports have demonstrated its presence in cytoplasm

*To whom correspondence should be addressed. Tel: +1 301 846 5256; Fax: +1 301 846 5405; Email: joanna.sztuba-solinska@nih.gov
Correspondence may also be addressed to Stuart F.J. Le Grice. Tel: +301 846 5256; Fax: +301 846 6013; Email: legrices@mail.nih.gov

(∼20%), inside viral particles, but also in KSHV-infected cell lines that are uninduced (15,16).

PAN RNA has evolved sophisticated structural and molecular strategies for avoiding cellular decay processes (1,2,15,17,18). In one such mechanism, a U-rich motif (the expression and nuclear retention element, or ENE), located near the 3′ terminus, sequesters the poly(A) tail to form a triple helix, thereby rendering it resistant to the RNA degradation machinery (2,19,20). The viral ORF57 protein, also known as the mRNA transcript accumulation protein (Mta), further contributes to PAN RNA stability by binding to the 5′ Mta responsive element (MRE) motif (ORF57-responsive element or ORE) to increase the PAN RNA half-life (1,18).

There are numerous instances in virtually all classes of viruses in which the secondary structures of non-coding RNAs, or the non-coding segments of coding RNAs, contribute an important regulatory function. Hence, resolving the structure of PAN RNA, from which it almost certainly derives its multifunctionality, is of utmost significance. Detailed structural maps of cellular lncRNAs, such as Xist (21) and SRA (22), have illuminated mechanisms by which individual RNA motifs execute subtle molecular functions. In this study, we used the comprehensive and quantitative RNA probing technology SHAPE-MaP to examine structural and functional features of PAN RNA in nuclear and cytoplasmic compartments of infected cells and within viral particles. A comparative analysis of mutational profiles obtained by probing deproteinized PAN RNA (*ex vivo* and *ex virio*) and PAN RNA present in nuclear and cytoplasmic compartments (*in cellulo*) and inside viral particles (*in virio*) allowed us to identify sites of ribonucleoprotein (RNP) interactions. Moreover, direct mapping of partially purified recombinant KSHV proteins on polyadenylated, *in vitro* transcribed PAN RNA allowed us to match individual binding sites to specific proteins. This study provides the first report in which the structure of a viral lncRNA and its associated protein binding sites are mapped with single-nucleotide precision and the results linked to specific cellular or viral environments.

MATERIALS AND METHODS

Cell lines and culture conditions

Primary effusion lymphoma cells (BCBL-1) were grown in RPMI-1640 medium supplemented with 10% fetal bovine serum, 1% Penicillin/Streptomycin, 1% L-Glutamine and 2.5 mg/l Plasmocin (Invivogen) at 37°C in 5% CO_2. Induction of lytic infection was carried out by treating 3×10^7 BCBL-1 cells per 25 ml of medium with 0.3 mM sodium butyrate for 48 h prior to probing experiments.

Detection of local RNA secondary structure changes induced by SNPs

The RNAsnp web server is freely available at http://rth.dk/resources/rnasnp/ (23). PAN RNA sequence was submitted in FASTA format together with the single-nucleotide polymorphism (SNP) description in *X pos Y* format, where *X* is the wild-type nt, *Y* is the mutant and *post* is the nt position

(Supplementary Figure S3). Mode 2 was selected for operation with the default parameters -W 200 and -L 120.

Virus production and purification

KSHV/BCBL-1 cells were grown as described above. Cultures were grown in Corning Cell STACK and split 1:10 every 3–4 days. Virus lytic phase was induced by supplementing complete media with 3 mM sodium butyrate (Sigma-Aldrich) at a final volume of 20 l in a GE Healthcare 20/50EHT WAVE Bioreactor. Culturing conditions for induced cultures were 16 rpm with a rock angle of 7°. After 4 days, cultures were harvested and cells collected by filtration through a 5.0 and 0.5 μm OPTICAP XL capsule filter (Millipore). Culture supernatants were stored overnight at 4°C. Virions were purified by sucrose density gradient ultracentrifugation (Beckman continuous-flow rotor model CF32Ti) at 4°C in a 25–50% (wt/vol) sucrose (RNase free) density gradient in TNE buffer (0.01 M Tris–HCl pH 7.2, 0.1 M NaCl and 1 mM ethylenediaminetetraacetic acid (EDTA) in Milli-Q water). The sample was pumped at 5 l per hour with the centrifuge running at 30 000 rpm and centrifugation was continued for 30 min to allow banding of the virus. The rotor was decelerated to 3000 rpm, and the gradient was displaced by pumping 55% sucrose into the rotor at 25 ml/min. Fractions (25 ml each) were collected and analyzed for sucrose density using a refractometer (Bausch and Lomb). Virus-containing fractions, as determined by the sucrose density (25-42%), were pooled, diluted ∼1:3 with 100 mM HEPES pH 8, 100 mM NaCl, 10mM $MgCl_2$ buffer and centrifuged at 30 000 rpm for 1 h to pellet the virus. The resulting viral pellet was resuspended in 100 mM HEPES pH 8, 100 mM NaCl, 10 mM $MgCl_2$ buffer to a final 5000-fold concentration relative to the original volume of cell culture fluid. A second centrifugation of the 5000× virus concentrates at 60 000 rpm for 10 min was performed to provide viral pellets for analysis.

Cloning of KSHV-encoded ORFs

Each KSHV ORF was amplified by polymerase chain reaction (PCR) as described previously (24). Open reading frames (ORFs) were cloned with a C-terminal FLAG epitope tag and an N-terminal TEV protease cleavage site (ENLYFQG) followed by a Kozak translation initiation sequence. Amplification was carried out using Phusion DNA polymerase (New England Biolabs, Beverley, MA, USA) and primers containing flanking Gateway recombination sequences, attB1 and attB2 (Life Technologies, Carlsbad, CA, USA) under the manufacturer's standard conditions with an extension time of 30 s/kB. Each purified PCR product (QiaQuick PCR Purification Kit, Qiagen, Valencia, CA, USA) was recombined in the Gateway Donor vector pDonr253 using a BP reaction. Entry clones were sequenced throughout the cloned regions. Spliced genes were cloned by overlap PCR in which separate amplicons with 25 bp overlapping ends were produced for each exon and fused together in a subsequent PCR. KSHV ORF entry clones were recombined using Gateway LR into expression vectors for *Escherichia coli* (ORF57) and baculovirus-infected Sf9 cells (ORF26, ORF59, ORF73/LANA). *E. coli* and

baculovirus clones include a hexa-histidine maltose binding protein tag (His6-MBP). *E. coli* expression subclones were directly transformed into *E. coli* BL21(DE3). Baculovirus clones were converted to bacmid DNAs using the Bac-to-Bac system (Life Technologies). Bacmid DNA was used to transfect insect cells (Supplementary Figure S1).

Purification of KSHV-encoded ORFs

Baculovirus-insect expression. Baculovirus stocks were prepared in Sf-9 cells grown in HyClone SFX insect medium and titrated using an end-point dilution assay. The day before infection, 1 l of H5 cells were set in 3 l Erlenmeyer in SFX medium at a cell density of 8.5×10^5 cells/ml at 27°C. The culture was incubated at 100–110 rpms. The following day the culture was counted to determine the proper multiplicity of infection and then infected at a multiplicity of infection (MOI) of 3. Following the incubation at 21°C for 3 days, the supernatant was collected by pelleting the cells by centrifugation at $2000 \times g$ for 10 min. Viable cell counts and average cell size were recorded before infection and at 72 h post-infection using a Cedex Cell Analysis System.

E. coli *expression.* Seed cultures were inoculated from glycerol stocks (BL21* pRare) and grown in 50 ml of MDAG-135 medium (non-inducing) in 250 ml liter baffled shake flasks for 17 h at 37°C until mid-log phase growth. A total of 4-l baffled shake flasks containing 1 l of buffered rich medium were inoculated with 2% v/v of the seed culture. The cultures were grown for 4.5 h at 37°C until the optical density (OD) A600 reached ∼5–6, chilled to 20°C and induced with Isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.5 mM. The cultures were grown overnight for an additional 19.5 h and the cells were collected by centrifugation and stored at −80°C.

Lysis. The cell pellets (∼20 g wet weight) were thawed and cells were resuspended in a final buffer of 20 mM HEPES, pH 7.3, 300 mM NaCl, 2 mM β-ME at ∼100 OD/ml. Protease inhibitor cocktail was used at 1 ml per 10 000 ODs. Cells were lysed mechanically by three passes at 10 000 psi (4000 psi for insect cells) through a high-pressure instrument. Lysates were clarified by ultracentrifugation (70K × g, 30 min) and stored at −80°C until purification.

Purification. Clarified lysate was thawed, adjusted to 50 mM imidazole and loaded at 1 ml/min onto a 5 ml IMAC column. The equilibration buffer (EB) for the column was 20 mM HEPES, pH 7.3, 300 mM NaCl, 2 mM β-ME, 50 mM imidazole. The column was washed to baseline with EB and proteins eluted with a 20 column-volume gradient from 50 to 500 mM imidazole. Elution fractions were analyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie-staining. Positive fractions were pooled, dialyzed to 1× PBS, pH 7.2, concentrated to ∼1.0 mg/ml and a sub-sample tested for macroscopic stability through one freeze (−80°C)/thaw cycle before creating final aliquots for storage at −80°C. Quality control on final samples consisted of positive identification of at least two tryptic peptides by mass-spectrometry.

RNA integrity and concentration analysis

Analyses of total nuclear, cytoplasmic, viral and *in vitro* transcribed RNAs were performed using a 2100 Bioanalyzer LabChip® Kit according to the manufacturer's instructions (Agilent, CA, USA). RNA concentrations were estimated using an Agilent RNA 6000 Pico Kit by comparison to a RNA ladder of known concentration, and total RNA integrity was confirmed by the presence of sharp peaks representing 18S and 28S rRNA (Supplementary Figure S2). Additionally, the abundance or lack of tRNA which should be present only in cytoplasmic fractions, was measured to verify nuclear/cytoplasmic separation.

The efficiency of nuclear/cytoplasmic fraction separation

Confirmation of nuclear and cytoplasmic fractionation was measured via quantitative RT PCR using Power SYBR Green RNA-to-C_T 1-Step Kit (Applied Biosystems), targeting the nuclear lncRNA, MALAT1 (Supplementary Table S1) in both fractions, according to manufacturer's protocol (Supplementary Figure S2). Serial dilutions of *in vitro* synthesized PAN RNA transcript were used to generate a standard curve and derive the concentration of unknown samples.

In vitro transcription of PAN RNA

The pcDNA3.1 plasmid containing KSHV PAN RNA sequence was kindly provided by Joseph Ziegelbauer (NIH, HIV and AIDS Malignancy Branch). Transcription templates were amplified by PCR using primers listed in Table S1, and RNAs synthesized using the T7-MEGAscript system (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. RNAs were fractionated over a denaturing 8 M urea/6% polyacrylamide gel, and purified by excision followed by electroelution and ethanol precipitation. Purified RNAs were dissolved in sterile water and stored at −20°C. PAN RNA was polyadenylated using Poly(A) Polymerase Tailing Kit (Epicentre) per manufacturer's protocol. Polyadenylated PAN RNA was purified using MEGAclear Transcription clean-up kit per manufacturer's instruction (ThermoFisher). The efficiency of poly(A) tailing was assessed using the Agilent 2100 Bioanalyzer system.

In vitro RNA structure probing

A total of 5 pmol of *in vitro* transcribed PAN RNA were heated at 95°C for 3 min and slowly cooled to 4°C. The volume was adjusted to 150 μl in a final buffer of 50 mM Tris–HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$. Samples were incubated at 37°C for 15 min. Folded RNA was divided into two equal portions (72 μl each) treated with 8 μl of 100 mM 1-methyl-7-nitroisatoic anhydride (1M7) [25] in anhydrous dimethyl sulfide (DMSO) (+) or DMSO (−). Tubes were incubated at 37°C for 5 min, and RNA was precipitated at −20°C with 60 ng/μl of glycogen, 0.3 M sodium acetate (pH 5.2) and three volumes of cold ethanol. Precipitated RNA was collected by centrifugation and resuspended in 10 μl of water.

In vitro RNP interaction analysis

A total of 5 pmol of *in vitro* transcribed polyadenylated PAN RNA was heated at 95°C for 3 min and slowly cooled to 4°C. The volume was adjusted to 10 μl in a final buffer of 50 mM Tris–HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$. Samples were incubated at 37°C for 15 min. Equimolar amounts of each KSHV ORF was added to the folded RNA and incubated at 37°C for 15 min. In a separate control reaction to assess binding specificity, RNA was incubated with bovine serum albumin (data not shown). Protein–RNA solution was divided into two equal portions and treated with 100 mM 1M7 in anhydrous DMSO (+) or DMSO alone (−) as described above.

In cellulo RNA structure probing

Approximately 10^7 BCBL-1 cells were washed once with phosphate buffered saline (PBS) and resuspended in 900 μl of fresh growth medium. To modify a sample, 100 μl of 100 mM 1M7 in 100% DMSO was added (10 mM final concentration) and cells were immediately mixed and incubated at 37°C for 5 min. The negative control sample was prepared similarly, with the exception that only DMSO was added. Medium was removed and cells were washed once with PBS before isolation of nuclear and cytoplasmic fractions using Nuclei EZ Prep Kit (Sigma) per manufacturer's protocol. Nuclear and cytoplasmic RNA fractions were extracted using TRIzol (Thermo Fisher) per manufacturer's protocol. DNase I digestion (Roche) was performed for 30 min at room temperature with addition of RNase inhibitor (Promega). RNA was purified using NucleoSpin RNA Clean-up XS (MACHEREY-NAGEL).

Ex vivo RNA structure probing

To preserve native secondary structures, RNA for *ex vivo* analysis was extracted using a procedure that avoids the use of harsh chemical denaturants. Approximately 10^7 BCBL-1 cells were washed and pelleted in ice-cold PBS, and directed to isolation of nuclear and cytoplasmic fractions using Nuclei EZ Prep Kit (Sigma) per manufacturer's protocol. The nuclear fraction (∼200 μl) was resuspended in 2 ml of lysis buffer (40 mM Tris pH 7.9, 25 mM NaCl, 6 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5% Triton X-100, 1000 units/ ml RNasin (Promega) and 450 units/ml DNase I (Roche), and rotated at 4°C for 5 min. Nuclear sample was next treated with 500 μg/ml proteinase K in the presence of 40 mM Tris pH 7.9, 200 mM NaCl, 6 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5% Triton X-100, 1.5% SDS and rotated at 20°C for 45 min. RNA was extracted twice with a phenol/chloroform/isoamyl alcohol solvent (24:24:1) pre-equilibrated with 1x folding buffer (100 mM HEPES pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$), followed by one extraction with chloroform. RNA was precipitated at −20°C with 60 ng/μl of glycogen, 0.3 M sodium acetate pH 5.2 and three volumes of cold ethanol. RNA was pelleted, resuspended in 1× folding buffer and incubated at 37°C for 15 min. Approximately 5 μg of RNA were added to a 1/9 volume of 100 mM 1M7 in 100% DMSO (final concentration of 10 mM) and incubated at 37°C for 5 min. Negative control RNA was prepared in the same way but substituting DMSO for 1M7. Modified and control RNAs were purified using NucleoSpin RNA Cleanup XS (MACHEREY-NAGEL). The cytoplasmic fraction (∼4 ml) was first concentrated to 200 μl using Amicon Ultra-3K columns (Millipore), and treated as described for the nuclear fraction.

In virio RNA structure probing

Viral particles (∼1.29 × 10^8 DNA copies/μl used per single RT reaction) were resuspended in 180 μl of 1× folding buffer (100 mM HEPES pH 8.0, 100 mM NaCl and 10 mM $MgCl_2$) and equilibrated at room temperature for 15 min. For *in virio* (+) reactions, 20 μl 100 mM 1M7 in 100% DMSO was added to each suspension and mixtures were incubated at 37°C for 5 min. Negative controls were prepared likewise, except DMSO alone was added to each suspension. RNA was extracted from virions using TRIzol (Thermo Fisher) as *per* manufacturer's protocol. DNase I digestion (Roche) was performed for 30 min at room temperature in the presence of an RNase inhibitor (Promega). RNA was purified using a NucleoSpin RNA Clean-up XS kit (MACHEREY-NAGEL).

Ex virio RNA structure probing

Viral particles were lysed with 1% SDS and proteinase K (1 mg/ml final) in 0.1M Tris–HCl pH 7.5 for 30 min at 37°C as described previously (26). RNA was extracted twice with phenol/chloroform/isoamyl alcohol pre-equilibrated with 1× folding buffer (100 mM HEPES, pH 8.0, 100 mM NaCl and 10 mM $MgCl_2$), once with chloroform and precipitated at −20°C with 60 ng/μl of glycogen, 0.3 M sodium acetate (pH 5.2) and three volumes of cold ethanol. RNA was pelleted, resuspended in 1× folding buffer and incubated at 37°C for 15 min to equilibrate. Approximately 1 μg of viral RNA was then treated with 100 mM 1M7 (+) or neat DMSO (−) and subsequently purified as described for *in virio* RNA.

Denaturing control

A total of 5 pmol of PAN RNA *in vitro* transcript in 3 μl of sterile water, 5 μl 100% formamide and 1 μl 10× denaturing buffer (500 mM HEPES pH 8.0, 40 mM EDTA) were mixed and incubated at 95°C for 1 min to denature RNA. To 9 μl of denatured sample, 1 μl of 100 mM 1M7 was added and allowed to react at 95°C for 1 min and then placed on ice. The total volume of denaturing control was adjusted to 50 μl with sterile water, and purified using a G-50 spin column (Fisher Scientific).

PAN RNA SHAPE-MaP

To maximize the amount of usable sequence we could obtain in SHAPE-MaP experiments (27), the 1077 nt PAN RNA sequence was divided into four overlapping ∼280 nt zones and zone-specific RT and PCR primers were designed accordingly (Supplementary Table S1). RNA recovered from (+), (−) and denatured RNA (28) probing reactions were each divided into four aliquots and subjected

to mutagenic reverse transcription initiating from zone-specific RT primers, respectively, as previously described. The resulting cDNAs were purified over a G-50 spin column and amplified separately by PCR using Q5 high-fidelity DNA polymerase (NEB), zone-specific PCR primers and by cycling as follows: 1 cycle (98°C, 30 s), 20 cycles (98°C, 10 s; 50°C, 30 s; 72°C, 30 s) and 1 cycle (72°C, 2 min). Products of these reactions were purified using a PureLink Quick PCR micro kit (Thermo Fisher Scientific) and subjected to a second round of PCR to add indices and sequencing adapters under the following conditions: 1 cycle (98°C, 30 s), 10 cycles (98°C, 10 s; 50°C, 30 s; 72°C, 30 s) and 1 cycle (72°C, 2 min). The resulting amplicons were purified from a 2% agarose gel by electroelution and their concentrations measured by Q-PCR using KAPA Universal Library Quantitation kit (KAPA Biosystems). Sequences of all primers used in RT and the two PCR reactions are provided in Supplementary Table S1.

Sequencing and SHAPE profile generation

Purified, individually-indexed amplicon libraries were pooled and sequenced on an Illumina MiSeq instrument to generate zone- and reaction-specific 300 nt single-read datasets. These datasets were automatically assembled into separate FASTQ files for each of the respective zones and reactions (i.e. Z1-plus, Z2-plus, Z3-plus, Z4-plus, Z1-minus, etc.). Custom Python scripts (available upon request) were used to trim primer hybridization site sequences from the respective datasets and separately merge the 'plus', 'minus' and 'dn' FASTQ files into master files (PANallPlus.fastq, PANallMinus.fastq and PANallDN.fastq) that were inputted into ShapeMapper via the configuration file (27,29). The ShapeMapper pipeline was otherwise executed as previously described, using all default parameters except 'randomlyPrimed', which was set to 'off'. Also, the folding module of *ShapeMapper* was not used.

Structure modeling

Minimum free-energy secondary structure models were generated using RNAstructure (v 5.7) from reactivity values obtained by SHAPE-MaP and resetting the default parameters as follows: maximum pairing distance <none>, SHAPEslope = 2.6, SHAPEintercept = −0.8 (30). The models presented in this work are the lowest energy conformers generated by the software altered only by manual removal of unlikely single- or tandem-basepaired motifs. As previously described (21,27), local median SHAPE reactivity and Shannon entropy values were calculated over a centered, 55-nt sliding window. To determine whether one or more pseudoknots might be present in PAN RNA, we inputted our reactivity data into *ShapeKnots* (31), reset the maximum pairing distance, SHAPEslope and SHAPEintercept parameters as described above, and iteratively folded the RNA in 600-nt windows offset by 100-nt increments. No potential pseudoknots were identified (data not shown).

Correlation analysis

Correlation analysis of SHAPE-MaP data were performed using JMP version 11.0.

Identification of regions that experienced conformational changes

Reactivity values obtained by probing RNA in the nucleus or *ex vivo* were averaged over a 21 nt sliding window, and the differences between these averages calculated. In cases where the absolute differences were greater than the median, the ratio of average *ex vivo* to *in vivo* nuclear reactivity values were plotted against nucleotide number. This analysis closely resembles that conducted previously (21).

Identification of protein binding sites

ΔSHAPE values were calculated by comparing reactivity profiles obtained from PAN RNA probed *ex vivo* and *in cellulo* as described previously (27,29). The standard error in the SHAPE reactivity measurements was used to perform a modified Z -factor test (32), which compares the magnitude of ΔSHAPE with the associated *ex vivo* and *in cellulo* measurement errors, identifying nucleotides for which the magnitudes of the errors are too large for the ΔSHAPE values to be significant. The Z -factor test was formulated such that the underlying *ex vivo* and *in cellulo* SHAPE reactivities must differ by more than 1.96 standard deviations (Z-factor >0), ensuring that the 95% confidence intervals of each measurement do not overlap. The standard score at each nucleotide was calculated to identify the largest ΔSHAPE values, which are expected to reflect the most stable protein–PAN RNA interactions. The requirement of each standard score to be ≥1 was set, meaning that individual ΔSHAPE values must be at least one standard deviation from the mean ΔSHAPE. To identify protein–PAN RNA interaction sites, we filtered ΔSHAPE values by Z-factor and standard score simultaneously. Stable protein binding sites were defined as having at least 3-nt in a 5-nt window with Z-factors of >0 and absolute standard scores of ≥1. The residues that fulfilled the above requirements, but showed slightly lower standard score (0.75 ≥|S| ≥1) were also considered.

PAN RNA Domain III 3D modeling

The secondary structural model of *ex vivo* nuclear PAN RNA depicted in Figure 1 was generated using RNAstructure from reactivity values obtained via SHAPE-MaP. Using the same software, this model was exported to dot-bracket notation in text format, from which the sequence and base-pairing data pertaining to Domain III nt 793–1077 were extracted. These data were inputted into RNAComposer, an automated RNA structure modeling server that assembles models from segments of the test RNA homologous to established structural motifs, followed by several rounds of energy minimization (33) http://rnacomposer.cs.put.poznan.pl/). Three-dimensional (3D) models predicted by this software were inspected in PyMol (The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA) and found to have steric conflicts and inconsistencies attributable to a kink imposed at the base of H22 (nt 1039) and overly rigid single stranded regions, particularly nt 982–999 and nt 1005–1038. Conflicts were resolved by imposing a total of five sets of distance constraints among these residues, thereby effectively breaking the artificial uniformity of stacking interactions in
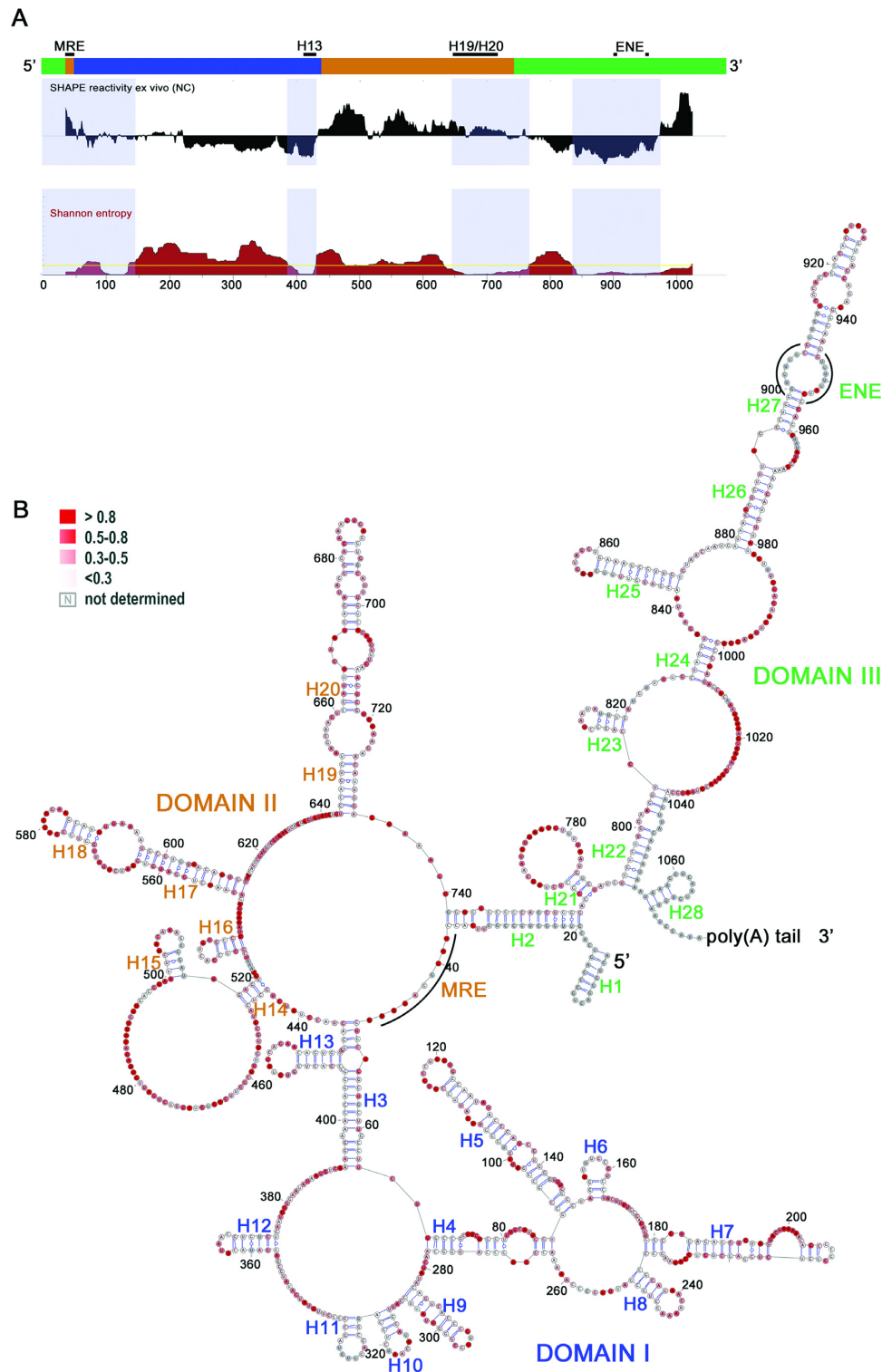
Figure 1. Architecture of the KSHV PAN RNA. (**A**) Location of three main domains: i (blue), ii (orange), iii (green) and selected motifs in PAN RNA structure is indicated (top). *Ex vivo* 1M7 reactivities obtained for nuclear PAN RNA shown as the median reactivity over 55-nt sliding windows (middle). Shannon entropy values for the *ex vivo* secondary structure model smoothed over 55-nt sliding windows (bottom). High values indicate regions that probe many possible conformations, and low values indicate a well-defined structure. Yellow line indicates the global median. Gray shading marks well-defined structures with low SHAPE and low Shannon entropy. (**B**) Secondary structure model of PAN RNA color-coded according to SHAPE-MaP reactivity. Red notations are predicted to fall into single-stranded regions, whereas bases indicated in white correspond to constrained residues. Gray nucleotides correspond to residues for which no reactivity values were obtained, due to either pausing during reverse transcription or the position of primer hybridization. Bases are labeled every 20 nts. Helix numbering (H) is from the 5′ to 3′ end, and helices are differentiated if they are separated by either (i) a junction, (ii) an internal loop >12 nt in total or (iii) an asymmetric internal loop with zero residues present on one side and >6 nt on the other side. The three major domains (i–iii), together with *cis*-acting motifs (MRE core, ENE and poly(A) tail) are indicated.

these single-stranded regions, reducing the amount of space the modeled versions of these regions occupied and increasing conformational flexibility at junctions between these regions and adjacent double-stranded motifs. Models generated with these restraints imposed did not contain conflicts of the originals. To examine how a poly(A) tail might be accommodated in the PAN ENE, a series of Domain III dot-bracket files were generated in which a 44 nt poly(A) tail was added to the Domain III 3′-terminus and 5 nt segments of this tail were hybridized to the distal 5-Us of ENE (nt 947–951) as per the PAN ENE crystal structure (20) in 5 nt increments. The version in which poly(A) nt 21–25 hybridized to ENE yielded the most plausible result and was included in Figure 5.

RESULTS

Nuclear PAN RNA adopts a branched secondary structure

We applied SHAPE-MaP to interrogate the secondary structure of PAN RNA isolated from the nuclear fraction (Supplementary Figure S2) of infected cell extracts and following mild extraction to remove bound proteins (nuclear *ex vivo*). Briefly, RNA is exposed to an acylating agent that selectively modifies unpaired nucleotides (22,34–36), after which it is reverse transcribed under conditions where modified nucleotides are likely to produce a mutation at the complementary position in the cDNA product. cDNA libraries are subsequently amplified by PCR and subjected to deep sequencing. The collective mutation frequencies at individual nucleotide positions of the probed RNA, or mutational profile, directly reflect nucleotide reactivity toward the acylating agent and therefore the likelihood that it is single or double-stranded in the context of the global RNA secondary structure. This information is used in RNA secondary structure prediction algorithms to assemble maximum likelihood models of RNA structure.

SHAPE-MaP offers numerous advantages compared to other RNA probing techniques (27,29). Importantly, because 1M7 is cell permeable, the structures of intracellular (or intra-virion) RNAs can be interrogated. In addition, because the cDNA library is amplified by PCR, the sensitivity of SHAPE-MaP far exceeds other methods. Quantitation of SHAPE-MaP results is highly precise, given the binary nature of mutagenesis and the fact that mutations can easily be counted and compared to the total number of sequence reads. This precision also allows researchers to calculate variability in reactivity values and, consequently, Shannon entropies, which inversely correlate with the frequency that an RNA motif is present among conformers in a population (30,37). Regions of RNA sequence with high Shannon entropies are likely to sample many different conformations, while those with low entropies are expected to exist in a single conformational state. In general, nucleotides within RNA motifs whose structure is related to function exhibit low reactivity values and Shannon entropies (32,38). To measure the persistence of RNA structural motifs among conformers, reactivity values and Shannon entropies are generally averaged over a 50–100 nt sliding window.

Our analysis indicated that *ex vivo* nuclear PAN RNA adopted a branched secondary structure that can be roughly

divided into three domains defined by 5′ (Domain I; nt 48–435), internal (Domain II; nt 38–48, 436–741) and 3′ nucleotides (Domain III; nt 1–37, 742–1077), within which we defined 28 hairpin (H) motifs (Figure 1A and B). The previously identified ORF57 binding site, i.e. the MRE motif, spans the junction between all three domains (nt 1–84) and was originally predicted to fold into three adjacent hairpins (1,18). Our analysis indicates that the MRE core region (nt 38–46, UAUGGAUUU) adopts an unstructured, single-stranded conformation, possibly providing an ORF57 landing pad (Figure 1A and B). This motif is also characterized by low average SHAPE reactivity and Shannon entropy, indicating it persists among conformers and therefore likely serves an important structure-dependent function (Figure 1B). Downstream of the MRE core, Domain I forms a multi-furcated series of stem loops (H3-H13) assembled around three junctions. Although average Shannon entropy values calculated for this region are relatively high, H13 is a notable exception. As this motif is structurally contiguous with the MRE core sequence, it may help maintain the single-stranded character of this element. Domain I also has low average SHAPE reactivity and the structure presented is predicted to exist in each of the three lowest free energy models.

Domain II comprises a series of hairpins (H14–H20) arranged around a largely unstructured six-way junction (U436–U740) that also houses the MRE core (Figure 1A), and is the least structured of the three domains. With the notable exceptions of the core sequence and the H19/H20 stem loop, Domain II is characterized by high average SHAPE reactivity and Shannon entropy values, indicating structural variability (Figure 1B). The H19-H20 motif was predicted to exist in the 10 lowest free energy models calculated.

Domain III assumes an extended multi stem-loop structure (H1–H2, H21–H28) terminating in a reactive apical loop (nt 925–929) (Figure 1A). This domain included an asymmetric internal U-rich loop within H27 (nt 901–905, nt 947–953) which, in conjunction with G•C base pairs of the lower stem (nt 898–900, 954–956) form the functional core of the previously defined ENE motif (20,39). Mutational analysis and X-ray crystallographic studies show that the ENE motif interacts with the 3′ poly(A) tail of PAN RNA to form a triple helix that contributes to its stability (19,39). Here, we place the ENE motif within the larger context of the global PAN structure. We were unable to directly assess reactivity of the ENE loop to 1M7, as reverse transcriptase frequently terminated prematurely at or near template nt 900–905 and 947–951. This impediment to reverse transcription is indicative of a strong RNA tertiary interaction such as the triple helix predicted to form within the ENE motif. The majority of Domain III is highly structured and characterized by low average SHAPE reactivity and Shannon entropy, suggesting the function of this PAN segment is structure-dependent (Figure 1B).

PAN sequence polymorphisms do not affect its structural organization

To assess whether our nuclear PAN structural model was susceptible to change arising from naturally occurring sequence polymorphisms, we performed secondary structure

Table 1. Single-nucleotide polymorphisms identified within PAN sequence and their structural context (disrupted pair, structurally conservative mutations, single-stranded regions)

Sample #	SNP	Position	Structure	Junctional bp
ZM095	C→U	215	ds, G•C to G•U	Yes
VG-1	A→G	264	ds, A•U to G•U	Yes
ZM027, ZM091, ZM095, ZM0102, ZM0106, ZM0108, ZM0114, ZM0116, ZM0117, ZM0118, ZM0121, ZM0123, ZM0124, ZM0128, ZM0130, VG-1	C→U	425	ds, G•C to G•U	Yes
ZM0123	AUAG del	486-489	ss	
ZM095, ZM0106, ZM0114, ZM0117, ZM0123, ZM0124, ZM0128, VG-1	A→C	505	ss	
ZM095, ZM0106, ZM0116, ZM0118, ZM0123, ZM0124, ZM0128, VG-1	U→A	526	ss	
ZM0123	G→A	558	ds G•U to A•U	No
ZM095, ZM0106, ZM0116, ZM0118, ZM0124, ZM0128	C→G	577	ds C•G to G•G	Yes
VG-1	A→G	594	ss	
ZM095, ZM0106, ZM0116, ZM0118, ZM0123, ZM0124, ZM0128, VG-1	A insert	597	ss	
ZM095, ZM0106, ZM0116, ZM0118, ZM0123, ZM0124, ZM0128, VG-1	U→C	644	ss	
VG-1	C→A	679	ds C•G to A•G	Yes
ZM095, ZM0106, ZM0116, ZM0118, ZM0123, ZM0124, ZM0128, VG-1	U→A	716	ds, U•A to A•A	No
ZM0123	U→C	771	ss	
ZM095, ZM0106, ZM0116, ZM0118, ZM0124, ZM0128	G→U	792	ss	
VG-1	A→C	803	ds, A•U to C•C	Yes
VG-1	A→U	814	ss	
ZM095, ZM0106, ZM0116, ZM0118, ZM0123, ZM0124, ZM0128, VG-1	C→U	852	ds, C•G to U•G	Yes
ZM0123	C→A	876	ss	
ZM095, ZM0106, ZM0116, ZM0118, ZM0124, ZM0128	G→U	895	ds, G•U to U•U	Yes
VG-1	C→A	924	ds, C•G to A•G	Yes
ZM0121	A→G	962	ss	
VG-1	U→A	965	ss	

predictions on PAN sequences derived from KSHV-positive cell lines (VG-1 and BCBL-1) and Kaposi's sarcoma skin lesions of 15 different Zambian patients (40). The results of this experiment are given in Table 1. A total of 12 mutations, including 1 deletion (nt 485–489) and 1 insertion (nt 597) localized to unstructured, single-stranded regions, while 11 SNPs localized to predicted base pairs. In the latter group, five SNPs are predicted to preserve base pairing (e.g. G•C to G•U at nt 215, A•U to G•U at nt 264), while the remaining six localized to junctional base pairs closing hairpins or internal loops, likely having only a minimal effect on the motif. Only the U716A SNP within the stem of H20 has the potential to affect the predicted structural element.

We also used the RNAsnp web server to evaluate the effect of SNPs on local folding of PAN RNA structure *de novo* (23). RNAsnp features both global and local folding methods to compute the ensemble of secondary structures. The analysis predicted one possible structural change within H20 (nt 660–718) induced by C679A and U716A SNPs identified in the VG-1 cell line (*P*-value 0.02; Supplementary Figure S3). The other 21 polymorphisms did not impose significant structural changes, indicating that

the majority of SNPs would have a negligible effect on PAN RNA structure.

Probing nuclear PAN RNA *in cellulo* reveals protein binding sites

By exploiting the cellular permeability of 1M7, we modified cellular PAN RNA *in situ* (*in cellulo*), extracted the nuclei and further processed this RNA via SHAPE-MaP. By contrasting these results with those from our *ex vivo* experiment, we could assess how the nuclear environment affects PAN RNA structure (Figure 2A). Values >1 likely indicate regions of RNA protected from acylation by nuclear proteins, while those <1 reflect an increase in reactivity, perhaps due to RNA structure relaxation invoked by protein binding.

In Domain II, residues comprising the MRE core (nt 41–55), which were largely unstructured *ex vivo*, were comparatively unreactive *in cellulo*, implicating association of one or more proteins with this region. Similar observations were made for sequence segments nt 88–107, 109–111, 118–129, 194–210, 473–491, 511–513, 571–595, 597–600, 723–732, 763–773, 830–834, 836–864, 928–934, 937–953, 989–
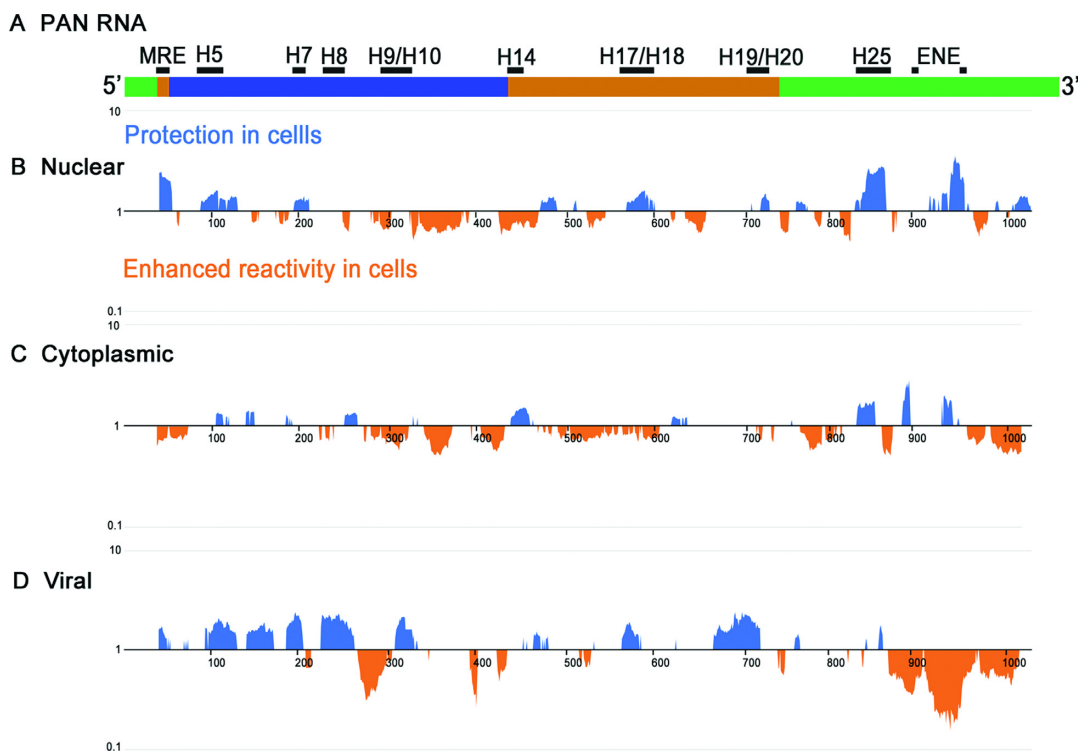
Figure 2. The modulating effect of (**B**) nuclear, (**C**) cytoplasmic and (**D**) viral environments on PAN RNA structure (**A**), location of three main domains: i (blue), ii (orange), iii (green) and selected motifs in PAN RNA structure. Charts represent the ratio between positive (blue) and negative (orange) reactivity differences within regions of substantial reactivity change. Blue and orange peaks indicate regions where protections or enhancements, respectively, are most pronounced. Red bars represent the position of RT stops. Bases are labeled every 100 nts.

992, 1011–1028 across all three domains. Conversely, regions in which *in cellulo* reactivity values are greater than those obtained *ex vivo* are mostly unstructured, suggesting the overall RNA structure is minimally impacted by protein binding.

While simple difference mapping can be used to identify protected regions on PAN RNA, local conformational sampling among RNA motifs is not factored in these calculations. To address this issue, and as a more stringent method for detecting protein binding sites, we calculated ∆SHAPE values as previously described (27,29). Because ∆SHAPE values factor both absolute reactivity differences as well as variation in reactivity value sets over a smaller sliding window (3 versus 21 nt), results of this analysis are more precise than when considering reactivity differences alone.

Most sites whose ∆SHAPE values satisfied the established statistical criteria were also within the regions of protection delineated in Figure 2A, were deemed protein binding sites and mapped onto the nuclear *ex vivo* RNA structural model (Figure 3A). These sites include nt 97–99 (H5) in Domain I, nt 44–48 (MRE core) and 481–484 in Domain II, and nt 835–837 and 853–855 (H25) in Domain III. It should be noted that ∆SHAPE analysis relies on measuring reactivity differences, so in the case of protein contacts that involve transient interactions or proteins that recognize regions that probe multiple conformations, the impact of binding upon the overall SHAPE measurement might be reduced. Thus, our ∆SHAPE analysis also includes the weaker/transient RNP contacts (i.e. Z-factor

>0, standard score 0.75–0.95) that were identified within nt 538–540 (H16), 651–653 (H19), 964–966 and 978–980 (H26) (41).

Protein binding sites in cytoplasmic PAN RNA

It has been reported previously that up to 20% of PAN RNA is exported from the nucleus to the cytoplasm (1). In addition, gene expression profiling has indicated three putative ORFs near the 5′ terminus of PAN that may be translated, potentially explaining the cytoplasmic localization of this lncRNA (42). SHAPE-MaP was therefore applied to cytoplasmic PAN RNA to interrogate potential alterations to its structure and its association with cellular proteins (Supplementary Figure S2).

The overall topologies of cytoplasmic and nuclear PAN RNA were comparable, sharing three branched domains housing well-defined motifs and less structured, dynamic segments (Supplementary Figures S4 and 5). The SHAPE-MaP reactivity profile for *ex vivo* cytoplasmic PAN RNA exhibited good correlation with the *ex vivo* nuclear profile (Spearman's $\rho = 0.69$) (Supplementary Figure S6A). While small differences were noted in Domain I (G256-G306, H4), Domain II (G445-C618, H14-H17) and Domain III (G755-U835, H21-H24; U912-C939, H27; C983-A989, A1005-C1011, H24; C1039-U1069, H22), most motifs, including the MRE and ENE, remained structurally unchanged.

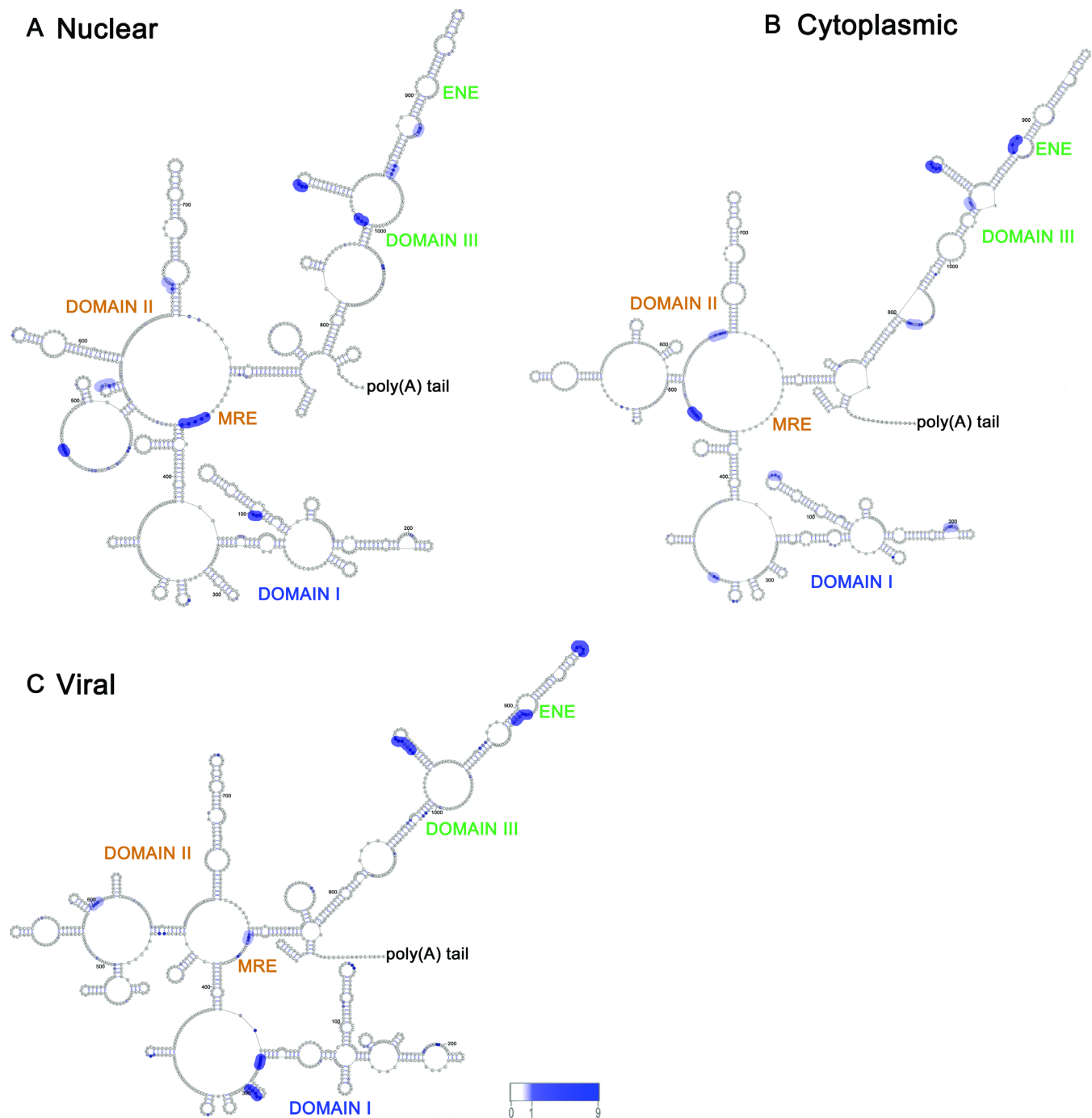Regions of low SHAPE reactivity and Shannon entropy (Supplementary Figure S4A) include all of Domain I, a

Figure 3. Protein binding sites detected by ΔSHAPE analysis indicated on secondary structure of (**A**) *ex vivo* nuclear, (**B**) *ex vivo* cytoplasmic and (**C**) *ex virio* PAN RNA. Nucleotides protected *in cellulo* are colored dark blue for strong and stable interactions (Z-factor >0, |S| ≥1) and light blue for weaker, transient interactions (Z-factor >0, 0.75 ≥|S| ≥1). The blue gradient scale under each structure indicates the standard score values.

small portion of Domain II (nt 720–757) and the majority of Domain III (nt 758–790, 835–990). Previously, it was shown that coding regions within the HIV-1 genome are prone to being unstructured (43). However, it seems that this tendency is not absolute, since the proposed ORFs hosted within the PAN transcript (PAN1.1, nt 39–92, PAN1.2, nt 215–349, PAN1.3, nt 272–349) (42) are entirely sequestered within structured Domain I. Most of Domain II (nt 440–

625, 665–715) and segments of Domain III (nt 791–834, 991–1020) were characterized by high SHAPE reactivity and Shannon entropy values, supporting the observation that the most prominent conformational rearrangements relative to the nuclear PAN RNA structure are localized to these regions (Figure 1 and Supplementary Figure S4).

Reactivity difference mapping between cytoplasmic PAN RNA probed *ex vivo* and *in cellulo* indicated multiple pro-

tected regions in Domains I (nt 109–116, 141–146, 148–152, 256–270), II (nt 446–467, 630–639, 645–647) and III (nt 842–862, 894–902, 941–951) (Figure 2B). Conversely, regions in which *in cellulo* reactivity values are greater than those obtained *ex vivo* are mostly unstructured and roughly correlate with corresponding regions observed for PAN RNA probed in nucleus (Figure 2A and B).

In the ΔSHAPE analysis, protein binding sites in cytoplasmic PAN RNA were identified within nt 454–456 of Domain II, and nt 852–855 and 892–894 of Domain III (H25-H26) (Figure 3B). More transient binding was mapped to nt 120–122 (H5), 198–200 (H7), 331–333 (H11) and 342–344 in Domain I (Z-factor >0, median standard scores 0.78–0.85), nt 639–643 in Domain II (Z-factor >0, median standard score 0.88) and nt 835–837 and 1035–1037 in Domain III (Z-factor >0, median standard scores 0.78–0.96). In general, binding patterns appear similar in Domains I and III of nuclear and cytoplasmic PAN RNA, respectively. The most notable, and arguably the most physiologically significant difference in binding sites is observed for the MRE core. In nuclear PAN RNA, the binding site located within the MRE comprises the largest contiguous segment satisfying the statistical requirements of both ΔSHAPE and reactivity difference calculations. In stark contrast we found no evidence of protein binding at this site in PAN RNA isolated from the cytoplasm (Figure 3A and B).

PAN RNA maintains its structure within KSHV virus particles

Although PAN RNA has been detected in KSHV virions, its structure and functionality in this context have not been addressed (15,16). Our studies indicate that the *ex virio* secondary structure of PAN RNA was similar to its cellular counterparts, although some significant differences were notable (Figure 1; Supplementary Figures S7 and 8). The *ex virio* and *ex vivo* nuclear SHAPE-MaP profiles for PAN RNA were moderately correlated (Spearman's ρ = 0.54) (Supplementary Figure S6B). Although the branched domains were distinguishable in the *ex virio* structure, significant rearrangements were evident in all three (G48–U60 and U164–A230, C435–C624 and U791–U836, respectively). However, structurally well-defined and persistent motifs identified in PAN from in different cellular contexts and having low SHAPE reactivity and Shannon entropy values (i.e. the MRE core, H5, H8, H10–H12, H19–H20 and H25–H27) were also found in PAN extracted from virions (Supplementary Figure S7).

Comparative SHAPE analyses of the HIV RNA genome probed *in virio* and *ex virio* revealed pronounced, localized reactivity differences produced by nucleocapsid and/or tRNA binding (26,44). A similar SHAPE-MaP analysis of PAN RNA in virions delineated multiple sites of protection within Domains I (nt 48–52, 100–131, 143–173, 188–206, 227–264, 311–332), II (nt 38–47, 468–475, 568–587, 672–724) and III (nt 764–769, 859–863) (Figure 2C). Notably, the region in Domain III spanning nt 870–1017 is characterized almost exclusively by *in virio* reactivity values greater than those obtained *ex virio*. A potential explanation for this observation is that protein binding in this environment selectively destabilizes PAN ENE and flanking helices.

More precise protein contact sites were revealed by a parallel ΔSHAPE analysis (Figure 3C); specifically, within nt 278–283 and 297–301 (H9) of Domain I and nt 849–855 (H25), 923–927 and 954–956 (H27) of Domain III. Unlike in the nuclear and cytoplasmic RNAs, some *in virio* sites identified by ΔSHAPE do not correlate with regions of protection depicted in Figure 2C. Weaker and perhaps transient RNP binding sites were also identified in sequence segments nt 38–41 and 579–581, many of which co-locate with those identified intracellularly, suggesting that protein association with PAN RNA is maintained throughout virus assembly and packaging.

KSHV proteins bind discrete PAN RNA motifs

Four KSHV proteins, designated by their open reading frames, have been implicated in PAN RNA binding, namely ORF26 (13), ORF57 (1), ORF59 (45) and ORF73/LANA (46). According to the sequence-based RBscore web server, which uses a machine learning approach to identify RNA binding motifs (47), each of these proteins possess significant RNA binding character (data not shown).

To map precise binding sites for these KSHV proteins on PAN RNA, we incubated partially purified recombinant protein with polyadenylated, *in vitro* transcribed RNA and calculated ΔSHAPE profiles for each. The reactivity profile for the PAN RNA transcript correlated well with *ex vivo* nuclear (Spearman's ρ = 0.58) and *ex vivo* cytoplasmic (Spearman's ρ = 0.67) datasets (Supplementary Figures S6C and D, S9 and S10), although small differences were evident within the 5′ MRE core, which in the *in vitro* transcript is partially base paired (nt G41-G42) and other areas (Domain I: 284–308, 325–338; Domain II: 500–519; Domain III: 754–830, 1006–1077). Also, we did not observe premature termination of reverse transcription within the ENE of the *in vitro* transcript, suggesting that triple helix formation in polyadenylated PAN RNA likely requires the presence of yet to be defined accessory proteins.

PAN RNA was previously identified as one of the targets of ORF57 (1). A CLIP assay utilizing anti-ORF57 antibodies indicated that ORF57–PAN contact sequences clustered within the 5′ half (nt G42–G48, U60–G80), 3′ half (nt U888–U966) and in the middle of PAN (nt G353–C402) (48). These results were later confirmed by two other groups (49,50), and it was proposed that ORF57 binds to the nascent PAN RNA transcript, multimerizes and forms multiple RNP contacts. In support of these observations, our ΔSHAPE analysis identified multiple ORF57–PAN RNA interactions, including sites of weaker binding within the 5′ MRE core (nt 45–47) and strong binding within nt 120–122 (H5) and nt 338–341 in Domain I, nt 481–483, 589–591 (H18), 602–604 (H17) in Domain II and nt 775–780 and 949–951 (21) in Domain III (Figure 4). An additional, possibly transient interaction site was noted within nt 435–437 (H13, Z-factor >0, median standard score 0.8).

The 3′ PAN Domain III (nt 750–1062) was also identified as an interacting region for ORF73/LANA, or latency-associated nuclear antigen, which acts as the origin binding protein and chromatin anchor of the extrachromosomal viral genome (46,51). ΔSHAPE analysis verified previous findings by identifying the LANA–PAN interaction
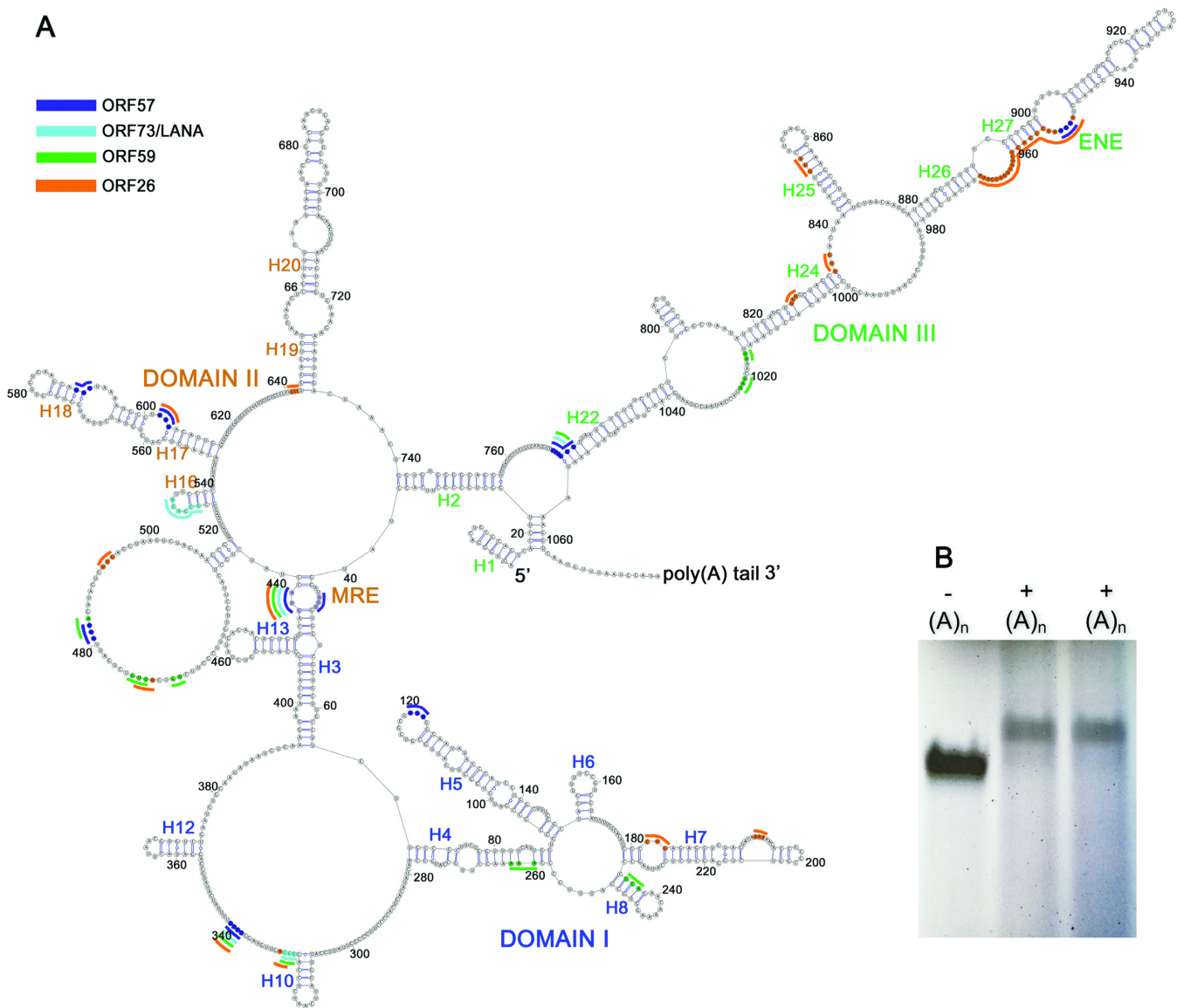
Figure 4. Secondary structure of polyadenylated *in vitro*-transcribed PAN RNA with mapped KSHV ORFs interaction sites. (**A**) RNP sites for each ORFs are color-coded as follows ORF26, orange), ORF57, blue, ORF59, green and ORF73/LANA, turquois. (**B**) Non-denaturing agarose gel indicating efficiency of PAN RNA poly(A) tailing.

site at nt 776–778 (H22). Additional RNP contact sites specific for LANA were mapped to nt 326–328 and 530–535, while other sites were shared with ORF57 (nt 338–341, 435–437, 602–604) (Figure 4). The broadest LANA–PAN interaction site mapped to G528–C541 (H16). This stem-loop motif was also identified as an RNP contact in nuclear, but not cytoplasmic PAN RNA (Figure 3A and B).

ORF59 is a KSHV DNA polymerase processivity factor that dimerizes in the cytoplasm, binds the polymerase and translocates it to the origin of lytic DNA replication in the nucleus (41). ORF59 and PAN RNA were found to colocalize to same segments on the KSHV genome (13,14), and it has been proposed previously that ORF59 is involved in the targeting mechanism of PAN to the KSHV episome (52). ΔSHAPE analysis indicates multiple transient ORF59–PAN contacts within Domain I (nt 235–237,

260–261, 327–329, 338–341), although the most extensive ORF59–PAN contact points were identified within Domains II (nt 466–468, 471–473, 481–483) and III (nt 776–778 and 1016–1023) (Figure 4). Of these, the nt 338–341 and 1016–1023 RNP sites were identified in ΔSHAPE analysis to be specific for cytoplasmic PAN RNA (Figure 3).

Finally, although little is known regarding the role ORF26–PAN RNA interaction, we used SHAPE-MaP and ΔSHAPE analysis to map ORF26 binding sites. ORF26 binds near the ENE (nt 948–971) and other regions including nt 182–184, 199–201, 327–329, 338–341, 435–437 in Domain I and weaker interactions within nt 470–472, 492–494, 602–604, 642–644 in Domain II and nt 826–828, 835–837, 849–851 in Domain III (Figure 4). It should be noted that for ORF57, ORF59 and ORF73/LANA the physiological importance of PAN RNA binding in the cytoplasm is un-

clear, since these viral gene products are thought to primarily localize to the nucleus (53).

DISCUSSION

KSHV has been classified as a direct carcinogen (54) with an estimated 44 247 new cases and 26 974 deaths worldwide in 2012 (http://globocan.iarc.fr/). In the HAART era, Kaposi's sarcoma remains the second most frequent tumor in HIV-infected patients worldwide and has become the most common cancer in Sub-Saharan Africa (55). Despite the global health burden of KSHV-associated cancers, we still lack a full understanding of KSHV replication and pathogenesis.

Upon infection, KSHV establishes latency by expressing viral products that modulate expression of cellular genes and allow the virus to subvert host immunity (56). The latent to lytic switch is marked by production of a highly abundant lncRNA, PAN, that interacts with cell- and virus-encoded factors to regulate the immune response gene expression (13–15). In this study, we determined the structure of PAN RNA in both the nucleus and cytoplasm, as well as inside viral particles and mapped sites of protein binding to this lncRNA.

Our studies indicated that PAN RNA folds into three branched domains, each of which contains both well-defined and less-constrained structural motifs (Figure 1; Supplementary Figures S4 and 7). The character of PAN RNA structure, i.e. the percentage of nucleotides base-paired, number of helices, number of junction regions, closely resemble those of previously characterized SRA and HOTAIR lncRNAs, supporting the validity of our model (57) (22,35) (Supplementary Table S2). The structures of at least those PAN RNA motifs with low reactivity and Shannon entropy values are likely to be functionally crucial, as they are maintained in and among conformeric mixtures probed *ex vivo* and *ex virio*. This is in stark contrast to transcriptome-wide analyses of messenger RNA showing that coding RNAs are relatively unstructured in a cellular context (58), perhaps due to processing by RNA helicases, single stranded RNA binding proteins and other factors that promote conformational sampling (59).

In our characterization of *ex vivo* nuclear PAN RNA, we mapped the previously identified 5′ MRE region and 3′ ENE (1,2,18–20) and placed them in a larger structural context (Figure 1; Supplementary Figures S4 and 7). ENE interacts with the PAN poly(A) tail to form a triple helix which has been characterized by X-ray crystallography (20). This motif has been reported to inhibit degradation and promote accumulation of PAN RNA in the nucleus of infected cells, although it appears this function may be context specific. The triple-helical structure of PAN ENE has a counterpart in MALAT1, an important cancer-related lncRNA, which also promotes its nuclear retention (60). An ENE deletion has also been shown to reduce steady-state levels of PAN RNA in a recombinant expression systems (2). When the 79-nt ENE was inserted near the 3′-end of a heterologous polyadenylated β-globin transcript, the level of chimeric RNA was increased 5-fold. Insertion of a larger fragment of PAN containing ENE increased expression 10-fold, suggesting that the spatial relationship between ENE and the poly(A) tail may be crucial for triple helix forma-

tion and ENE function. More specifically, it seems likely that triple helix formation would be favored in a structural context in which the trajectory of the tail emerging from the 3′ end of PAN RNA Domain III was toward, not away from, the ENE motif. A 3D model derived from our secondary structural map of PAN RNA Domain III suggests that the structural context of this domain may indeed produce a poly(A) trajectory conducive to triple helix formation. This model, together with the PAN ENE structure resolved by X-ray crystallography, are depicted in Figure 5.

Domains I and III both contain additional elements with well-defined secondary structures whose complexity is comparable to those of other viral and lncRNAs (21,38). Domain II is less structured, which perhaps serves to provide access to proteins or support more compact folding in adjacent regions. The predicted structural models for *ex vivo* cytoplasmic, *ex virio* and *in vitro* transcribed PAN RNA differ to some extent within Domain II, as well as small portions of Domains I and III (Figures 1 and 4; Supplementary Figures S4, 5, 7 and 8). This is not surprising, since the regions that differ among isolates are relatively unstructured and characterized as having high Shannon entropy values associated with frequent conformational sampling.

Importantly, structured regions of PAN RNA also appear to be conserved among patient isolates, suggesting they are likely linked to function (Table 1 and Supplementary Figure S3). The native structure of PAN RNA may also be influenced by post translational modifications, of which N6 methyladenosine (m^{6}A) is the most common. Two algorithms that identify consensus target sequences for m^{6}A modification independently predict methylation of adenosines 7 and 733, adjacent to helices H1 and H19, respectively (61,62). Since post-transcriptional mRNA and lncRNA modifications have been shown to cause structural remodeling and affect protein–RNA interactions (63,64), these modifications may also have an influence on PAN RNA structure and function.

Comparing SHAPE reactivity values obtained from PAN RNA probed in the presence and absence of protein, together with more stringent ΔSHAPE analyses, demonstrates extensive interactions between this lncRNA and cellular/viral proteins (Figures 2 and 3). One particularly intriguing result was that protein binding at the MRE core appears exclusive to PAN RNA localized to the nucleus, as we found no evidence that this motif of cytoplasmic and virion-associated RNA was protected from acylation. This is consistent with previous findings demonstrating that ORF57, the reported binding partner of the MRE, is itself primarily nuclear, suggesting that ORF57-MRE binding may also contribute to nuclear retention of PAN RNA (Figures 2A, B and 3A, B) (65,66). Other ORF57–PAN interactions were also evident in our experiments, including near the ENE (Figure 4). Together, our findings are consistent with, and offer refinements of, a previous report that identified multiple ORF57 binding sites (1). There is ample evidence that both the poly(A) interaction with the ENE motif and ORF57 binding to the MRE promote PAN RNA stability in cells (1,2,67). It has been shown that ENE-lacking or ENE-mutant PAN transcripts are more likely to be degraded in the absence of ORF57 and that ORF57 increases wild-type PAN RNA levels in a dose-dependent fashion,
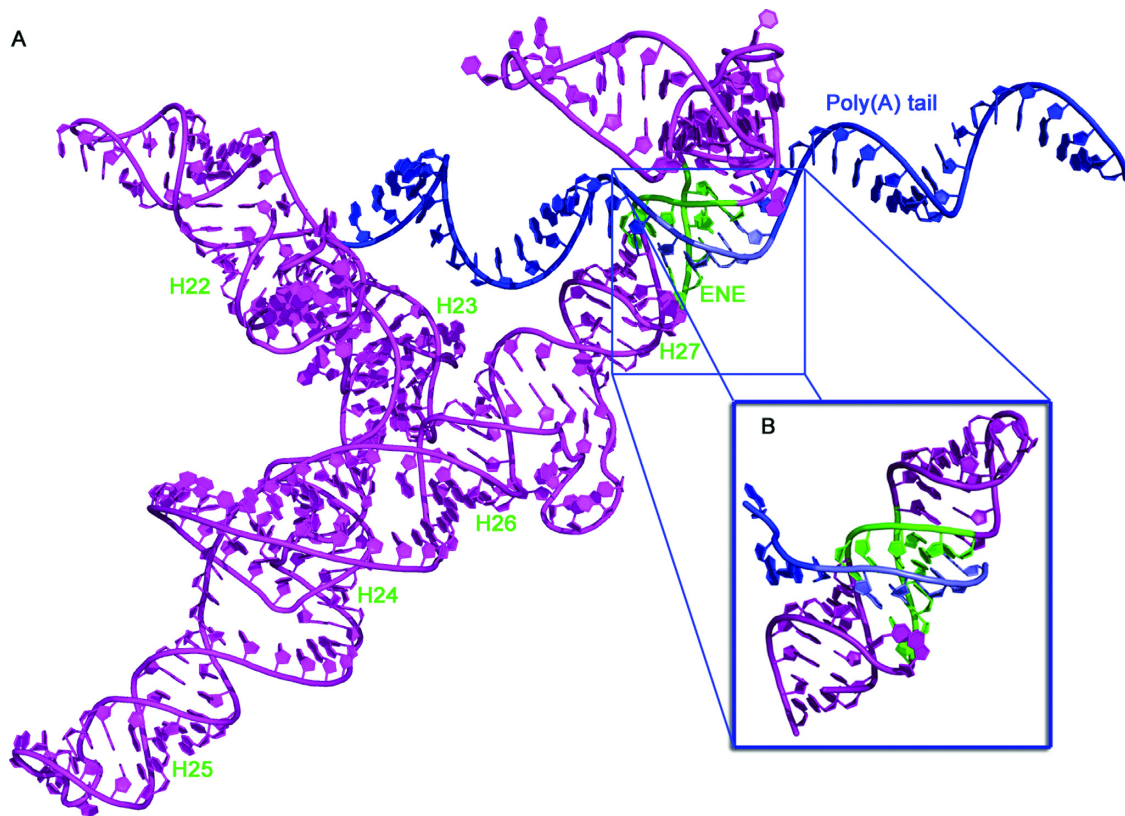
Figure 5. Three-dimensional projection model of the PAN RNA Domain III. (**A**) The ENE motif (green), 3′ poly(A) tail (blue) and A9 residues involved in triple helix formation are indicated (light blue). Surrounding helices (H) are labeled accordingly (green). (**B**) The insert indicates the crystal structure of PAN ENE:A9 complex obtained previously (20).

while ENE-lacking transcripts show an even greater response (67). This suggests the intriguing possibility that the two mechanisms work cooperatively, perhaps via ORF57 binding to both the MRE and ENE, as our data suggests, and local complexes facilitate the higher-level assembly of a more compact and stable PAN RNA–protein holo-complex (1). PABPC1, a cytoplasmic-nuclear shuttling host protein, also binds the 9-nt MRE core and according to Massimelli *et al.* it antagonizes ORF57-mediated stabilization of PAN (1,68). However, Borah *et al.* (17) suggest that PABPC1 binds to PAN poly(A) tail and stabilizes and enhances its nuclear abundance. The nature of the interplay among these proteins, their association with common and unique binding sites, and their effect on PAN RNA stability during infection remains an intriguing question.

We also identified multiple LANA–PAN binding sites, including single-stranded regions in Domain I, a small hairpin within Domain II (G528–C541; H16) and previously proposed nucleotides near the PAN 3′ terminus (Figure 4) (46). LANA spirally coats the KSHV episome and host DNA via sequence specific and non-specific interactions, forming speckles that contain a plethora of cellular molecules, including RNA binding proteins and an RNA helicase (28,69). It has been suggested that the association of PAN RNA with LANA promotes dissociation of LANA from viral episome and sequesters the protein during lytic reactivation (46). The energetics of this process would be most efficient if PAN RNA could bind more than

one LANA molecule at a given time—a notion consistent with our results. Interestingly, LANA–PAN contact sites within H16 overlap with the RNP contact area identified in nuclear PAN RNA (Figures 3A and 4), suggesting that this interaction may be specific to (or most commonly occur in) the nucleus.

The DNA polymerase processivity factor ORF59, a cytoplasmic-nuclear shuttling protein (41,70), was also found to bind PAN RNA. Our analysis detected multiple ORF59–PAN contact sites (Figure 4), some of which overlapped those detected in cytoplasmic (nt 338–341, 1016–1023; Figure 3B) and nuclear (nt 481–483; Figure 3A) PAN RNA fractions, respectively. ORF59 and PAN RNA co-localize to the same regions on the KSHV genome, and it was proposed previously that ORF59 anchors PAN RNA to specific regions of the KSHV episome (52). A similar process has been described for the human cytomegalovirus lncRNA and viral DNA polymerase accessory subunit UL44 (71). Prior studies also indicate that PAN RNA (13) and ORF59 (C. Rossetto, personal communication) associate with histones, suggesting a potential cooperative role in regulating KSHV gene expression.

Viral RNA packaging by KSHV (72) and murine gammaherpesvirus-68 (MHV-68) does not correspond to mRNA expression levels in infected cells (73,74), suggesting a mechanism by which they are selectively recognized by packaging proteins. ORF26 minor capsid protein may mediate PAN RNA packaging (16), perhaps via interactions

with PAN RNA contact sites identified in the present study (Figure 4). The largest such site comprises the U-rich ENE loop and nearby stems, and overlaps with an RNP contact detected by ΔSHAPE analysis of viral PAN RNA (Figure 3C). It has been suggested previously that release of PAN from the virion core immediately following viral entry may suppress the host cell immune response prior to the onset of lytic replication (13).

Other RNP contacts, many of which are common to cytoplasmic, nuclear and viral PAN RNA, were formed within Domain I (H5, H7), single-stranded junctions of Domains I and II, and near the functional ENE core (H25-H27) in Domain III (Figures 2 and 3). Most of these were localized to UG-rich sequences in single-stranded RNA segments and near the junctions of apical or internal loops.

Taken together, our data pave the way for a comprehensive reconstruction of the protein–lncRNA interaction network in various biological contexts. PAN RNA is a highly multifunctional viral transcript that mediates its effects through locally structured RNA segments and interactions with several viral and host proteins. Additional studies of PAN and other KSHV lncRNAs will further elucidate the complex molecular biology of this important pathogen.

ACCESSION NUMBERS

The consensus sequences of each KS sample from Zambian patients, and their annotation, can be found at the following accession numbers: KT271454 (ZM027), KT271455 (ZM091), KT271456 (ZM095), KT271457 (ZM102), KT271458 (ZM106), KT271459 (ZM108), KT271460 (ZM114), KT271461 (ZM116), KT271462 (ZM117), KT271463 (ZM118), KT271464 (ZM121), KT271465 (ZM123), KT271466 (ZM124), KT271467 (ZM128), KT271468 (ZM130). The PAN RNA sequence can be found at the accession number: U50139.1.

The accessions numbers for KSHV ORFs are: AFU082 93.1 (ORF26), AFU08278.1 (ORF57), AFU08321.1 (ORF59), AFU08336.1 (ORF73/LANA).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Massimelli,M.J., Kang,J.G., Majerciak,V., Le,S.Y., Liewehr,D.J., Steinberg,S.M. and Zheng,Z.M. (2011) Stability of a long noncoding viral RNA depends on a 9-nt core element at the RNA 5′ end to interact with viral ORF57 and cellular PABPC1. *Int. J. Biol. Sci.*, **7**, 1145–1160.
2. Conrad,N.K. and Steitz,J.A. (2005) A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts. *EMBO J.*, **24**, 1831–1841.
3. Cawley,S., Bekiranov,S., Ng,H.H., Kapranov,P., Sekinger,E.A., Kampa,D., Piccolboni,A., Sementchenko,V., Cheng,J., Williams,A.J. *et al.* (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, **116**, 499–509.
4. Iyer,M.K., Niknafs,Y.S., Malik,R., Singhal,U., Sahu,A., Hosono,Y., Barrette,T.R., Prensner,J.R., Evans,J.R., Zhao,S. *et al.* (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.*, **47**, 199–208.
5. Cech,T.R. and Steitz,J.A. (2014) The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, **157**, 77–94.
6. Fitzgerald,K.A. and Caffrey,D.R. (2014) Long noncoding RNAs in innate and adaptive immunity. *Curr. Opin. Immunol.*, **26**, 140–146.
7. Imamura,K., Imamachi,N., Akizuki,G., Kumakura,M., Kawaguchi,A., Nagata,K., Kato,A., Kawaguchi,Y., Sato,H., Yoneda,M. *et al.* (2014) Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol. Cell*, **53**, 393–406.
8. Nishitsuji,H., Ujino,S., Yoshio,S., Sugiyama,M., Mizokami,M., Kanto,T. and Shimotohno,K. (2016) Long noncoding RNA #32 contributes to antiviral responses by controlling interferon-stimulated gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 10388–10393.
9. Fortes,P. and Morris,K.V. (2016) Long noncoding RNAs in viral infections. *Virus Res.*, **212**, 1–11.
10. Tycowski,K.T., Guo,Y.E., Lee,N., Moss,W.N., Vallery,T.K., Xie,M. and Steitz,J.A. (2015) Viral noncoding RNAs: more surprises. *Genes Dev.*, **29**, 567–584.
11. Sun,R., Lin,S.F., Gradoville,L. and Miller,G. (1996) Polyadenylylated nuclear RNA encoded by Kaposi sarcoma-associated herpesvirus. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 11883–11888.
12. Zhong,W.D., Wang,H., Herndier,B. and Ganem,D. (1996) Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 6641–6646.
13. Rossetto,C.C. and Pari,G.S. (2011) Kaposi's sarcoma-associated herpesvirus noncoding polyadenylated nuclear RNA interacts with virus- and host cell-encoded proteins and suppresses expression of genes involved in immune modulation. *J. Virol.*, **85**, 13290–13297.
14. Rossetto,C.C. and Pari,G. (2012) KSHV PAN RNA associates with demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome. *PLoS Pathog.*, **8**, e1002680.
15. Rossetto,C.C., Tarrant-Elorza,M., Verma,S., Purushothaman,P. and Pari,G.S. (2013) Regulation of viral and cellular gene expression by Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA. *J. Virol.*, **87**, 5540–5553.
16. Bechtel,J., Grundhoff,A. and Ganem,D. (2005) RNAs in the virion of Kaposi's sarcoma-associated herpesvirus. *J. Virol.*, **79**, 10138–10146.
17. Borah,S., Darricarrere,N., Darnell,A., Myoung,J. and Steitz,J.A. (2011) A viral nuclear noncoding RNA binds re-localized poly(A) binding protein and is required for late KSHV gene expression. *PLoS Pathog.*, **7**, e1002300.
18. Sei,E. and Conrad,N.K. (2011) Delineation of a core RNA element required for Kaposi's sarcoma-associated herpesvirus ORF57 binding and activity. *Virology*, **419**, 107–116.
19. Conrad,N.K., Mili,S., Marshall,E.L., Shu,M.D. and Steitz,J.A. (2006) Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element. *Mol. Cell*, **24**, 943–953.
20. Mitton-Fry,R.M., DeGregorio,S.J., Wang,J., Steitz,T.A. and Steitz,J.A. (2010) Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science*, **330**, 1244–1247.

21. Smola,M.J., Christy,T.W., Inoue,K., Nicholson,C.O., Friedersdorf,M., Keene,J.D., Lee,D.M., Calabrese,J.M. and Weeks,K.M. (2016) SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 10322–10327.

22. Novikova,I.V., Hennelly,S.P. and Sanbonmatsu,K.Y. (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.*, **40**, 5034–5051.

23. Sabarinathan,R., Tafer,H., Seemann,S.E., Hofacker,I.L., Stadler,P.F. and Gorodkin,J. (2013) The RNAsnp web server: predicting SNP effects on local RNA secondary structure. *Nucleic Acids Res.*, **41**, W475–W479.

24. Labo,N., Miley,W., Marshall,V., Gillette,W., Esposito,D., Bess,M., Turano,A., Uldrick,T., Polizzotto,M.N., Wyvill,K.M. *et al.* (2014) Heterogeneity and breadth of host antibody response to KSHV infection demonstrated by systematic analysis of the KSHV proteome. *PLoS Pathog.*, **10**, e1004046.

25. Turner,R., Shefer,K. and Ares,M. Jr (2013) Safer one-pot synthesis of the 'SHAPE' reagent 1-methyl-7-nitroisatoic anhydride (1m7). *RNA*, **19**, 1857–1863.

26. Seif,E., Niu,M. and Kleiman,L. (2015) In virio SHAPE analysis of tRNA(Lys3) annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. *Retrovirology*, **12**, 1–10.

27. Smola,M.J., Rice,G.M., Busan,S., Siegfried,N.A. and Weeks,K.M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.*, **10**, 1643–1669.

28. Hellert,J., Weidner-Glunde,M., Krausze,J., Lunsdorf,H., Ritter,C., Schulz,T.F. and Luhrs,T. (2015) The 3D structure of Kaposi sarcoma herpesvirus LANA C-terminal domain bound to DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 6694–6699.

29. Smola,M.J., Calabrese,J.M. and Weeks,K.M. (2015) Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry*, **54**, 6867–6875.

30. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.

31. Hajdin,C.E., Bellaousov,S., Huggins,W., Leonard,C.W., Mathews,D.H. and Weeks,K.M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5498–5503.

32. Siegfried,N.A., Busan,S., Rice,G.M., Nelson,J.A. and Weeks,K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*, **11**, 959–965.

33. Popenda,M., Szachniuk,M., Antczak,M., Purzycka,K.J., Lukasiak,P., Bartol,N., Blazewicz,J. and Adamiak,R.W. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.*, **40**, e112.

34. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.

35. Somarowthu,S., Legiewicz,M., Chillon,I., Marcia,M., Liu,F. and Pyle,A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Mol. Cell*, **58**, 353–361.

36. Nainar,S., Feng,C. and Spitale,R.C. (2016) Chemical tools for dissecting the role of lncRNAs in epigenetic regulation. *ACS Chem. Biol.*, **11**, 2091–2100.

37. Martin,J.S. (2014) Describing the structural diversity within an RNA's ensemble. *Entropy*, **16**, 1331–1348.

38. Mauger,D.M., Golden,M., Yamane,D., Williford,S., Lemon,S.M., Martin,D.P. and Weeks,K.M. (2015) Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 3692–3697.

39. Conrad,N.K., Shu,M.D., Uyhazi,K.E. and Steitz,J.A. (2007) Mutational analysis of a viral RNA element that counteracts rapid RNA decay by interaction with the polyadenylate tail. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 10412–10417.

40. Olp,L.N., Jeanniard,A., Marimo,C., West,J.T. and Wood,C. (2015) Whole-genome sequencing of Kaposi's sarcoma-associated herpesvirus from Zambian Kaposi's sarcoma biopsy specimens reveals unique viral diversity. *J. Virol.*, **89**, 12299–12308.

41. McDowell,M.E., Purushothaman,P., Rossetto,C.C., Pari,G.S. and Verma,S.C. (2013) Phosphorylation of Kaposi's sarcoma-associated herpesvirus processivity factor ORF59 by a viral kinase modulates its ability to associate with RTA and oriLyt. *J. Virol.*, **87**, 8038–8052.

42. Arias,C., Weisburd,B., Stern-Ginossar,N., Mercier,A., Madrid,A.S., Bellare,P., Holdorf,M., Weissman,J.S. and Ganem,D. (2014) KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog.*, **10**, e1003847.

43. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.

44. Wilkinson,K.A., Gorelick,R.J., Vasa,S.M., Guex,N., Rein,A., Mathews,D.H., Giddings,M.C. and Weeks,K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.

45. Chan,S.R., Bloomer,C. and Chandran,B. (1998) Identification and characterization of human herpesvirus-8 lytic cycle-associated ORF 59 protein and the encoding cDNA by monoclonal antibody. *Virology*, **240**, 118–126.

46. Campbell,M., Kim,K.Y., Chang,P.C., Huerta,S., Shevchenko,B., Wang,D.H., Izumiya,C., Kung,H.J. and Izumiya,Y. (2014) A lytic viral long noncoding RNA modulates the function of a latent protein. *J. Virol.*, **88**, 1843–1848.

47. Miao,Z. and Westhof,E. (2016) RBscore&NBench: a high-level web server for nucleic acid binding residues prediction with a large-scale benchmarking database. *Nucleic Acids Res.*, **44**, W562–W567.

48. Kang,J.G., Pripuzova,N., Majerciak,V., Kruhlak,M., Le,S.Y. and Zheng,Z.M. (2011) Kaposi's sarcoma-associated herpesvirus ORF57 promotes escape of viral and human interleukin-6 from microRNA-mediated suppression. *J. Virol.*, **85**, 2620–2630.

49. Sei,E., Wang,T., Hunter,O.V., Xie,Y. and Conrad,N.K. (2015) HITS-CLIP analysis uncovers a link between the Kaposi's sarcoma-associated herpesvirus ORF57 protein and host pre-mRNA metabolism. *PLoS Pathog.*, **11**, e1004652.

50. Ma,Y., Liu,P., Majerciak,V., Zhu,J. and Zheng,Z.M. (2016) CLIP-seq to Identify KSHV ORF57-Binding RNA in Host B Cells. *Curr. Protoc. Microbiol.*, **41**, doi:10.1002/cpmc.3.

51. Uppal,T., Banerjee,S., Sun,Z., Verma,S.C. and Robertson,E.S. (2014) KSHV LANA–the master regulator of KSHV latency. *Viruses*, **6**, 4961–4998.

52. Rossetto,C.C. and Pari,G.S. (2014) PAN's Labyrinth: Molecular biology of Kaposi's sarcoma-associated herpesvirus (KSHV) PAN RNA, a multifunctional long noncoding RNA. *Viruses*, **6**, 4212–4226.

53. Sander,G., Konrad,A., Thurau,M., Wies,E., Leubert,R., Kremmer,E., Dinkel,H., Schulz,T., Neipel,F. and Sturzl,M. (2008) Intracellular localization map of human herpesvirus 8 proteins. *J. Virol.*, **82**, 1908–1922.

54. Ignatovich,I.A., Dizhe,E.B., Akif'ev,B.N., Burov,S.V., Boiarchuk,E.A. and Perevozchikov,A.P. (2002) [Delivery of "suicide" thymidine kinase gene of herpes virus in the complex with cationic peptide into human hepatoma cells in vitro]. *Tsitologiia*, **44**, 455–462.

55. La Ferla,L., Pinzone,M.R., Nunnari,G., Martellotta,F., Lleshi,A., Tirelli,U., De Paoli,P., Berretta,M. and Cacopardo,B. (2013) Kaposi's sarcoma in HIV-positive patients: the state of art in the HAART-era. *Eur. Rev. Med. Pharmacol. Sci.*, **17**, 2354–2365.

56. Gregory,S.M., Wang,L., West,J.A., Dittmer,D.P. and Damania,B. (2012) Latent Kaposi's sarcoma-associated herpesvirus infection of monocytes downregulates expression of adaptive immune response costimulatory receptors and proinflammatory cytokines. *J. Virol.*, **86**, 3916–3923.

57. Borodavka,A., Singaram,S.W., Stockley,P.G., Gelbart,W.M., Ben-Shaul,A. and Tuma,R. (2016) Sizes of long RNA molecules are determined by the branching patterns of their secondary structures. *Biophys. J.*, **111**, 2077–2085.

58. Rouskin,S., Zubradt,M., Washietl,S., Kellis,M. and Weissman,J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.

59. Herschlag,D. (1995) RNA chaperones and the RNA folding problem. *J. Biol. Chem.*, **270**, 20871–20874.

60. Brown,J.A., Bulkley,D., Wang,J., Valenstein,M.L., Yario,T.A., Steitz,T.A. and Steitz,J.A. (2014) Structural insights into the

stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.*, **21**, 633–640.

61. Zhou,Y., Zeng,P., Li,Y.H., Zhang,Z. and Cui,Q. (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.*, **44**, e91.

62. Liu,Z., Xiao,X., Yu,D.J., Jia,J., Qiu,W.R. and Chou,K.C. (2016) pRNAm-PC: predicting N(6)-methyladenosine sites in RNA sequences via physical-chemical properties. *Anal. Biochem.*, **497**, 60–67.

63. Liu,N., Parisien,M., Dai,Q., Zheng,G., He,C. and Pan,T. (2013) Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA*, **19**, 1848–1856.

64. Liu,N., Dai,Q., Zheng,G., He,C., Parisien,M. and Pan,T. (2015) N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, **518**, 560–564.

65. Majerciak,V., Yamanegi,K., Nie,S.H. and Zheng,Z.M. (2006) Structural and functional analyses of Kaposi sarcoma-associated herpesvirus ORF57 nuclear localization signals in living cells. *J. Biol. Chem.*, **281**, 28365–28378.

66. Majerciak,V. and Zheng,Z.M. (2015) KSHV ORF57, a protein of many faces. *Viruses*, **7**, 604–633.

67. Sahin,B.B., Patel,D. and Conrad,N.K. (2010) Kaposi's sarcoma-associated herpesvirus ORF57 protein binds and protects a nuclear noncoding RNA from cellular RNA decay pathways. *PLoS Pathog.*, **6**, e1000799.

68. Massimelli,M.J., Majerciak,V., Kruhlak,M. and Zheng,Z.M. (2013) Interplay between polyadenylate-binding protein 1 and Kaposi's sarcoma-associated herpesvirus ORF57 in accumulation of polyadenylated nuclear RNA, a viral long noncoding RNA. *J. Virol.*, **87**, 243–256.

69. Kaul,R., Verma,S.C. and Robertson,E.S. (2007) Protein complexes associated with the Kaposi's sarcoma-associated herpesvirus-encoded LANA. *Virology*, **364**, 317–329.

70. Chen,Y., Ciustea,M. and Ricciardi,R.P. (2005) Processivity factor of KSHV contains a nuclear localization signal and binding domains for transporting viral DNA polymerase into the nucleus. *Virology*, **340**, 183–191.

71. Rossetto,C.C., Tarrant-Elorza,M. and Pari,G.S. (2013) Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLoS Pathog.*, **9**, e1003366.

72. Lin,X., Li,X., Liang,D. and Lan,K. (2012) MicroRNAs and unusual small RNAs discovered in Kaposi's sarcoma-associated herpesvirus virions. *J. Virol.*, **86**, 12717–12730.

73. Ebrahimi,B., Dutia,B.M., Roberts,K.L., Garcia-Ramirez,J.J., Dickinson,P., Stewart,J.P., Ghazal,P., Roy,D.J. and Nash,A.A. (2003) Transcriptome profile of murine gammaherpesvirus-68 lytic infection. *J. Gen. Virol.*, **84**, 99–109.

74. Cliffe,A.R., Nash,A.A. and Dutia,B.M. (2009) Selective uptake of small RNA molecules in the virion of murine gammaherpesvirus 68. *J. Virol.*, **83**, 2321–2326.