Conservation of a Packaging Signal and the Viral Genome RNA Packaging Mechanism in Alphavirus Evolution[∇]

Dal Young Kim,^{1†} Andrew E. Firth,^{2†} Svetlana Atasheva,¹ Elena I. Frolova,¹ and Ilya Frolov^{1*}

*Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama,¹ and
Department of Pathology, University of Cambridge, Cambridge, United Kingdom²*

Received 30 March 2011/Accepted 1 June 2011

Alphaviruses are a group of small, enveloped viruses which are widely distributed on all continents. In infected cells, alphaviruses display remarkable specificity in RNA packaging by encapsidating only their genomic RNA while avoiding packaging of the more abundant viral subgenomic (SG), cellular messenger and transfer RNAs into released virions. In this work, we demonstrate that in spite of evolution in geographically isolated areas and accumulation of considerable diversity in the nonstructural and structural genes, many alphaviruses belonging to different serocomplexes harbor RNA packaging signals (PSs) which contain the same structural and functional elements. Their characteristic features are as follows. (i) Sindbis, eastern, western, and Venezuelan equine encephalitis and most likely many other alphaviruses, except those belonging to the Semliki Forest virus (SFV) clade, have PSs which can be recognized by the capsid proteins of heterologous alphaviruses. (ii) The PS consists of 4 to 6 stem-loop RNA structures bearing conserved GGG sequences located at the base of the loop. These short motifs are integral elements of the PS and can function even in the artificially designed PS. (iii) Mutagenesis of the entire PS or simply the GGG sequences has strong negative effects on viral genome packaging and leads to release of viral particles containing mostly SG RNAs. (iv) Packaging of RNA appears to be determined to some extent by the number of GGG-containing stem-loops, and more than one stem-loop is required for efficient RNA encapsidation. (v) Viruses of the SFV clade are the exception to the general rule. They contain PSs in the nsP2 gene, but their capsid protein retains the ability to use the nsP1-specific PS of other alphaviruses. These new discoveries regarding alphavirus PS structure and function provide an opportunity for the development of virus variants, which are irreversibly attenuated in terms of production of infectious virus but release high levels of genome-free virions.

Alphaviruses are a group of small, enveloped viruses which are widely distributed on all continents. Under natural conditions, these viruses are transmitted between vertebrate hosts by mosquito vectors. In mosquitoes, alphaviruses develop a persistent, life-long infection, characterized by the high concentrations of virus accumulated in the salivary glands (44), which promote virus transmission to amplifying hosts during the blood meal. In vertebrates, alphaviruses induce an acute infection and high-titer viremia, which is required for infection of new mosquitoes during blood ingestion. Thus, to support continuous virus circulation, this transmission cycle requires the presence of infectious virus at high titers in both vertebrates and invertebrates. To achieve this, alphaviruses have developed very efficient means of genome replication, structural protein synthesis, and interference with development of the antiviral response (2, 3, 13, 14). These processes enable viral accumulation in both mosquitoes and the blood of infected vertebrate organisms.

Alphaviruses enter the cells by receptor-mediated endocytosis, and after fusion of viral and endosomal membranes, nucleocapsids are released into the cytoplasm. Following their

further disassembly, free viral genomes begin to serve as templates for translation of the nonstructural polyproteins (nsPs) and, later, for synthesis of the negative-strand RNA intermediates, forming the double-stranded RNAs (dsRNAs). The dsRNA-nsP complexes initially assemble at the plasma membrane (10, 15), and at later time points postinfection, some of them are transported into the cytoplasm on the surface of endosomes (11, 15, 38). These complexes synthesize large quantities of genomic and subgenomic RNAs. The latter RNA is efficiently translated into structural proteins that package genomic RNA into infectious virions. This RNA packaging process remains poorly understood. It is unclear how the viral genome is selectively packaged into viral particles, while the more abundant subgenomic RNA or cellular messenger RNAs and tRNAs are left behind. It is generally believed that such selective packaging is determined by the binding of capsid protein to particular sequences in the viral RNA termed the packaging signals (PSs) (45, 46). In most of the alphaviruses, with the possible exception of Aura virus (36), PSs were expected to be found in the nonstructural (ns) polyprotein-coding sequence. They have been studied by analyzing the sequences retained in defective interfering (DI) RNAs (5, 23, 37, 47), which accumulate during virus passaging at a high multiplicity of infection (MOI) in cultured cells. In some cases, the identified sequences were additionally analyzed by assessing the efficiency of binding of some RNA fragments to certain capsid-specific peptides *in vitro* (45, 46). For Ross River virus (RRV), putative PSs were also defined by expressing fragments

* Corresponding author. Mailing address: Department of Microbiology, BBRB 373/Box 3, University of Alabama, 1530 Third Avenue South, Birmingham, AL 35294-2170. Phone: (205) 996-8957. Fax: (205) 996-4008. E-mail: ivfrolov@UAB.edu.

† These authors contributed equally to this work.

[∇] Published ahead of print on 15 June 2011.

of RNA genomes in cells producing complete sets of viral structural proteins and later analyzing their packaging into viral particles. RNA fragments demonstrating higher affinity for capsid and/or 3- to 4-fold-higher encapsidation efficiencies than other sequences have been proposed as virus-specific PSs (8). However, analysis of PS structure and function has not been performed in the context of replicating viruses, nor have the existing data been applied toward modification of virus replication.

In previous studies, we and others have designed a wide variety of chimeric alphaviruses, encoding nsPs and RNA promoter elements of one virus and the structural proteins derived from a heterologous alphavirus. All of these chimeras were viable and replicated in tissue culture to titers above 10^9 PFU/ml without adaptation of either capsid or nsP-coding sequences (3, 4, 7, 12, 18, 30, 31, 41, 42). These data brought into question the validity of the hypothesis regarding specificity of packaging of viral genomes into nucleocapsid and suggested that packaging appears to be determined by close intracellular compartmentalization of newly synthesized RNA and capsid protein rather than specific capsid RNA interactions. Therefore, in this study, we performed a detailed investigation into the existence of PS in the Venezuelan equine encephalitis virus (VEEV) genome, identified the main functional and structural components of this PS, and demonstrated a common recognition motif that is present in the PSs of some other alphaviruses, belonging to very distant, geographically isolated serocomplexes. Our data demonstrate the following. (i) An nsP1-specific RNA sequence located between nucleotides (nt) 856 and 1150 of the VEEV genome determines its packaging into infectious virions. (ii) Functioning of this sequence in RNA packaging is mostly independent of its position in the viral genome. (iii) Short GGG sequences located in the loops of the predicted RNA stem-loop structures play a critical role in the function of the VEEV PS in RNA packaging. (iv) While in different alphaviruses, the PS is composed of 4 to 6 stem-loops, containing GGG motifs, the presence of more than one stem-loop is required to achieve the wild-type (wt) level of packaging efficiency. (v) An artificially designed, GGG-containing PS can be as efficient as the wt PS in VEEV RNA packaging. (vi) PSs in other alphaviruses, except the SFV clade, demonstrate similar organization. (vii) The chikungunya virus (CHIKV)-specific PS and, most likely, the PSs of other members of the SFV clade are located in the nsP2-coding sequence. (viii) However, capsid proteins of the latter viruses retain an ability to recognize heterologous, nsP1-specific PSs of other alphaviruses.

MATERIALS AND METHODS

Cell cultures. The BHK-21 cells were kindly provided by Paul Olivo (Washington University, St. Louis, MO) and the Vero cells by Charles Rice (Rockefeller University, New York, NY). These cell lines were maintained at 37°C in alpha minimum essential medium (α MEM) supplemented with 10% fetal bovine serum (FBS) and vitamins. Mosquito C7/10 cells were obtained from Henry Huang (Washington University, St. Louis, MO) and propagated in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated FBS and 10% tryptose phosphate broth (TPB).

Plasmid constructs. The original plasmids (i) with the VEEV TC-83 genome (pVEEV), (ii) with the Sindbis virus (SINV) genome, having the green fluorescent protein (GFP) gene under the control of an additional subgenomic promoter (pSIN/GFP), and (iii) with the genome of chimeric virus encoding VEEV nsPs and CHIKV structural proteins (pVEE/CHIKV) were described elsewhere

(9, 17, 32). Plasmid containing CHIKV Thai strain nsP1- and nsP2-coding sequences were obtained from Scott Weaver (UTMB). Plasmid encoding Semliki Forest virus (SFV) nsPs was obtained from Invitrogen. Standard recombinant DNA techniques were used for all plasmid constructions. Maps and sequences are available from the authors upon request. DNA fragments with large numbers of mutations were synthesized from oligonucleotides using PCR-based approaches. The genome fragments of other alphaviruses analyzed, i.e., eastern equine encephalitis virus (EEEV) strain BeAr436087, SINV strain AR339, and CHIKV, were also synthesized by PCR, cloned into the viral genome-containing plasmids, and sequenced before rescuing the viruses.

RNA transcriptions. Plasmids were purified by centrifugation in CsCl gradients and linearized by different restriction enzymes, which utilized the restriction sites located downstream of the poly(A) sequence. RNAs were synthesized by SP6 RNA polymerase (Ambion) in the presence of a cap analog (New England BioLabs) as described elsewhere (34). The yield and integrity of transcripts were analyzed by gel electrophoresis under nondenaturing conditions. RNA concentration was measured on a Fotodyne imager, and transcription reactions were used for electroporation without additional purification.

RNA transfections. In all of the experiments, the transfections were performed by electroporation of the *in vitro*-synthesized RNA using previously described conditions (24). To rescue the viruses, 1 μ g of *in vitro*-synthesized viral genome RNA was electroporated into the cells (24), and then they were seeded into 100-mm dishes and incubated until cytopathic effects (CPEs) were observed. Virus titers were determined using a standard plaque assay on BHK-21 cells (22). In parallel, to assess the RNA infectivity, 10-fold dilutions of electroporated BHK-21 cells were seeded in 6-well Costar plates containing subconfluent, naïve BHK-21 cells. After 1 h of incubation at 37°C in a 5% CO₂ incubator, cells were overlaid with 2 ml of MEM-containing 0.5% ultrapure agarose (Invitrogen) supplemented with 3% FBS. Plaques were stained with crystal violet after 2 days incubation at 37°C, and infectivity was determined in PFU per μ g of transfected RNA.

Analysis of virus replication. BHK-21 or C7/10 cells in 6-well Costar plates (5×10^5 cells/well) were infected at multiplicities of infection (MOIs) indicated in the figure legends. After a 1-h-long incubation at 37°C or 30°C (appropriate for BHK-21 and C7/10 cells, respectively), cells were washed with phosphate-buffered saline (PBS) and covered by 1 ml of complete medium and further incubated at the appropriate temperature in the CO₂ incubator. At the time points indicated in the figures, medium was replaced, and virus titers in the harvested samples were determined by plaque assay on BHK-21 cells (22). All of the experiments were repeated two or three times. All of the data were highly reproducible, and figures present the results of one of the representative experiments.

Virus purification. BHK-21 cells (8×10^6) in 150-mm dishes were infected with recombinant viruses at an MOI of 20 PFU/cell for 1 h, washed with PBS and overlaid with 15 ml of complete medium. At 6 h postinfection, cells were washed with PBS to remove the serum and overlaid with 15 ml of serum YP-SFM medium. At 24 h postinfection, medium was harvested, HEPES buffer pH 7.4 was added to 0.05 M, and, after centrifugation at $12,000 \times g$ for 20 min, viral particles were concentrated to final volume 1 ml using Amicon Ultra 100K centrifugal filters (Millipore). Virus particles were pelleted from concentrated samples at 50,000 rpm and 4°C for 1 h in a TLA-55 rotor on a TL-100 ultracentrifuge (Beckman). These pellets were used for both RNA isolation and protein analysis. For the latter analysis, viral particles were dissolved in the gel-loading buffer and proteins were analyzed on SDS-10% PAGE.

RNA analysis. RNA was isolated from pelleted viral particles using RNeasy minikit (Qiagen). Purified RNA was separated on a nondenaturing 1% agarose gel in TAE buffer and stained with ethidium bromide. Cellular RNA, mostly represented by 28S and 18S ribosomal RNAs, and *in vitro*-synthesized viral RNAs were used as molecular weight standards.

RT-qPCR. BHK-21 cells were infected with recombinant viruses at an MOI of 10 PFU/cell. Total RNA was isolated from infected cells at 4 and 7 h postinfection using TRIzol (Invitrogen) and was additionally purified by RNeasy minikit (Qiagen). cDNA was synthesized on 1 μ g of total RNA using QuantiTect reverse transcription (RT) kit (Qiagen). The nsP2 gene-specific primers were used to quantify viral genomic RNA. The E2 gene-specific primers were used to quantify total viral RNA. Quantitative PCR (qPCR) was performed using SsoFast EvaGreen Supermix (Bio-Rad) in the CFX96 real-time PCR detection system (Bio-Rad) for 40 2-step cycles (5-s-long denaturing step at 98°C, and 5-s-long annealing and an extension step at 60°C). Data were analyzed with CFX Manager Software. Specificity of qPCR was confirmed by melting curve analysis of the amplified products. Results of quantification were normalized to the amount of β -actin mRNA in the same samples. Each qPCR was performed in triplicate, and means and standard deviations (SD) were calculated.

Computational analysis. Alphavirus sequences were obtained from GenBank, most recently in February 2011. Nucleotide sequences with coverage (>98%) of the nonstructural polyprotein coding sequences were identified by applying NCBI tblastn (1) to the polyprotein amino acid sequences from GenBank RefSeqs NC_001449 (VEEV), NC_003899 (EEEV), NC_003908 (WEEV), NC_001547 (SINV), NC_001544 (RRV), NC_004162 (CHIKV), NC_003417 (MAYV), and NC_003215 (SFV). Sequences were grouped into clades according to the tblastn analysis. Patent sequences and constructs were removed, and sequences annotated as defective (search terms “nonfunctional,” “truncated,” and “defective” in GenBank records) were inspected and removed from analyses where appropriate. Within each clade, sequences were translated, aligned, and back-translated to nucleotide sequence alignments using EMBOSS (35) and Clustal (21). The synonymous site conservation statistic was calculated as described elsewhere (6). In order to map the conservation statistic onto the coordinates of a specific sequence in each alignment, all alignment columns with gaps in a chosen reference sequence (i.e., the afore-mentioned GenBank RefSeqs) were removed prior to calculation of conservation. RNA structures were predicted using a combination of Vienna RNA RNAfold and alidot (16), pknotsRG (33) and manual inspection. Computer predictions of the foldings of RNA fragments used for the experimental analyses were also verified with the mfold Web Server of the RNA Institute, College of Arts and Sciences, State University of New York at Albany (48).

RESULTS

Conserved sequences in alphavirus genomes are not limited to promoter elements. In previous studies, our group and others have designed a variety of chimeric alphavirus genomes, which contain the RNA promoter elements and nsP-coding sequences derived from one alphavirus, and the structural genes derived from heterologous, distantly related alphaviruses (3, 4, 7, 12, 18, 30, 31, 41, 42). All of these chimeric viruses were viable and capable of efficient *in vitro* replication. Representative examples of the chimeras and their replication in mammalian cells are presented in Fig. 1A and B. In all of the experiments, chimeras accumulated in the medium to titers exceeding 10^9 PFU/ml. The resulting infectious titers and replication rates of some of these (see SIN/VEEV as an example) were even higher than those of the parental viruses used as sources of the nsP-coding sequence.

VEEV, CHIKV, EEEV, and SINV belong to four genetically divergent groups of alphaviruses, demonstrating amino acid identity in the nonstructural and structural proteins of ~65% and ~45%, respectively. The identity at the nucleotide level is almost undetectable, except in the sequences representing RNA promoter and replication enhancer elements. Nevertheless, all of the chimeras designed to date demonstrated efficient release of infectious virions (3, 4, 7, 12, 18, 30, 31, 41, 42). Taken together, the accumulated data suggest that packaging of recombinant viral RNA could be determined primarily by the close cocompartmentalization in the cells of newly synthesized capsid proteins and transcribed viral genomes. Another possibility is that the analysis of viral genomes was not thorough enough, and all or most of the alphavirus genomes contain a conserved yet uncharacterized RNA packaging signal(s) recognized by both homologous and heterologous capsid proteins during particle formation.

To distinguish between these two possibilities, we have performed an analysis of RNA fragments demonstrating reduced variability at synonymous sites in the P1234-encoding sequence of the viruses that belong to the VEEV, WEEV, EEEV, and SINV clades. The results are presented in Fig. 2. The first, third, and fourth fragments with reduced variability correspond to (i) the 5' promoter and 51-nt CSE, a replication

FIG. 1. Chimeric alphaviruses demonstrate efficient *in vitro* replication. (A) Schematic representation of the chimeric genomes. (B) Replication of the chimeric viruses in BHK-21 cells. Cells were infected with the indicated viruses at an MOI of 20 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells.

enhancer, (ii) an extended stem-loop readthrough-stimulating element, and (iii) the 5' end of the subgenomic RNA promoter sequence (Fig. 2A and B). The second identified fragment was located in the sequence corresponding to codons 260 to 370 of nsP1 (nt 825 to 1155 of the VEEV genome). This could either be a previously undetected element of the RNA promoter(s) or possibly a conserved PS present in the genomes of the indicated alphaviruses. More detailed analysis of this RNA fragment in the viruses of the VEEV clade and its putative folding are presented in Fig. 2C and D, respectively.

The VEEV nsP1 gene contains an RNA element that is required for RNA packaging. In order to determine the function of the identified conserved RNA element in the VEEV

FIG. 2. Synonymous site conservation in the alphavirus nonstructural polyprotein-coding sequence. (A) Map of the alphavirus genome fragment encoding the nonstructural polyprotein. (B) Conservation at synonymous sites in a 25-codon sliding window for an alignment of 76 VEEV, EEEV, WEEV, and SINV sequences. The red line depicts the probability that the degree of conservation within a given window could be obtained under a null model of neutral evolution at synonymous sites, while the brown line shows the absolute amount of conservation as represented by the ratio of the observed number of substitutions within a given window to the number expected under the null model. (C) Conservation at synonymous sites in a 9-codon sliding window for an alignment of 36 VEEV sequences. The inset shows an expanded view of the conserved regions within the VEEV nsP1-coding sequence: the positions of phylogenetically conserved predicted stem-loops are indicated by horizontal lines, numbered according to SL2, SL3, SL4, SL5, and SL8 in panel D. Filled circles indicate the positions of the GGG motifs in each stem-loop. (D) The computer-predicted (m-Fold) folding of the RNA fragment corresponding to nt 856 to 1150 of the VEEV TC-83 genome.

genome, we needed to analyze the effects on virus replication of (i) the nucleotide sequence, (ii) RNA folding, and (iii) its position in the genomic RNA. Before investigating these characteristics, we conducted initial experiments to test whether the proposed VEEV-specific nucleotide sequence indeed played a detectable role in infectious virus release. To answer this question, we synthesized the entire fragment corresponding to nt 798 to 1133 of the VEEV genomic RNA, in which we introduced 135 mutations, and then cloned it into a cDNA of the VEEV genome to replace the natural sequence (Fig. 3A). The mutations were introduced in a way that conserved the encoded amino acid sequence, and no rare codons were used. However, the introduced mutations completely destroyed the predicted RNA secondary structure, shown in Fig. 2D. For this variant and most of the other constructs used in this study, we cloned an additional GFP-coding sequence under the control of a second subgenomic promoter into the recombinant viral genome (Fig. 3A). GFP expression was used as a convenient

means of indirectly monitoring the levels of viral genome replication and accumulation of subgenomic RNA. The designed VEEV/PS⁻/GFP construct was viable, and in an infectious center assay, its *in vitro*-synthesized RNA demonstrated essentially the same infectivity as that of the VEEV/GFP RNA containing the wt sequence of nsP1 (data not shown). This was an indication that no additional adaptive mutations were required for virus replication. However, the introduced mutations had a strong negative effect on the rates of infectious virus replication. VEEV/PS⁻/GFP formed smaller plaques and demonstrated an obvious delay in virus release and lower growth rates in both BHK-21 and mosquito C7/10 cells (Fig. 3B and C). At all time points postinfection, titers of the mutant were ~100-fold lower than those of control VEEV/GFP. The infected cells expressed the same levels of GFP (data not shown) indicating similar RNA replication rates. To confirm this result, in additional experiments, we measured the levels of intracellular genomic and subgenomic RNA accumulation

FIG. 3. Mutations in the putative PS of the VEEV genome have a strong negative effect on genomic RNA packaging. (A) Schematic representation of recombinant VEEV genomes. The filled box in the nsP1-coding sequence of VEEV/PS⁻/GFP indicates the presence of 135 synonymous mutations in the RNA fragment presented in Fig. 2C and D. (B and C) Replication of VEEV/GFP and VEEV/PS⁻/GFP viruses in BHK-21 and mosquito C7/10 cells, respectively. Cells were infected at the indicated MOIs, and at the indicated time points, medium was replaced, and virus titers were determined by a plaque assay on BHK-21 cells. (D) BHK-21 cells were infected with VEEV/PS⁻/GFP and VEEV/GFP viruses at an MOI of 10 PFU/cell. Total RNA was isolated as described in Materials and Methods. Concentrations of viral genomic and genomic plus subgenomic RNAs were determined by qPCR and normalized to the concentrations of β -actin mRNA. Data presented in this figure were normalized to the concentration of VEEV/GFP-specific RNAs determined at 4 h postinfection. (E) BHK-21 cells were infected at an MOI of 10 PFU/cell with VEEV/GFP or VEEV/PS⁻/GFP variants, and then incubated in serum-free VP-SF medium. Medium was harvested at 18 h postinfection, when all of the cells remained attached to plastic and before they developed profound CPEs. Then samples were concentrated under identical conditions on Millipore centrifugal filter units and viruses were pelleted by ultracentrifugation (see Materials and Methods for details). Samples corresponding to equal volumes (4 ml) of medium were analyzed by SDS-PAGE and stained by Coomassie brilliant blue. (F) RNAs isolated from the viral particles were analyzed by agarose gel electrophoresis under nondenaturing conditions. Samples corresponded to equal amounts of the original medium (1.8 ml). VEEV/GFP stock harvested from the infected cells and the concentrated sample had 40-fold-higher infectious titers (PFU/ml) than those of VEEV/PS⁻/GFP.

by qPCR. At various time points postinfection, no significant difference in RNA concentrations was found in cells infected with VEEV/GFP or VEEV/PS⁻/GFP at the same MOI (Fig. 3D). Thus, the mutations in the studied fragment affected packaging and release of infectious virions rather than RNA synthesis. The most probable explanation was that the mutations affected the efficiency of PS function. Of note, in our previous studies, this fragment did not function in the VEEV genome as part of any RNA promoters or enhancers (19, 20, 25), but was present in the helper RNAs, which were capable of self-packaging into virions (40).

VEEV/PS⁻/GFP replicated in BHK-21 cells to final titers that were consistently 30- to 100-fold lower than those of VEEV/GFP. In spite of these differences in infectious titers, cells infected by either virus released into the medium equal amounts of viral proteins, most likely in the form of viral particles, which could be purified by ultracentrifugation (Fig. 3E). This difference in PFU/particle ratio was determined by a strong increase in the packaging of both viral subgenomic and, possibly, other RNAs instead of genomic RNA of VEEV/PS⁻/GFP (Fig. 3F). In the samples of concentrated VEEV/GFP virions, the subgenomic RNAs were barely detectable; however, virions isolated from the medium of VEEV/PS⁻/GFP-infected cells contained high levels of subgenomic RNA instead of their viral genomes (Fig. 3F). Our qPCR data were in agreement with those described above, and the particle-to-genome-equivalent ratio was more than 20-fold higher for purified VEEV/PS⁻/GFP than for VEEV/GFP samples (data not shown).

VEEV PS function is position-independent. Our following experiments addressed the question of whether the function of the VEEV PS RNA fragment is dependent on its position in the viral genome. We designed VEEV genomes in which the natural PS sequence was inactivated by the above-described, numerous mutations, and additional test sequences (corresponding to nt 774 to 1159 of the VEEV genome) were cloned between the nsP4 gene and the subgenomic promoter driving GFP expression (Fig. 4A). The subgenomic promoter in the nsP4-coding sequence upstream of the newly inserted test PS sequences was inactivated by 10 synonymous mutations. Thus, the designed viruses VEEV/mut/wt/GFP and VEEV/mut/mut/GFP contained a mutated PS in the natural PS position and wt or mutated PS sequences, respectively, downstream of the nsP4 gene. The control construct, VEEV/wt/mut/GFP, had the same design. It contained the natural, wt PS in its original location and the mutated PS downstream of the nsP4 gene. All of the constructs were viable and *in vitro*-synthesized RNAs were equally infectious in the infectious center assay (data not shown). Replication rates of VEEV/mut/wt/GFP were close to those of VEEV/wt/mut/GFP, in which the wt PS was in its natural position. In the repeated experiments we detected 2- to 3-fold differences in the titers of VEEV/mut/wt/GFP and VEEV/wt/mut/GFP, suggesting that the sequences surrounding the PS might have some effect on its function. However, at all time points postinfection, VEEV/mut/wt/GFP titers were almost 2 orders of magnitude higher than titers of VEEV/mut/mut/GFP, in which both PS sequences were mutated (Fig. 4B). These results suggest that the VEEV-specific PS is functional regardless of its position in the viral genome.

Positioning of the PS between nsP4 and the subgenomic

FIG. 4. VEEV PS function is mostly independent of its position in the viral genome. (A) Schematic representation of the viral genomes. Filled boxes indicate positions of the mutated VEEV PS. The open box in the VEEV/mut/wt/GFP genome indicates the position of the wt PS cloned downstream of the nsP4 gene. (B) Replication of the recombinant viruses in BHK-21 cells. Cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells.

promoter makes its function in RNA packaging independent of the protein sequence encoded by the PS in its natural location and, thus, opens an opportunity for a wide range of modifications, which are no longer limited to synonymous mutations but can also include extended deletions and insertions.

VEEV PS function is determined by GGG sequences in the RNA loops. The computer prediction of RNA folding suggested the presence of numerous stem-loop (SL) structures in the fragment comprising nt 825 to 1155 of the VEEV genome. The distinguishing feature of the predicted RNA secondary structure presented in Fig. 5A, is the presence of a GGG sequence in the same position of the loops in SLs 2, 4, 5, and 8 (and in a different position in the loop of SL3). The same positioning of GGG in numerous loops could certainly be a coincidence; however, the computer analysis suggested the existence of GGG in the same loop position of 4 to 6 SLs in corresponding nsP1-specific fragments of the EEEV, WEEV, and SINV genomes (see below). Their stem-loops were predicted to have different sequences and foldings but contained the apparently conserved GGG motif at the loop bases.

Based on these data, we hypothesized that the short GGG sequences, specifically in the distinct position of RNA loops, might be critical elements of the VEEV PS and, possibly, the PSs of other alphaviruses. We then tested whether the GGG motifs are important for virus replication/packaging. New mutations were introduced into the wt VEEV PS of the VEEV/mut/wt/GFP genome so as to destroy the GGG triplets (Fig. 5A and B). These mutations did not affect the predicted RNA

FIG. 5. GGG motifs play a critical role in VEEV PS function. (A) The computer-predicted (m-Fold) folding of the RNA fragment corresponding to nt 856 to 1150 of the VEEV TC-83 genome and the predicted (m-fold) folding of the same fragment with mutated GGG motifs. The introduced mutations are indicated by red color font. (B) Schematic representation of the viral genomes. Filled black boxes indicate positions of the mutated VEEV PS. The open box in the VEEV/mut/wt/GFP genome indicates the position of the wt PS cloned downstream of the nsP4 gene. The red box indicates the position of the

secondary structure of the PS (Fig. 5A), but nonetheless had a strong negative effect on virus replication rates (Fig. 5C). The VEEV/mut/3xGm/GFP mutant replicated as inefficiently as the VEEV/mut/mut/GFP variant, in which the test fragment contained 135 mutations, leading to a very different sequence and overall predicted folding. Thus, the GGG sequences appear to represent key elements of the VEEV PS.

The GGG-containing stem-loops have additive effects on PS function. In order to dissect the functions of predicted stem-loops in viral RNA packaging, we sequentially deleted them in the VEEV/mut/wt/GFP genome (Fig. 6) and assessed the effects of these deletions on virus replication. Analysis of the virus growth curves presented in Fig. 6F, suggest that it is unlikely that any one particular stem-loop plays the sole or dominant role in virus packaging. Their presence in the fragment appeared to have an additive effect on the rates of infectious virus release. Based on the selection of deletions tested, presence of stem-loops 2, 3, and 4 or 3, 4, and 5 in VEEV/mut/wt/ Δ 2/GFP and VEEV/mut/wt/ Δ 1+3/GFP, respectively (Fig. 6B and D), was sufficient to mediate packaging at the same level achieved by the entire fragment. Only the deletion of all of the stem-loops except 3 and 4 in VEEV/mut/wt/ Δ 2+3/GFP (Fig. 6E) caused a strong decrease in infectious virus release; however, in repeated experiments, it remained at a level higher than that of VEEV/mut/mut/GFP, suggesting some residual activity for this remaining short fragment in RNA packaging and, ultimately, in virus replication. In SL3, the GGG motif is in a different position than in the other loops, and it is unlikely that it functions in the same mode as the others. However, there is enhanced conservation in the SL3 region (Fig. 2C), which suggests that it might play some role in virus replication, such as preserving the overall secondary structure of the RNA fragment. We did not perform the experiments necessary to completely rule out this possibility, but a lack of this stem-loop in the artificially designed PS (see Fig. 10) supports the hypothesis that SL3 is not a critical functional element.

The VEEV capsid protein is capable of recognizing the PSs of heterologous alphaviruses. The above-described data demonstrate that chimeric viruses encoding VEEV-specific structural proteins are capable of efficient replication (Fig. 1 and aforementioned references). The high replication rates of VEEV capsid-encoding chimeric viruses strongly indicated that during the encapsidation process, the VEEV capsid protein might be capable of efficient interaction with heterologous PSs in the nsP-coding sequence. To test this hypothesis, we redesigned the VEEV/mut/wt/GFP variant by replacing the wt VEEV PS with the corresponding sequences derived from other alphaviruses. VEEV/mut/SIN/GFP, VEEV/mut/EEE/GFP, and VEEV/mut/SFV/GFP contained fragments of the SINV, EEEV, and SFV genomes corresponding to the VEEV nsP1 PS-containing fragment (Fig. 7A). The fragments tested

PS with mutated GGG motifs. (C) Replication of the recombinant viruses in BHK-21 cells. Cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells.

FIG. 6. Deletions of the predicted individual stem-loops have no deleterious effect on VEEV PS function. (A) Computer-predicted (mFold) folding of the RNA fragment corresponding to nt 856 to 1150 of the VEEV TC-83 genome. Red lines indicate deletions of the stem-loops used for construction of the PS mutants presented in other panels. (B, C, D, and E) Computer-predicted foldings of the RNA fragments with deleted sequences corresponding to different stem-loops. (F) Schematic representation of the VEEV genome used for testing the PS deletion mutants. The filled black box indicates the presence of the mutated VEEV PS. The white box downstream of the nsP4-coding sequence represents the position of the tested sequences. Replication of the recombinant viruses in BHK-21 cells. Cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells.

corresponded to nt 731 to 1213, 726 to 1185, and 797 to 1212 of the SINV, EEEV, and SFV genomes, respectively. All of these variants were viable and expressed the same levels of GFP, indicating similar levels of RNA synthesis, but differed in quantity of infectious virus produced. SINV- and EEEV-derived fragments promoted virus release as efficiently as did the natural VEEV-specific fragment (Fig. 7B). However, the presence of the SFV-specific fragment did not cause an increase in the rates of infectious virus production. It remained at the level previously seen in the experiments with VEEV/mut/mut/GFP (Fig. 6F, 5C, and others). These results correlated with differences in RNA sequences and predicted secondary structures

(Fig. 8A and data not shown). The VEEV-, EEEV- and SINV-derived RNA fragments contained 4 to 6 stem-loops, in which GGG sequences were located at the exact same positions as in the VEEV PS stem-loops. The SFV-specific fragment, in contrast, was predicted to contain no stem-loop structures with GGG sequences. More extensive computer analysis of other members of the SFV clade also predicted a lack of GGG motifs in RNA loops. Taken together, the results of this analysis correlate with the above-presented data and indicate that the GGG motifs function in RNA packaging of numerous alphaviruses, except those in the SFV clade.

FIG. 7. EEEV and SINV, but not SFV, nsP1-derived sequences are functionally utilized by the VEEV capsid protein in RNA packaging. (A) Schematic representation of the recombinant genomes. Filled black boxes indicate the mutated PS in the VEEV nsP1-coding sequence. White, red, green, and blue boxes indicate the presence of VEEV-, SFV-, EEEV-, and SINV-specific nsP1 sequences, respectively, corresponding to the VEEV-specific PS. (B) Replication of the recombinant viruses in BHK-21 cells. Cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells.

The SINV PS is located in the same nsP1-specific fragment of the viral genome. The data described above suggest that in spite of differences in the overall folding, VEEV, EEEV, and SINV appear to have a PS sequence, recognized by VEEV capsid, in the same position within their genomes. However, this does not necessarily mean that the same PSs can be recognized by the homologous capsid proteins. For safety reasons, EEEV could not be used in the experiments, but we could perform extensive mutagenesis of the SINV genome and test the effects of these mutations on virus replication. The computer-predicted, putative SINV-specific PS, containing six loops with apical GGG sequences, is presented in Fig. 8A. We have synthesized a new 478-nt-long fragment of the SINV genome between nt 733 and 1212 and cloned it into SINV/GFP to replace the original sequence (Fig. 8B). More than 100 mutations were introduced into this sequence with virtually all of the codons replaced by alternative ones. These mutations were designed in such a way as to preserve the existing codon frequency and amino acid sequence in the SINV nsP1. As we described above for VEEV with mutated PS, the introduced mutations had no effect on the infectivity of the *in vitro*-synthesized RNA or on the GFP expression in virus-infected cells (data not shown), but caused a strong negative effect on virus replication (Fig. 8C). At all time points postinfection, titers of SINV/PS⁻/GFP were 2 to 3 orders of magnitude lower than those of SINV/GFP containing the wt nsP1-coding sequence.

FIG. 8. Mutations in the SINV nsP1-specific PS have a strong negative effect on SINV replication. (A) Computer-predicted (m-Fold) folding of the RNA fragment corresponding to nt 753 to 1166 of the SINV genome. (B) Schematic representation of SINV genomes containing the wt nsP1 sequence and the mutated sequence at nt 734 to 1211, indicated by the white box. (C) Replication of the recombinant viruses in BHK-21 cells. Cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells.

These data suggest that in the SINV genome, the PS recognized by SINV capsid is located at the same position as in the VEEV genomic RNA. The results also correlated with those from previous studies, in which the SINV-specific PS was analyzed in the context of DI RNAs (5, 27). Deletions of the SINV PS in DI RNAs had a detectable negative effect on their ability to be packaged into virions.

Members of the SFV clade contain a packaging signal in the nsP2 gene, but not in the nsP1-coding sequence. The inability of SFV-specific nsP1 sequence to support RNA packaging mediated by VEEV capsid (Fig. 7), created some ambiguity in the interpretation of results, because structural proteins derived from another virus that belongs to the SFV clade, CHIKV, were very efficient in the packaging of a VEE/CHIKV

genome. The latter chimeric virus replicated in tissue culture to titers approaching almost 10^{10} PFU/ml (Fig. 1A). Thus, CHIKV-specific capsid protein appeared to be efficient in recognizing the VEEV PS. However, neither SFV- nor CHIKV-specific fragments corresponding to the VEEV PS, demonstrated the presence of GGG motifs in any of the computer-predicted stem-loops (data not shown). Thus, taken together, the data suggested that capsid proteins of viruses belonging to the SFV clade, appeared to be capable of interacting with the PS(s) of other alphaviruses, but likely recognized their own, virus-specific PS in another fragment of the genome. This hypothesis was supported by computer analysis of the SFV-clade genomes (Fig. 9A), which predicted the presence of a nucleotide sequence with a lower rate of substitutions at synonymous sites in the nsP2 gene, but not in the nsP1-coding RNA fragment. In the following experiments, we tested two hypotheses: (i) CHIKV capsid can recognize the VEEV-specific PS, and (ii) the CHIKV-specific PS is located in the nsP2-but not the nsP1-coding sequence.

First, we designed a recombinant, chimeric virus VEE/PS⁻/CHIKV, which expressed CHIKV structural genes and had the natural VEEV PS inactivated by the above-described 135 mutations (Fig. 9B). This virus was viable, but its replication rates were lower than those of VEE/CHIKV, having the wt VEEV-specific PS (Fig. 9B). This result demonstrated that the CHIKV capsid protein is capable of utilizing the VEEV-specific PS.

The next constructs were based on using a chimeric VEE/CHIKV genome with mutated natural VEEV PS, but they also contained additional CHIKV-specific sequences. VEE/mut/nsP1/CHIKV and VEE/mut/nsP2/CHIKV contained nt 826 to 1182 and nt 2501 to 3078 of the CHIKV genome, respectively (Fig. 9C). Both sequences were cloned downstream of the VEEV nsP4 gene and upstream of the subgenomic promoter. The VEE/mut/mut/CHIKV chimera, encoding two mutated VEEV PSs was designed as an appropriate negative control. All three constructs were viable, but differed in virus replication rates. The sequence comprising nt 826 to 1182 of CHIKV nsP1, which corresponds to the VEEV PS-containing fragment, did not stimulate packaging of the recombinant genome. It remained at the level of VEEV/mut/mut/CHIKV (Fig. 9C). However, the nsP2-specific CHIKV fragment had a strong stimulatory effect on virus replication rates, indicating its activity in the packaging of RNA into CHIKV structural proteins. The possibility that this fragment contains the CHIKV-specific PS was indirectly supported by our previous data on Ross River virus (RRV), which is also a member of the SFV clade (8). Inclusion of the indicated sequence in the subgenomic RNA significantly increases its packaging into RRV-specific viral particles. It was also interesting to test whether VEEV capsid protein can utilize CHIKV-specific fragments as PSs. To test this, we cloned the above-described fragments of CHIKV nsP1- or nsP2-coding sequences into VEEV/mut/wt/GFP to replace the wt PS (Fig. 9D). Both designed viruses, VEEV/mut/nsP1ch/GFP and VEEV/mut/nsP2ch/GFP, replicated as inefficiently as VEEV/mut/mut/GFP control (Fig. 9D), indicating that VEEV capsid protein cannot use either CHIKV nsP1 or nsP2 fragments as PSs.

A designed artificial VEEV PS efficiently mediates RNA packaging. The experimental data gathered demonstrated that

GGG motifs in the loops of the predicted stem-loop structures might be the key RNA elements that determine specificity of genomic RNA packaging and mediate its selective encapsidation. However, it remained unclear whether the entire stem-loops are involved in this encapsidation process, or whether the packaging is primarily determined by the presence of the GGG motif in the particular position of the folded RNA fragment. In order to shed more light on this question, we attempted to design an artificial PS, containing 4 stem-loops with GGG sequences (Fig. 10A). Similar to the above-presented experiments, we cloned this artificial fragment downstream of the nsP4-coding sequence in the VEEV genome with a mutated natural PS (Fig. 10A). In spite of different nucleotide sequences in the stems, the new stem-loops in the proposed fragment were expected to have similar stability and folding as those in the natural VEEV PS (compare foldings presented in Fig. 10A and 2D), and, most importantly, they were designed to have GGG motifs in the exact same positions of the loops (Fig. 10A). Due to strong modifications of the PS-containing fragment, the new sequence was designed to contain complementary sequences at the 5' and 3' ends, forming a highly stable stem. This modification was expected to reduce the possibility of secondary structure formation between the newly designed PS and the surrounding sequences in the VEEV/mut/artPS/GFP genome.

The *in vitro*-synthesized VEEV/mut/artPS/GFP RNA was as infectious as all other above-described constructs. Rescued virus formed plaques of the same size as VEEV/mut/wt/GFP, which has a natural PS in the same position as the artificial PS in the VEEV/mut/artPS/GFP genome (data not shown). Most importantly, VEEV/mut/artPS/GFP demonstrated replication rates that were almost the same as those of VEEV/mut/wt/GFP and higher than those of VEEV/mut/mut/GFP (Fig. 10B). These data supported the notion that the VEEV PS function is determined by the presence of GGG sequences in the stem-loops of folded RNA, but not by the stem sequences.

DISCUSSION

Replication of alphavirus genomes proceeds in the cytoplasm of infected cells, where these viruses produce high quantities of the structural proteins and viral genomic RNA required for virion formation. Efficient synthesis of virus-specific macromolecules is certainly essential for viremia development, but it is not the only prerequisite for alphavirus replication to high titers. Packaging of their newly synthesized genomes into nucleocapsids occurs in the presence of the more abundant viral subgenomic RNA, and the cellular mRNAs and tRNAs; however, all of these appear to be poorly recognized by the alphavirus capsid protein. Levels of subgenomic RNAs in released virions are barely detectable and, while no studies of cellular RNA packaging have been performed to date, it is highly unlikely that they are packaged into virions. These data imply the existence of specific interactions between the 5' two-thirds of the alphavirus genomic RNA and capsid proteins, which lead to selective encapsidation of viral genomes but not subgenomic or other RNA species.

Previous studies of SINV DI RNA replication and copackaging suggested that a 132-nt-long (SINV PS132) segment of the SINV genome (nt 945 to 1076) is required for efficient

FIG. 10. A designed, artificial PS stimulates infectious VEEV release. (A) Schematic representation of VEEV genomes, and the computer-predicted (m-Fold) folding of the newly designed PS RNA fragment. Filled black boxes indicate positions of the mutated VEEV PS, the white box indicates the position of the wt VEEV PS cloned downstream of the nsP4 gene, and the red box indicates the position of the newly designed artificial PS. (B) BHK-21 cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells.

RNA packaging (45). Two copies of this sequence were present in the SINV DI RNA genome (26, 28, 39), and their deletion had a negative effect on RNA encapsidation, but not on RNA replication. Further studies strongly suggested that

the interaction of this PS with SINV capsid is mediated by a short amino acid sequence located between amino acids (aa) 96 and 107 of the SINV capsid protein (29, 43). These data provided a plausible explanation for specific packaging of SINV genomic RNA. However, the direct effect of mutations in the PS on virus replication had not been investigated. Moreover, in the genome fragments corresponding to SINV PS132, distantly related alphaviruses, such as VEEV, demonstrate low levels of identity at the nucleotide level and are predicted to have very different secondary structures (Fig. 2D and 8A). Nevertheless, the previously generated chimeric alphaviruses, which encode heterologous structural and nonstructural genes (3, 4, 7, 12, 18, 30, 31, 41, 42), release high levels of infectious virions, suggesting efficient encapsidation of their genomes. Replication rates of these chimeras were found to be close to or even exceeding those of the parental viruses. A nonspecific, background encapsidation mechanism is very difficult to reconcile with the observation of the exclusive presence of genomic RNA in virions, and therefore, to identify putative, universal PSs in the alphavirus genomes, we applied a more advanced *in silico* analysis of sequence conservation and complemented it by experimental analysis of predicted sequence function in specific RNA packaging. The RNA elements with the greatest level of sequence conservation in viruses of VEEV, EEEV, WEEV, and SINV clades were found in the nsP1-coding sequence of the viral genomes (Fig. 2A). The experiments that followed demonstrated that the *in silico* analysis generated reliable information, and serial synonymous mutations in this fragment in the VEEV genome (nt 798 to 1133) had strong negative effects on virus growth rates and final titers. The introduced mutations did not affect the efficiency of viral RNA replication or viral particle release (Fig. 3). Assembled virions accumulated in the medium to essentially the same concentrations regardless of whether the wt or mutated sequence was present in the viral genome. However, the majority of virions released from VEEV/PS⁻/GFP-infected cells, contained not the genomic, but the subgenomic RNA. Similar mutagenesis of the corresponding sequence in the SINV genome also strongly affected rates of infectious virus release (Fig. 8). Thus, taken together the data suggested that the studied RNA element plays a critical role in the replication of evolutionarily diverse alphaviruses.

FIG. 9. Viruses belonging to the SFV clade utilize different sequences as PSs. (A) Conservation at synonymous sites in a 25-codon sliding window for the SFV clade. The red line depicts the probability that the degree of conservation within a given window could be obtained under a null model of neutral evolution at synonymous sites, while the brown line shows the absolute amount of conservation as represented by the ratio of the observed number of substitutions within a given window to the number expected under the null model. Arrow 1 indicates the RNA fragment utilized as a PS in the genomes of alphaviruses in clades other than the SFV clade. Arrow 2 indicates a sequence with higher conservation at synonymous sites in the nsP2-coding fragment of the SFV-clade alphaviruses. (B) Schematic representation of chimeric viral genomes and replication of the rescued viruses in BHK-21 cells. The filled box in the nsP1-coding sequence indicates the mutated VEEV-specific PS. Cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells. (C) Schematic representation of chimeric viral genomes and replication of the rescued viruses in BHK-21 cells. The filled box in the nsP1-coding sequence indicates the mutated VEEV-specific PS. Gray boxes 1 and 2 located downstream of the nsP4-coding sequence represent the CHIKV nsP1- and nsP2-specific sequences indicated in panel A (see also text for details). Cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells. (D) Schematic representation of viral genomes and replication of the rescued viruses in BHK-21 cells. The filled box in the nsP1-coding sequence indicates the mutated VEEV-specific PS. Gray boxes 1 and 2 located downstream of the nsP4-coding sequence represent the CHIKV nsP1- and nsP2-specific sequences indicated in panel A. The same CHIKV-specific sequences were used in chimeric viruses presented in panel C. Cells were infected with the indicated viruses at an MOI of 0.01 PFU/cell. At the indicated time points, medium was replaced, and virus titers were determined by plaque assay on BHK-21 cells.

Other experiments were based on the application of an experimental system, which we have previously tested while studying the functional elements of alphavirus promoters (25). Within the VEEV genome, we separated the functional PS and its nsP1-coding sequence (Fig. 4). We were then able to perform a wide range of genetic manipulations without affecting the nsP1 protein sequence. First, these studies demonstrated that addition of the wt nt 774 to 1159 to a VEEV genome with a mutated PS made the virus replication more efficient in terms of infectious virion release. Second, the corresponding, nsP1-specific sequences, derived from the EEEV and SINV genomes, but not the SFV genome, were utilized by the VEEV capsid protein and, thus, were able to function as PSs in packaging of the VEEV genomes with mutated natural PSs. These results were in agreement with the computer predictions of the fragments' folding and the indirect data from previous studies, which suggested that a conserved peptide in alphavirus capsid proteins plays a critical role in the specificity of RNA packaging (29). Based on conservation of this peptide, it was reasonable to expect that it might interact with the same or at least similar sequences or motifs in viral RNAs.

The only conserved feature that we could distinguish in the nsP1-specific fragments of the VEEV, EEEV, WEEV, and SINV genomes was the presence of GGG motifs in the same position within numerous computer-predicted stem-loops. Two different lines of evidence demonstrated the importance of GGG for RNA packaging. First, mutagenesis of these GGG sequences alone was found to have as strong negative effect on VEEV replication (Fig. 5) as did complete alteration of both the nucleotide sequence and the RNA secondary structure (Fig. 3). Second, an artificial PS with fewer stem-loops, containing different nucleotide sequences, but still possessing GGG motifs in the same relative locations within the loops, was active in viral genome packaging (Fig. 10). These were compelling pieces of evidence that GGG sequences play a critical role in PS function. The activity of the GGG-containing stem-loops is determined to some degree by their number in the packaging signal, and more than one stem-loop is required for efficient genomic RNA encapsidation. Deletion of individual GGG-containing stem-loops did not cause noticeable changes in the rates of infectious virus release, but deletion of multiple predicted stem-loops resulted in a detectable decrease in the rates of infectious virus production. Thus, while there is certainly a minimum number of stem-loops required for genome packaging, in the case of VEEV (and probably other alphaviruses), the term "minimum PS" is likely to be inapplicable, because different combinations of the stem-loops can efficiently function in RNA encapsidation. The independent and repetitive nature of the GGG-containing SLs suggests that each SL-specific GGG motif appears to interact with a different molecules capsid protein. Clustering of the SLs in one small genomic fragment may be a mechanism to bring several capsomers into close proximity to initiate their dimerization and nucleate virion formation. However, this hypothesis requires further experimental support.

Interestingly, the CHIKV-specific capsid protein was capable of VEE/CHIKV chimeric genome packaging, and mutations in the VEEV PS affected the release of this chimeric virus from infected cells (Fig. 9B). However, viruses belonging to the SFV clade, e.g., SFV, CHIKV, and RRV, appear to be excep-

tions from the above-described, general packaging mechanism. (i) The computer analysis of the genome sequences in the SFV clade viruses did not suggest the existence of an nsP1-specific fragment with lower rates of substitutions at synonymous sites (Fig. 9A). (ii) The predicted folding of the RNA fragment between nt 750 and 1250 of the SFV and CHIKV genomes did not reveal any GGG-containing loops. (iii) SFV- and CHIKV-specific nsP1 fragments had no positive effect on infectious virus release when they were cloned into the VEEV genome having a mutated natural PS (Fig. 7 and 9C). Thus, the data suggested that CHIKV and SFV capsid proteins may recognize sequences that differ from the VEEV PS. Further investigation revealed that the genomes of the SFV clade viruses did contain a region with a lower rate of substitutions at synonymous sites, but it was contained within the nsP2 gene rather than the nsP1 gene (Fig. 9A). Its position correlated with that of a fragment we previously identified in the RRV genome as capable of increasing RNA packaging into RRV virions (8). Similarly to VEEV, the nucleotide sequence analysis was supported by the experimental evidence. It was found that nt 2501 to 3078 of the CHIKV genome stimulated packaging of VEE/CHIKV chimeric genomic RNA, which had the natural VEEV PS inactivated (Fig. 9C). Based on these results, we hypothesize that a nsP1-specific PS was present in an ancestral alphavirus. While the GGG motif-containing SLs remained in the nsP1 region of most alphaviruses during their evolution, they disappeared in the genomes of SFV-like viruses and were replaced by a PS located in the nsP2 gene. The latter PS appears to have a similar mode of interaction with capsid protein, because CHIKV capsid is still capable of using both the nsP1 fragment of VEEV and the nsP2 fragment of CHIKV genomes. However, it should be noted that viruses from the SFV clade do not contain GGG sequences in the predicted loops of the nsP2-specific PS. Their equivalent may be a GUG(G) motif (where the 5' G and U belong to the apical two base-pairings of an SL), but its functional role has yet to be experimentally demonstrated. Interestingly, VEEV-specific capsid protein appears to be incapable of interacting with both nsP1- and nsP2-specific fragments of the CHIKV genome, which are indicated in Fig. 9A. Cloning of these fragments into the VEEV genome with mutated natural PS did not have noticeable positive effect on the infectious titers of the released virus (Fig. 9D).

Taken together, the results of this study demonstrate the following. (i) The genomes of SINV, VEEV, EEEV, and most likely other alphaviruses, except for those in the SFV clade, have PSs that can be recognized by capsid proteins of heterologous alphaviruses. (ii) The GGG sequences located in the loops of PS-specific stem-loops represent a key element of the functional PS. (iii) Mutagenesis of the entire PS or just the GGG sequences has a strong negative effect on viral genome packaging and leads to release of viral particles containing mostly SG RNAs. (iv) RNA packaging appears to be determined to some extent by the number of GGG-containing stem-loops, and more than one stem-loop is required for efficient RNA encapsidation. (v) Viruses of the SFV clade contain a PS in the nsP2 gene, but their capsid protein retains the ability to use the nsP1-specific PS of other alphaviruses. It should also be noted that in this study, we have likely identified the most important PS. However, this does not necessarily mean that it is the only sequence utilized during genomic RNA packaging.

In a previous study of RRV sequences that increase RNA packaging (8), we found that other viral genome fragments can also have detectable stimulatory effects on RNA encapsidation. It was not as efficient as that located between nt 2700 and 3100 of nsP2 in the RRV genome, but might also contribute to the RNA packaging process. Similarly, less efficiently functioning sequences can likely be found in other alphaviruses.

This new information regarding alphavirus PS structure and its mode of function provides an interesting opportunity to develop alphavirus variants, which are strongly attenuated in terms of infectious virus production in the cells of both vertebrate and invertebrate origin. These would produce virions that primarily lack the viral genome but remain capable of serving as efficient immunogens. Such a viral phenotype would likely be stable, because it is supported by a large number of synonymous mutations that destroy both the RNA secondary structure and GGG motifs, but would not interfere with viral RNA replication and protein expression.

ACKNOWLEDGMENTS

We thank Maryna Akhrymuk for excellent technical assistance. We also thank Scott Weaver for providing a plasmid containing CHIKV nsP1 and nsP2.

This work was supported by grants from the National Institute of Allergy and Infectious Diseases (R01AI070207 to D.Y.K., S.A., and I.F. and R01AI073301 to E.I.F.). A.E.F. was supported by the Wellcome Trust (grant number 088789).

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Atasheva, S., A. Fish, M. Fornerod, and E. I. Frolova. 2010. Venezuelan equine Encephalitis virus capsid protein forms a tetrameric complex with CRM1 and importin alpha/beta that obstructs nuclear pore complex function. *J. Virol.* **84**:4158–4171.
- Atasheva, S., N. Garmashova, I. Frolov, and E. Frolova. 2008. Venezuelan equine encephalitis virus capsid protein inhibits nuclear import in Mammalian but not in mosquito cells. *J. Virol.* **82**:4028–4041.
- Atasheva, S., et al. 2009. Chimeric alphavirus vaccine candidates protect mice from intranasal challenge with western equine encephalitis virus. *Vaccine* **27**:4309–4319.
- Bredenbeek, P. J., I. Frolov, C. M. Rice, and S. Schlesinger. 1993. Sindbis virus expression vectors: packaging of RNA replicons by using defective helper RNAs. *J. Virol.* **67**:6439–6446.
- Firth, A. E., and J. F. Atkins. 2009. A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. *Virol. J.* **6**:14.
- Frolov, I., E. Frolova, and S. Schlesinger. 1997. Sindbis virus replicons and Sindbis virus: assembly of chimeras and of particles deficient in virus RNA. *J. Virol.* **71**:2819–2829.
- Frolova, E., I. Frolov, and S. Schlesinger. 1997. Packaging signals in alphaviruses. *J. Virol.* **71**:248–258.
- Frolova, E. I., et al. 2002. Roles of nonstructural protein nsP2 and alpha/beta interferons in determining the outcome of Sindbis virus infection. *J. Virol.* **76**:11254–11264.
- Frolova, E. I., R. Gorchakov, L. Pereboeva, S. Atasheva, and I. Frolov. 2010. Functional Sindbis virus replicative complexes are formed at the plasma membrane. *J. Virol.* **84**:11679–11695.
- Froshauer, S., J. Kartenbeck, and A. Helenius. 1988. Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes. *J. Cell Biol.* **107**:2075–2086.
- Garmashova, N., et al. 2007. Analysis of Venezuelan equine encephalitis virus capsid protein function in the inhibition of cellular transcription. *J. Virol.* **81**:13552–13565.
- Garmashova, N., R. Gorchakov, E. Frolova, and I. Frolov. 2006. Sindbis virus nonstructural protein nsP2 is cytotoxic and inhibits cellular transcription. *J. Virol.* **80**:5686–5696.
- Garmashova, N., et al. 2007. The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. *J. Virol.* **81**:2472–2484.
- Gorchakov, R., N. Garmashova, E. Frolova, and I. Frolov. 2008. Different types of nsP3-containing protein complexes in sindbis virus-infected cells. *J. Virol.* **82**:10088–10101.
- Hofacker, I. L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**:3429–3431.
- Kim, D. Y., et al. 2011. Design of chimeric alphaviruses with a programmed, attenuated, cell type-restricted phenotype. *J. Virol.* **85**:4363–4376.
- Kuhn, R. J., D. E. Griffin, K. E. Owen, H. G. Niesters, and J. H. Strauss. 1996. Chimeric Sindbis-Ross River viruses to study interactions between alphavirus nonstructural and structural regions. *J. Virol.* **70**:7900–7909.
- Kulasegaran-Shylini, R., S. Atasheva, D. G. Gorenstein, and I. Frolov. 2009. Structural and functional elements of the promoter encoded by the 5' untranslated region of the Venezuelan equine encephalitis virus genome. *J. Virol.* **83**:8327–8339.
- Kulasegaran-Shylini, R., V. Thivyanathan, D. G. Gorenstein, and I. Frolov. 2009. The 5'UTR-specific mutation in VEEV TC-83 genome has a strong effect on RNA replication and subgenomic RNA synthesis, but not on translation of the encoded proteins. *Virology* **387**:211–221.
- Larkin, M. A., et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948.
- Lemm, J. A., R. K. Durbin, V. Stollar, and C. M. Rice. 1990. Mutations which alter the level or structure of nsP4 can affect the efficiency of Sindbis virus replication in a host-dependent manner. *J. Virol.* **64**:3001–3011.
- Levis, R., B. G. Weiss, M. Tsiang, H. Huang, and S. Schlesinger. 1986. Deletion mapping of Sindbis virus DI RNAs derived from cDNAs defines the sequences essential for replication and packaging. *Cell* **44**:137–145.
- Liljeström, P., S. Lusa, D. Huylebroeck, and H. Garoff. 1991. *In vitro* mutagenesis of a full-length cDNA clone of Semliki Forest virus: the small 6,000-molecular-weight membrane protein modulates virus release. *J. Virol.* **65**:4107–4113.
- Michel, G., O. Petrakova, S. Atasheva, and I. Frolov. 2007. Adaptation of Venezuelan equine encephalitis virus lacking 51-nt conserved sequence element to replication in mammalian and mosquito cells. *Virology* **362**:475–487.
- Monroe, S. S., et al. 1982. Sequence analysis of cDNAs derived from Sindbis viruses and of defective interfering particles. *J. Virol.* **41**:153–162.
- Monroe, S. S., and S. Schlesinger. 1984. Common and distinct regions of defective interfering RNAs of Sindbis virus. *J. Virol.* **49**:865–872.
- Monroe, S. S., and S. Schlesinger. 1983. RNAs from two independently isolated defective interfering particles of Sindbis virus contain a cellular tRNA sequence at their 5' ends. *Proc. Natl. Acad. Sci. U. S. A.* **80**:3279–3283.
- Owen, K. E., and R. J. Kuhn. 1996. Identification of a region in the Sindbis virus nucleocapsid protein that is involved in specificity of RNA encapsidation. *J. Virol.* **70**:2757–2763.
- Paessler, S., et al. 2003. Recombinant sindbis/Venezuelan equine encephalitis virus is highly attenuated and immunogenic. *J. Virol.* **77**:9278–9286.
- Paessler, S., et al. 2006. Replication and clearance of Venezuelan equine encephalitis virus from the brains of animals vaccinated with chimeric SIN/VEE viruses. *J. Virol.* **80**:2784–2796.
- Petrakova, O., et al. 2005. Noncytopathic replication of Venezuelan equine encephalitis virus and eastern equine encephalitis virus replicons in Mammalian cells. *J. Virol.* **79**:7597–7608.
- Reeder, J., P. Steffen, and R. Giegerich. 2007. pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic Acids Res.* **35**:W320–W324.
- Rice, C. M., R. Levis, J. H. Strauss, and H. V. Huang. 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of a temperature-sensitive marker, and *in vitro* mutagenesis to generate defined mutants. *J. Virol.* **61**:3809–3819.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**:276–277.
- Rumenapf, T., et al. 1995. Aura alphavirus subgenomic RNA is packaged into virions of two sizes. *J. Virol.* **69**:1741–1746.
- Schlesinger, S., and M. J. Schlesinger (ed.). 1986. The Togaviridae and Flaviviridae. Plenum Press, New York, NY.
- Spuul, P., G. Balistreri, L. Kaariainen, and T. Ahola. 2010. Phosphatidylinositol 3-kinase-, actin-, and microtubule-dependent transport of Semliki Forest virus replication complexes from the plasma membrane to modified lysosomes. *J. Virol.* **84**:7543–7557.
- Strauss, J. H., and E. G. Strauss. 1994. The alphaviruses: gene expression, replication, evolution. *Microbiol. Rev.* **58**:491–562.
- Volkova, E., R. Gorchakov, and I. Frolov. 2006. The efficient packaging of Venezuelan equine encephalitis virus-specific RNAs into viral particles is determined by nsP1-3 synthesis. *Virology* **344**:315–327.
- Wang, E., et al. 2007. Chimeric Sindbis/eastern equine encephalitis vaccine candidates are highly attenuated and immunogenic in mice. *Vaccine* **25**:7573–7581.
- Wang, E., et al. 2008. Chimeric alphavirus vaccine candidates for chikungunya. *Vaccine* **26**:5030–5039.
- Warrier, R., B. R. Linger, B. L. Golden, and R. J. Kuhn. 2008. Role of sindbis virus capsid protein region II in nucleocapsid core assembly and encapsidation of genomic RNA. *J. Virol.* **82**:4461–4470.
- Weaver, S. C., and I. Frolov. 2005. Togaviruses, p. 1010–1024. *In* B. W. J.

- Mahy and V. t. Meulen (ed.), Virology, vol. 2. ASM Press, Salisbury, United Kingdom.
45. **Weiss, B., U. Geigenmuller-Gnirke, and S. Schlesinger.** 1994. Interactions between Sindbis virus RNAs and a 68 amino acid derivative of the viral capsid protein further defines the capsid binding site. *Nucleic Acids Res.* **22**:780–786.
46. **Weiss, B., H. Nitschko, I. Ghattas, R. Wright, and S. Schlesinger.** 1989. Evidence for specificity in the encapsidation of Sindbis virus RNAs. *J. Virol.* **63**:5310–5318.
47. **Weiss, B., R. Rosenthal, and S. Schlesinger.** 1980. Establishment and maintenance of persistent infection by Sindbis virus in BHK cells. *J. Virol.* **33**:463–474.
48. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.