Multiple capsid protein binding sites mediate selective packaging of the alphavirus genomic RNA

Rebecca S. Brown, Dimitrios G. Anastasakis, Markus Hafner & Margaret Kielian

The alphavirus capsid protein (Cp) selectively packages genomic RNA (gRNA) into the viral nucleocapsid to produce infectious virus. Using photoactivatable ribonucleoside crosslinking and an innovative biotinylated Cp retrieval method, here we comprehensively define binding sites for Semliki Forest virus (SFV) Cp on the gRNA. While data in infected cells demonstrate Cp binding to the proposed genome packaging signal (PS), mutagenesis experiments show that PS is not required for production of infectious SFV or Chikungunya virus. Instead, we identify multiple Cp binding sites that are enriched on gRNA-specific regions and promote infectious SFV production and gRNA packaging. Comparisons of binding sites in cytoplasmic vs. viral nucleocapsids demonstrate that budding causes discrete changes in Cp-gRNA interactions. Notably, Cp’s top binding site is maintained throughout virus assembly, and specifically binds and assembles with Cp into core-like particles in vitro. Together our data suggest a model for selective alphavirus genome recognition and assembly.

1 Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA. 2 Laboratory of Muscle Stem Cells and Gene Regulation, National Institute of Arthritis and Musculoskeletal and Skin Disease, Bethesda, MD 20892, USA. ✉email: margaret.kielian@einsteinmed.org

https://doi.org/10.1038/s41467-020-18447-z
Alphaviruses are enveloped, highly organized RNA viruses that include a number of important human pathogens such as Chikungunya virus (CHIKV), Ross River virus (RRV), and Venezuelan Equine Encephalitis virus (VEEV). These viruses are transmitted by mosquito vectors and can infect a wide variety of mammalian and avian species. Alphaviruses are organized into complexes based on genetic and antigenic relationships. While the general features of structure and lifecycle are shared, alphaviruses can differ in properties such as receptor usage, tissue tropism, and pathogenesis. Within the Semiliki Forest virus (SFV) complex, CHIKV and RRV cause severe polyarthritis and have emerged to produce epidemics that affect millions of people globally.

To date, there are no specific antiviral treatments or licensed vaccines for any alphavirus.

Alphaviruses infect host cells by endocytic uptake and membrane fusion, thereby delivering the nucleocapsid (NC) core into the cytoplasm where replication takes place. NC disassembly is thought to be mediated by interactions with ribosomes, resulting in the uncoating of the RNA genome. The viral genome (gRNA) is an ~11.5 kb single-stranded plus-sense RNA that is 5'-capped and contains a 3'-polyA tail. It is translated to produce the viral nonstructural proteins (nsP1-4), which assemble a membrane-associated replication complex (reviewed in refs. 1, 2). An antisense complement of the gRNA is synthesized and this negative-sense RNA serves as the template for producing new gRNAs. In addition, an internal promoter drives the production of a smaller RNA that is identical in sequence to the last ~1/3 of the gRNA. This subgenomic RNA (sgRNA) is translated to produce the viral structural proteins: capsid protein (Cp), 6K/TF, and the E1, E2, and E3 envelope proteins. Cp specifically interacts with the gRNA to assemble the NC (see below). Budding of progeny virus at the plasma membrane releases particles with icosahedral symmetry, containing 240 copies each of Cp, E1, and E2, and 1 gRNA.

Alphaviruses have a robust growth cycle and produce particles with high specific infectivity, indicating strong quality control of particle assembly and genome packaging. However, it is not clear how the specific recognition and assembly of the viral RNA into NC occurs, given that Cp must select gRNA in the presence of a 3–4-fold molar excess of sgRNA produced from the replication complexes. In addition, while virus infection inhibits transcription of most cellular RNAs, abundant host RNAs still remain in the infected cell, but are excluded from the viral particle. Serial passage under high multiplicity conditions can lead to the production of defective viral genomes that can be packaged into defective interfering (DI) particles. Such DI particles are missing portions of the gRNA, and are dependent on wild-type helper virus for their propagation. Mapping, sequence analysis, and mutagenesis studies of DI particles support the presence of a packaging signal (PS) in the gRNA that promotes its recruitment into virus particles (reviewed in Mendes and Kuhn). The nature and location of the PS appear to vary depending on the virus complex, with the VEEV complex and Sindbis virus (SINV) complex having a PS located in the nsP1 region. In contrast, the PS for the SFV complex including CHIKV and RRV maps within the nsP2 region. The evidence for a role of the PS in genome packaging is strongest for VEEV, where mutation of the sensitive and specific binding site on the gRNA, rule out a role for the SFV and CHIKV PS, highlight changes in gRNA-gRNA interactions during virus biogenesis, and identify a site on the gRNA that Cp preferentially binds and assembles with in vitro. Together our data support a model for alphavirus gRNA packaging and assembly that is mediated by Cp binding to multiple sites on the gRNA.

Results

AVI-tagged Cp is a tool to stringently and efficiently retrieve Cp. The alphavirus Cp is composed of an N-terminal polybasic RNA-binding domain that mediates gRNA packaging, and a C-terminal protease domain that forms the outer NC shell (reviewed in Mendes and Kuhn). The charged nature of the Cp (pI ~10) promotes non-specific, electrostatically driven interactions, but wash conditions sufficient to disrupt them lead to poor protein retrieval by a variety of polyclonal and monoclonal Cp antibodies (RSB, unpublished observations). Moreover, prior studies showed that tagging Cp with proteins such as mCherry caused aberrant Cp localization and NC morphology (e.g. Zheng and Kielian). In order to map Cp’s binding sites on the gRNA, we devised a Cp retrieval strategy that takes advantage of the high affinity interaction of biotin with Streptavidin. The 13 residue minimal biotin acceptor peptide (mAVI tag; Fig. 1a) is specifically and efficiently biotinylated by the biotin ligase BirA. We engineered the mAVI tag into the SFV infectious clone after Cp residue 95 (SFV-mAVI; Fig. 1a), a site previously shown to be permissible for small peptide insertions. As a host cell for our studies we constructed a Vero cell line stably expressing humanized BirA in the cytoplasm (Vero BirA cells). Virus growth experiments showed that the mAVI tag was well tolerated, causing a modest decrease in maximal titers (Supplementary Fig. 1a). We observed no difference in virus or NC morphology by transmission electron microscopy (Supplementary Fig. 1b), or in the expression or localization of biotinylated Cp-mAVI vs. Cp-WT in infected cells (Fig. 1b and Supplementary Fig. 1c–e). No change was detected in SFV-mAVI vs. WT specific infectivity as calculated by comparing the number of infectious particles to total particles (PFU:E2 ratio; Fig. 1c), thus indicating accurate packaging of the gRNA by Cp-mAVI. Using the biotin handle on Cp, we optimized retrieval conditions with Streptavidin beads to efficiently and specifically pull-down Cp. Under these stringent conditions (including 0.5% SDS and washing with 0.5 M NaCl), Cp was retrieved from infected BirA cells only when biotin was present in the culture medium (Fig. 1d).

Cp binds many discrete sites across the genome. To define the Cp binding sites on the gRNA in infected cells at nucleotide resolution, we combined our optimized Cp retrieval system with the sensitive and specific PAR-CLIP method. PAR-CLIP uses a photoactivatable ribonucleoside to crosslink labeled RNAs to their binding proteins, with up to 1000-fold higher crosslinking efficiency compared to traditional CLIP-seq approaches. Crosslinking induces T-to-C mutations during cDNA library generation, enabling precise identification of a protein’s specific binding sites on the RNA. Vero BirA cells were infected with SFV-mAVI for 1.5 h. The cells were then cultured in the presence of excess biotin to ensure robust Cp-mAVI biotinylation and the photoactivatable ribonucleoside 4-thiouridine (4SU) to label nascent RNAs. 4SU had no effect on virus growth. Because alphaviruses inhibit Pol II transcription early in infection, 4SU labeling strategy strongly biases labeling toward viral RNAs over cellular RNAs. At 7 h post-infection cells were

PAR-CLIP (photoactivatable ribonucleoside crosslinking and immunoprecipitation) to comprehensively map Cp interactions on the SFV genome with nucleotide precision. Our results identify Cp binding sites on the gRNA, rule out a role for the SFV and CHIKV PS, highlight changes in gRNA-gRNA interactions during virus biogenesis, and identify a site on the gRNA that Cp preferentially binds and assembles with in vitro. Together our data support a model for alphavirus gRNA packaging and assembly that is mediated by Cp binding to multiple sites on the gRNA.
irradiated with UV (λ 368 nm) to crosslink RNAs with bound proteins, lysed, and RNAs digested with RNaseT1 to produce footprints protected by RNA-binding proteins. The total cellular pool of Cp-mAVI-biotin was then retrieved with Streptavidin beads, and crosslinked RNAs were 5′-end labeled with γ-32P-ATP and subjected to SDS-PAGE followed by transfer to a nitrocellulose membrane. The resulting Cp-RNA adducts were only detected upon UV irradiation and were the only UV-dependent crosslinked products that were retrieved (Fig. 1e and Supplementary Fig. 1f).

The RNAs crosslinked to Cp were purified and converted into cDNA libraries and sequenced using the Illumina MiSeq Platform (see Methods section for details). From two biological replicates we obtained 1,384,633 and 3,213,621 sequence reads of which 121,119 and 284,837 mapped to the viral genome, respectively. For further analysis we only considered the 105,920 and
233,188 sequence reads, respectively, that contained the diagnostic T-to-C mutation introduced during cDNA library construction of 4SU-labeled and crosslinked RNA. This allowed us to (a) remove background sequences from co-purifying, non-crosslinked fragments from abundant RNAs and (b) identify the crosslinking site at nucleotide resolution. Comparison of the crosslinked sequence reads revealed an excellent correlation for read density of the gRNA between the two biological replicates (Pearson correlation coefficient $r = 0.8123$). We proceeded to combine the two datasets and count normalized read density across every nucleotide in the gRNA (Fig. 2a). Cp binding sites were defined using an arbitrary cutoff of read density (Experimental Methods). Altogether we defined 58 high-confidence sites with a median length of 24 nt (Supplementary Data 1).
Approximately ~100 additional Cp interaction sites could be observed on the gRNA, but with significantly lower read densities (Supplementary Data 2).

Analysis of the nucleotide composition of Cp’s top 58 binding sites revealed that these sequences were relatively U-rich and A-poor (Supplementary Fig. 2a). In contrast, sequences negative for Cp binding in either replicate were relatively A-rich. Previous studies show that PAR-CLIP can identify binding sites of all compositions. For example, binding sites for the HIV-Gag protein are not U-rich33, while CNBP and DHX36 selectively bind G-rich sites34,35. Thus our results suggest that the SFV Cp does in fact preferentially bind U-rich sequences in the cell. A computational search36 for sequence motifs enriched in the top 58 Cp binding sites revealed that GC- and UG-based motifs were common, but did not identify a specific sequence motif from these 58 sites (Supplementary Fig. 3a).

The published DI-RNA studies mapped the SFV PS to a 266-nt region located within the nsP2 gene (nt:2726–2991; here termed Full PS), suggesting a role for this region in gRNA packaging19,20. Our PAR-CLIP data showed that Cp bound a 35-nt region toward the 3’ end of this region (nt: 2892–2926; here termed Cp-PS binding site), with weaker Cp binding also observed within nt: 2815–2856 (Fig. 2c). This is the first direct demonstration of in cell interaction of Cp within the Full PS region. However, the Cp-PS binding site was not the top binding site on the gRNA, with 20 other sites showing higher binding (Fig. 2a, b).

We re-analyzed the Cp binding sites on the gRNA using the Cp-PS binding site as a lower cutoff. Within these 21 high affinity sites there was a stronger trend toward Cp binding U-rich and A-poor sequences (Supplementary Fig. 2a). Sequence motif analysis of the top 21 sites showed an enrichment for UUG and UGG-trinucleotide motifs (Supplementary Fig. 3a). Secondary structure predictions for the top 21 sites predicted all 21 RNAs would adopt stem-loop structures, but of varying sizes (Supplementary Fig. 3b)37.

Manual inspection of Cp’s highest affinity sites revealed that Cp’s top two binding sites (nt: 5988–6016 and nt: 4563–4592; Fig. 2b, sites #1 and #2) are very similar in sequence and contain UUG and UGG motifs (Fig. 2d, bolded). Furthermore, sites #1 and #2 share an identical stretch of seven-nucleotides not present at any other location within the gRNA (Fig. 2d, underlined). Sequence alignment analysis showed that site #1’s sequence is hyperconserved across several other viruses in the SFV complex compared to their overall sequence identity (Fig. 2e), while site #2 and 12 other top binding sites are hyperconserved in CHIKV as well (Supplementary Fig. 4a). Closer inspection of site #1 and #2’s predicted secondary structures revealed similar stem-loop structures containing tandem G:U wobble base pairs within the stem (Supplementary Fig. 3b). Tandem G:U wobble base pairs were also predicted to form in the stems of sites #8, #14, and #15 (Supplementary Fig. 3b). While these results are interesting, we note that such predictions are limited because they do not account for the sequence’s placement within the context of the gRNA (~11.5 kb), or for the effects of Cp binding or the complex cellular environment. In keeping with this, secondary structure predictions using longer fragments centered around each of the 17 input sequences produced variations in the predicted stem-loop structures (as shown in Supplementary Fig. 3c, d for Sites#1 and #2). Thus direct experimental evidence will be needed to determine the structures adopted by these Cp binding sites in the gRNA within the infected cell and virus particle.

Cp preferentially binds sequences unique to the genomic RNA.

The sgRNA is identical in sequence to the last ~1/3rd of the gRNA and is present in 3–4X molar excess over it in infected cells9,10. It has therefore been hypothesized that Cp binding sites specific to the gRNA, such as the PS, dictate selective packaging in infected cells. Inspection of our PAR-CLIP data suggested that Cp binding was biased toward the first ~2/3rd of the genome (Fig. 2a). To formally test this, we summed the number of observed reads at each nucleotide position and categorized them as mapping to the gRNA-specific region vs. mapping to the sgRNA-overlapping region (Fig. 3a). We performed a $\chi^2$ analysis to test whether the observed summed read distributions were statistically different from our null hypothesis, in which random Cp binding would result in ~2/3 of the summed reads mapping within the gRNA-specific region and ~1/3 within the sgRNA region. The results showed that for each replicate, observed read distributions mapping to gRNA-specific regions were significantly higher than expected ($p < 0.0001$; Fig. 3a). This result was not due to a nucleotide bias as the gRNA-specific and sgRNA sequences have nearly identical compositions (Fig. 3b). The enriched binding of Cp to gRNA-unique sequences suggests a direct molecular rationale for selective Cp recognition and packaging of gRNA vs. sgRNA.

Cp binding sites promote infectious virus production.

We set out to assess the significance of Cp’s top binding sites on genome packaging by mutating them and measuring the effect on infectious virus production. All mutations were designed to strictly maintain amino acid sequence identity, disrupt predicted secondary structures, and minimize rare codon usage and changes in dinucleotide frequencies (Supplementary Data 3). We first mutated the Full PS previously identified by DI-RNA studies (see Supplementary Fig. 3e for predicted effects on PS secondary structure). To our knowledge, while such mutational analyses have been performed for the SINV and VEEV PS18, this is the first direct test of PS function for any SFV complex virus. We infected Vero cells with the SFV WT or Full PS mutant and measured infectious virus production over time. The Full PS mutant did not affect virus growth compared to WT (Fig. 4a).
This is consistent with our PAR-CLIP data, which demonstrated that Cp does not extensively interact with the PS region, and more efficiently binds ~20 other sites on the gRNA.

To test if other alphaviruses in the SFV complex were dependent on the previously identified PS, we mutated the region of the CHIKV gRNA corresponding to the SFV DI-RNA-derived Full PS (Supplementary Fig. 4b and Supplementary Data 3). Similar to our results with SFV, we observed no effect of mutation of the CHIKV Full PS on infectious CHIKV production (Fig. 4b).

We next focused on a series of mutational analyses of the top gRNA-specific binding sites. Mutation of Cp binding sites #1 and #2 (Double mutant) had a minor effect on infectious virus production (~4-fold reduction at 5 hpi; Fig. 4c). Mutagenesis of sites #1–4 plus the Cp-PS binding site (Quintuple mutant) caused a more pronounced effect on infectious virus production (~8-fold reduction at 5 hpi; Fig. 4c); growth at later time points was comparable to that of WT. We then tested if packaging could occur through cooperation among multiple Cp binding sites on the gRNA. Such a multi-site packaging mechanism has been described for some plus-sense RNA viruses. Using the 35-nt Cp-PS binding site as a threshold, we mutated 17 gRNA-specific binding sites on the gRNA (Supplementary Data 3). Other than the site within the PS, none of these have been previously implicated in any stage of the alphavirus life cycle. The 17-site mutant produced significantly less infectious virus compared to WT at early time points of virus production (Fig. 4c; ~74-fold reduction at 5 hpi, and ~6-fold reduction at 6 hpi). Combining the 17-site mutant with the full PS mutant produced no additional decrease in virus production. The 17-site mutant phenotype became more moderate at late times of virus production (Fig. 4c; ~3-fold reduction at 8 and 10 hpi), suggesting that these sites are most important at early time points in virus assembly when Cp levels are limiting. This is consistent with a multi-site packaging mechanism where high affinity binding sites on the gRNA have a greater impact on gRNA recognition and NC assembly at early infection times when Cp levels are lower.

To determine if the 17-site mutant specifically impacted gRNA packaging and virus assembly vs. aspects of virus replication, we first checked viral protein expression levels by western blot. We observed comparable levels of both nsP2 and Cp between WT vs. 17-site mutant-infected cells (Fig. 4d). We next used RT-qPCR to measure cellular gRNA and sgRNA levels over the course of virus infection, using primers that either specifically amplify the gRNA or amplify both gRNA and sgRNA due to their overlapping sequences. No significant difference was observed in cellular gRNA or gRNA+sgRNA levels between WT and 17-site mutant-infected cells from 3 to 7 hpi (Fig. 4e). Thus, although the gRNA mutations significantly reduce infectious virus production, this was not due to defects in viral RNA replication or protein expression.

We next asked whether the reduction in infectious particles could be due to a decrease in genome packaging in the 17-site mutant. We infected cells, purified SFV WT or 17-site mutant viruses, and quantitated viral gRNA and [gRNA+sgRNA] levels in the released particles. There was a significant decrease in gRNA in the 17-site mutant virus collected at 5 hpi (>10-fold reduction; Fig. 4f). There was no significant difference in gRNA levels at 7 hpi, or in [gRNA+sgRNA] levels for either time point (Fig. 4f). The ~10-fold reduction in gRNA levels in the 17-site mutant virus at 5 hpi correlated well with the ~10-fold reduction in infectious particles observed in a parallel sample at 5 hpi (Fig. 4g).

At these time points and multiplicities we were unable to directly quantitate particle numbers by quantitating E2 in released viral particles. Since it was previously demonstrated that...
a decrease in alphavirus gRNA packaging can result in increased packaging of the sgRNA\textsuperscript{11,18}, we analyzed if there was more sgRNA packaged into 17-site mutant particles. Results showed that even though the gRNA levels were significantly reduced in the mutant virus particles at 5 hpi, their [gRNA + sgRNA] levels were not significantly different from those of WT virus (Fig. 4f). Comparing the ratio of [gRNA + sgRNA] to gRNA between the WT and 17-site mutant particles showed a significantly higher ratio (~4-fold) for the 17-site mutant compared to WT (Fig. 4h), suggesting a higher proportion of sgRNA in the mutant particles. These data thus suggest that the decrease in gRNA and infectious particles observed for the 17-site mutant is at least in part caused by increased non-specific packaging of the sgRNA.

Together, our results rule out a significant role of the previously proposed DI-RNA-derived PS in infectious virus production for the SFV complex. Instead, they support a mechanism in which alphaviruses package their gRNAs through Cp interaction with multiple high affinity sites present in the gRNA-specific region of the genome.

**Changes in Cp-gRNA interactions during virus exit.** Our PAR-CLIP results were obtained by analysis of the total cellular Cp in infected cells, a pool that includes both unassembled Cp and assembled NC (Fig. 5a). To examine different stages of virus assembly, we compared these total cellular Cp results with
PAR-CLIP analyses performed on cellular and viral NC produced in 4SU-labeled cells (Fig. 5a). For cellular NC, UV-crosslinking was performed on infected cells as above, followed by cell lysis and subsequent NC purification by sucrose gradient sedimentation (Supplementary Fig. 2b). For viral NC, released virus particles were pelleted, UV-crosslinked, and then lysed. Biotinylated Cp-RNA adducts from viral and cellular NCs were retrieved with Streptavidin beads (Supplementary Fig. 2c), as described above. As observed for the total cellular Cp, retrieval was strictly dependent on UV irradiation (Fig. 5b). PAR-CLIP libraries were generated from two biological replicates for each sample type, and reads were processed as above. Biological replicates correlated strongly (Pearson correlation coefficient of \( r = 0.8655 \) for cellular NC and \( r = 0.9459 \) for viral NC) and thus we averaged the duplicate libraries and plotted normalized read density across the gRNA sequence (Fig. 5c, d).

Overall, Cp-gRNA interactions appear similar across virus assembly states, with many of the same binding sites maintained over the course of assembly (Fig. 5c, d, arrowheads). Cp’s highest affinity site in the total cellular Cp sample (Fig. 2b, #1) is also the highest affinity site in cellular NCs (Fig. 5c, termed NC#1) and viral NCs (Fig. 5d, termed V#1), demonstrating that Cp preferentially binds this specific RNA sequence across virus assembly states. Additionally, the third best bound site in the total cellular Cp population (Fig. 2b, #3) was also the third best bound site in cellular NC (Fig. 5c, NC#3) and the second-best bound site in viral NC (Fig. 5b, V#2). Cp’s top four binding sites in the total cellular Cp population also showed strong binding within cellular NC and viral NC (Figs. 2b and 5c, d and Supplementary Data 2), further emphasizing Cp’s affinity for these RNA sequences. In addition, all of the assembly states displayed enriched binding of Cp to gRNA-specific sequences (Fig. 5d).

Distinct changes were observed between cellular NC and viral NC (Fig. 5c, black arrows). For example, Cp’s second highest affinity site in cellular NC is located at nt: 2997–3016 (Fig. 5c, NC#2). Upon virion budding, this binding site drops below the top 50 binding sites within the viral NC. This pattern of strong Cp-gRNA binding events in cellular NC that become less prominent in viral NC is also observed for cellular NC sites NC#4 (nt:6096–6115), NC#5 (nt:4954–4972), and NC#8 (nt:5569–5706), among others (Fig. 5c, arrows). In contrast, certain sites such as site #1 are bound in all of the assembly states we assayed, suggesting these sites may play critical roles in NC assembly and structure.

**Specific binding of Cp and gRNA site #1 in vitro.** Cp’s high promiscuity for nucleic acids and anionic molecules in vitro induces its assembly into spherical core-like particles (CLPs) that strongly resemble viral NCs in overall structure and symmetry.\(^{21,23,24}\) CLP assembly is largely driven by electrostatic interactions with little discrimination among anionic cargoes\(^{25,40}\). Because of this, it is unclear whether specific RNA sequences can drive CLP and NC assembly. To date, no discrete RNA sequence has been shown to be specifically recognized by full-length Cp both in infected cells and in vitro. Our PAR-CLIP data showed that Cp’s top binding site on the gRNA across all stages of virus assembly was the same sequence (#1 in Figs. 2a and 5c, d). This suggested that site #1 may encompass a distinct RNA motif recognized by Cp. To assess this, we tested whether Cp specifically binds and/or assembles with site #1 in vitro. We recombinantly expressed and purified N-terminally Strep-tagged SFV Cp (Supplementary Fig. 5a). To specifically test binding, Cp was immobilized on Streptactin beads to prevent CLP assembly. Immobilized Cp was then incubated with \(32^P\)-labeled RNAs representing site #1 or a mutated version identical to that generated in the virus mutant, either alone (first lanes) or in the presence of increasing amounts of an unlabeled random RNA pool (Fig. 6a; upper panels). Both RNAs bound to Cp, but the mutant site #1 RNA was more efficiently competed by random RNA. However, when the binding experiment was performed using Cp preincubated with poly(I:C) to shield non-specific electrostatic interactions, Cp specifically bound site #1 RNA even in the presence of increasing concentrations of random RNA (Fig. 6b; lower panel, left side). In contrast, under the same conditions Cp inefficiently bound the mutant RNA, and its binding was entirely competed by random RNA (Fig. 6b; lower panel, right side). Thus, Cp specifically and directly binds gRNA site #1 both in cells and in vitro, demonstrating that site #1 encompasses a motif that Cp specifically recognizes.

We next assessed whether Cp could specifically assemble with site #1 into CLPs. We found that poly(I:C) alone induced CLP formation, in keeping with the ability of long anionic polymers such as heparin to promote CLP formation\(^{25}\). Therefore, to shield non-specific electrostatic interactions we instead used a 10 nucleotide DNA oligo (here termed 10mer) that is below the length needed to induce CLP formation\(^{21}\). We confirmed that CLPs did not form with the 10mer by negative stain electron microscopy (EM), which showed amorphous material but no detectable CLPs (Fig. 6b; upper panels). The addition of site #1 RNA produced abundant spherical CLPs of \(\sim 40 \, \text{nm} \) in diameter (Fig. 6b; lower), consistent in appearance to that of previously published CLPs\(^{21,23}\). We then quantitated CLP formation by sucrose gradient sedimentation and detection of Cp by ELISA (Fig. 6c and Supplementary Fig. 5b). As predicted, Cp alone or incubated with 10mer (1:20 Cp:10mer) showed most Cp remained unassembled at the top of the gradient (Fig. 6c). Cp incubated with site #1 RNA alone
(1:1 Cp:RNA) or in the presence of 10mer DNA (1:20:1 Cp:10mer:RNA) showed a single peak centered toward the middle of the gradient (Fig. 6c), consistent with previous sucrose gradient sedimentation analyses. Thus, Cp robustly assembles into CLPs with site #1 RNA even in the presence of an excess of electrostatic competitor nucleic acid.

We then tested the effect of mutation of site #1’s sequence on CLP assembly (Fig. 6d). Cp robustly assembled into CLPs with site #1 RNA in the presence of 10mer at physiological salt concentration (Fig. 6d, 150 mM NaCl panel). However, under the same conditions CLP formation with the mutant RNA was impaired (Fig. 6d, 150 mM NaCl panel). This decrease in CLP
formation was more apparent when the assembly reaction’s salt concentration was raised to shield non-specific electrostatic interactions. Site #1 induced some CLP formation even at 300 mM NaCl, while the mutant only assembled CLPs at 150 mM NaCl (Fig. 6d). Thus, site #1 RNA specifically promotes CLP assembly in vitro even when general Cp electrostatic interactions are masked, suggesting that site #1 on the gRNA may also act to promote NC assembly in infected cells.

Discussion

We employed PAR-CLIP and biotin-based Cp retrieval as an experimental approach uniquely suited to define the interactions of the alphavirus Cp with the genomic RNA. Our approach identified a set of high-confidence Cp binding sites on the SFV gRNA, including an RNA sequence that specifically bound and assembled with Cp both in infected cells and in vitro. The Cp binding sites were preferentially distributed on the first 2/3 of the gRNA, suggesting a molecular rationale for the previously observed selective packaging of gRNA over sgRNA10–12. An overall preferential binding to gRNA-specific sequences was maintained across different NC assembly states. Such continuous engagement with gRNA-specific sequences can promote gRNA packaging and high infectivity.

A previous study used CLIP to identify SINV Cp interactions with the gRNA in infected cells12. Comparison with our results showed that the SINV and SFV Cp top binding sites do not overlap, and that SINV Cp’s major binding sites are located within the last ~1/3rd of the gRNA sequence, overlapping with the sgRNA. Mutation of the SINV Cp binding sites does not affect gRNA packaging or NC assembly, but instead affects the stability of the incoming gRNA. These differences may reflect biological differences between SINV and SFV, but could also be due to differences in experimental approaches.

Alphavirus gRNA packaging signals were originally identified through analysis of DI-RNAs19,20,31. Elegant studies showed that appending the SINV PS onto reporter or helper RNAs could increase their packaging in replicon systems and their direct binding to Cp in vitro, although the increase was at most 5-fold19,42. Conversely, SINV or VEEV with mutated PSs are impaired but viable18. While our data are the first to show that the alphavirus Cp directly binds the DI-RNA-derived PS in cell culture, we also found that genome packaging within the SFV complex does not require the PS. Instead, our data argue that packaging involves preferential Cp binding to multiple gRNA-specific sites. Although the relative importance of the PS appears to vary between alphavirus complexes, it seems clear that the PS is not strictly required for gRNA packaging/NC assembly. Given that recent studies indicate that DI-RNAs and defective viral genomes can have multifaceted functions (reviewed in Vignuzzi and Lopez33), the prior DI-RNA studies may have more complex interpretations than previously suspected.

We found that while mutation of multiple high affinity Cp binding sites did not significantly affect viral replication or protein expression, it significantly reduced infectious virus production. This correlated with reduced gRNA but elevated sgRNA levels in virus particles. Inhibition of infectious virus production was strongest at early stages of virus assembly, in keeping with increasing Cp concentrations at later times of infection driving Cp engagement with mutated binding sites. Based on our results we propose a model for selective genome packaging via Cp’s interactions with multiple sites selectively located on the gRNA (Fig. 7). At early stages of virus assembly, gRNA levels are high but intracellular Cp levels are relatively low, leading to preferential Cp binding to high affinity sites on the gRNA (Fig. 7a vs. ib). This stage may involve Cp dimers22, and could be represented by a proposed 90S NC assembly intermediate (not shown), which has a lower CpgRNA ratio than cellular NC

44–46. For some alphaviruses such as SINV and VEEV, this early stage may involve high affinity binding of Cp to the PS. As Cp levels accumulate in the cell, Cp begins engaging with lower affinity sites on the gRNA, enabling compaction of the gRNA molecule through charge neutralization16,38,47 (Fig. 7ii). More Cp would then be recruited until completion of the NC

48 (Fig. 7iii). E2-Cp interactions promote budding of the NC to produce the enveloped viral particle, a process that results in changes in Cp-RNA interactions (Fig. 7iv). Additional unknown factors may influence genome packaging and NC assembly, including specific subcellular locations or host factors.

From a broader perspective, a multi-site packaging mechanism could be favorable for many RNA viruses given that RNA polymerases generally have low fidelity. Their relatively high mutational frequency produces genetic diversity important for RNA virus adaption to environmental changes, but can also decrease virus fitness49,50. A multi-site packaging mechanism could help to ensure correct genome packaging even if deleterious mutations arose in several Cp binding sites. For example, retroviruses such as HIV-1 have high mutation rates and a defined packaging signal known as the Ψ element (reviewed in D’Souza and Summers31). However, Ψ mutants only modestly decrease (~3–5-fold) infectious virus production or genome packaging33,32. Multi-site packaging mechanisms have been described for RNA viruses such as parechovirus133 and hepatitis B virus34, as reviewed in Twarock and Stockley25. In addition, multi-site packaging may promote successful co-assembly of Cps and large ssRNAs while decreasing non-productive assembly pathways39. Thus, the multi-site-based packaging that we describe here for alphaviruses may be a representative of a more general mechanism utilized by many RNA viruses to selectively package their genomes.

Our data represent the first demonstration of a discrete RNA sequence, site #1, with which alphavirus Cp specifically binds and assembles both in cells and in vitro. While multiple binding sites such as site #1 were maintained across virus assembly, we also observed discrete differences in CpgRNA interactions between the cellular and viral NCs. This could reflect changes induced by different chemical environments, or could be induced by envelope protein association and virus budding. Changes in NC architecture during various stages of virus exit have been previously observed using morphological and biochemical techniques56–59.
Our data thus suggest that such changes in the NC structure also significantly affect Cp-gRNA binding. If E2 interactions/budding affect Cp-gRNA interactions, this would suggest that changes in NC architecture are propagated from the outer, envelope-protein interacting face of the NC shell to the inner RNA core of the virus particle. Further functional analyses are needed to determine the potential importance of such changes in virus Cp-RNA interactions during exit.

**Methods**

Note, the included Source Data files contain full images of gels and spreadsheets of data for graphs, etc.
**Fig. 6** Cp specifically binds and assembles with site #1 in vitro. a Upper panel: 333 nM 32P-labeled site #1 (left) or mutant (right; mutated residues highlighted in red) RNAs were bound to 13.3 pmole Cp immobilized on Streptactin beads in a 30-μL reaction. Increasing concentrations (4–1000 nM) of random RNA were added as a competitor. Bound RNAs were extracted and subjected to Urea-PAGE followed by autoradiography. Lower panel: as above, but immobilized Cp was first preincubated with 20 μg/mL of poly(I:C) to shield non-specific electrostatic interactions. Results are representative of independent experiments. b Negative stain EM of CLP assembly reactions of Cp + 10mer or Cp + 10mer + Site #1 RNA. Dashed boxes in the right panel are ×3 magnification of the corresponding left panel. Scale bar is 80 nm. Images represent one experiment. CLP assembly reactions were incubated for 30 min at 25 °C (150 mM NaCl) and then analyzed by sucrose gradient sedimentation. Cp from aliquots of fractions was detected by ELISA. Individual data points from technical duplicates are plotted and the line represents their mean. Results are representative of independent experiments for Cp alone, Cp + 10mer (1:120), Cp + Site #1 (1:1), and Cp + 10mer + Site #1 (1:20:1), and n = 2 independent experiments for Buffer alone. d As in e, except the indicated salt concentrations were used in CLP assembly reactions comparing CLP assembly with site #1 or mutant RNAs. Individual data points from technical duplicates are plotted and the line represents their mean. Results are representative of n = 4 independent experiments for 250 mM NaCl, and n = 2 independent experiments for 150, 300, and 350 mM NaCl. See also Supplementary Fig. 5.

**Fig. 7** A multi-site genome packaging model for alphaviruses. (i) Cp (green) specifically interacts with high affinity sites (blue rectangles) on the gRNA (black lines). Such high affinity interactions (low Kd) may facilitate longer dwell times for Cp on the gRNA (slow koff), resulting in stable binding. (ii) Cp can also non-specifically interact with low affinity sites (high Kd) on the gRNA or other RNAs such as the sgrNA or cellular RNAs (not shown), but such low affinity interactions may lead to short dwell times for Cp on those RNAs (fast koff) and less stable binding. (iii) As Cp levels increase in the cell, Cp begins binding multiple high affinity sites on the gRNA and stays stably bound (low Kd). This may enable further Cp recruitment, including Cp dimer formation, and lead to Cp-Cp oligomerization and gRNA compaction. (iii) NC assembly is completed, resulting in 1 copy of the gRNA surrounded by a shell comprised of 240 copies of Cp arranged with icosahedral symmetry. (iv) Virus assembly and budding occur at the plasma membrane where Cp directly binds the cytoplasmic domain of the E2 envelope protein (coloring as in Fig. 5a). Virus assembly and budding cause discrete changes in Cp:gRNA interactions within the viral NC.

**Cells.** Baby hamster kidney (BHK-21) cells (gift from Dr. Ari Helenius) were cultured in Dulbecco’s modified Eagle’s medium (DMEM, HyClone) containing 4 mM l-glutamine, 100 U penicillin/mL, 100 μg streptomycin/mL, 1% tryptose phosphate broth, and 5% FBS at 37 °C. Vero cells (ATCC; gift from Dr. Kartik Chandran) were cultured in DMEM containing 4 mM l-glutamine, 100 U penicillin/mL, 100 μg streptomycin/mL, and 10% FBS at 37 °C. Vero cells stably expressing HA-BirA were generated by transfecting cells with the expression plasmid described below using Lipofectamine 2000 (Invitrogen). 48 h post-transfection, cells were split to 25% confluence and selected with 1 mg G418/mL (Sigma-Aldrich) for 2 weeks. G418-resistant cells were then seeded sparsely and individual clones isolated using cloning cylinders (Corning). Clones were screened for their ability to biotinylate Cp-mAVI by western blot and fluorescence microscopy using streptavidin probes. Cell lines were authenticated by morphologic evaluation and were checked for mycoplasma contamination (MycoAlert™ PLUS Mycoplasma Detection Kit, Lonza).

**Antibodies.** The following antibodies and SA probes were used for western blots and immunofluorescence as indicated: monoclonal anti-β tubulin (E7; WB-1:3000; Developmental Studies Hybridoma Bank), rabbit polyclonal anti-α tubulin (WB-1:5000; Abcam, 18251), rabbit polyclonal anti-E2/1 (WB-1:4000 and IF-1:5000), monoclonal anti-Cp (42-1; WB-1:4000 and IF-1:5000), monoclonal anti-E2 (WB-1:2 dilution of the clarified hybridoma supernatant), Streptavidin Alexa Fluor 680 conjugate (WB-1:10,000; ThermoFisher Scientific, A-31556), and Streptavidin Alexa Fluor 568 conjugate (IF-1:700; ThermoFisher Scientific, A-21076). Rabbit polyclonal anti-nsp2 (WB-1:3000) was a gift from Dr. Andres Merits. Rabbit polyclonal anti-Cp was generated (Converse) against the C-terminal domain of SFV NC (aa 267, expressed and purified by affinity chromatography). A 240-μL reaction was incubated for 30 min at 25 °C (150 mM NaCl) and then analyzed by sucrose gradient sedimentation. Cp from aliquots of fractions was detected by ELISA. Individual data points from technical duplicates are plotted and the line represents their mean. Results are representative of independent experiments for Cp alone, Cp + 10mer (1:120), Cp + Site #1 (1:1), and Cp + 10mer + Site #1 (1:20:1), and n = 2 independent experiments for Buffer alone. d As in e, except the indicated salt concentrations were used in CLP assembly reactions comparing CLP assembly with site #1 or mutant RNAs. Individual data points from technical duplicates are plotted and the line represents their mean. Results are representative of n = 4 independent experiments for 250 mM NaCl, and n = 2 independent experiments for 150, 300, and 350 mM NaCl. See also Supplementary Fig. 5.

**Viruses, mutants, and plasmids.** A minimal biotin acceptor motif (mAVI tag; LNDIFEAQKIEWH)65,66 was inserted after Cp residue 95 by overlapping PCR using restriction sites XbaI and NsiI in the SFV infectious clone pSF6-SFV4. Overlapping PCR primers: 5′-CTGAAGACATCCTTGGGCCCCGCAATCG ATGTCGACaaaagcagacagaac-3′ and 5′-GTGGCCATTGGGATTGGGGGTCGC TGAGATGTGTCGAGGATCTTTGCTG-3′. Individual Cp binding site mutants were constructed by overlapping PCR and using the nearest unique restriction sites present in pSF6-SFV4. Double and Quintuple Cp binding site mutants were constructed by restriction digest and subcloning from individual mutants. The 17-site mutant, Full PS mutant, and 17-site + Full PS mutant were custom synthesized (Epoch Life Science, Inc), and subcloned by restriction digest into pSF6-SFV4. The Chikungunya Full PS mutant was custom synthesized (Epoch Life Science, Inc) and subcloned by restriction digest into the pSinRep5-181/25 Chikungunya infectious clone (kindly provided by Dr. Terrence Dermody67). All infectious clone mutants were verified by sequencing the entire region affected by the cloning approach (Genewiz, South Plainfield, NJ) and analyzed by restriction digest to check for plasmid rearrangements. A humanized BirA construct was provided by Vasily Ogryzko65 and subcloned into pcDNA3.1- with an N-terminal HA tag added by PCR.

**Virus stocks.** SFV WT and mAVI virus stocks were generated by electroporating BHK cells with in vitro transcribed (IVT) viral RNA and collecting the cell media at 24 h post-electroporation. Virus-containing media were clarified by centrifugation at 10,621xg 4°C for 10 min, and 10 mL HEPES pH 8.0 was added to the
supernatant before aliquoting and freezing. Virus stocks for growth comparisons of SFV WT, Full PS, and the indicated Cp binding site mutants were generated as described above except that the cell media were harvested at 22 h post-electroporation. CHIKV WT and Full PS mutant stocks were generated by coinfection with the WT at a 5:1 ratio to compare Cp expression levels by western blot (Supplementary Fig. 2c). Streptavidin Dynabeads (SA-DB; Dynabeads MyOne Streptavidin C1, Thermofisher Scientific) were equilibrated with PBS while samples thawed on ice. A final concentration of 0.5% SDS was added to the samples, mixed, and then incubated with the SA-DB by vortexing for 30 min at 4 °C. The SA-DB were washed three times with high salt (10 mM Tris-Cl pH 7.4, 0.5 M NaCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, 0.5 mM DTT, and a complete protease inhibitor tablet (Roche; 1 tablet/10 mL)) on ice for 5 min each, once with PBS, and twice with 100 µM RNAseT1 at 1 µL/for 15 min at room temperature before cooling on ice for 5 min. The SA-DB were washed again with high salt RIPA buffer for 5 min, then washed twice with dephosphorylation buffer (50 mM Tris-Cl pH 7.9, 100 mM NaCl, 10 mM MgCl2, and 1 mM DTT), and treated with calf intestine phosphatase at 0.5 U/µL for 10 min at 37 °C. The SA-DB were washed once with lysis buffer, twice with PKN buffer (50 mM Tris-Cl pH 7.5, 30 mM NaCl, 10 mM MgCl2, 5 mM DTT) lacking DTT, then resuspended in complete PKN buffer. In all, 1 U/1 mL of PKN and 0.75 µCi of [32P]-ATP were added to the SA-DB for 30 min at 37 °C with gentle mixing every 5 min. After 30 min, 100 µM cold ATP was added for 5 min at 37 °C. The SA-DB were washed five times with PKN buffer without DTT and then once with PBS. Samples were boiled for 5 min in SDS sample buffer four sequential times to maximize elution before pooling eluates. The eluates were subjected to SDS-PAGE and then transferred to nitrocellulose membranes for autoradiography. The film was overlaid on top of the membrane and the corresponding cDNA crosslinked adducts were excised from the membrane. The membrane was treated with proteinase K for 90 minutes at 55 °C, and the RNA was extracted with phenolchloroform and precipitation.

Cellular and viral Ncs: PAR-CLIP on cellular and viral Ncs was performed as described above, but with the following differences. Two 10 cm plates of Vero + BirA cells were infected with SFV mAVI at an MOI = 10. At 7.5 hpi, the virus-containing media were collected and centrifuged at 10,000 rpm for 5 min at 4 °C. The virus was then purified over a 20% sucrose cushion by centrifugation at 35,000 rpm (Beckman Coulter SW41 rotor) for 4 h at 4 °C. The pellet was resuspended on ice for 3 h in TN buffer and then transferred to a 35-mm dish. The virus was UV irradiated or mock irradiated with 368 nm light at 0.15 J/cm², and then lysed in lysis buffer, denatured with 0.5% SDS, and processed as described under Total cellular Cp population. By this method, the virus population was confirmed by UV-irradiating the samples with PBS and then UV-irradiated with 368 nm light at 0.15 J/cm² or mock-irradiated. Cells were pelleted, lysed in lysis buffer, centrifuged at 13,000 rpm for 10 min at 4 °C, and then treated with 25 mM EDTA for 20 min to dissociate polysomes. Samples were loaded onto 7.5–20% (wt/wt) sucrose gradients in TN buffer + 2 mM EDTA and 0.1% NP-40, and centrifuged at 41,000 rpm (Beckman Coulter SW41 rotor) for 2 h at 4 °C. In total, 1 mL fractions were collected and aliquots were analyzed by SDS-PAGE followed by western blot using an anti-Cp antibody to identify the NC fractions. The NC fractions were pooled, denatured with 0.5% SDS, retrieved with SA-DB, and processed as described under Total cellular Cp population.

cDNA library construction and sequencing. The extracted RNA was ligated to the 3′ adenylated adapter using RnlI (5′-2491K27Q) ligation at 4 °C overnight. A size marker containing synthetic 19- and 35-nt oligos was used as a positive control for all ligation steps. After ligation, the reaction was denatured at 90 °C for 1 min and electrophoresed on a 15% denaturing Urea-PAGE. The gel was then exposed to a phosphoimager screen for an hour at −20 °C. The image was printed to its original size and aligned on to the gel where the successfully ligated product was excised, shredded, and incubated with 0.3 M NaCl at 60 °C for 45 min. The sample was filtered and then the RNA was precipitated in ethanol at −20 °C for at least an hour. The precipitated RNA was pelleted and dissolved in water. The 5′ adapter was ligated to the sample using RnlI for 1 h at 37°C. The sample was then processed the same way as the 3′ adapter ligation reaction. The ligated sample was reverse transcribed to make cDNA using Superscript III reverse transcriptase for 2 h at 50°C. A pilot PCR was performed to optimize PCR cycle number to prevent over amplification of the library. This entailed taking a 10-µL aliquot from the pilot PCR after every three cycles between cycles 12 and 30 for agarose gel analysis. The lowest number of cycle number to generate a visible PCR product by ethidium-bromide staining was used for the final PCR. The final PCR product was purified away from linker-linker (3′ adapters ligated to 5′ adapters) byproducts using a 3% Pippin Prep (Sage Science) before sequencing on an Illumina MiSeq machine. Oligonucleotides used (5′ to 3′):

- RNA PCR Index Primer 9
  - CAGGGGATACAGGCCGCAATACAGGATCTGCTGAGTCTGGAGTCTGGACGCCGAGATGCTCAG
  - RNA PCR Index Primer 10
  - CAAGGAGGACCGGCCGATAACAGGATCACTGATGCTGAGTCTGGAGTTCTGGACGCCGAGATGCTCAG
  - RT Primer GCCGCTCTGAGCAGGCGGACTCTTGCACAGGACGCGAATTTCAA
  - 3′ Barcode adapter 29.31 5′-APPNNNTAAGGATTTACTGCTGGGTGGCCCAA
  - GG-L;
3′ Barcode adapter 29.32 5′-appNCTGATGTTGAATCTCAGGGTCCAA GG-L; 3′ barcode adapter 29.33 5′-appNTAGTCGTTGATTCTCAGGGTCCAA GG-L; L = aminoliniker; 3′-amino modifier C7.

RT-qPCR. To quantify viral RNA in infected cells, Vero cells were mock infected or infected with SFV WT or 17-site mutant along with increasing concentrations of random 29 nucleotide RNA (4–1000 nM) in each well with or without 20 μg/ml polyclonal precipitation. Binding was performed for 30 minutes at 30°C and the plate was then placed on a 96-well magnetic rack. The beads were washed three times with wash buffer (25 mM Tris pH 7.5, 150 mM KCl, 0.5 mM MgCl2, 0.01% (v/v) Tween 20, 1 mg/mL BSA, and 1 mM DTT). Binding reactions were prepared in 96-well plates. A total of 13.3 pmol of Cp was immobilized on beads per 30 μl reaction and concentrations of 333 nM site 1 or its mutant along with increasing concentrations of random 29 mer RNA (4–1000 nM) was added in each well with or without 20 μg/ml polyclonal precipitation. Binding was performed for 30 minutes at 30°C and the plate was then placed on a 96-well magnetic rack. The beads were washed three times with wash buffer (25 mM Tris pH 7.5, 150 mM KCl, 0.5 mM MgCl2, 0.01% (v/v) Tween 20). In all, 40 μl of elution buffer (10 mM Tris pH 7.0, 1 mM EDTA, 1% SDS) was added to the beads before heating at 70°C for 10 min. The supernatant from each well was collected in a new plate and the bound RNA was purified away from Cp by phenol/chloroform isolation. The purified RNA was resolved by 15% denaturing Urea-PAGE.

PAR-CLIP data analysis. Adapters were removed using Cutadapt 2.4 and the remaining sequences were mapped to the SFV4 genome using Bowtie2.1.1. Only mapped sequences that contained the characteristic T-to-C mutation were further used. Initial analyses compared sequence reads (uncollapsed reads) vs. unique reads stemming from the barcoded adapters (collapsed reads) to assess PCR duplicates. No significant difference was observed in Cp’s top binding sites between the uncollapsed vs. collapsed reads, which was expected because we carefully optimized PCR cycle number during cDNA library construction (as described above). Therefore, we continued our analyses with the uncollapsed reads. Reads were normalized to the nucleotide position with the highest read frequency within the library, and then biological replicates were averaged and renormalized to produce the final normalized read density. High confidence binding sites were manually defined with the following criteria: the binding site must be at least 10 nt long, have at least one position containing an average normalized read density ≥5.0, and both replicate libraries must have ≥10 reads per replicate library, and 3′ G bias due to RNaseT1 digest. Low confidence binding sites were defined as those with at least an average normalized read density ≥1.0 and ≥10 reads for each replicate library.

Statistical analysis. Statistical analyses were performed using GraphPad Prism 7.04 (GraphPad Software). The specific statistical tests used and the number of replicates used for each figure are specified in the figure legends. In all graphs, four asterisks indicate a P-value of <0.0001 and “ns” indicates no significant effect. P-values for all statistical tests are listed in the Source Data.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The PAR-CLIP datasets are publicly available in the NCBI GEO repository under GEO accession number GSE163631. The source RDB file used to create the image in Fig. 1a is publicly available in the RCSB Protein Data Bank as PDB accession number 6MTX. Source data are provided with this paper.

Received: 1 August 2019; Accepted: 24 August 2020; Published online: 17 September 2020

References
1. Kuhn, R. J. In Fields Virology. Vol. 1 (eds. D. M. Knipe & P. M. Howley) Ch. 22, 629–650 (Lippincott, Williams and Wilkins, 2013).
2. Weaver, S. C., Winegar, R., Manger, I. D. & Forrester, N. L. Alphaviruses: population genetics and determinants of emergence. Antivir. Res. 94, 242–257 (2012).
3. McFarland, A., Jaffar-Bandjee, M. C. & Gasque, P. Arthritogenic alphaviruses—an overview. Nat. Rev. Rheumatol. 8, 420–429 (2012).
4. Levi, L. I. & Vignuzzi, M. Arthritogenic alphaviruses: a worldwide emerging threat? Microorganisms 7, 133 (2019).
Acknowledgements

We thank all of the members of the Kielian lab and Dimitrios Zattas for helpful discussions and comments on the manuscript. We thank Matthew Angelidis for technical assistance. We thank the Einstein Analytical Imaging Facility and facility members Xhieni Nishuku and Timothy Mendez for training and technical assistance on the electron microscope. We thank Susan Buhl and Matthew Scharff for technical assistance and use of their plate reader. This work was supported by grants to M.K. from the National Institute of General Medical Sciences (RO1-GM057545) and the National Institute of Allergy and Infectious Diseases (RO1-AI075647) and by Cancer Center Core Support Grant NIH/NCI P30-CA13330. R.S.B. was supported by an NIH NRSA (F32-GM122450) postdoctoral fellowship and the Charles H. Revson Senior Fellowship in Biomedical Sciences. M.H. and D.G.A. are supported by the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences, the National Institute of Allergy and Infectious Diseases, National Institute of Arthritis and Musculoskeletal and Skin Diseases, or the National Institutes of Health. This study was made possible (in part) by funds granted by the Charles H. Revson Foundation. The statements made and views expressed, however, are solely the responsibility of the authors.

Author contributions

R.S.B. and M.K. conceived the project. R.S.B. and D.G.A. made the PAR-CLIP libraries. D.G.A. performed the in vitro binding assay. R.S.B. performed all other experiments. M.K. and M.H. supervised the research. R.S.B. and M.K. wrote the original draft. R.S.B., D.G.A., M.H., and M.K. reviewed and edited the draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-18447-z.

Correspondence and requests for materials should be addressed to M.K.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020