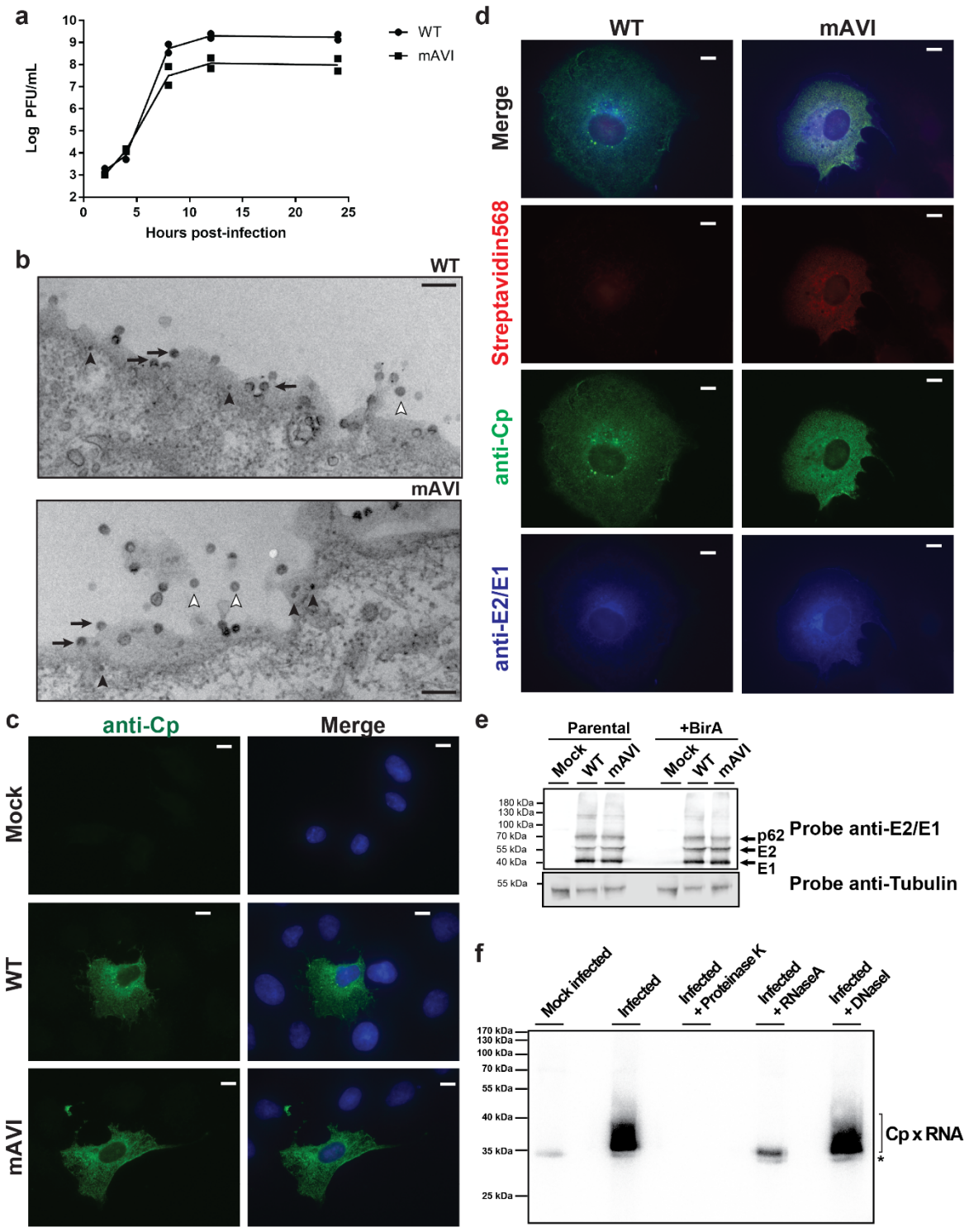
**SUPPLEMENTARY INFORMATION**

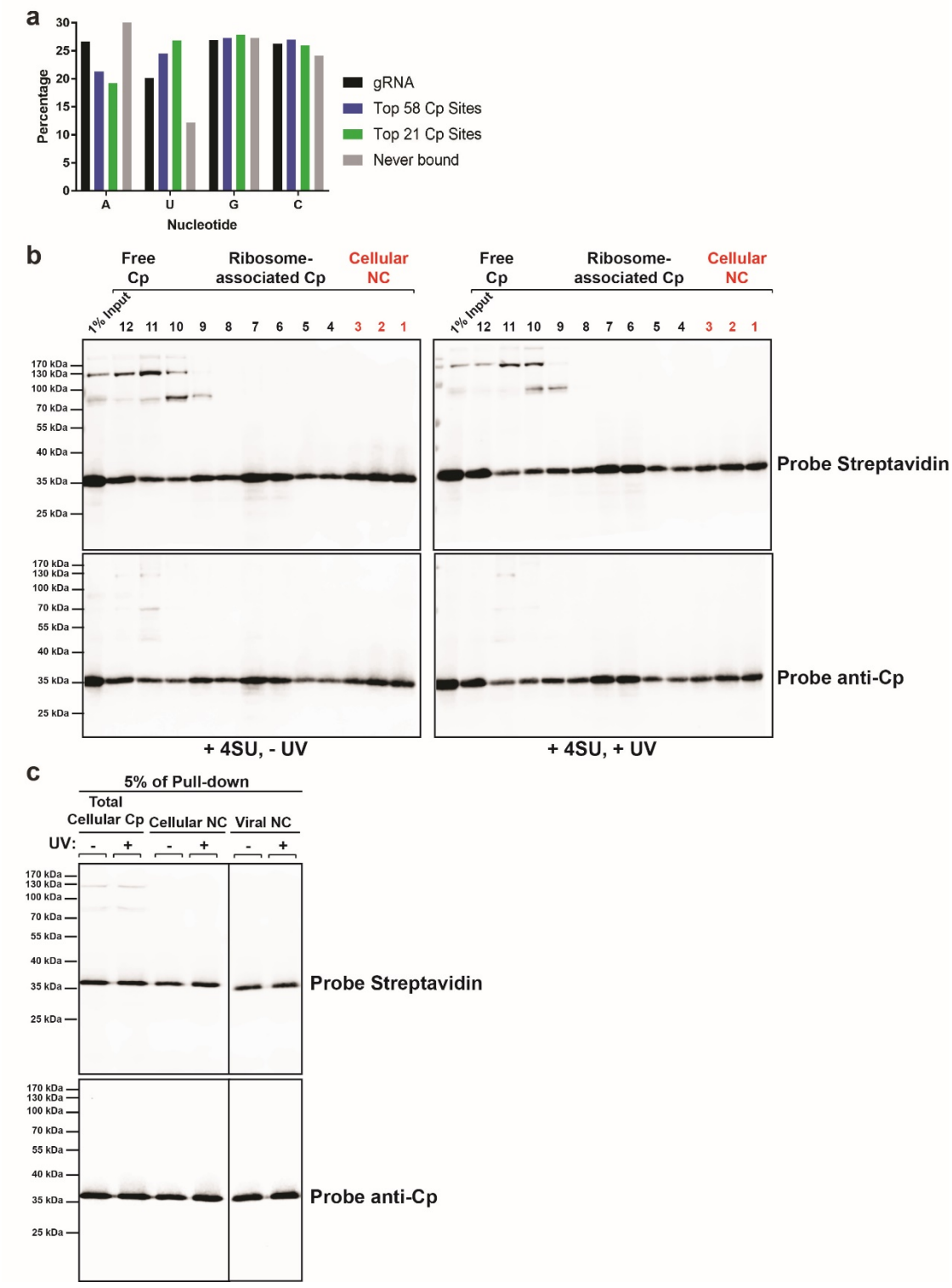
# MULTIPLE CAPSID PROTEIN BINDING SITES MEDIATE SELECTIVE PACKAGING OF THE ALPHAVIRUS GENOMIC RNA

**Brown et al.**




**Supplementary Figure 1. SFV/Cp mAVI virus is comparable to SFV/Cp WT. Related to Figure 1.** (a) Single-step growth curve comparing SFV WT vs. mAVI. Cells were infected at an MOI=10 and the media collected at the indicated time points and titered by plaque assay. Individual points from n=2 biological replicates are plotted. Source data are provided as a Source Data file. (b) Transmission electron microscopy

of Vero BirA cells infected for 7.5 h with SFV WT (upper) or SFV mAVI (lower). Scale bar represents 200 nm. Black arrows denote NCs of budding viral particles, white arrowheads denote budded viral particles, and black arrowheads denote cellular NCs. The images are representative examples from n=1 experiment. (c) Immunofluorescence of Vero cells mock infected or infected with SFV WT or SFV mAVI for 7.5 h and stained with a monoclonal antibody against Cp. Scale bar represents 10 μm. The images are representative examples from n=2 biological replicates. (d) Immunofluorescence of Vero BirA cells infected with SFV WT or SFV mAVI for 7.5 h and stained with a monoclonal antibody against Cp (green), a polyclonal antibody against E2/E1 (blue), and a Streptavidin Alexa-568 probe (red). Scale bar represents 10 μm. The images are representative examples from n=2 biological replicates. (e) The lysates from Figure 1B were analyzed by western blot using a polyclonal antibody to envelope proteins E2/E1 and a polyclonal antibody to tubulin. The images are representative examples from n=2 biological replicates. Source data are provided as a Source Data file. (f) PAR-CLIP was performed on Vero BirA cells mock infected or infected with SFV mAVI. SA-DB-retrieved samples were $^{32}$P-end labeled and either mock treated or treated with Proteinase K, RNaseA, or DNaseI to identify the nature of the lower band (asterisk). The lower band is a UV-independent (Figure 1e) and proteinaceous species. The image represents n=1 experiment. Source data are provided as a Source Data file.




**Supplementary Figure 2. PAR-CLIP sample preparation. Related to Figures 2 and 5.** (a) Nucleotide composition of the genome (gRNA), the top 58 Cp binding sites, the top 21 Cp binding sites, and the sites never bound by Cp (zero reads). (b) Vero cells were infected at an MOI=10 with SFV mAVI and incubated with 4SU. At 7 hpi, cells

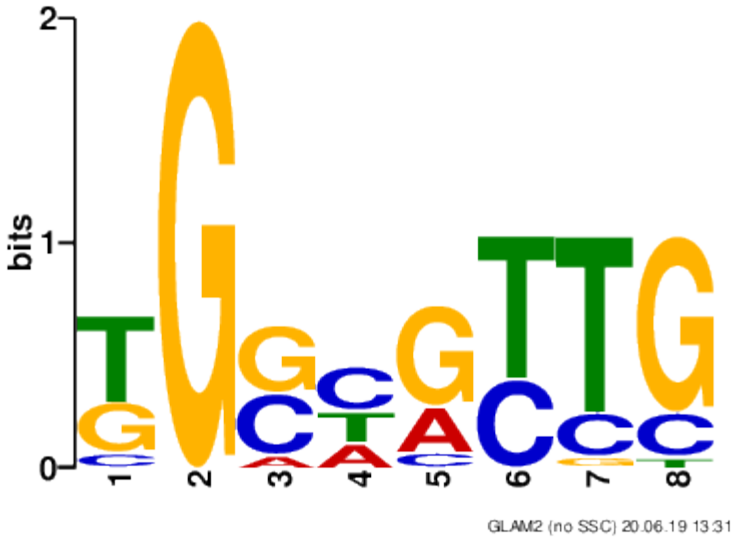
were UV irradiated or mock irradiated, lysed, processed, and the cellular NCs were fractionated by 7.5-20% (wt/wt) sucrose gradient sedimentation. Aliquots of each fraction were subjected to SDS-PAGE and western blot analysis using a polyclonal antibody to Cp and a Streptavidin Alexa-680 probe. Cellular NC fractions (red) were pooled, denatured with 0.5% SDS, and then incubated with SA-DB to isolate CpxRNA complexes. The images are representative examples from n=2 biological replicates. Source data are provided as a Source Data file. (c) Cp PAR-CLIP libraries from the total cellular, cellular NC, and viral NC populations were generated using similar starting protein concentrations. 5% of each of the SA-DB-retrieved samples was analyzed by western blot using a monoclonal antibody to Cp and a Streptavidin Alexa-680 probe. Samples are from the same gel/blot, but the lanes were rearranged post-image acquisition for lane consistency and therefore a black line was added between cellular and viral NC. The images represent n=2 independent experiments. Source data are provided as a Source Data file.

**a GLAM2 motif analysis of top 21 and top 58 binding sites.**

## Cp's top 21 sites. Best Motif Found:

| NAME | START | SITES | END | STRAND | MARGINAL SCORE |
|---|---|---|---|---|---|
| 1 | 11 | c g g . g t c g | 17 | + | 1.53 |
| 2 | 16 | t g g a g t t g | 23 | + | 8.50 |
| 3 | 16 | t g g t g t t g | 23 | + | 8.87 |
| 4 | 5 | t g c t g t t c | 12 | + | 6.53 |
| 5 | 9 | g g c a a c t c | 16 | + | 3.38 |
| 6 | 25 | t g g c c t t g | 32 | + | 6.26 |
| 7 | 10 | g g g a g c g g | 17 | + | 1.65 |
| 8 | 11 | t g g . a t t g | 17 | + | 5.53 |
| 9 | 11 | g g c c g c t g | 18 | + | 7.67 |
| 10 | 11 | t g c . g t t t | 17 | + | 1.30 |
| 11 | 10 | t g c t g c t g | 17 | + | 7.96 |
| 12 | 15 | t g c t a t t g | 22 | + | 7.51 |
| 13 | 5 | g g c c g t c g | 12 | + | 6.23 |
| 14 | 5 | t g g a g t t g | 12 | + | 8.50 |
| 15 | 23 | g g g . g c c g | 29 | + | 2.97 |
| 16 | 37 | t g a c a c t g | 44 | + | 4.72 |
| 17 | 3 | t g c c g t c c | 10 | + | 4.89 |
| 18 | 24 | g g g . a t t c | 30 | + | 2.52 |
| 19 | 11 | c g a c g t t g | 18 | + | 3.64 |
| 20 | 7 | g g g t a c t g | 14 | + | 6.09 |
| 21 | 14 | t g c c c c t g | 21 | + | 5.35 |

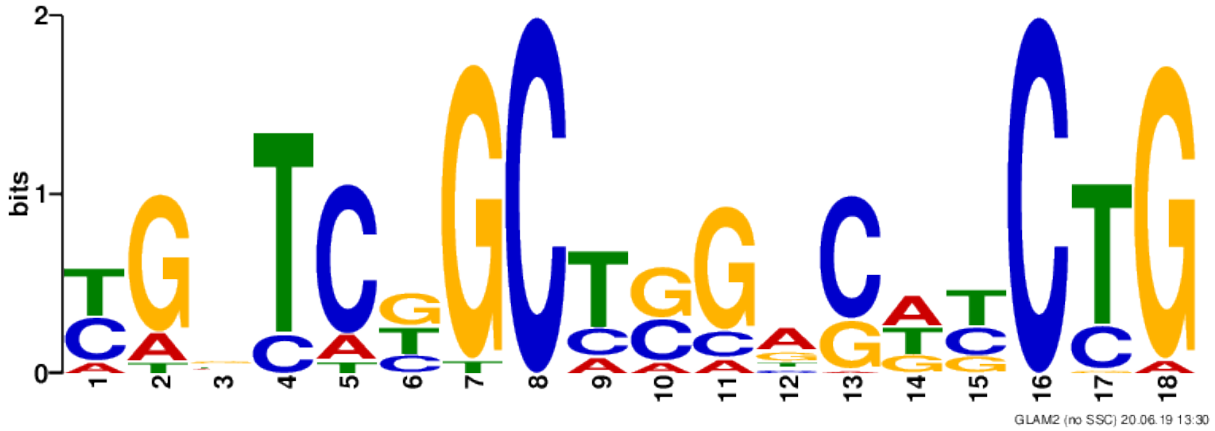
Score: **81.8401**



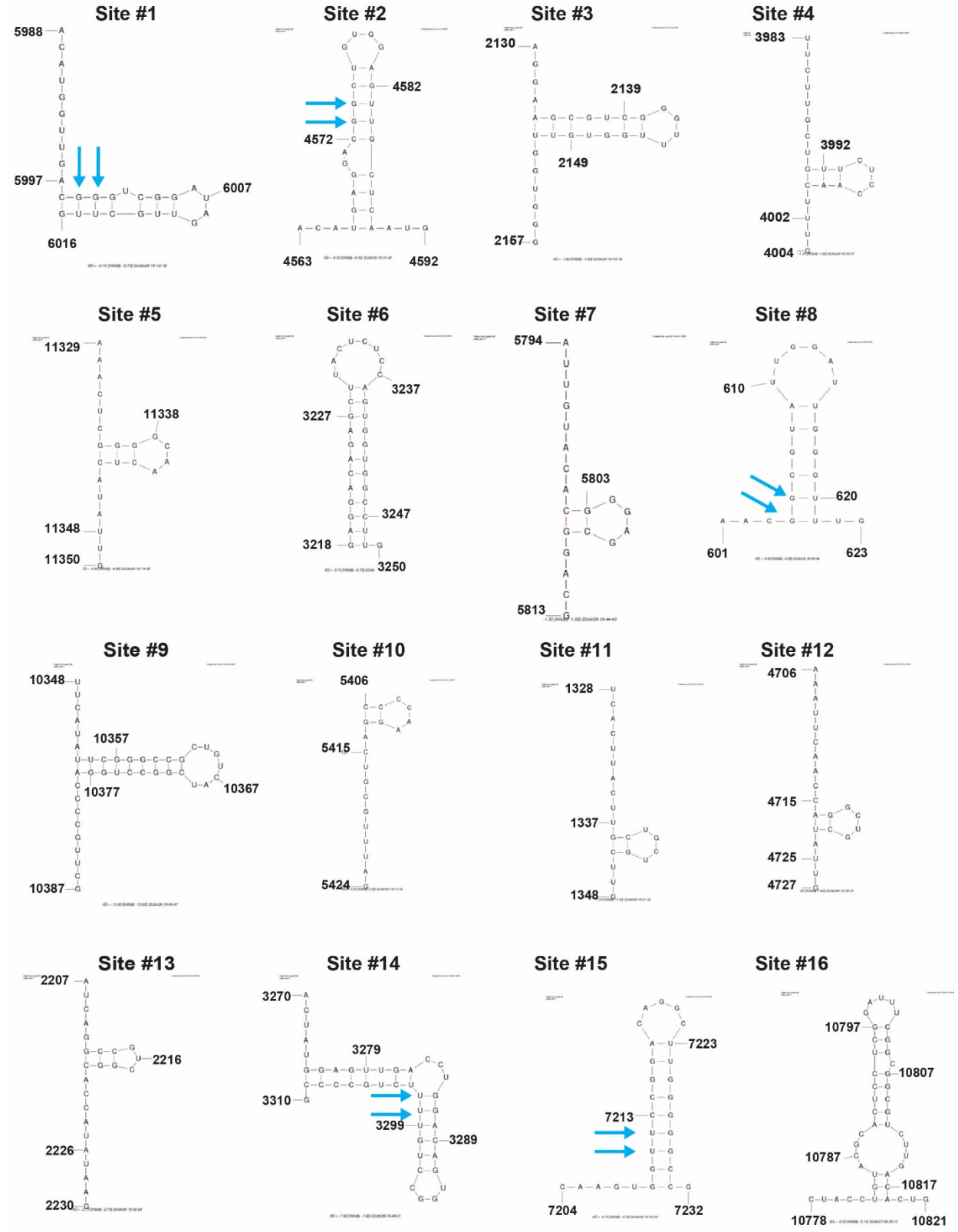
## Cp's top 58 sites. Best Motif Found:

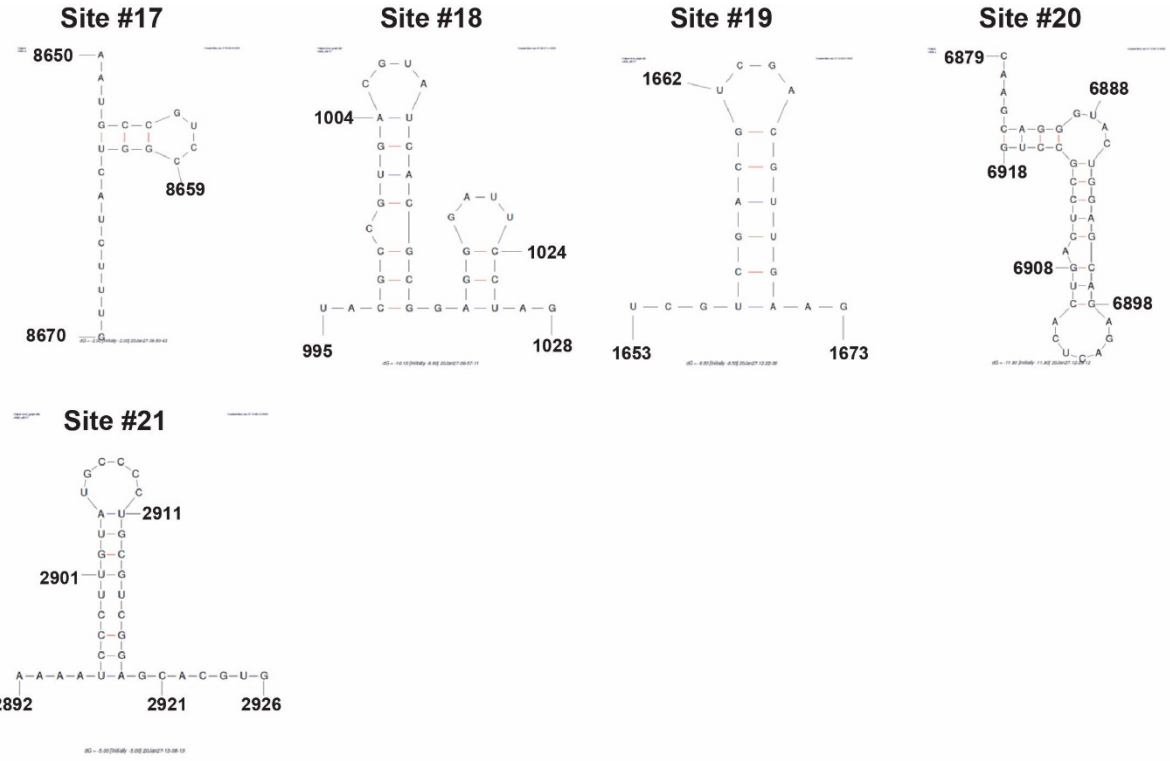
| NAME | START | SITES | END | STRAND | MARGINAL SCORE |
|---|---|---|---|---|---|
| 1 | 4 | t g g . t t g . a c g g g . t . c g | 17 | + | 3.01 |
| 2 | 10 | c g g . c t g . t g g a g . t . t g | 23 | + | 12.2 |
| 3 | 4 | a a g . c g . . t c g g g t t . t g | 17 | + | 5.58 |
| 4 | 8 | t g t t c t . . c c a a c t t . t g | 22 | + | 6.36 |
| 6 | 19 | c a g . t g g . t g g c c . t . t g | 32 | + | 5.38 |
| 7 | 3 | t g t . a c . . a c g g g a g . c g | 16 | + | 1.21 |
| 8 | 11 | t g g . a t . . t . g g g t t . t g | 23 | + | 4.23 |
| 9 | 23 | c g g . c c . . t g g a c . c c c g | 36 | + | 7.41 |
| 11 | 5 | t t a . c t . . t g c t g . c . t g | 17 | + | 1.27 |
| 13 | 4 | a g g . c c g . t c g g c a c . c a | 18 | + | 3.67 |
| 14 | 11 | t g a . c c . . t g g a c a g . t g | 24 | + | 10.4 |
| 15 | 16 | a g g . c t . . t g g g g g c . c g | 29 | + | 5.21 |
| 16 | 32 | c g t . c t . . t . g a c a c . t g | 44 | + | 8.73 |
| 17 | 6 | c g t c c g g . t c a t c t t . t g | 21 | + | 4.56 |
| 18 | 3 | c g c . c g . . t . g a c . g . t a | 14 | + | 1.13 |
| 19 | 5 | c g a . c g . . t c g a c g t . t g | 18 | + | 11.2 |
| 20 | 29 | t g a . c t . . c . c g c . c . t g | 40 | + | 2.47 |
| 21 | 22 | c g t . c g g . a g c a c . g . t g | 35 | + | 7.75 |
| 22 | 7 | t g a . c t . . t g g a c . t . g g | 19 | + | 4.73 |
| 24 | 15 | t g c t c t g . t a c a c . t . c g | 29 | + | 5.01 |
| 25 | 5 | c g a . t t g . t g g c c g t . c g | 19 | + | 5.45 |
| 26 | 17 | c g t t c g . . t c a t c . c . c g | 30 | + | 2.89 |
| 28 | 24 | t g t . c c . . t g g a c a c . t g | 37 | + | 11.5 |
| 31 | 11 | c g t . c g t . t c g a g . g . t g | 24 | + | 5.00 |
| 32 | 17 | c g g . c t g . c . g c c t g . t g | 30 | + | 6.40 |
| 34 | 6 | t g a . c g . . t a c a c . c c t g | 19 | + | 5.01 |
| 35 | 5 | t g a . c t g . t . a t c t t . c g | 18 | + | 3.91 |
| 36 | 9 | t t g . c g . . c c c g g a t . t g | 22 | + | 2.69 |
| 38 | 7 | c g g . c c g . a a g a c t t . t g | 21 | + | 9.35 |
| 39 | 2 | a g t t c c g . t g g a c g c . t g | 17 | + | 9.65 |
| 40 | 23 | c g c . c c g . t g g a a . c . t g | 36 | + | 4.94 |
| 41 | 7 | a a c . a t . . a g g a c t g . t g | 20 | + | 0.690 |
| 42 | 41 | t g g . a g g . c c g a g . c . t g | 54 | + | 7.83 |
| 43 | 17 | c g c . c g . . t a c g g a t c t g | 31 | + | 3.01 |
| 45 | 8 | c g t t c g g . c . g a c t t . t g | 22 | + | 8.80 |
| 46 | 6 | t a c . c t g c t g g a c . t . t g | 20 | + | 5.57 |
| 48 | 1 | t a t . c c g . c g g t c g t . c g | 15 | + | 5.54 |
| 52 | 9 | t t a . c g g . c c g c c g g . t g | 23 | + | 4.75 |
| 53 | 15 | t g g . a c . . t c g g c a g . t g | 28 | + | 6.94 |
| 54 | 4 | t g t . c g g . t g g t c a c c t g | 19 | + | 11.1 |
| 55 | 2 | c a a . c g . . c g g a c . c . t g | 14 | + | 6.96 |
| 57 | 41 | t g g . a g g . c c g a g . c . t g | 54 | + | 7.83 |
| 58 | 10 | c a g . c g g . a . g c c a t . t g | 23 | + | 4.82 |

Score: **145.701**

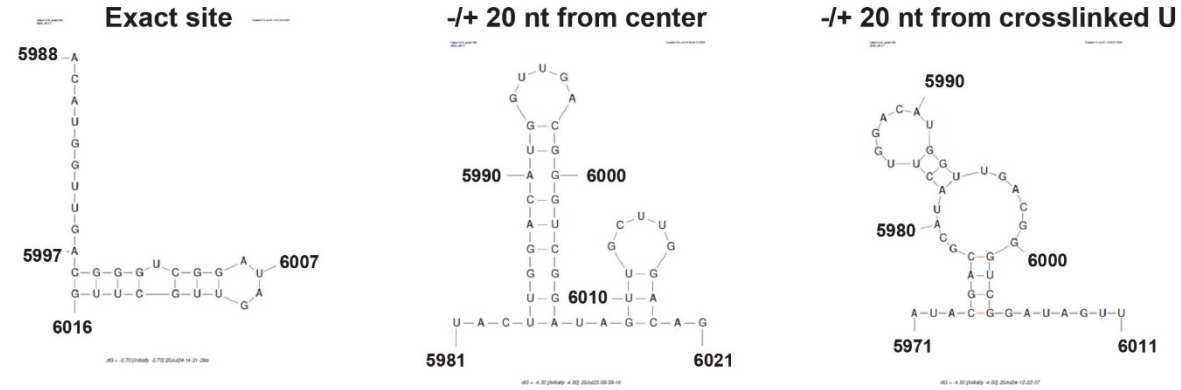

bits
2
1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
GLAM2 (no SSC) 20.06.19 13:30

**b mFold secondary structure predictions of top 21 binding sites.**

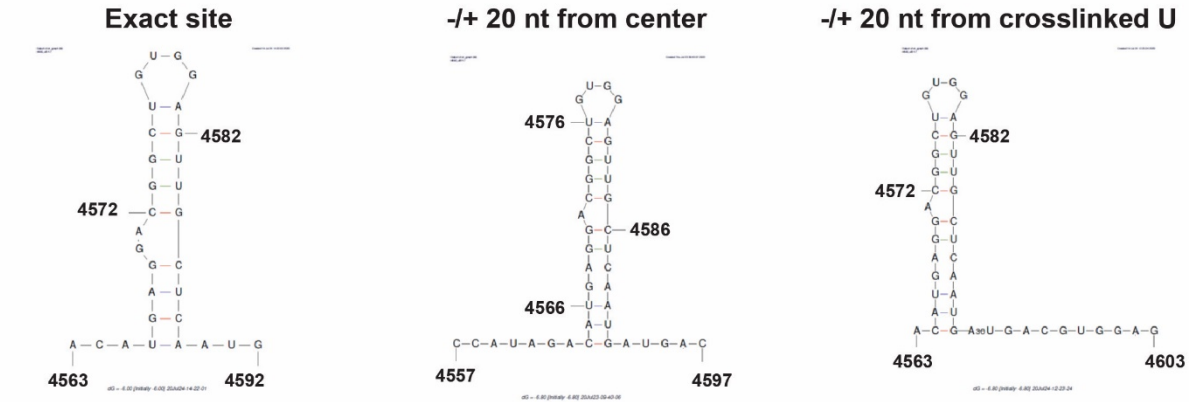




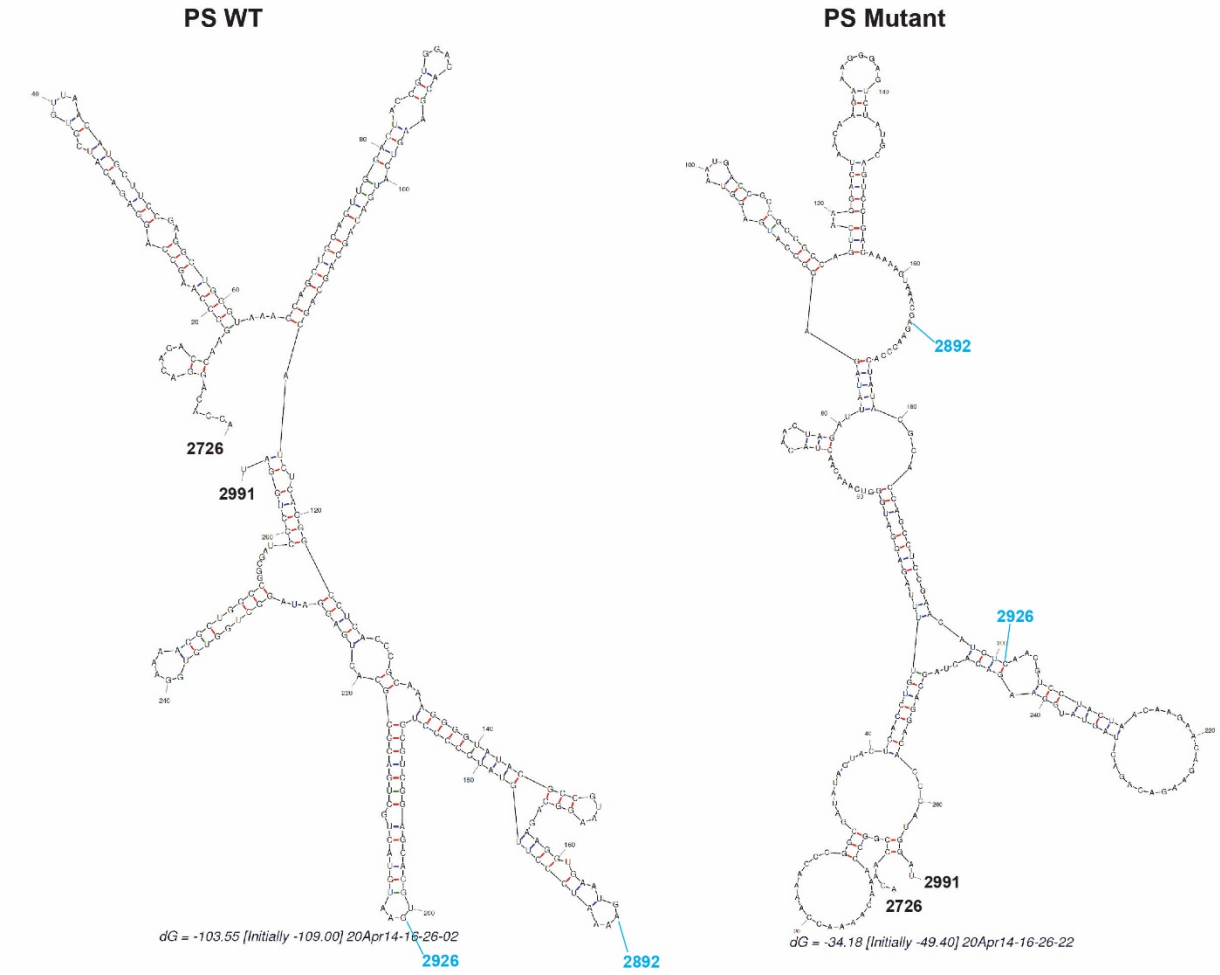

**c Comparison of different mFold secondary structure predictions for site #1.**




**d Comparison of different mFold secondary structure predictions for site #2.**



**e mFold secondary structure predictions of SFV PS WT vs. Mutant.**




**Supplementary Figure 3. Sequence and structural motif analyses for Cp's top binding sites. Related to Figures 2 and 6.** (a) GLAM2 was used with its default settings to discover potential gapped sequence motifs within Cp's top 21 and top 58 binding sites. (b) RNA secondary structure predictions for Cp's top 21 binding sites using mFold under default settings. Input sequence was the exact sequence identified as the binding site. Genome nucleotide positions are labeled. Blue arrows indicate tandem G:U wobble base pairs present in sites #1, 2, 8, 14, and 15. (c) mFold secondary structure predictions for site #1 RNA using different input sequences: the exact sequence identified as the binding site (left), -/+20 nt from the center nucleotide within the binding site (middle), and -/+20 nt from the U that crosslinked most frequently (right). (d) As in (c) but for site #2 RNA. (e) RNA secondary structure prediction for the WT PS sequence (nt:2726-2991) (left) compared to the mutated sequence of the Full PS mutant (right). Blue numbers mark the Cp-PS binding site's 5' and 3' ends.

**a**

```
Overall sequence identity between SFV and CHIKV_AF15561 is ~67%
Site 1
27/29 93%
SFV             ACATGGTTGACGGGTCGGATAGTTGCTTG
AF15561         ACATGGTGGACGGGTCGGAGAGTTGCTTG
                ******* ********** *********

Site 2
21/30 70%
SFV             ACATGAGGACGGCTGTGGAGTTGCTCAATG
AF15561         AGATGCGGACCCAAGTGGAGCTGCTGGATG
                * *** ****    ******.**** .***

Site 3
20/28 71%
SFV             AGGAAGCGTCGGGTTTGGTGTTGGTGGG
AF15561         AAGAAGCTGCAGGACTGGTACTGGTGGG
                *.*****  *.** .****..*******

Site 4
15/22 68%
SFV             TTCTTGCTGTTCTCCAACTTTG
AF15561         TTTTTCCTATTTAGCAATTTTG
                **.** **.**.  ***.****

Site 5
12/22 54.5%
SFV             AAAC-TCGGGGCAACTCATATTG
AF15561         GGACGTAGGAGATGTTCAAAGTG
                ..** * **.*  ..*** * **

Site 6
23/32 72%
SFV             AGGACAGAGCTTACTCTCCAGTGGTGGCCTTG
AF15561         AAGACAAAGCATACTCACCCGAAGTAGCCCTG
                *.****.*** ***** ** * .**.***.**

Site 7
11/20 55%
SFV             ATTGTACACGGGAGCGGA-----CG
AF15561         ACTGTATTTAATGGCAGAGACCCCG
                *.****. ... .**.**     **

Site 8
17/23 74%
SFV             AACGGCGTATTGGATTGGGTTTG
AF15561         AGTGGCGTACTGGGTAGGGTTCG
                *..******.***.* *****.*

Site 9
30/40 75%
SFV             TTCATATTCGGGCCGCTGTCATCGGCCTGGACCCCGTTCG
AF15561         TTCATTGTGGGGCCAATGTCTTCAGCCTGGACACCTTTTG
                *****  * *****..**** **.******** ** **.*
```

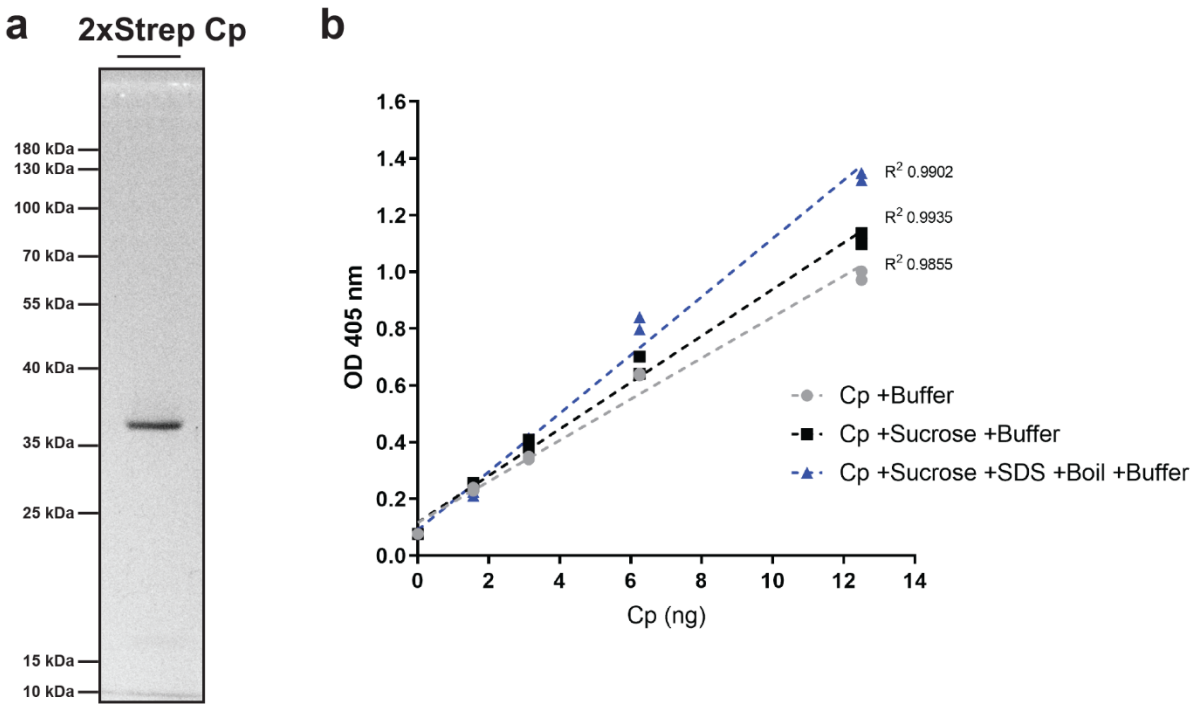
```
Site 10
11/19 58%
SFV             CC-----------------------CCAAGGACT-----------------------
AF15561         ACACAGCTATTTCCCTTCAGGCACCGCCAAGTACCACCATGGAACTGAGCCATCCACCG
                 *                         ***** **.
SFV             ------------GCGTTTAG
AF15561         ATCTCCTTCGGAGCACCAAG
                           * ... **

Site 11
15/21 71%
SFV             TCACTTACTTGCTGCTGCTTG
AF15561         ACACTAACCTGCTGCTGTCTA
                 **** **.********..*.

Site 12
12/22 55%
SFV             AAATTCAACCAGGCTGCTATTG
af15561         CGTTTTCACCAGACGGCAGTGG
                 . **. *****.* ** .* *

Site 13
16/24 67%
SFV             ATCAG-GCCGTCGGCACCATATAAG
AF15561         ATTCGCCCCGCCTGC-CCATACAAA
                **. *  ***.* ** *****.**.

Site 14
30/42 71%
SFV             ACTATGGAGTTGACCTGGACAGTGGCCTGTTTTCTGCCCCG
AF15561         TGTATGGGGTGGATCTAGACAGTGGGCTATTCTCTAAACCG
                  *****.** **.**.******** **.**.***.  ***

Site 15
16/29 55%
SFV             CAAGTGGTTCCGGACAGGCTTGGGGGCCGA
AF15561         CAGATGGCAACGAACAGGGCTAATAGATGA
                **..***.  **.***** .*.. .* .**

Site 16
28/44 66%
SFV             CTACCTGTACGCACTCCTCGGATTTCGGCGGCGTCTTGACACTG
AF15561         CAGCCTGCACCCACTCCTCAGACTTTGGGGGCGTAGCCATCATT
                * .****.** ********.**.**.** *****    *.  *

Site 17
11/21 52%
SFV             AATGCCGTCCGGTCATCTTTG
AF15561         AGGGCCG---GGCTATTTGT-
                *. ****   **..**.* *

Site 18
24/34 71%
SFV             TACGCCGTGACGTATCACGCGGAGGGATTCCTAG
AF15561         TATGCGGTAACCCACCACGCAGACGGATTCTTGA
                **.** **.** .*.*****.** ******.*..
```

```
Site 19
17/21 81%
SFV             TCGACGTCGACGTTGAAG
AF15561         TCGAAATCGACGTGGAAC
                **** .******* ***

Site 20
26/40 65%
SFV             CAAGCAGGGTACTGGAGCAGAGACTCACTGACTCCGCCTG
AF15561         CTAGCCGGGTGTTGGAAGATCGTCTGACAAAATCCGCATG
                * *** ****..****. *  * ** ** .* ***** **

Packaging Sequence Binding Site
Site 21
24/35 69%
SFV             AAAATCCCTTGTATGCCCCTGCGTCGGAGCACGTG
AF15561         AAAACCCACTCTATGCATCAACATCAGAGCACGTC
                ****.** .* ***** .* .*.**.********
```

**b**

```
Full Packaging Sequence
197/266 74%

SFV             ACCACAGGACAGACCAAGCCCAAGCCAGGAGACATCGTGTTAACATGCTTCCGAGGCTGGG
AF15561         ACTACAGGCTCAACGAAACCTGACCCTGGAGACCTCGTGTTAACGTGCTTCAGAGGGTGGG
                **.***** . .** **.**..* ** ****** **********.****** **** ****

SFV             TAAAGCAGCTGCAGTTGGACTACCGTGGACACGAAGTCATGACAGCAGCAGCATCTCAGGG
AF15561         TTAAACAACTGCAAATTGACTATCGTGGACACGAGGTCATGACAGCAGCCGCATCCCAAGG
                * **.**.*****. * *****.**********.************** *****.**.**

SFV             CCTCACCCGCAAAGGGGTATACGCCGTAAGGCAGAAGGTGAATGAAAATCCCTTGTATGCC
AF15561         GTTAACTAGAAAAGGAGTTTACGCAGTTAGGCAAAAAGTTAACGAAAACCCACTCTATGCA
                 .* **. * *****.** ***** ** *****.**.** **.*****.** .* *****

SFV             CCTGCGTCGGAGCACGTGAATGTACTGCTGACGCGCACTGAGGATAGGCTGGTGTGGAAAA
AF15561         TCAACATCAGAGCACGTCAACGTACTCCTAACGCGTACGGAAGGTAAACTGGTATGGAAGA
                .* .*.**.******** **.***** **.*****.** **.*.**..*****.*****.*

SFV             CGCTGGCCGGCGATCCCTGGAT
AF15561         CACTCTCTGGTGACCCGTGGAT
                *.**  *.**.**.** *****
```

**Supplementary Figure 4. Sequence alignments between SFV and CHIKV. Related to Figures 2 and 4.** Multiple sequence alignment (MUSCLE)[1] was used to align the SFV and CHIKV genome sequences. (a) The top 21 SFV Cp binding sites aligned with the corresponding CHIKV sequences. Sequences highlighted in yellow are hyperconserved between SFV and CHIKV compared to the overall sequence identity (~67%) between SFV and CHIKV genomes. (b) The SFV full PS (nt:2726-2991) aligned with the corresponding CHIKV sequence. The sequence is highlighted in yellow because it is hyperconserved.




**Supplementary Figure 5. Purified Cp and Cp ELISA. Related to Figure 6.** (a) 300 ng of purified 2xStrep-tagged Cp was subjected to SDS-PAGE and Coomassie staining. The gel is representative of n=5 independent purifications. Source data are provided as a Source Data file. (b) ELISA signals corresponding to different Cp treatment conditions. A 2-fold dilution series of Cp was incubated either 30% sucrose, or 30% sucrose and 1% SDS plus boiling for 5 min before diluting in binding buffer. Individual absorbance readings from technical duplicates are plotted. Lines represent linear regression. ELISA signals between the various treatments were comparable and linear ($R^2$>0.98) within the same protein range (~1-12.5 ng). The graph is representative of n=2 independent experiments. Source data are provided as a Source Data file.

## Supplementary References

1. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-w641 (2019).