Exploiting functional regions in the viral RNA genome as druggable entities

Dehua Luo[1, 2, 3, 4†], Yingge Zheng[1, 2, 3, 7†], Zhiyuan Huang[1, 2, 8†], Zi Wen[1, 2, 8], Lijun Guo[1, 2, 3], Yingxiang Deng[1, 2, 3], Qingling Li[1, 2, 3], Yuqing Bai[1, 4], Shozeb Haider[9], Dengguo Wei[1, 2, 3, 4, 5, 6, 7 *]

1. National Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, Hubei, China
2. Hubei Hongshan Laboratory, Interdisciplinary Sciences Institute, Huazhong Agricultural University, Wuhan, Hubei, China
3. National Reference Laboratory of Veterinary Drug Residues (HZAU) and National Safety Laboratory of Veterinary Drug (HZAU), MOA Key Laboratory for Detection of Veterinary Drug Residues, MOA Laboratory for Risk Assessment of Quality and Safety of Livestock and Poultry Products, Wuhan, Hubei, China
4. College of Life Science and Technology, Huazhong Agricultural University, Wuhan, Hubei, China
5. Shenzhen Institute of Nutrition and Health, Huazhong Agricultural University, Wuhan, Hubei, China
6. Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China
7. Frontiers Science Center for Animal Breeding and Sustainable Production, Wuhan, Hubei, China
8. College of Informatics, Huazhong Agricultural University, Wuhan, Hubei, China
9. UCL School of Pharmacy, University College London, London WC1N 1AX, UK

† These authors contributed equally to this work.

Correspondence: dgwei@mail.hzau.edu.cn

ABSTRACT

RNA-targeting compounds or small interfering RNAs (siRNAs) offer a potent means for controlling viral infections. An essential prerequisite to their design depends on identifying conserved and functional viral RNA structures in cells. Techniques that probe RNA structures *in situ* have been developed recently including SHAPE-MaP, which has been helpful in analyzing the secondary structures of RNA. In this study, we report on the application of SHAPE-MaP to the Porcine Epidemic Diarrhoea Virus (PEDV) RNA genome to categorize different functional regions including potential quadruplex forming sequence and target sites of small interfering RNA (siRNA). Our results show that these structures can be exploited to inhibit viral proliferation and that SHAPE-MaP is an effective method to the identification of secondary structures in RNA genomes.

Keywords: SHAPE-MaP; RNA; PEDV; viral genome; siRNA; G-quadruplex

INTRODUCTION

RNA-targeting drugs are expected to address the challenges posed by "undruggable" protein targets and significantly expand the landscape of targetable macromolecules[1–5]. The U.S. Food and Drug Administration (FDA) has approved six siRNAs[6], nine antisense oligonucleotides (ASOs)[7–9], and the RNA-targeting compound Risdiplam[10,11] for treating complicated diseases such as spinal muscular atrophies. The selection of RNAs that are functional, conserved and accessible to be targeted *in vivo* is fundamental to RNA-targeting drug development. Unlike RNAs *in vitro*, the folding of RNAs in cells is affected by the intricate intracellular environment[12,13]. Thus, *in situ* probing for the accessibility of viral RNAs by antivirals is essential for the identification of targetable RNAs.

In recent years, the integration of chemical probing methods and high-throughput sequencing technologies has been employed to analyse the structure of the viral RNA genome in infected cells[12–26]. These methods have confirmed the widespread presence of structurally organized RNA elements within the viral genome, which are composed of various types of RNA structures such as hairpin structure, RNA single-strand, RNA pseudoknot and RNA G-quadruplex (G4), collectively forming a complex and dynamic global genome architecture[21,27]. Of these, RNA G4 has shown considerable promise because of the high-stability and modulation by small molecules[27–32]. Furthermore, RNA interference (RNAi) have been demonstrated to inhibit expression of genes and viral replication[6,9,33–37]. In spite of the advances, identifying RNA elements or regions that serve as targets within complex genomic structural contexts remains challenging.

SHAPE mutational profile (SHAPE-MaP) stands out in the *ex vivo* RNA structure probing technologies for its ability to quantify nucleotide flexibility and solvent accessibility at single-nucleotide resolution[18,20,38]. In short, the SHAPE employs 2"-hydroxyl-selective reagents to label RNA ribose 2'-OH groups, thereby enabling the identification of RNA secondary structures. Following this, the mutational profiling (MaP) is used to introduce non-complementary nucleotide mutations at SHAPE adducts during reverse transcription[20]. In the viral long-stranded RNA genome, SHAPE reactivities are used as constraints in the prediction of RNA structures based on nearest-neighbour rules[39,40], which may result in thousands of possible conformations. A Shannon entropy value, which reflects the probability of formation of each base pair across all possible structures in the ensemble, is then assigned to each nucleotide[20,41]. Therefore, SHAPE reactivity provides information about the pairing probability

and accessibility of nucleotides, while Shannon entropy reflects the likelihood of alternative conformations in a given region. RNA motifs with low SHAPE reactivity and low Shannon entropy are assumed to be well fold structures and potential targets[21]. In this study, we categorized RNAs with distinct features in the viral genome according to the SHAPE reactivity and Shannon entropy, and investigate their potential as targets (Figure 1).

Porcine epidemic diarrhoea virus (PEDV) is an α-coronavirus with a positive single-stranded RNA genome of 28 kb[42,43]. It results in severe diarrhoea and vomiting, leading to the death of ~90 percent of infected piglets. The spread of PEDV is world-wide and its infection is of primary concern in pig husbandry[42]. The epizootic re-emergence of PEDV has also been reported to infect human intestinal cells and displays potential susceptibility to cross-species infection[44].

In this study, we adapted the SHAPE-MaP method to detect the *in situ* structure of the PEDV RNA genome in infected cells. More specifically, the combination of SHAPE reactivity and Shannon entropy categorized the PEDV RNA genome into different functional types (Figure 1), based on the distinctive profiles, including the 5' untranslated region (5' UTR), frameshift stimulation element (FSE), and 3' untranslated region (3' UTR). We particularly focused on guanine-rich regions, which could potentially fold into a stable G4 structure and sites suitable for binding small interfering RNA (siRNA). Our results show that the folding of G4 in a dynamic single-stranded region with high SHAPE reactivity and high Shannon entropy inhibited viral proliferation, while siRNA targeting persistent single-stranded regions with high SHAPE reactivity and low Shannon entropy exhibited greater success rates in inhibiting viral replication. Finally, this work proposes a strategy for selecting targetable RNAs based on the *in situ* folding characteristics of the viral genome.

MATERIAL AND METHODS

Cell culture and PEDV infection

Vero E6 cells were cultured in 10 cm culture dishes in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37°C in an atmosphere of 5% CO_2. The cells were infected at an MOI of 0.1 with PEDV/AJ1102 (GenBank accession: MK584552.1) when the cell confluence reached 95% or greater.

In situ RNA probing and purification

During the logarithmic phase of viral genome replication after infection (12 h, Figure S1), infected cells were washed 3 times with 10 ml cold PBS and then resuspended in PBS containing 2-methylnicotinic acid imidazolide (NAI) at a final concentration of 100 mM. Two independent replicates were performed for *ex vivo* experiments. An equal amount of DMSO was added to another sample as a control. The samples were then incubated at room temperature with continuous shaking for 15 minutes, after which the reaction was terminated by the addition of DTT at a final concentration of 0.5 M. The supernatant was then removed, and the cells were lysed by adding 6 ml of TRIzol. The RNA was extracted by adding 1.2 ml of chloroform:isoamyl alcohol (24:1). Three milliliters of the upper aqueous phase was transferred to a new tube, 6 ml of anhydrous ethanol was added, and the mixture was incubated overnight at -20°C. The RNA was pelleted at 18,000 × g for 20 min at 4°C, washed twice with 80% EtOH, and then the pellet was air-dried. Finally, RNA was dissolved in RNase-free water and then stored at -80°C until further use.

Isolation and fragmentation of poly(A)PEDV genomic RNA

We isolated poly(A) RNA with VAHTS mRNA capture beads (Vazyme #N401) according to the manufacturer's instructions. Five micrograms of total RNA was diluted to 50 µl with RNase-free water and purified in two rounds with mRNA capture beads. Poly(A) RNA was fragmented to a median size of 150-200 nt by incubation at 94°C for 8 min in fragmentation buffer (65 mM Tris–HCl; 4 mM $MgCl_2$; 95 mM KCl; pH 8.0) and then purified with an RNA Clean & Concentrator-5 kit (Zymo Research, R1015) following the manufacturer's instructions.

Reverse transcription of modified PEDV genomic RNAs

20 µl of poly(A) RNA (~500 ng) was added to separate 0.65 ml RNase-free tubes, 2 µl of 200 ng/µl random primer was added to each tube, and the mixture was incubated at 65°C for 5 min. Sixteen microliters of 2.5× MaP buffer (125 mM Tris, pH 8.0; 187.5 mM KCl; 15 mM $MnCl_2$; 25 mM DTT and 1.25 mM dNTPs) were added to each tube, and the mixture was incubated at 25°C for 2 min. Then, 2 µl of SuperScript II reverse transcriptase (200 U/µl; Life Technologies, cat. no. 18064-014) was added, and the mixture was incubated at 42°C for 3 h. The reactions were heat-inactivated by incubation at 75°C for 15 min, after which the mixture was placed on ice. cDNA was purified from the reactions using G-25 micro spin columns (GE Healthcare, cat. no. 27-5325-01) following the manufacturer's instructions. The second-strand cDNA synthesis reaction was then performed immediately. The volume of 1st strand cDNA was adjusted to 56 µl with nuclease-free H_2O, 6.5 µl of NEBNext Second Strand Synthesis Reaction Buffer (10×; New England Biolabs, cat. no. E6111S) and 3.5 µl of NEBNext Second Strand Synthesis Enzyme Mix (New England Biolabs, cat. no. E6111S) were added, and the mixture was then incubated at 16°C for 2.5 h.

SHAPE-MaP sequencing library construction and quality control

Sequencing libraries were generated using a VAHTS® Universal V8 RNA-seq Library Prep Kit for MGI (Vazyme #NRM605) according to the manufacturer's protocol. Libraries concentrations were measured using a Qubit dsDNA HS Assay Kit (Thermo Fisher, Cat. No. Q32851). One microliter of the final library product was removed, and the library quality was analyzed using an Agilent DNA1000 Kit (Agilent, Cat. No. 5067-1504).

SHAPE-MaP data analysis

SHAPE-MaP data analysis was performed using RNA Framework v2.6.9[45] (https://github.com/dincarnato/RNAFramework), referring to Manfredonia *et al*[13], as follows: a) Data pre-processing and alignment. Reads were processed and aligned to the PEDV reference genome using the rf-map module with parameters: -ctn -cmn 0 -cqo -cq5 20 -b2. Reads with internal Ns were discarded, and low-quality bases (Phred < 20) were excluded before alignment. b) Nucleotide modification counting. The index of the PEDV genome was built using Bowtie2 software [46]. Individual nucleotide mutation signals were assessed with the rf-count module with parameters: -m -rd, generating a binary RC file with sequence data, mutation counts, coverage, and total aligned fragments. c) Normalization. The rf-norm module

was applied for data normalization with parameters: -sm 3 -nm 3 -rb AC -n 500 -mm 1, utilizing the scoring method developed by Siegfried *et al*[20].

PEDV RNA secondary structure modeling guided by SHAPE data

The rf-fold module, employing a previously established windowed approach[13, 20], predicts the RNA secondary structure of the entire PEDV genome using the normalization step's XML file, with a default pseudo-free energy intercept and slope values of -0.6 (kcal/mol) and 1.8 (kcal/mol), respectively, with the following parameters: -fw 3000 -fo 300 -wt 200 -pw 1000 -po 250 -dp -sh -nlp -md 600. This means that RNAstructure will generate a 3000 nt window and slide in 300 nt increments over the genome in 1000 nt partition windows, predicting the most reasonable secondary structure of the genome under the condition that the maximum base pairing distance does not exceed 600 nt. SHAPE reactivities were incorporated in the form of soft constraints using the Vienna RNA package [38]. For each window, the first and last 100 nt were ignored to avoid terminal biases. Additionally, RF Fold can provide pairing probabilities between bases, Shannon entropy of individual bases, and potential pseudoknot structures. The Shannon entropy is commonly used to measure the diversity of RNA secondary structures:

$$H_i = -\sum_{j=1}^{J} P_{i,j} \log_{10} P_{i,j}$$

where P represents the pairing probability between bases.

Identification of characterized regions in viral genomes

Based on the previous research methods for identifying well-folded regions[13, 20, 25], modifications have been made as follows:

a) For the identification of low SHAPE-low Shannon regions, we first calculated the local medians of SHAPE reactivities and Shannon entropies in sliding, centered, 50 nt windows and subtracted the global median for visualization. Then, a window of 50 nt was slid along the genome with a 1 nt step. Windows with >75% of the bases below both the global SHAPE and Shannon median (calculated on the full PEDV genome) were selected, and windows less than 10 nt apart were merged. Following the initial identification of well-folded regions, entire structures exhibiting consistent conformations in both datasets and

having at least 50% of their bases in well-folded regions were further selected. Subsequently, well-folded RNA structures within cells were determined.

b) For the identification of low SHAPE-high Shannon regions, a window of 50 nt was slid along the PEDV genome. Windows with >75% of the bases below the global SHAPE median and above the global Shannon median (based on the full PEDV genome) were selected, and windows less than 10 nt apart were merged.

c) For the identification of high SHAPE-low Shannon regions, instead, a window of 25 nt was slid along the PEDV genome. Windows with >75% of the bases were above the global SHAPE median and below the global Shannon median (calculated on the full PEDV genome), and >50% of the bases were predicted to be single-stranded in the MEA structure; windows less than 10 nt apart were selected and merged.

d) For the identification of high SHAPE-high Shannon regions, a window of 25 nt was slid along the PEDV genome. Windows with >75% of the bases above both the global SHAPE and Shannon median (based on the full PEDV genome) were selected, and windows less than 10 nt apart were merged.

G-quadruplex forming sequence prediction and conservativeness analysis

The G4 sequence was defined as $G_{\geq 2}N_{1\text{-}6}G_{\geq 2}N_{1\text{-}6}G_{\geq 2}N_{1\text{-}6}G_{\geq 2}$, where G refers to guanine and N refers to any nucleotide, including guanine. Four independent G4 prediction tools (QGRS-mapper, G4Catchall, ImGQfinder and pqsfinder) were employed to analyse the potential G4 forming sequences (PQSs) in the PEDV genome[47–50]. Conservation analysis was performed using Molecular Evolutionary Genetics Analysis (MEGA) software (version 6.0; available at www.megasoftware.net). 761 PEDV genome sequences were retrieved from the National Center for Biotechnology Information website (NCBI, https://www.ncbi.nlm.nih.gov/). The graphical representation of the alignment sequence was generated by WebLogo 2 (http://weblogo.berkeley.edu)[51].

Verification of G-quadruplex structure formation by native PAGE

Cy2-tagged RNA was annealed in buffers (100 mM potassium arsenate, pH 7.0) containing different concentrations of KCl (0, 50 mM, 100 mM). The marker used in this experiment was a mixture of P15 (UAAUACGACUCACUA), M17 (UAAUACGACUCACUAUAUA), and M20 (UAAUACGACUCACUAUACGA). Acrylamide at a 20% concentration was used to separate the gel. The concentration of the RNA sample was 100 nM, and electrophoresis was

conducted at 100 V for 2 h. Native PAGE was carried out under a controlled temperature (4.0°C) with a vertical electrophoresis instrument (DYCZ-22A; Bio-Rad, Beijing, China). After electrophoresis, RNA oligomers were scanned by a Pharos FX molecular imager for visualization.

Circular dichroism (CD) spectroscopy

The compound was dissolved in DMSO at a concentration of 20 mM. RNAs (20.0 μM) were dissolved in 10 mM Tris-HCl buffer (pH 7.0) containing varying concentrations of KCl (0, 50, 100 mM). The CD experiment was performed at 25°C using a JASCO-810 spectrophotometer (Jasco, Easton, MD, USA) with a quartz cell path length of 1.0 mm. CD spectra were collected from 320 to 200 nm with a scanning speed of 50 nm/min. All CD spectra were baseline corrected for the signal contribution of the buffer and represent the mean of at least two repeats.

Nuclear magnetic resonance (NMR) spectroscopy

^{1}H NMR spectra were recorded at 298 K using an 800-MHz Brock Avance DRX-800 spectrometer equipped with a low-temperature triple-resonance reverse autotuning and matching probe. The presumption of water is achieved by excitation engraving. The RNA samples were dissolved in DEPC solution containing 200 mM KCl and 10% D_2O at a final concentration of 0.5 mM. The sample was heated at 95°C for 5.0 min and cooled slowly to 25°C. The facility was set for 100 ms diffusion time (TD), 50 ms eddy recovery time (TE), and 2 SEC relaxation delay (TR).

CD melting studies

CD-melted RNA samples were dissolved in buffer containing 10 mM Tris-HCl and 100 mM KCl at a concentration of 20 mM. The samples were heated at 95°C for 10 min and then slowly annealed to room temperature. CD melting was measured on a Jasco-810 spectrophotometer equipped with a water bath temperature control attachment. CD melting curves were recorded using a heating rate of 0.2°C/min, and absorbance values were collected at 1°C intervals. The thermal stability of G4 RNA was measured by recording the CD value at 264 nm in relation to temperature. The required temperature range was set from 4.0 ℃ to 98 ℃.

RNA stop assay

P15 (300 nM) and template RNA (600 nM) were annealed in buffer at 25°C [50 mM HEPES, 20 mM NaCl, 1.0 mM KCl, 5 mM $MgCl_2$ and 4 mM DTT (pH 7.0)], heated at 95°C for 5.0 min and slowly cooled to 4°C. Subsequently, NTPs (final concentration, 200 μM) and 3Dpol (0.02 mg/reaction) were added to the system, and the reaction was carried out at 33°C for 20 min. An equal amount of stop buffer (95% formamide) was added, and the reaction was stopped after heating at 90°C for 4.0 min. The products were loaded and separated on a 20% denatured polyacrylamide gel. Finally, the gel was scanned using a Pharos FX molecular Imager (Bio-Rad) operated in fluorescence mode.

EGFP expression repression by RNA G-quadruplex stabilization

We constructed pEGFP-C1 vector with wild-type PQS1 sequences or G4-mutated PQS1mut sequences fused to the N terminus of an EGFP reporter gene. Vero cells were transfected with Lipofectamine 3000 (Invitrogen, USA) according to the manufacturer's protocol and incubated with media supplemented with various concentrations of Braco-19 for 48 hours. The cells were then washed with PBS and fixed at 25°C for 15 minutes with 4% paraformaldehyde. The samples were washed three times with PBS for 5 minutes each and then fixed with precooled methanol at -20°C for 10 minutes. Observations were conducted using a two-photon laser scanning confocal microscope (Olympus, fv1000mp).

Construction of the RNA G4-mutated recombinant virus

The pBAC-CMV-PEDV template was obtained from Prof. Xiao Shaobo (College of Veterinary Medicine, Huazhong Agricultural University). The primers used were chemically synthesized (Table S1). The sequence of the G-rich region (3109-3125 nt) was mutated (AAGGTTACAGGTGGTTGG to AAAGTTACAGGTGGTTGG). Two pairs of primers (PQS1Mut-F1 and PQS1Mut-R1, PQS1Mut-F2 and PQS1Mut-R2) were used to amplify the fragment containing the mutant site. The PQS1mut fragment containing mutation sites was amplified through overlap PCR. A scaffold oligo was used as a template, PQS1m-sgRNA-a and PQS1m-sgRNA-a were used as primers to synthesize guide DNA, and then, guide RNA (sgRNA) was obtained by reverse transcription. pBAC-CMV-PEDV was cut with the Cas9 enzyme and sgRNA. The PQS1mut fragment was cloned and inserted into a linearized pBAC-CMV-PEDV vector to generate the pBAC-PEDV-PQS1mut mutant plasmid. Vero cells were placed on 6-well plates, and plasmid transfection was performed when the cells grew to 60%

to 80% confluence according to the manufacturer's protocol (Lipofectamine 3000, Thermo, L3000015). If no lesions appeared, the freeze–thaw solution was added to the six-well plate with 80% confluence after three repeated freeze–thaw cycles for further culture. When lesions appeared, successful mutation of the mutation site was confirmed by reverse transcription sequencing, and the PEDV-PQS1mut mutant strain was obtained.

Determination of the antiviral activity of Braco-19 against PEDV by RT–qPCR

Braco-19 at different concentrations (0.5, 1, 2 and 5 μM) were inoculated into monolayer Vero cells and incubated for 2 h. PEDV-WT and PEDV-PQS1mut were inoculated at an MOI of 0.1 and incubated for 2 h. The DMEM was replaced with DMEM containing Braco-19, and the mixture was incubated for 16 h. cDNA was synthesized from 100 ng of total RNA per sample using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Switzerland) according to the manufacturer's instructions. Quantitative PCR (qPCR) was conducted following the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines [52] using Taq Pro Universal SYBR qPCR Master Mix (Vazyme, Q712) according to the manufacturer's protocol. mRNA expression was normalized to beta-actin levels with primers β-actin-F/R (Supplementary Table S1). PEDV genomic RNA was quantified at the nonstructural protein 12 (*nsp12*) gene using primers qPCR-nsp12-F/R (Supplementary Table S1).

Western blot analysis to detect PEDV N gene expression

In this study, protein samples were separated by SDS–PAGE using a 12.5% polyacrylamide gel and then transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore, USA). The membrane was blocked with 5% BSA in Tween-PBS buffer and incubated overnight at 4°C with primary antibodies. After three washes, it was incubated with HRP-conjugated secondary antibodies (Beyotime, China). Following three more washes, protein bands were detected using enhanced chemiluminescence (Bio-Rad, USA).

CCK-8 assays

Vero cells in 96-well plates were exposed to different compound concentrations for 12 hours, followed by the addition of 10 μl CCK-8 reagent (Beyotime, China) and a 2-hour incubation at 37°C. Cell proliferation was evaluated by measuring the absorbance at 450 nm with a microplate reader.

Indirect immunofluorescence assay (IFA)

Cells in 24-well plates were treated with Braco-19 and then infected with PEDV at 37°C for 2 hours. After washing with serum-free DMEM, cells were cultured with or without compounds in DMEM. Post-infection, cells were fixed, permeabilized, and washed before blocking and incubating with primary (mouse anti-PEDV N protein antibody) and Alexa Fluor 488-secondary antibody. Nuclei were DAPI-stained, and fluorescence images were captured using an Olympus IX73 microscope.

Isothermal Titration Calorimetry (ITC)

ITC experiments were conducted on an AutoiTC100 titration calorimeter (MicroCal) at 25°C. All solutions (buffer, RNA, and ligand) were degassed prior to experimentation. A 10 μM solution of RNA 5' UTR-SL5 was introduced into the cell, and a 250 μM compound solution was added to the rotating syringe (750 rpm). A total of 40 μL of compound was added to the RNA in 20 injections, with a 150 s interval between injections and an initial injection of 0.4 μL to account for diffusion from the syringe into the cell during equilibration. This initial injection was excluded from the data fitting. Additionally, dilution experiments were performed with buffer (100 mM HEPES (pH 8.0), 100 mM NaCl, and 10 mM $MgCl_2$) in the cell, and the compound remained in the syringe. We utilized Microcal ITC analysis software to analyze the raw ITC data using the two-site binding model. We subtracted the dilution data from the raw interaction data of the compound with RNA prior to analysis. Following the fitting of these experimental data, the enthalpy change during the process was determined.

siRNA antiviral assay

siRNAs were synthesized in GENERAL BIOL and transfected into 8×10^4 Vero E6 cells in 24-well plates using 10 pmol of each siRNA and Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen, 13778075) according to the manufacturer's protocol. The transfected cells were infected with PEDV (MOI = 0.1) after 24 h. After 16 hours of PEDV infection, minimal cytopathic effects (CPEs) on the Vero cells were observed under the microscope. To quantify RNA levels of PEDV in infected cells, the total RNAs from the infected Vero cells were extracted with TRIzol Reagent (Invitrogen, 15596026). For RT-qPCR, 100 ng of total RNAs were converted into cDNA using HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, R212). The qPCR was performed with ChamQ Universal SYBR qPCR Master Mix (Vazyme, Q711) on a StepOnePlus real-time PCR machine (Applied Biosystems). The primers targeting

the nucleocapsid (*N*) gene of PEDV and the *β-actin* gene of Vero cells were used for qPCR. For the 50% tissue culture infectious dose ($TCID_{50}$) assay, monolayers of Vero cells seeded in 96-well plates were washed twice with DMEM. The supernatants of infected cells were serially diluted tenfold. Cells were then inoculated with eight replicates of the appropriate dilutions of the virus suspension. The $TCID_{50}$ was calculated by the Reed-Muench method after culturing the cells for 48 h at 37°C in 5% CO_2.

Statistical analysis

The experimental data were analyzed by Student's t test or one-way or two-way analysis of variance using GraphPad Prism 8 software. *P < 0.05 was considered to indicate statistical significance.

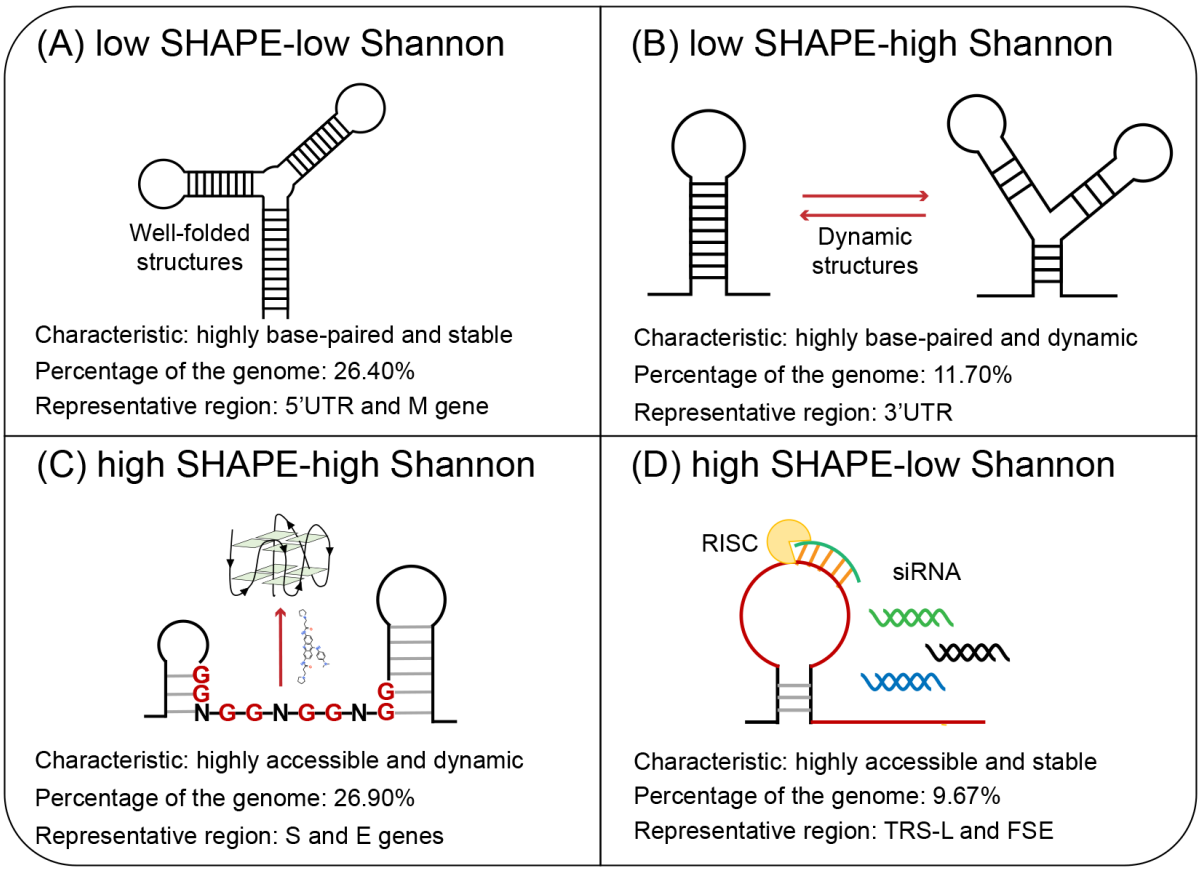

Figure 1. Four regions with different characteristics in the PEDV genome. (A) Regions with low SHAPE reactivity and low Shannon entropy, account for 26.40% of the PEDV genome. **(B)** Regions with low SHAPE reactivity and high Shannon entropy, account for 11.70% of the PEDV genome. **(C)** Regions with high SHAPE reactivity and high Shannon entropy, account for 26.90% of the PEDV genome. **(D)** Regions with high SHAPE reactivity and low Shannon entropy, account for 9.67% of the PEDV genome.

RESULTS

Mapping RNA structures in the PEDV genome

SHAPE-MaP was used to characterize the structural features of the PEDV RNA genome in infected cells (Figure S1-S2A). The high correlations of mutational signatures for biological replicates (*ex vivo* SHAPE R^2=0.955; Figure S2B) and the high mutation rate with NAI treatment (Figure S2C) indicate the high quality of the sequencing data and the experimental workflow. Hence, two replicates were combined for downstream analyses.

As the 5'UTR with conserved functions across coronaviruses has been extensively reported[53–55], we first examined the structure of this region in the PEDV genome for its stability *ex vivo*. The PEDV 5' UTR structure modelled with SHAPE reactivity constraints contains four stem loops (SL1, SL2, SL4, and SL5; Figure S2D) similar to other alpha-coronaviruses[53–55]. Our SHAPE dataset showed that ~95.6% of highly reactive bases (SHAPE reactivity ≥ 0.7) are not involved in canonical base pairing or are located adjacent to helical termini or bulges/loops (Figure S2D). This result suggested that our dataset is reliable as a constraint for predicting the intracellular genomic structure of PEDV.

To characterize the intracellular folding features of the PEDV genome, median SHAPE reactivities and Shannon entropies were calculated in sliding, centered windows (window = 50 nt, step = 1 nt), and a full-genome SHAPE structural map was constructed accordingly (Figure 2). The global median SHAPE reactivity of the genome is 0.233. Specifically, within the 1-8000 nt region at the 5' end, the median reactivity is 0.173; it increases to 0.268 in the middle segment of the genome (8000-24000 nt) and then decreases to 0.194 at the 3' end (24000-28000 nt). These results indicate that the folding of PEDV genomic RNA within infected cells is nonuniform, with the 5' and 3' ends being more compactly structured, while the central region is more flexible (Figure 2).

Within the coding regions of the ORF1a, ORF3, M, and N genes, both SHAPE reactivity and Shannon entropy are below the global median, suggesting that these regions are stably folded in the genome (Figure S3). In contrast, the coding regions of the S and E genes exhibit more relaxed and variable conformations, with both SHAPE reactivity and Shannon entropy exceeding the global median (Figure S3). Among the known functional regions, the 5' UTR (position: 1-310 nt), which is involved in translation initiation and genome packaging, exhibited lower SHAPE reactivity and lower Shannon entropy compared to the global median,

indicating a well-folded overall structure. The frameshifting stimulation element (FSE) (position 12605-12690), which is involved in mediating ribosomal frame shifting, exhibited high SHAPE reactivity and low Shannon entropy values. The 3' UTR (position 27702-28044 nt), which is responsible for genome cyclization, mRNA stability, and initiation of negative-strand synthesis, showed low SHAPE reactivity and high Shannon entropy values. These functional regions in the PEDV genome exhibiting distinct SHAPE reactivity and Shannon entropy profiles suggest that these regions may adopt different structural folding patterns to perform specific functions. (Figure S3 and Figure 1).

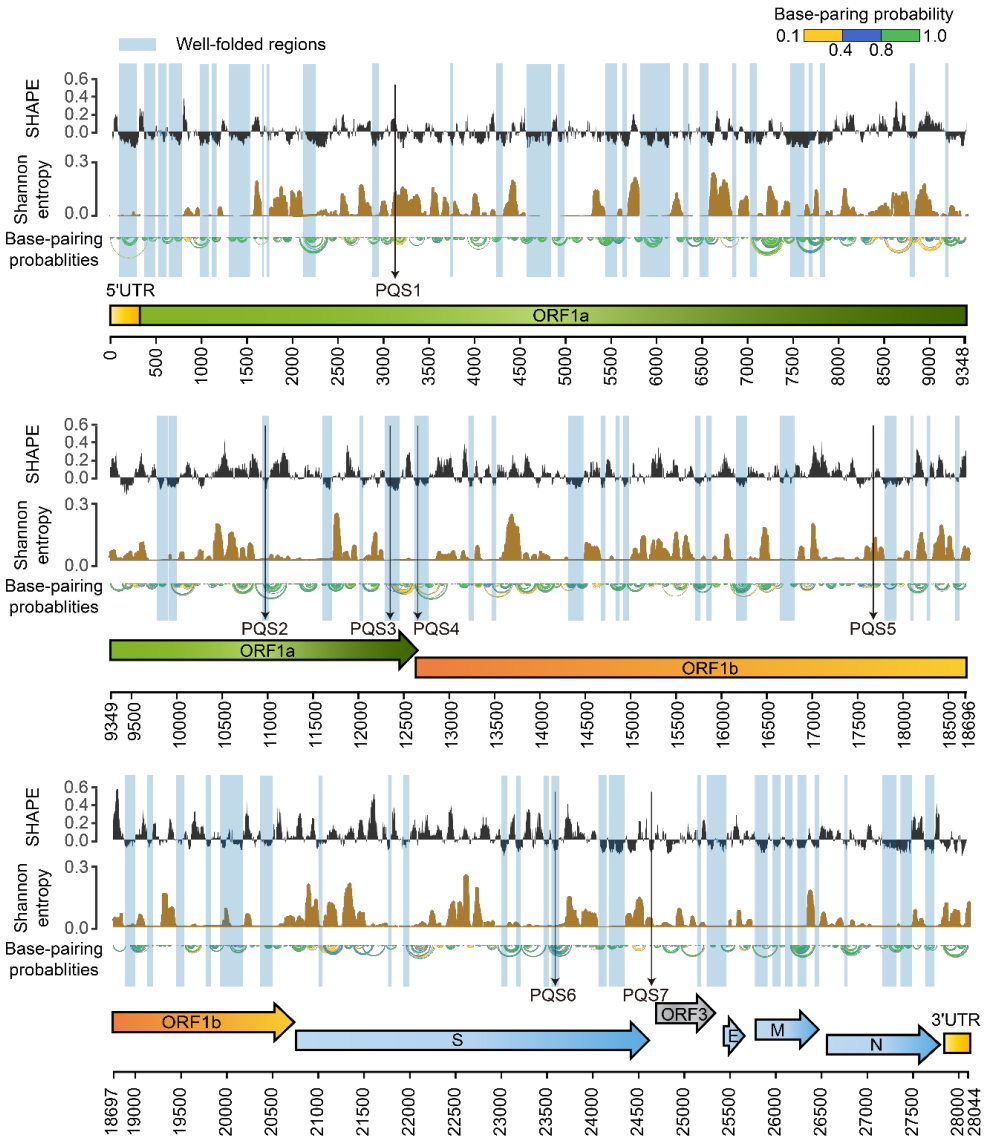

Figure 2. SHAPE structural map of the PEDV genome in infected cells. From top to bottom: SHAPE reactivity, Shannon entropy, base-pairing probabilities, translation reading frames are indicated by arrows, and genomic coordinates are marked at the bottom. Well-folded regions with low SHAPE reactivity and low Shannon entropy are shaded in blue. Potential G4 forming sequences (PQSs) are marked with a black arrow.

Functional categorization of PEDV genome reveals potentially targetable RNA

To further analyse other functional regions through SHAPE reactivity and Shannon entropy, we combined these two parameters for the classification of structural features (see Methods). Low SHAPE regions were defined as windows in which >75% of the bases exhibited SHAPE reactivities below the global median (calculated on the full PEDV genome), and low Shannon entropy regions were defined according to the same rule, following the Manfredonia's methods [13]. Windows in which >75% of their bases exhibiting values above the global median were picked as high SHAPE or high Shannon regions. Finally, four categories of RNAs in the PEDV genome were defined according to the relative values of SHAPE reactivity and Shannon entropy. (Figure 1, Table S2 and Table S3).

Regions with high base-pairing rates (low SHAPE) accounted for 38.05% of the PEDV genome, including the 5' UTR and 3' UTR (Figure 1A, B). Of these the low SHAPE-low Shannon regions (26.40%) maintain stable conformations throughout the viral life cycle and are assumed to represent the well-folded skeleton and the potentially functional regions of the viral genome (Figure 1A). 60 well-folded structures were further identified in these regions, which exhibited consistent conformations in two replicates (Figure S4-Figure S13). The low SHAPE-high Shannon regions (11.65%), which exhibit variable conformations, might perform different functions through conformational transitions (Figure 1B). Designing drugs targeting these regions would require identifying the specific conformations that are responsible for the functions.

Regions with low base-pairing rates (high SHAPE) accounted for 15.82% of the PEDV genome, including the TRS-Ls of the 5' UTRs (positions 57-72 nt) and FSE regions (positions 12605-12690 nt) (Figure 1C, D). RNAs in the high SHAPE-high Shannon regions (6.23%) are prone to conformational shifts, and their high accessibility suggests that they are more susceptible to interact with small molecule drugs (Figure 1C). In contrast, RNAs in the high SHAPE-low Shannon regions (9.59% of the PEDV genome) maintain a stable single-stranded state within the cell, facilitating pairing with exogenous complementary sequences (Figure 1D). Consequently, these regions are considered to be ideal targets for designing antiviral oligonucleotides [13,33,34].

Not all folded functional structures are potential antiviral targets

Although well-folded structures are thought to be potential targets for antiviral molecules[13,25,56,57], there have been relatively limited experimental validations. The 5'UTR-SL5 of coronaviruses is required for defective interfering (DI) RNA replication[58] and important for viral packaging[59]. In the PEDV genome, the 5'UTR-SL5 with four-way junction structure exhibits well-folded characteristics, similar to that of SARS-CoV-2 (Figure S14). Seven compounds (Table S7, compounds 1-7), previously reported to interact with the 5' UTR-SL5 of SARS-CoV-2 [57], were found to bind with the 5' UTR-SL5 of PEDV *in vitro*, and two compounds (compounds 1 and 4) exhibited dissociation constants (K_D) at the μM-level (Figure S15). However, none of these compounds inhibited the proliferation of PEDV at 50 μM in cells (Figure S16). This result suggests that compounds bound to well-folded functional structures *in vitro* do not necessarily exhibit antiviral activity in cells. The accessibility of compounds to potential RNA targets could be an important issue, suggesting that the targetability of the 60 well-folded RNA structures in the PEDV genome requires further validation.

Identification of PQSs in dynamic single-stranded regions

Dynamic RNAs in the high SHAPE-high Shannon regions can be induced by compounds to form stable structures that perform specific biological functions[11]. We identified 49 regions with high SHAPE reactivity and high Shannon entropy with at least 25 nucleotides in length (Table S3), spanning a total of 1748 bases (~6.2% of the PEDV genome). To predict the three-dimensional structures of sequences in these regions, we aligned these sequences with those of known structures in the PDB database. Fourteen sequences with length in the range 12-46 nucleotides were found to exhibit similarity >70% to sequences of RNA structures in the PDB database (Table S4). In particular, the RNA motif spanning positions 10756-10783 nt showed 75% sequence similarity to the structure of the Cas12i1 crRNA (PDB id: 7D3J, Figure S17A-C)[60]. Similarly, the motif spanning positions 11099-11119 nt exhibited 81% sequence similarity to the sequence of the CPEB3 nuclease P4 hairpin (PDB id: 2M5U, Figure 3A-C)[61]. Although the PDB database provides a wealth of three-dimensional RNA structure information as valuable templates for SHAPE-based RNA structure building, identifying and finding small molecules that can accurately recognise and bind to specific RNA structures to effectively stabilise their conformation remains a formidable challenge.

To investigate whether RNA motifs in high SHAPE-high Shannon regions can be stabilized to inhibit viral proliferation, we focused on the G4 structure in these regions. Nine putative PQSs were identified by at least three of the G4 prediction tools (Figure S18). Given the abundance of PEDV mutant strains in nature, further conservation analyses identified seven PQSs that are highly conserved (> 90%) among 761 PEDV strains as potential functional G4 candidates (PQS1-PQS7, Table S5). PQS2, PQS3, PQS4, and PQS6 form stable double-stranded structures with low SHAPE reactivity and low Shannon entropy, making them less likely to fold into stable G4 structures (Figure S19-S22). PQS5 and PQS7 did not fall in our four defined regions and were also omitted from further investigation. Conversely, PQS1 is located in the dynamic single-stranded region with high SHAPE reactivity and high Shannon entropy (Figure 3D and 3E), with relatively little competition from local hairpin structure folding. Therefore, PQS1 was selected for further structural and functional validation.

PQS1 is located in the nsp3 gene of the PEDV genome (Figure 3F) and is highly conserved in all 761 PEDV strains (98.37%, Figure S23A). Native polyacrylamide gel electrophoresis (PAGE) and circular dichroism (CD) measurements were conducted to further verify the formation of the PQS1 G4 structure. The migration rate of PQS1 noticeably accelerated with increasing K^+ concentration (Figure S23B), whereas the migration rate of the mutant sequences (PQS1m) that did not form G4 remained relatively constant. CD spectra revealed that PQS1 adopted a parallel G4 structure in the presence of K^+, with the corresponding typical features of positive CD band at approximately 265 nm and a negative peak near 240 nm (Figure S23C). ^{1}H nuclear magnetic resonance (NMR) was utilized to further characterize the G4 structure of PQS1. Chemical shifts in the 10-12 ppm range are considered a hallmark of G4 structures[62]. The ^{1}H NMR spectrum of PQS1 displayed prominent imino proton peaks in this region, confirming the formation of the PQS1 G4 structure (Figure 4A).

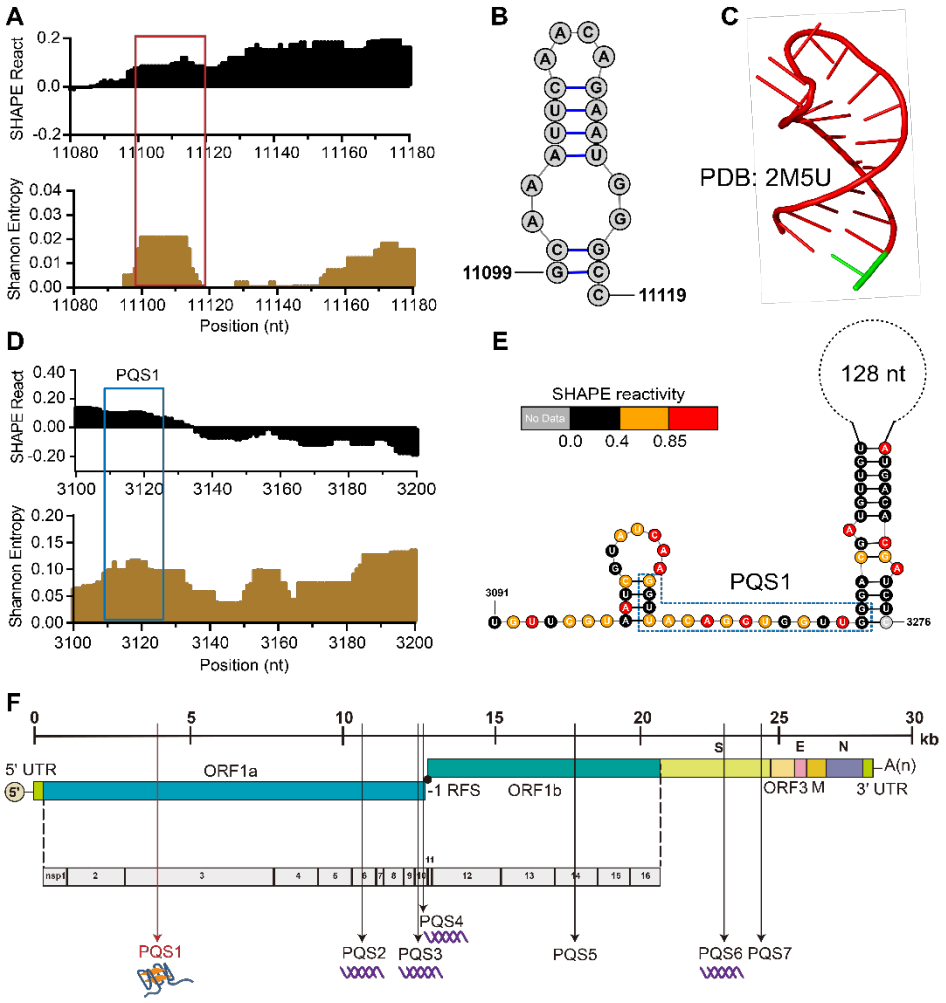

Figure 3. Potential RNA motifs with specific structures in high SHAPE-high Shannon regions. (A) SHAPE reactivity and Shannon entropy of RNA motif (11099-11119 nt) in the PEDV genome. **(B)** Secondary structure of the RNA motif (11099-11119 nt) predicted by RNAstructure. **(C)** Structure of CPEB3 nuclease P4 hairpin (PDB: 2M5U). The red section indicates nucleotides that are fully matched with the RNA sequence (11099-11119 nt) in the PEDV genome. **(D)** SHAPE reactivity and Shannon entropy for the regions containing PQS1. **(E)** Local secondary structure of the region containing PQS1 predicted with SHAPE reactivity constraints. PQS1 is marked with a blue dashed box. **(F)** Distribution of PQSs in the PEDV genome.

Stabilization of PQS1 as a G-quadruplex inhibits RNA synthesis, gene expression and viral proliferation

The addition of the G4-specific ligand Braco-19 (Figure S25A) significantly increased the melting temperature (Tm) of PQS1 by 10.8 °C, indicating that the RNA G4 structure of PQS1 is stabilized by Braco-19 (Figure 4B). Furthermore, the RNA stop assays and the EGFP expression inhibition experiments highlighted the effect of a stable G4 structure of PQS1 in inhibiting RNA replication (Figure S24A, B) and protein expression (Figure 4C and Figure S25B, C).

To explore whether PQS1 could be a potential drug target, we constructed a recombinant virus without PQS1 (AJ1102-PQS1mut) through a synonymous mutation (G3109A) that does not change the amino acid code (Figure S26A). The proliferation level of AJ1102-PQS1mut was significantly faster than that of the wild-type strain rAJ1102, suggesting that disruption of the folding of the PQS1 facilitates viral proliferation (Figure 4D). Another potential G4 sequence, PQS3, located in a stable double-stranded region with low SHAPE reactivity and low Shannon entropy, was used as a control. The PQS3 mutant recombinant virus (AJ1102-PQS3mut) proliferated at the same level as the wild type after synonymous mutation (G12322A) of its G4 sequence (Figure S26B).

The inhibitory effect of Braco-19 on the proliferation of the PEDV AJ1102 was quantified by RT-qPCR. Braco-19 inhibited PEDV proliferation with a median effective inhibitory concentration (EC50) of 3.60 μM (Figure S27A), which is well below the median cytotoxic concentration (CC50 > 100 μM) of Braco-19 (Figure S27B). In contrast, no inhibition was observed for Braco-19 against the mutant strain AJ1102-PQS1mut (Figure 4E). Western blot analysis was used to determine the effects of Braco-19 on viral protein expression in infected cells using the PEDV N protein antibody. The production of AJ1102-WT N protein was significantly reduced with increasing concentrations of Braco-19, while no change was observed in the production of AJ1102-PQS1mut N protein (Figure 4F), indicating that Braco-19 inhibited protein expression of PEDV in cells by targeting PQS1.

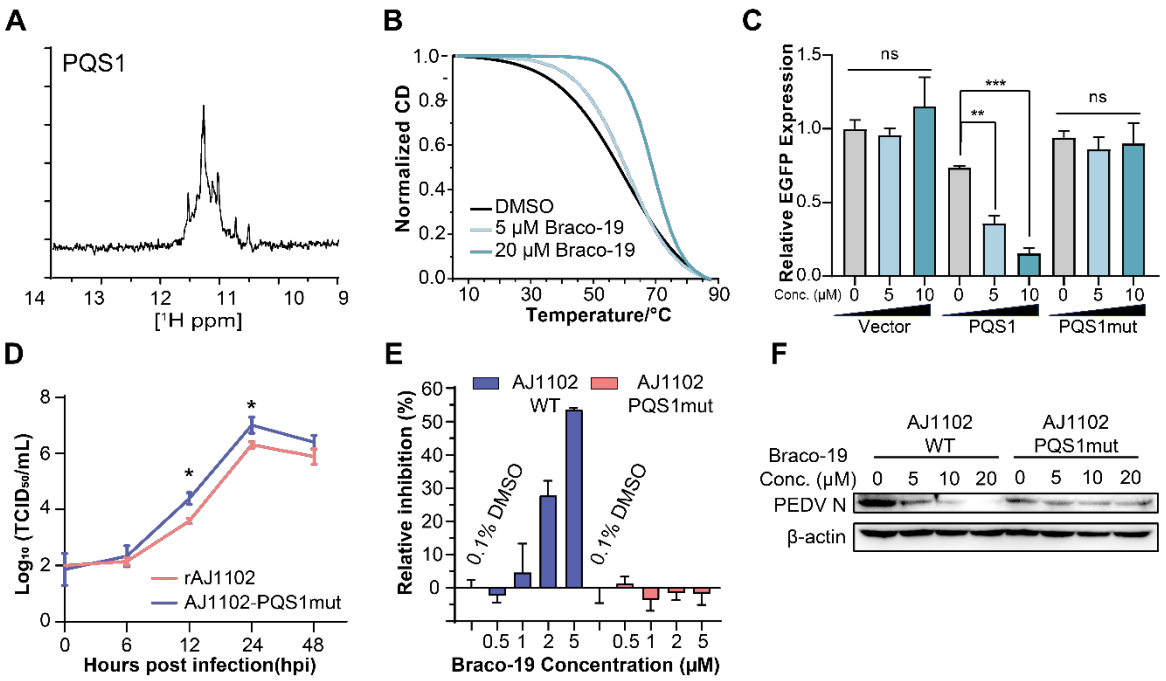

Figure 4. The G-quadruplex structure, biological functions of PQS1, and antiviral effects of Braco-19. (A) ^{1}H NMR analysis of PQS1. **(B)** CD melting profiles of PQS1. **(C)** Quantitative fluorescence signal using the corrected total cell fluorescence method for EGFP in cells transfected with plasmids containing the empty vector, PQS1 and PQS1mut. **(D)** Proliferation curve of the PEDV wild type (WT) strain and PQS1 mutant strain. (**E**) The relative inhibition rates of Braco-19 against AJ1102-WT and AJ1102-PQS1mut. (**F**) Western blot analysis of the effects of Braco-19 on the viral N protein expression of AJ1102-WT and AJ1102-PQS1mut.

Efficient antiviral siRNAs design targeting stable single-stranded regions

High SHAPE-low Shannon regions represent stable, accessible RNAs in cells, likely exposing long Watson–Crick edges of multiple residues available for base-pairing with complementary sequences (Figure 1C). Our analysis of fold characterisation (see Methods) identified 73 such regions (Table S3) with at least 25 nucleotides in length, spanning a total of 2696 bases (~9.6% of the PEDV genome). To test the suitability of these regions as targets for antiviral siRNAs, we designed four siRNAs targeting these regions that (i) have at least 10 consecutive unpaired nucleotides, and (ii) were highly conserved across the different PEDV strains (Table S4, Figure 5A-D). As a control, we designed four additional highly conserved siRNAs targeting duplex regions (Figure S28) and one non-targeting siRNA that does not target any sequence of the viral and host genomes (Table S4).

RT–qPCR results of infected cells at 16 h post viral exposure revealed that, four siRNAs targeting high SHAPE-low Shannon regions significantly inhibited viral proliferation in cells (Figure 5E) compared with the non-targeting siRNA control (siRNA-NC). Consistent with the RT-qPCR results, compared with the siRNA-NC, we observed that these four siRNAs had minimal cytopathic effects (CPEs) in Vero cells (Figure S29). Further $TCID_{50}$ analysis demonstrated that, three out of the four tested siRNAs targeting single-stranded regions reduced the $TCID_{50}$ by approximately 1.5log10 titer compared to the siRNA-NC (ss-1, ss-2, and ss-3; Figure 5F). In contrast, only one of the four siRNAs targeting duplexes could slightly prevent proliferation of PEDV in cells (ds4; Figure 5E, F and Figure S29). The target sequences of four single-stranded targeting siRNAs are more than 85% conserved across 761 known PEDV strains. Combining two or more of these siRNAs is expected to be effective in inhibiting nearly all known PEDV strains. These results showed that siRNAs targeting single-stranded regions with high SHAPE reactivity and low Shannon entropy can be more effective at exerting their antiviral effects.

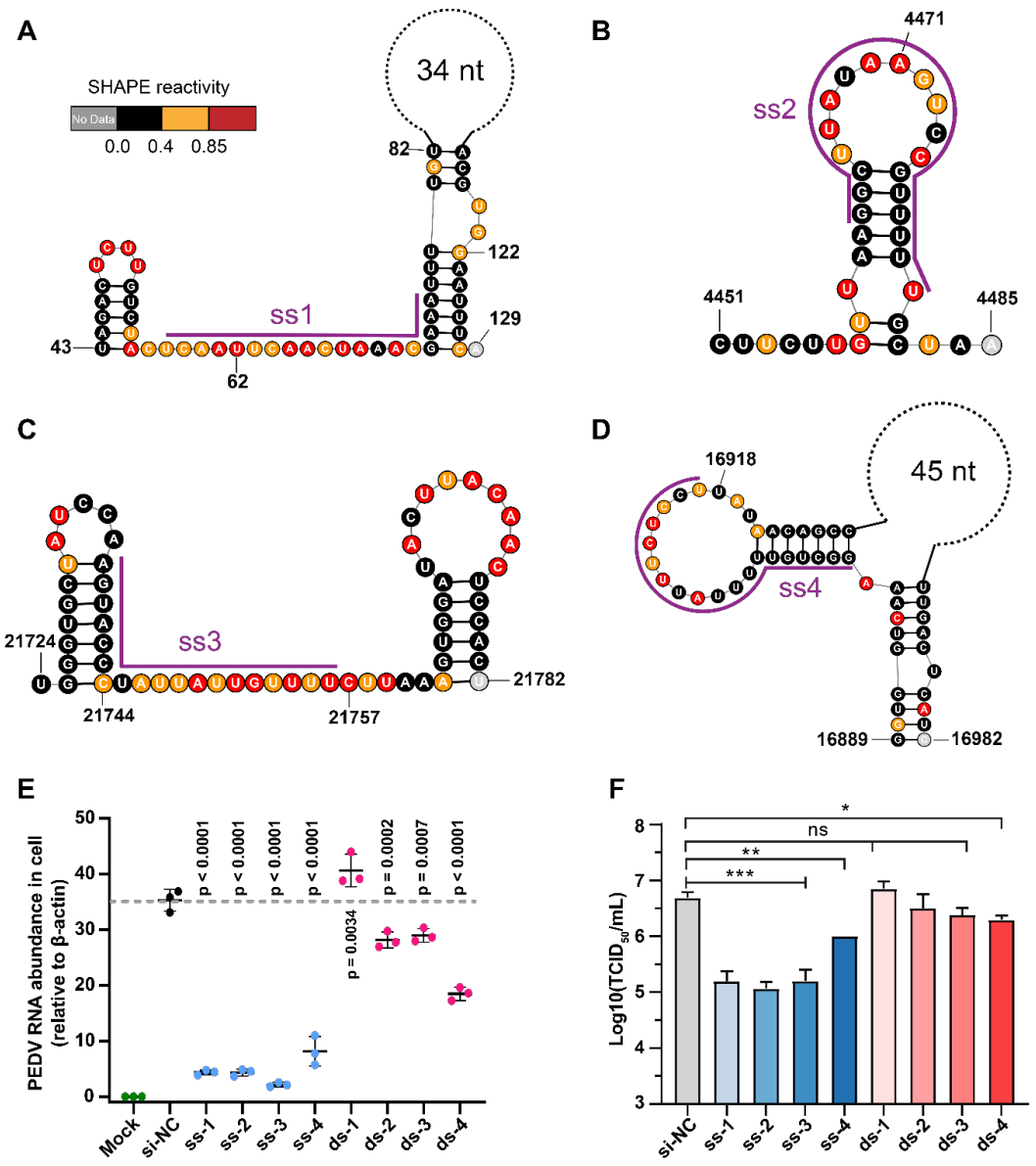

Figure 5. Secondary structure of target regions and antiviral effects of siRNAs. (A-D) Local secondary structures of the stable single-stranded regions targeted by siRNAs predicted by SHAPE reactivity as a constraint; (A) ss1; (B) ss2; (C) ss3; (D) ss4. **(E)** qPCR showing the relative abundance of the PEDV RNA genome in infected Vero cells. The four siRNAs targeting the high SHAPE-low Shannon regions and the four siRNAs targeting the duplex regions are labeled ss-1 to ss-4 and ds-1 to ds-4, respectively. si-NC was a control siRNA that did not target any viral or host sequences, and the mock group was not inoculated with virus. **(F)** $TCID_{50}$ assays for detecting virus titers. The presented results represent the means and standard deviations of data from three independent experiments. ns: no significant difference. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$, Duncan's multiple comparison test.

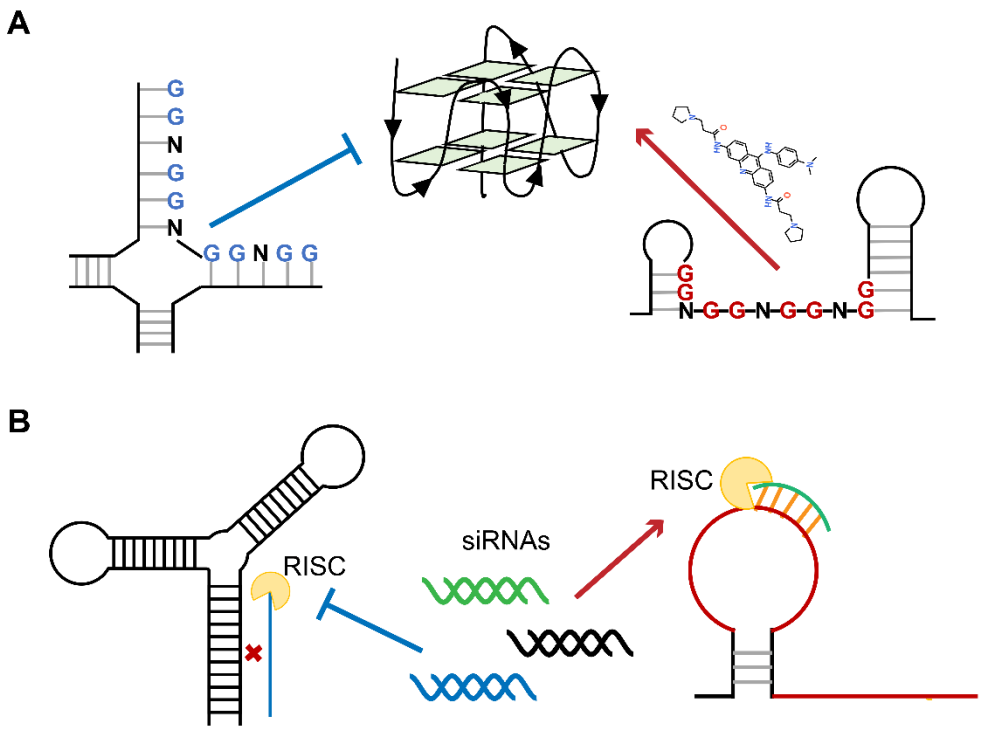

Figure 6. Single-stranded RNA regions in viral genomes as the targets for antiviral therapy. (A) PQSs in the single strands are easier to be induced into G4s by ligands than those in the paired regions. (**B**) Influence of the secondary structures on the binding efficiency of siRNAs.

DISCUSSION AND CONCLUSIONS

Several RNA-targeting antiviral strategies have highlighted the potential of therapeutic application[63]. However, it is challenging to identify these RNAs from the viral genome *in situ*. In SHAPE-MaP analysis, regions with low reactivity and low entropy values are more likely to form a single stable structure in cells, and are therefore considered candidates for potential functional RNAs and targets for antiviral therapy. Novel well-defined RNA structures that are critical for viral fitness in HIV, HCV, DENV and SINV have been identified by the analysis of low SHAPE and low entropy regions [20,23,64,65]. However, there are still no clinically approved small molecule drugs or oligonucleotides targeting stable folded RNA structures. Our data shows that two compounds that interacted with the well-folded 5'UTR-SL5 of PEDV *in vitro* at μM-level K_D, but did not show antiviral activity at 50 μM in cells (Figure S16). This result suggests that, on the one hand, the low accessibility of the regions with low SHAPE reactivity makes them difficult to access by compounds in cells, and may require compounds to have better K_D to RNA. On the other hand, the binding of compounds to the well-folded

structure does not necessarily alter its function. In contrast, the FDA- approved compound Risdiplam exerts its biological activity by stabilizing transient RNA structures[10,11,66], suggesting that it may be possible to inhibit viral proliferation by stabilizing variable conformations. We therefore focused our attention on highly accessible regions with high SHAPE reactivity in cells.

We identified 42 low SHAPE and high Shannon regions and 49 high SHAPE and high Shannon regions in the PEDV genome by SHAPE-MaP, representing conformationally variable RNAs in the viral genome. As the low SHAPE regions are highly structured, it is difficult for the G4 in these regions to outcompete the folding of long base paired structures (Figure 6A). We therefore selected a potential G4 sequence (PQS1) located in the high SHAPE-high Shannon region of nsp3 (PLpro), where a G4-disrupting mutation promotes viral proliferation. We found that the compound Braco-19 significantly stabilised the G4 structure of PQS1 *in vitro* and inhibited PEDV proliferation in cells. Importantly, Braco-19 did not inhibit the proliferation of the mutant strain without this G4 forming sequence. This suggests that the compound can selectively target the PQS1 of the high SHAPE-high Shannon region in cells. This approach can be extended to other RNA viruses and potentially to DNA viruses where RNA intermediates play critical roles in the viral life cycle.

Recently, the FDA has approved several siRNAs for therapeutic purposes[67]. siRNA-mediated RNAi is particularly suited to inhibit viruses with single-stranded RNA genomes due to its potential to specifically silence virtually any therapeutic target. When screening siRNA target sequences, sequence conservation and the functional importance of the target region are usually given priority. However, the RNA structure of the target region also has a significant impact on the efficiency of RNAi (Figure 6B)[13,16,33,34,68,69]. Manfredonia et al. proposed that the identification of persistent single-stranded regions in cells by SHAPE-MaP as a suitable target for antiviral oligonucleotides [13]. Here we identified 73 persistent single-stranded regions with high SHAPE reactivity and low Shannon entropy in the PEDV genome in cells and demonstrated higher antiviral success of siRNAs targeting these regions. Our data supports Manfredonia's hypothesis that fold characterisation based on SHAPE-MaP can lead to more efficient design of antiviral siRNAs. With the maturation and cost reduction of *in situ* RNA structure probing technologies, we believe that the structural feature classification presented here can be extended to other arbitrary RNA viral genomes.

DATA AVAILABILITY

SHAPE-MaP datasets generated in this study have been deposited at Zenodo.org (https://zenodo.org/doi/10.5281/zenodo.13743827) and include Gene Expression Omnibus (GSE271098). XML files with normalized SHAPE reactivities (SHAPE_react_rep1/2.react.xml); WIG files with Shannon entropies (SHAPE_react_rep1/2.react.xml) and the full secondary structure (PEDV_incell_secondary structure.ct).

SUPPLEMENTARY INFORMATION

The supplementary information contains Tables S1-S7 and Figures S1-S29

AUTHOR CONTRIBUTIONS

Dehua Luo: Designed of the study, investigation, formal analysis, validation, writing; **Yingge Zheng:** performed the EGFP reporter vector construction and PQSs mutant virus rescue; **Zhiyuan Huang, Zi Wen, Yingxiang Deng:** SHAPE data processing. **Lijun Guo, Qingling Li, Yuqing Bai:** data visualisation; **Shozeb Haider:** validation, writing - review and editing; **Dengguo Wei:** conceptualization, supervision, validation, writing – review and editing.

Acknowledgements

This work was supported by the National Key R&D Plan of China (grant no. 2021YFD1801102), the National Natural Science Foundation of China (22077043 and 32272491), the Fundamental Research Funds for the Central Universities (2662023PY005 to DW), the HZAU-AGIS Cooperation Fund (SZYJY2022016 to DW), the Natural Science Foundation of Hubei Province (2021CFA061 to DW, 2017CFB233 to SL), and funding from the National Key Laboratory of Agricultural Microbiology (AML2023B05 to DW). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

COMPETING INTEREST STATEMENT

Shozeb Haider is a reviewing editor at eLife. The other authors declare no competing interests

REFERENCES

1. K. D. Warner, C. E. Hajdin and K. M. Weeks, *Nat Rev Drug Discov*, 2018, 17, 547–558.
2. M. D. Disney, *J Am Chem Soc*, 2019, 141, 6776–6790.
3. A. Ursu, J. L. Childs-Disney, R. J. Andrews, C. A. O'Leary, S. M. Meyer, A. J. Angelbello, W. N. Moss and M. D. Disney, *Chem. Soc. Rev.*, 2020, 49, 7252–7270.
4. F. Li and A. Wang, *Trends Microbiol*, 2019, 27, 792–805.
5. S. Kovachka, M. Panosetti, B. Grimaldi, S. Azoulay, A. Di Giorgio and M. Duca, *Nat Rev Chem*, 2024, 8, 120–135.
6. Q. Tang and A. Khvorova, *Nat Rev Drug Discov*, 2024, 23, 341–364.
7. S. T. Crooke, B. F. Baker, R. M. Crooke and X.-H. Liang, *Nat Rev Drug Discov*, 2021, 20, 427–453.
8. S. Chen, S. N. Heendeniya, B. T. Le, K. Rahimizadeh, N. Rabiee, Q. U. A. Zahra and R. N. Veedu, *BioDrugs*, 2024, 38, 177–203.
9. C. F. Bennett, *Annu Rev Med*, 2019, 70, 307–321.
10. C. Sheridan, *Nature Biotechnology*, 2021, 39, 6–8.
11. H. Ratni, R. S. Scalco and A. H. Stephan, *ACS Med Chem Lett*, 2021, 12, 874–877.
12. Zhang, K. Huang, D. Xie, J. Y. Lau, W. Shen, P. Li, D. Wang, Z. Zou, S. Shi, H. Ren, Y. Wang, Y. Mao, M. Jin, G. Kudla and Z. Zhao, *Nat Commun*, 2021, 12, 5695.
13. I. Manfredonia, C. Nithin, A. Ponce-Salvatierra, P. Ghosh, T. K. Wirecki, T. Marinus, N. S. Ogando, E. J. Snijder, M. J. van Hemert, J. M. Bujnicki and D. Incarnato, *Nucleic Acids Res*, 2020, 48, 12436–12452.
14. R. A. Flynn, Q. C. Zhang, R. C. Spitale, B. Lee, M. R. Mumbach and H. Y. Chang, *Nat Protoc*, 2016, 11, 273–290.
15. O. Ziv, J. Price, L. Shalamova, T. Kamenova, I. Goodfellow, F. Weber and E. A. Miska, *Molecular Cell*, 2020, 80, 1067-1077.e5.
16. C. Cao, Z. Cai, X. Xiao, J. Rao, J. Chen, N. Hu, M. Yang, X. Xing, Y. Wang, M. Li, B. Zhou, X. Wang, J. Wang and Y. Xue, *Nat Commun*, 2021, 12, 3917.
17. R. G. Huber, X. N. Lim, W. C. Ng, A. Y. L. Sim, H. X. Poh, Y. Shen, S. Y. Lim, K. B. Sundstrom, X. Sun, J. G. Aw, H. K. Too, P. H. Boey, A. Wilm, T. Chawla, M. M. Choy, L. Jiang, P. F. de Sessions, X. J. Loh, S. Alonso, M. Hibberd, N. Nagarajan, E. E. Ooi, P. J. Bond, O. M. Sessions and Y. Wan, *Nat Commun*, 2019, 10, 1408.

18. M. J. Smola, T. W. Christy, K. Inoue, C. O. Nicholson, M. Friedersdorf, J. D. Keene, D. M. Lee, J. M. Calabrese and K. M. Weeks, *Proc Natl Acad Sci U S A*, 2016, 113, 10322–10327.
19. T. C. T. Lan, M. F. Allan, L. E. Malsick, J. Z. Woo, C. Zhu, F. Zhang, S. Khandwala, S. S. Y. Nyeo, Y. Sun, J. U. Guo, M. Bathe, A. Näär, A. Griffiths and S. Rouskin, *Nat Commun*, 2022, 13, 1128.
20. N. A. Siegfried, S. Busan, G. M. Rice, J. A. E. Nelson and K. M. Weeks, *Nat Methods*, 2014, 11, 959–965.
21. M. A. Boerneke, J. E. Ehrhardt and K. M. Weeks, *Annu Rev Virol*, 2019, 6, 93–117.
22. P. Li, Y. Wei, M. Mei, L. Tang, L. Sun, W. Huang, J. Zhou, C. Zou, S. Zhang, C.-F. Qin, T. Jiang, J. Dai, X. Tan and Q. C. Zhang, *Cell Host Microbe*, 2018, 24, 875-886.e5.
23. D. M. Mauger, M. Golden, D. Yamane, S. Williford, S. M. Lemon, D. P. Martin and K. M. Weeks, *Proc Natl Acad Sci U S A*, 2015, 112, 3692–3697.
24. S. L. Yang, L. DeFalco, D. E. Anderson, Y. Zhang, J. G. A. Aw, S. Y. Lim, X. N. Lim, K. Y. Tan, T. Zhang, T. Chawla, Y. Su, A. Lezhava, A. Merits, L.-F. Wang, R. G. Huber and Y. Wan, *Nat Commun*, 2021, 12, 5113.
25. N. C. Huston, H. Wan, M. S. Strine, R. de Cesaris Araujo Tavares, C. B. Wilen and A. M. Pyle, *Mol Cell*, 2021, 81, 584-598.e5.
26. E. Pollom, K. K. Dang, E. L. Potter, R. J. Gorelick, C. L. Burch, K. M. Weeks and R. Swanstrom, *PLoS Pathog*, 2013, 9, e1003294.
27. C. K. Kwok, G. Marsico, A. B. Sahakyan, V. S. Chambers and S. Balasubramanian, *Nat Methods*, 2016, 13, 841–844.
28. A. M. Fleming, Y. Ding, A. Alenko and C. J. Burrows, *ACS Infect Dis*, 2016, 2, 674–681.
29. P. Fang, C. Xie, T. Pan, T. Cheng, W. Chen, S. Xia, T. Ding, J. Fang, Y. Zhou, L. Fang, D. Wei and S. Xiao, *Nucleic Acids Res*, 2023, 51, 10752–10767.
30. C. Zhao, G. Qin, J. Niu, Z. Wang, C. Wang, J. Ren and X. Qu, *Angew Chem Int Ed Engl*, 2021, 60, 432–438.
31. E. Ruggiero and S. N. Richter, *Nucleic Acids Res*, 2018, 46, 3270–3283.
32. S.-R. Wang, Y.-Q. Min, J.-Q. Wang, C.-X. Liu, B.-S. Fu, F. Wu, L.-Y. Wu, Z.-X. Qiao, Y.-Y. Song, G.-H. Xu, Z.-G. Wu, G. Huang, N.-F. Peng, R. Huang, W.-X. Mao, S. Peng, Y.-Q. Chen, Y. Zhu, T. Tian, X.-L. Zhang and X. Zhou, *Sci Adv*, 2016, 2, e1501535.
33. S. M. Sagan, N. Nasheri, C. Luebbert and J. P. Pezacki, *Chemistry & Biology*, 2010, 17, 515–527.

34. J. Piasecka, E. Lenartowicz, M. Soszynska-Jozwiak, B. Szutkowska, R. Kierzek and E. Kierzek, *Molecular Therapy - Nucleic Acids*, 2020, 19, 627–642.
35. E. Bowden-Reid, S. Ledger, Y. Zhang, F. Di Giallonardo, A. Aggarwal, A. O. Stella, A. Akerman, V. Milogiannakis, G. Walker, W. Rawlinson, S. Turville, A. D. Kelleher and C. Ahlenstiel, *Antiviral Res*, 2023, 217, 105677.
36. E. M. Westerhout and B. Berkhout, *Nucleic Acids Research*, 2007, 35, 4322–4330.
37. A. Qureshi, V. G. Tantray, A. R. Kirmani and A. G. Ahangar, *Rev Med Virol*, 2018, 28, e1976.
38. R. Lorenz, D. Luntzer, I. L. Hofacker, P. F. Stadler and M. T. Wolfinger, *Bioinformatics*, 2016, 32, 145–147.
39. K. E. Deigan, T. W. Li, D. H. Mathews and K. M. Weeks, *Proc Natl Acad Sci U S A*, 2009, 106, 97–102.
40. D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker and D. H. Turner, *Proc Natl Acad Sci U S A*, 2004, 101, 7287–7292.
41. D. H. MATHEWS, *RNA*, 2004, 10, 1178–1190.
42. C. Lee, *Virol J*, 2015, 12, 193.
43. K. Jung, L. J. Saif and Q. Wang, *Virus Res*, 2020, 286, 198045.
44. Z. Niu, S. Zhang, S. Xu, J. Wang, S. Wang, X. Hu, L. Zhang, L. Ren, J. Zhang, X. Liu, Y. Zhou, L. Yang and Z. Song, *Viruses*, 2023, 15, 956.
45. D. Incarnato, E. Morandi, L. M. Simon and S. Oliviero, *Nucleic Acids Res*, 2018, 46, e97.
46. B. Langmead and S. L. Salzberg, *Nat Methods*, 2012, 9, 357–359.
47. O. Kikin, L. D'Antonio and P. S. Bagga, *Nucleic Acids Res*, 2006, 34, W676-682.
48. O. Doluca, *Journal of Theoretical Biology*, 2019, 463, 92–98.
49. A. Varizhuk, D. Ischenko, V. Tsvetkov, R. Novikov, N. Kulemin, D. Kaluzhny, M. Vlasenok, V. Naumov, I. Smirnov and G. Pozmogova, *Biochimie*, 2017, 135, 54–62.
50. J. Hon, T. Martínek, J. Zendulka and M. Lexa, *Bioinformatics*, 2017, 33, 3373–3379.
51. G. E. Crooks, G. Hon, J.-M. Chandonia and S. E. Brenner, *Genome Res*, 2004, 14, 1188–1190.
52. S. A. Bustin, V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele and C. T. Wittwer, *Clin Chem*, 2009, 55, 611–622.
53. S.-C. Chen and R. C. L. Olsthoorn, *Virology*, 2010, 401, 29–41.
54. D. Yang and J. L. Leibowitz, *Virus Res*, 2015, 206, 120–133.

55. R. Madhugiri, N. Karl, D. Petersen, K. Lamkiewicz, M. Fricke, U. Wend, R. Scheuer, M. Marz and J. Ziebuhr, *Virology*, 2018, 517, 44–55.
56. J. L. Childs-Disney, X. Yang, Q. M. R. Gibaut, Y. Tong, R. T. Batey and M. D. Disney, *Nat Rev Drug Discov*, 2022, 21, 736–762.
57. S. Sreeramulu, C. Richter, H. Berg, M. A. Wirtz Martin, B. Ceylan, T. Matzel, J. Adam, N. Altincekic, K. Azzaoui, J. K. Bains, M. J. J. Blommers, J. Ferner, B. Fürtig, M. Göbel, J. T. Grün, M. Hengesbach, K. F. Hohmann, D. Hymon, B. Knezic, J. N. Martins, K. R. Mertinkus, A. Niesteruk, S. A. Peter, D. J. Pyper, N. S. Qureshi, U. Scheffer, A. Schlundt, R. Schnieders, E. Stirnal, A. Sudakov, A. Tröster, J. Vögele, A. Wacker, J. E. Weigand, J. Wirmer-Bartoschek, J. Wöhnert and H. Schwalbe, *Angew Chem Int Ed Engl*, 2021, 60, 19191–19200.
58. C. G. Brown, K. S. Nixon, S. D. Senanayake and D. A. Brian, *J Virol*, 2007, 81, 7716–7724.
59. P. S. Masters, *Virology*, 2019, 537, 198–207.
60. B. Zhang, D. Luo, Y. Li, V. Perčulija, J. Chen, J. Lin, Y. Ye and S. Ouyang, *Nat Commun*, 2021, 12, 3476.
61. M. Skilandat, M. Rowinska-Zyrek and R. K. O. Sigel, *J Biol Inorg Chem*, 2014, 19, 903–912.
62. M. Adrian, B. Heddi and A. T. Phan, *Methods*, 2012, 57, 11–24.
63. J. Dai, X. Jiang, E. F. da Silva-Júnior, S. Du, X. Liu and P. Zhan, *Drug Discov Today*, 2024, 29, 104074.
64. K. M. Kutchko, E. A. Madden, C. Morrison, K. S. Plante, W. Sanders, H. A. Vincent, M. C. Cruz Cisneros, K. M. Long, N. J. Moorman, M. T. Heise and A. Laederach, *Nucleic Acids Res*, 2018, 46, 3657–3670.
65. E. A. Dethoff, M. A. Boerneke, N. S. Gokhale, B. M. Muhire, D. P. Martin, M. T. Sacco, M. J. McFadden, J. B. Weinstein, W. B. Messer, S. M. Horner and K. M. Weeks, *Proc Natl Acad Sci U S A*, 2018, 115, 11513–11518.
66. S. Campagne, S. Boigner, S. Rüdisser, A. Moursy, L. Gillioz, A. Knörlein, J. Hall, H. Ratni, A. Cléry and F. H.-T. Allain, *Nat Chem Biol*, 2019, 15, 1191–1198.
67. S. S. Ali Zaidi, F. Fatima, S. A. Ali Zaidi, D. Zhou, W. Deng and S. Liu, *J Nanobiotechnology*, 2023, 21, 381.
68. V. Patzel, S. Rutz, I. Dietrich, C. Köberle, A. Scheffold and S. H. E. Kaufmann, *Nat Biotechnol*, 2005, 23, 1440–1444.

69. M. Szabat, D. Lorent, T. Czapik, M. Tomaszewska, E. Kierzek and R. Kierzek, *Pathogens*, 2020, 9, 925.