Genetic Encoding of Fluoro-L-Tryptophans for Site-Specific Detection of Conformational Heterogeneity in Proteins by NMR spectroscopy

Haocheng Qianzhu,[1] Elwy H. Abdelkader,[2] Gottfried Otting,[2] Thomas Huber[1]

[1] Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[2] ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Table of contents:

Methods

a) Materials

No unexpected or unusually high safety hazards were encountered. 4-Fluoro-DL-tryptophan, 5-fluoro-DL-tryptophan, 6-fluoro-DL-tryptophan, and 7-fluoro-DL-tryptophan were purchased from Ambeed Inc., USA (catalogue number: A901033, A114530, A339425, and A158059) and were used assuming L-amino acid (4F-Trp, 5F-Trp, 6F-Trp, or 7F-Trp) content of 50%. 4-Fluoroindole, 5-fluoroindole, 6-fluoroindole, and 7-fluoroindole were purchased from AK Scientific Inc., USA (catalogue code: J92053, F683, E942, and S276).

b) Screening of functional G1PylRS enzymes recognizing 4F-Trp, 5F-Trp, and 6F-Trp

To carry out the selection, the tRNA synthetase plasmid pBK-G1RS carrying the *G1*PylRS library was transformed into *E. coli* DH10B cells harboring the reporter plasmid pBAD-H6RFP.[1] Following recovery from transformation, the culture was directly inoculated into a flask with 25 mL LB medium containing 100 mg/L carbenicillin and 50 mg/L kanamycin, supplied with 0.4% L-arabinose and 1 mM FTrp, which served as the sample for the first round of positive selection (**1P+**). Overnight expression at 37 ^{o}C led to a readily detectable level of RFP expression. Cells were resuspended in 5 mL PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) after harvesting. A 100-fold dilution yielded a concentration suitable for cell sorting by FACS on a BD FACSAria II or FACSAria Fusion cell sorter (BD Biosciences, USA; Figures S1).

Cells with high RFP levels were collected from the **1P+** sample (0.3% for 4F-Trp, 0.8% for 5F-Trp, and 1.1% for 6F-Trp as indicated by violet shades in Figure S1) and subjected to a following round of negative selection. Without the addition of F-Trp, the cells were regrown as sample **2N–**, from where cells with low RFP expression levels (47.7% for 4F-Trp, 38.3% for 5F-Trp, and 43.4% for 6F-Trp) were collected. These cells were aliquoted to inoculate media with positive (**3P+**) and negative (**3P–**) conditions. The cell populations with high RFP fluorescence in the **3P**+ samples in all three selection attempts were all obviously higher than the numbers in the **3P–** samples, indicating the successful accumulation of active *G1*PylRS variants specific for each F-Trps. The top 5.1% RFP fluorescent cells of each **3P+** sample were collected. Individually for the 4F-Trp, 5F-Trp, and 6-Trp selection experiments, an aliquot of 2,000 cells were allowed to recover on LB agar plates containing 100 mg/L carbenicillin and 50 mg/L kanamycin, and isolated colonies were analyzed using 96-well plates. Cells containing 60 enzyme candidates were inoculated into both positive (with 1 mM F-Trp) and negative (without non-canonical amino acid (ncAA)) growth conditions. The intensity of red fluorescence was measured as an indicator of the expression level of amber-interrupted RFP after expression overnight, using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland) and normalized by the OD_{600} of the cell culture. Sequences of four candidates from the 4F-Trp selection, six candidates from the 5F-Trp selection, and five candidates from the 6F-Trp selection were identified as novel *G1*PylRS mutants incorporating F-Trps. The respective amino acid mutation sets are listed in Table S1.

c) Substrate cross-specificity analysis of selected G1PylRS mutants

The tRNA synthetase plasmids pBK-G1RS carrying each of the 20 selected functional FTrpRS were individually co-transformed with the reporter plasmid pBAD-H6RFP into *E. coli* DH10B cells. Cells were grown in 150 μL LB medium containing 100 mg/L carbenicillin and 50 mg/L kanamycin on 96-well plates at 37 °C with 400 rpm shaking. 5 μL of each overnight culture were used to inoculate 150 μL LB medium grown on 96-well plates, supplemented with the same antibiotic, along with 0.4% L-arabinose under five different conditions of ncAA supplementation: without any ncAA; with 1 mM of 4F-Trp, 5F-Trp, 6F-Trp, or 7F-Trp. Each culture was prepared in triplicate. The plates were shaken at 400 rpm for 12 h at 37 °C and for 24 h at 25 °C. The average of the triplicates was calculated according to the measured intensity of red fluorescence and normalized by the OD_{600} of the cell culture using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland). Based on the cross-specific reactivity, F4W27 was subsequently used to incorporate 4F-Trp and 5F-Trp, and F7W36 was used to incorporate 6F-Trp and 7F-Trp.

d) *In vivo* protein expression and purification

The genes of F4W27 and F7W36 were cloned into a high-copy number pRSF plasmid to yield the new tRNA synthetase plasmids pRSF-G1F4W27 and pRSF-G1F7W36. The pCDF plasmids containing the amber codon interrupted gene of the protein of interest were co-transformed with the appropriate pRSF tRNA synthetase plasmid into *E. coli* B-95.ΔAΔ*fabR* cells. The transformed cells were grown at 37 °C in LB medium containing 25 mg/L kanamycin and 25 mg/L spectinomycin. 10 mL overnight culture was used to inoculate 1 L LB medium supplemented with 25 mg/L kanamycin and 25 mg/L spectinomycin. The cells were grown at 37 °C to an OD_{600} of 0.5, at which point either fluoroindole or F-Trp was added depending on the reporter protein. 1 mM fluoroindole was added to the culture for the expression of AncCDT-1, while 1 mM F-Trp and 4 mM tryptophan were supplied for producing the flaviviral

NS2B-NS3 proteases. The cells were grown at 37 °C for an additional 30 minutes. Subsequently, the temperature was reduced to 25 °C and protein expression was induced by the addition of 1 mM IPTG.

After expression for 16 h, the cells were harvested by centrifugation. Following resuspension in buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM imidazole), the cells were lysed using an Avestin Emulsiflex C5 system (Avestin, Canada) using two passes with a pressure of 10,000–15,000 psi. The cell lysates were centrifuged for 1 h at 30,000 *g*. The supernatant was loaded onto a 1 mL His GraviTrap column (Cytiva, USA). The column was washed with 20 column volumes buffer A and the protein was eluted with 5 column volumes buffer B (same as buffer A but with 500 mM imidazole). For the purified flaviviral NS2B-NS3 protease samples, the buffer was directly exchanged to the desired NMR buffer (20 mM MES, pH 6.5, 150 mM NaCl; 50 mM Tris-HCl, pH 7.5, 100 mM NaCl; 50 mM Tris-HCl, pH 8.5, 100 mM NaCl) using an Amicon ultrafiltration centrifugal tube (Merck Millipore, USA) with a molecular weight cut-off of 10 kDa. To prepare the AncCDT-1 samples free of any ligand molecules that may have bound to AncCDT-1 during protein expression, the purified AncCDT-1 mutants were denatured by adding urea to a final concentration of 8 M. After incubating the samples at 25 ^{o}C for 15 min, the AncCDT-1 samples were refolded by the step-wise removal of urea by dialysis against NMR buffer (50 mM HEPES, pH 7.5, 150 mM NaCl) containing 8, 4, and 2 M urea for 2 h each at 25 ^{o}C and finally against NMR buffer without urea for 16 h at 4 ^{o}C. Prior to NMR measurements, 10% D_2O and 0.1 mM trifluoroacetic acid (TFA) were added to all protein samples.

e) Intact protein mass spectrometric analysis

Intact protein analysis was performed on an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific, USA) connected to a Thermo Fisher Scientific UltiMate 3000 HPLC system equipped with ZORBAX 300SB-C3, 3.5 µm, 4.6 x 50 mm HPLC column (Agilent Technologies, USA). Approximately 50 pmol of sample was injected using a 500 µL/min linear gradient of solvent A (0.1% (v/v) formic acid in water) and solvent B (0.1% (v/v) formic acid in acetonitrile), ramping solvent B from 5% solvent B at the start to 80% after 12 min. Data were collected using an electrospray ionization (ESI) source in positive ion mode. Protein intact mass was determined by deconvolution using the program Xcalibur 3.0.63 (Thermo Fisher Scientific, USA).

f) NMR spectroscopy

^{19}F NMR spectra were recorded on a Bruker 600 MHz NMR spectrometer equipped with a 5 mm TCI cryoprobe with an inner coil tunable to ^{19}F. Samples were in 5 mm NMR tubes at 50–300 μM concentrations. Parameters used: 200 ms acquisition time, exponential window multiplication with 50 Hz line broadening prior to Fourier transformation. Each 1D NMR spectrum was recorded at 25 ^{o}C in about 10–60 minutes depending on the protein concentration. Chemical shifts were calibrated relative to internal TFA (-75.25 ppm).

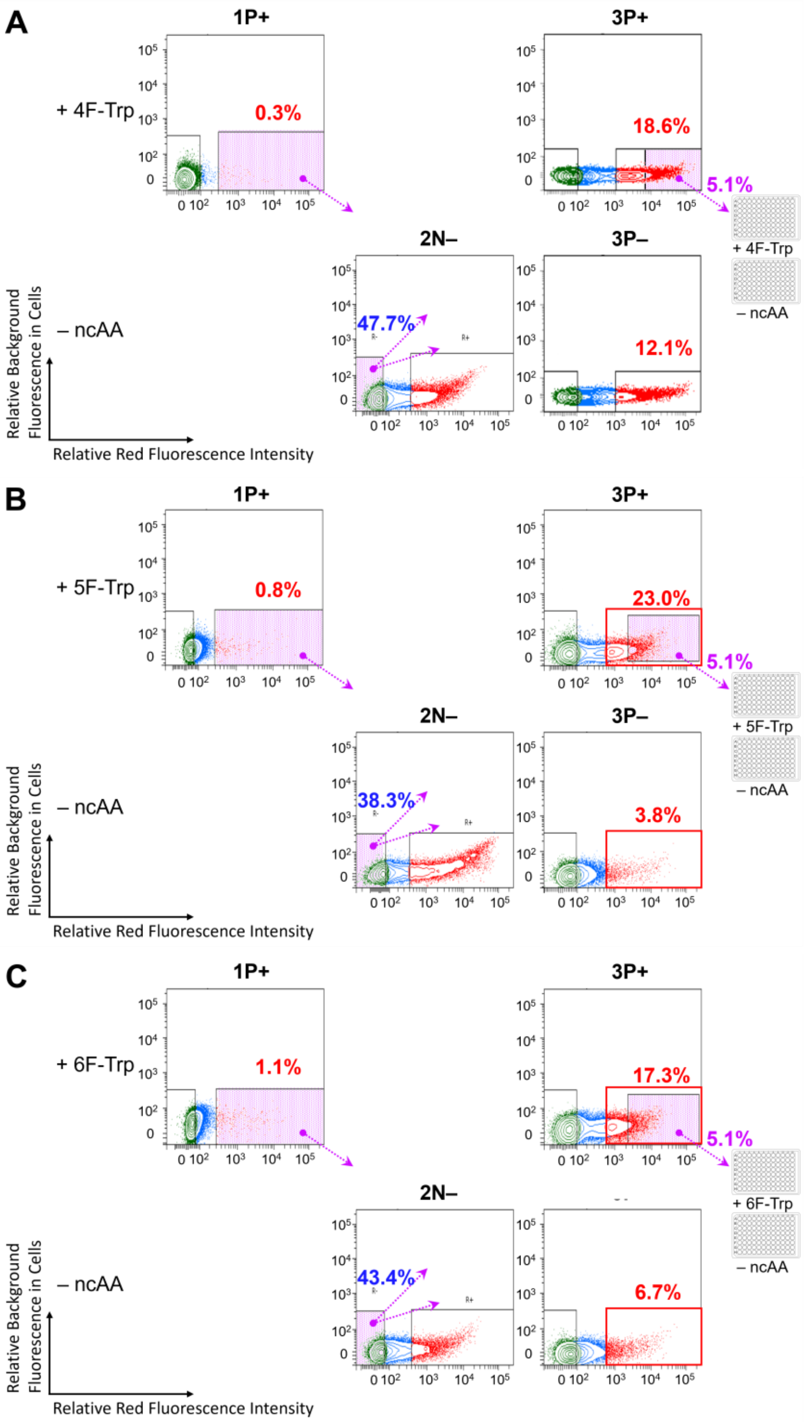

Figure S1. FACS experiments for selection of active and specific *G1*PylRS enzymes that recognize (A) 4F-Trp, (B) 5F-Trp, or (C) 6F-Trp. The horizontal axis indicates the relative intensity of red fluorescence. The vertical axis plots the level of background fluorescence in cells excited at 488 nm. **P** stands for positive selection rounds where cells were cultured in the presence of F-Trp; **N** indicates negative rounds without ncAA supplied. Violet shades and arrows identify, respectively, the cell populations collected and the following selection scheme, which the selected cells were subjected to following amplification by culturing.

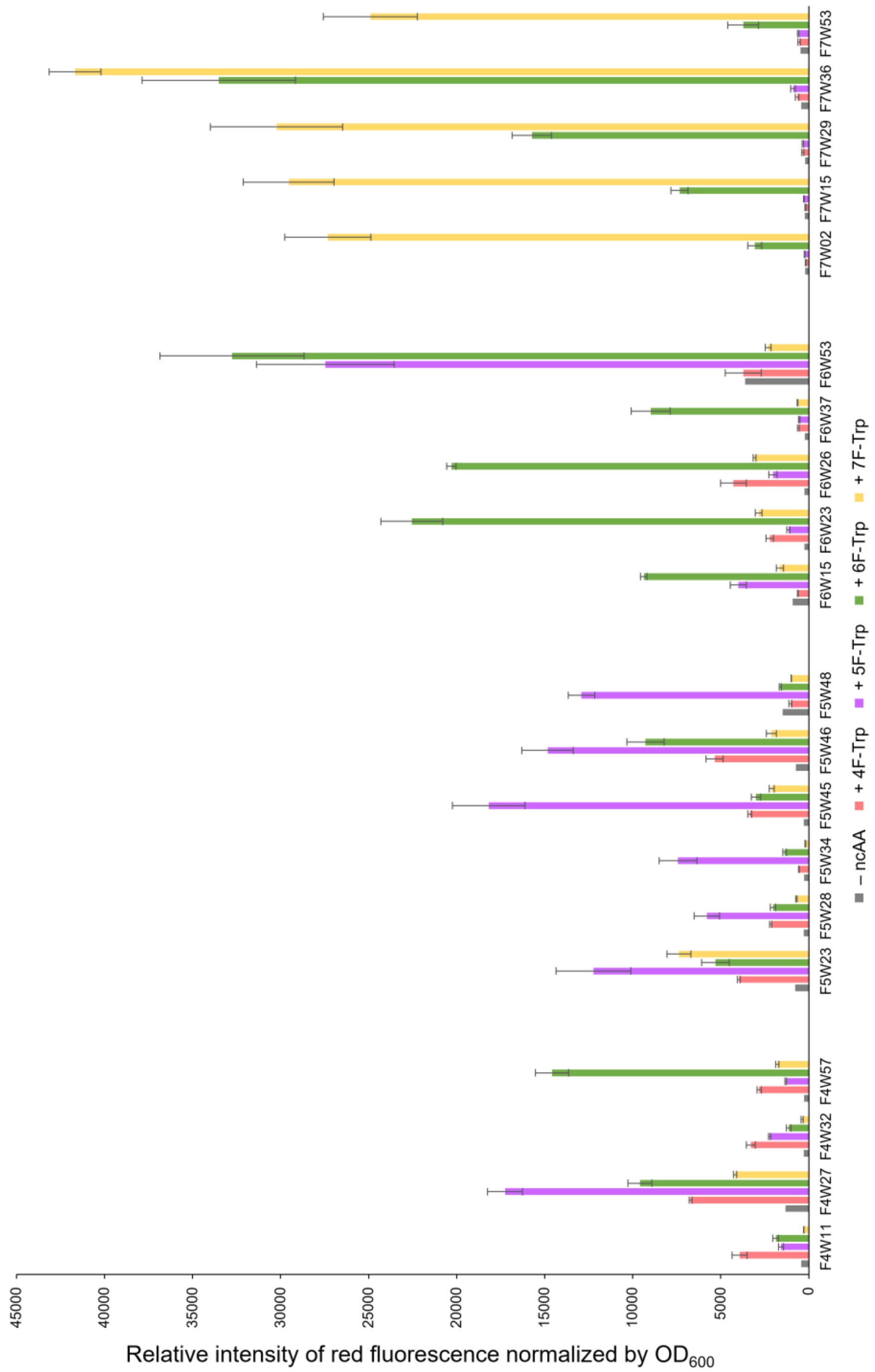

Figure S2. Levels of fluorescence of amber-interrupted RFP indicating the activity of 20 selected F-Trp tRNA synthetases. The names of the RS enzymes identify the F-Trp isomer they have been selected for (F4W–F7W) and the clone number. The fluorescence was measured in a plate reader, following bacterial cell growth in the presence of 1 mM 4F-Trp, 5F-Trp, 6F-Trp, or 7F-Trp, or without addition of any ncAA. Red fluorescence intensity was measured, normalized by OD, averaged across biological triplicates and standard deviation is indicated by the error bars. The figure indicates the degree to which the RS enzyme variants are active with F-Trp isomers they have not been selected for.

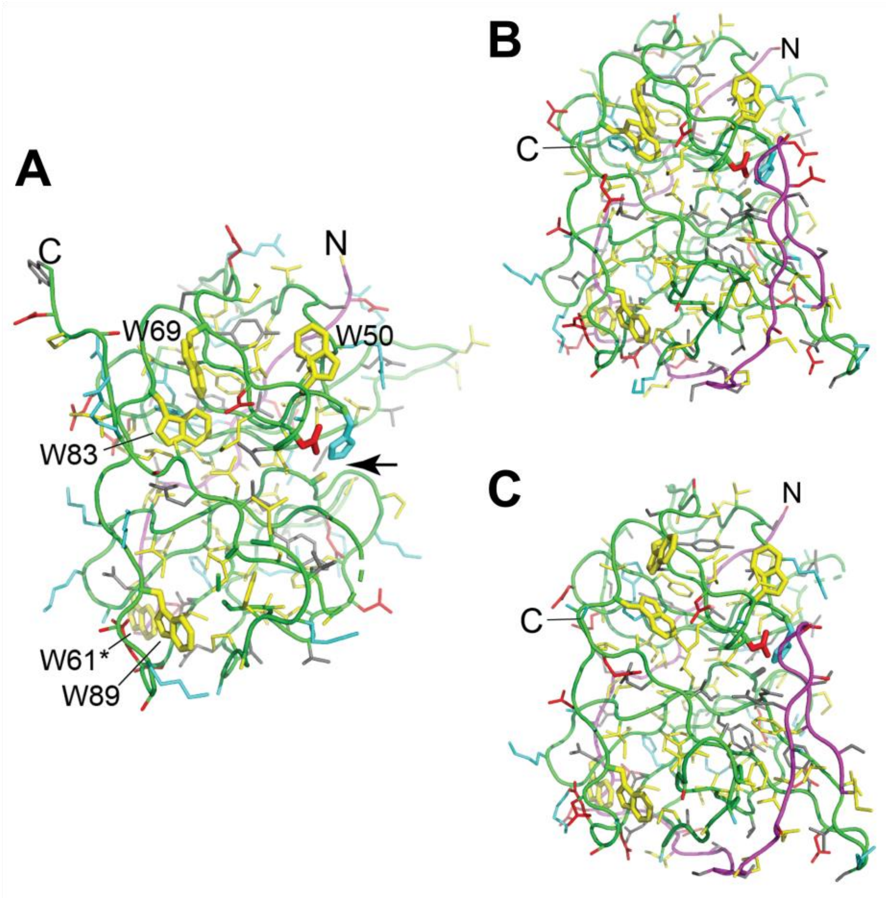

Figure S3. Crystal structures of different DENV4 NS2B-NS3 constructs.[2] The backbone of NS2B and NS3pro is shown in magenta and green, respectively. The N-terminus of NS2B and the C-terminus of NS3 are identified. The tryptophan side chains and the residues of the catalytic triad are shown in bold stick representations. The helicase domain of NS3 (not shown) is connected to the C-terminus of the protease domain via a flexible linker. In the orientation shown, the helicase domain is located "north" of NS3pro. (a) 5YVJ, showing the canonical orientation of the tryptophan side chains. The arrow points at the catalytic triad of NS3 (His51, Asp75, Ser135), except that this structure is of the inactive mutant S135A. In this construct, the C-terminal end of NS2B is covalently linked with the N-terminus of NS3 via a $Gly_4SerGly_4$ linker. (b) 5YWU, showing the side chain of Trp83 in a flipped orientation. The construct is the same as in (a) but bound to aprotinin (not shown). (c) 5YW1, showing the side chains of Trp69 and Trp83 in flipped orientations. In this construct, the linker between NS2B and NS3 contains a recognition site for the protease, leading to an enzymatically digested unlinked construct. The structure was determined in complex with aprotinin (not shown).

```
                          61*
gZIKVp      MSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMAGGGGSGGGGSG
gWNVp       -MTDMWIERTADITWESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWAGGGGSGGGGGG
gMVEVp      MATDMWLERAADVSWEAGAAITGTSERLDVQLDDDGDFHLLNDPGVPWKGGGGSGGGGGG
gJEVp       MATDMWLDRAADISWEMEAAITGSSRRLDVKLDDDGDFHLIDDPGVPWKGGGGSGGGGGG
gDENV2p     -MADLELERAADVRWEEQAEISGSSPILSITISEDGSMSIKNEEEEQTLGGGGSGGGGAG
gYFVp       -MDGLELRKLGEVSWEEEAEISGSSARYDVTLSEQGEFKLLSEEKVPWDGGGGSGGGGGD
              5                                               50
gZIKVp      ALWDVPAPKEVKNGE-TTDGVYRVMTRGLLGSTQVGVGVMQEGVFHTMWHVTKGSALRSG
gWNVp       VLWDTPSPKEYKKGD-TTTGVYRIMTRGLLGSYQAGAGVMVEGVFHTLWHTTKGAALMSG
gMVEVp      VFWDTPSPKVYPKGD-TTPGVYRIMARGILGRYQAGVGVMHEGVFHTLWHTTRGAAIMSG
gJEVp       VFWDTPSPKPCLKGD-TTTGVYRIMARGVLGTYQAGVGVMYENVFHTLWHTTRGAAIMSG
gDENV2p     VLWDVPSPPPVGKAE-LEDGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHK
gYFVp       VLWDIPTPKIIEECEYLEDGIYGIFQSTFLGASQRGVGVAQGGVFHTMWHVTRGAFLVRN
                   69          83   89
gZIKVp      EGRLDPYWGDVKQDLVSYSGPWKLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDG
gWNVp       EGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMIVVEPGKNVKNVQTKPGVFKTPEG
gMVEVp      EGRLTPYWGNVKEDRVTYGGPWKLDQKWNGVDDVQMIVVEPGKPAINVQTKPGIFKTAHG
gJEVp       EGKLTPYWGSVKEDRISYGGPWRFDRKWNGTDDVQVIVVEPGKPAVNIQTKPGVFRTPFG
gDENV2p     GKRIEPSWADVKKDLISYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGLFKTNTG
gYFVp       GKKLVPSWASVKEDLVAYGGSWKLDGRWDGEEEVQLIAAAPGKNVVNVQTKPSLFKVKNG

gZIKVp      -DIGAVALDYPAGTSGSPILDKSGRVIGLYGNGVVIKNGSYVSAITQGRR
gWNVp       -EIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYISAIVQGER
gMVEVp      -EIGAVSLDYPIGTSGSPIVNSNGEIIGLYGNGVILGNGAYVSAIVQGER
gJEVp       -EVGAVSLDYPRGTSGSPILDSNGDIIGLYGNGVELGDGSYVSAIVQGDR
gDENV2p     -TIGAVSLDFSPGTSGSPIVDKKGKVVGLYGNGVVTRSGAYVSAIANTEK
gYFVp       GEIGAVALDYPSGTSGSPIVNRNGEVIGLYGNGILVGDNSFVSAISQTEV
```

Figure S4. Multiple sequence alignment of flaviviral NS2B-NS3 proteases highlighting the conserved tryptophan residues. The $Gly_4SerGly_4$ linker between NS2Bc and NS3p is shown in blue. Strictly conserved tryptophan residues in NS2B (W61*) and NS3p (W5, W50, W69, W83, W89) are colored red. The sequence numbers refer to the numbering in ZIKVp.

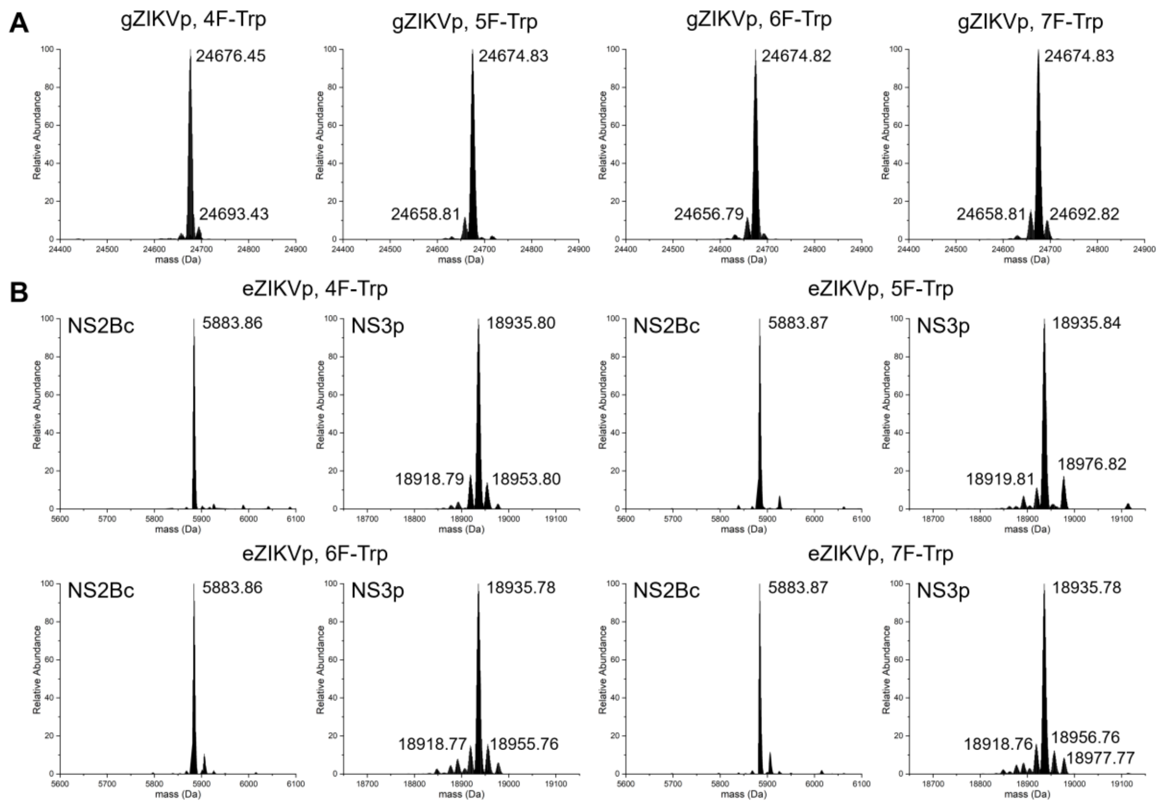

Figure S5. Intact protein mass spectrometric analysis of (A) gZIKVp and (B) eZIKVp with F-Trp isomers. The calculated mass of gZIKVp containing a single F-Trp residue (following loss of the N-terminal methionine) is 24676.44 Da. With the peptide sequence VKTGRK as the self-cleavable linker, the NS2B segment of the eZIKVp construct has an expected mass of 5884.46 Da, while the mass of singly fluorinated NS3p is 18936.33 Da. Minor species with –18 Da or +18 Da/+36 Da mass correspond to misincorporation of tryptophan at the amber position or replacement of Trp by F-Trp at native tryptophan codons.

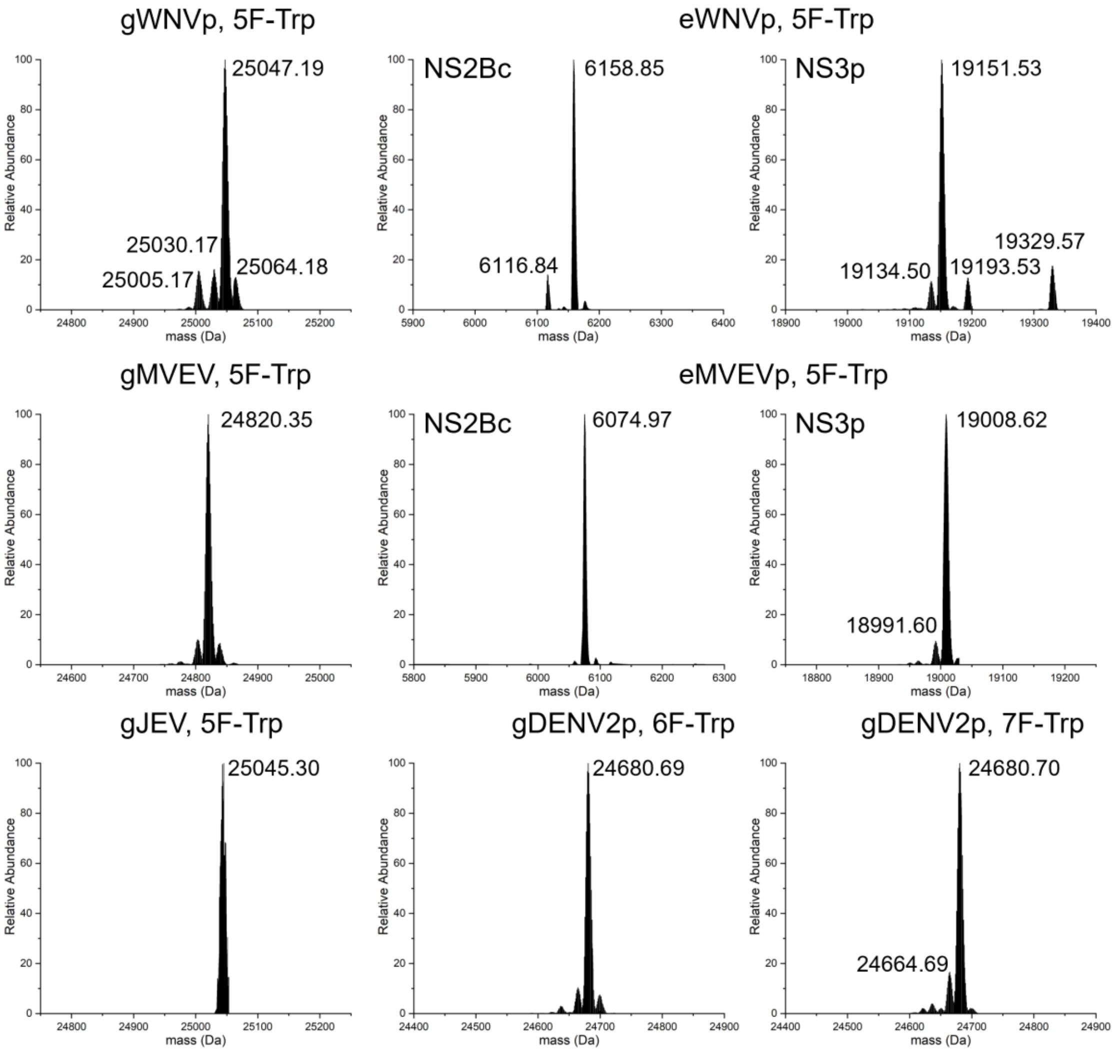

Figure S6. Intact protein mass spectrometric analysis of gWNVp, eWNVp, gMVEVp, eMVEVp, gJEVp, and gDENV2p with F-Trp isomers. The calculated mass of gWNVp / gMVEVp / gJEVp / gDENV2p containing a single F-Trp residue (following loss of the N-terminal methionine) is 25005.85 Da / 24820.76 Da / 25042.89 Da / 24681.68 Da, respectively. The expected mass of the NS2B segment of eWNVp / eMVEVp is 6117.65 Da / 6075.73 Da, and the singly fluorinated NS3p part of eWNVp / eMVEVp is 19152.65 Da / 19009.53 Da, respectively. N-terminal acetylation (+42 Da) occurred in gWNVp and eWNVp samples; gluconoylation (+178 Da) was also observed on the NS3p part of eWNVp. In all spectra, minor species with –18 Da or +18 Da mass correspond to misincorporation of tryptophan at the amber position or replacement with F-Trp into other tryptophan sites.

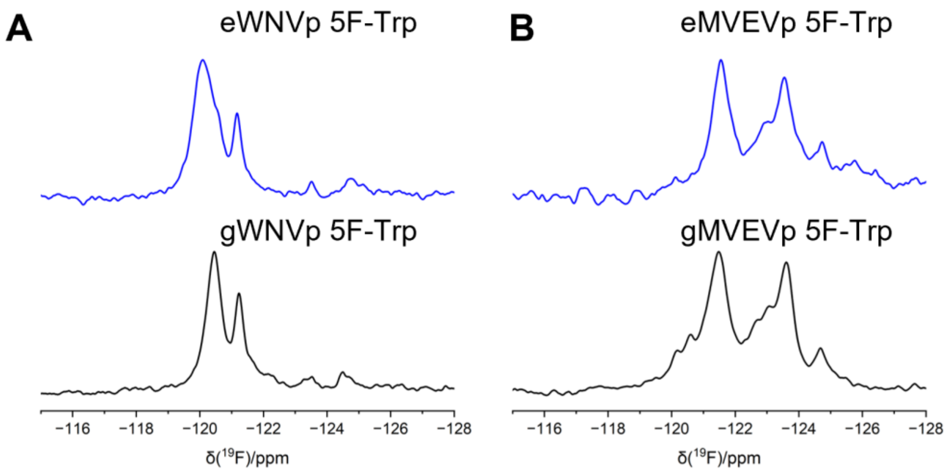

Figure S7. Linked and unlinked constructs of WNVp and MVEVp show the same heterogeneities. The linked constructs gWNVp and gMVEVp contained a G_4SG_4 linker connecting the C-terminal end of NS2B with the N-terminus of NS3. The unlinked constructs eWNVp and eMVEVp were designed with the requisite protease recognition site in the linker for self-cleavage. The samples were prepared with 5F-Trp replacing Trp83. ^{19}F NMR spectra were recorded on a 600 MHz NMR spectrometer at 25 °C in a buffer of 50 mM Tris-HCl, pH 7.5, 100 mM NaCl. (A) Spectra of eWNVp and gWNVp. (B) Spectra of eMVEVp and gMVEVp.

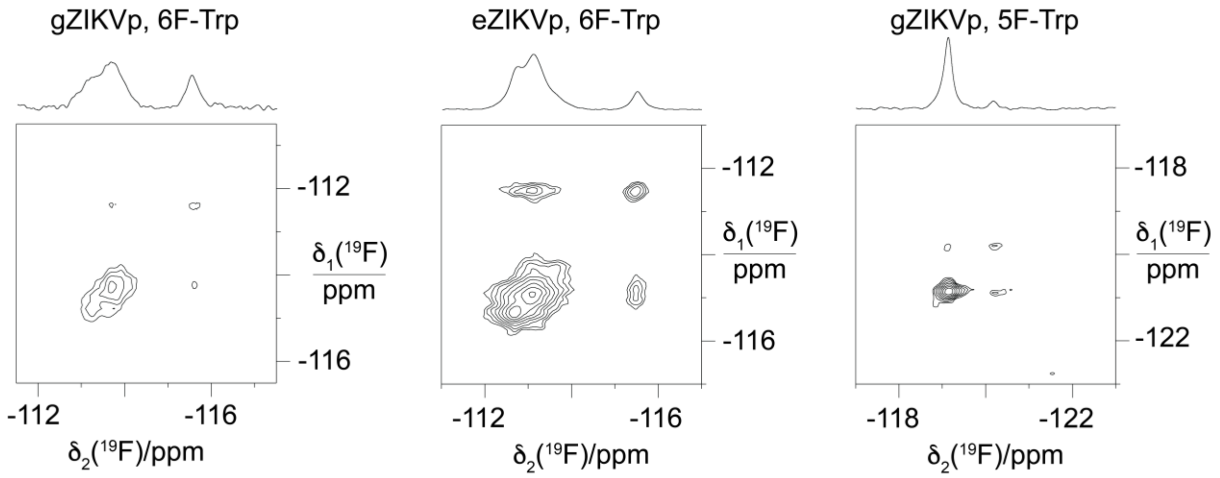

Figure S8. Exchange cross-peaks in the 2D ^{19}F–^{19}F NOESY spectra (50 ms mixing time) of different Zika virus NS2B-NS3 protease constructs containing F-Trp in position 83. The 1D ^{19}F NMR spectra are plotted at the top. The 2D spectra were recorded with 50 ms mixing time. The proteins were dissolved in buffer (20 mM MES, pH 6.5, 150 mM NaCl) and measured at 25 °C. Comparing the heights of the cross-peaks with the heights of the diagonal peaks yields the following estimates for the exchange lifetimes of the major species: 0.13 s for gZIKVp with 6F-Trp, 0.4 s for eZIKVp with 6F-Trp, and 0.5 s for gZIKVp with 5F-Trp. Integration of the 1D NMR spectra yields the following ratios of the minor conformation relative to the (overlapped total) signal of the major conformation: 29 % for gZIKVp with 6F-Trp, 13 % for eZIKVp with 6F-Trp, and 8 % for gZIKVp with 5F-Trp. The experiments show that the quantitative exchange rates and relative abundances of the different conformational species depend on the presence or absence of the $Gly_4SerGly_4$ linker between NS2B and NS3 as well as on the position of the fluorine atom in the indole side chain.

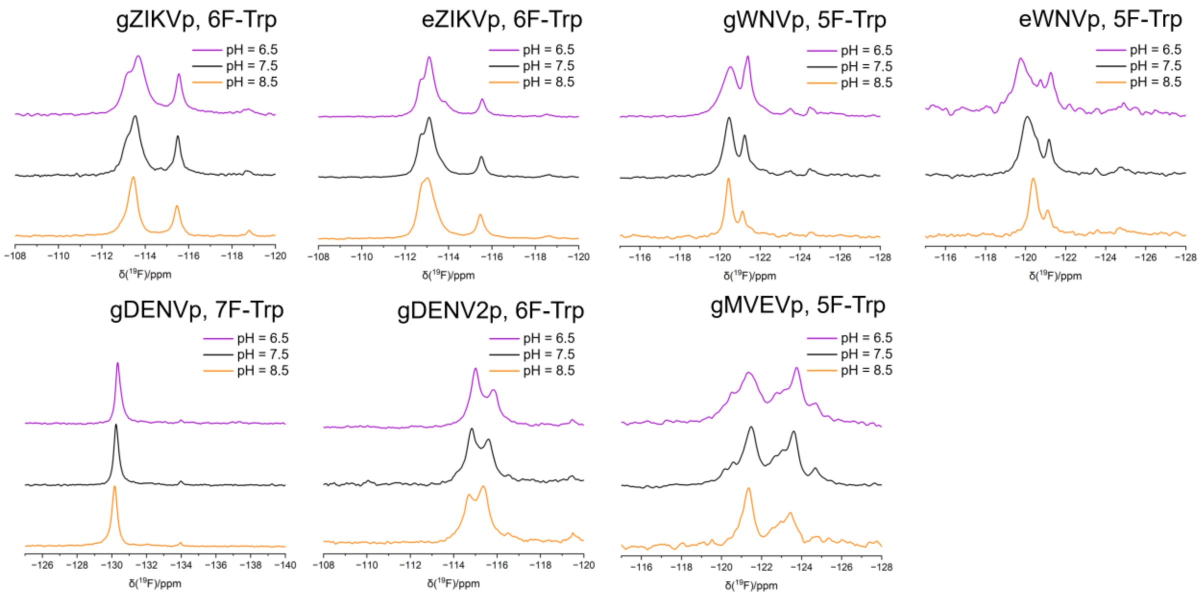

Figure S9. 1D ^{19}F NMR spectra showing the pH dependence of F-Trp signals, when Trp83 is replaced by different F-Trp isomers. The purple, black, and orange spectra are recorded at pH 6.5, 7.5, and 8.5 in corresponding buffers (20 mM MES, pH 6.5, 150 mM NaCl; 50 mM Tris-HCl, pH 7.5, 100 mM NaCl; 50 mM Tris-HCl, pH 8.5, 100 mM NaCl). Listed samples are with 5F-Trp in gZIKVp and eZIKVp, 5F-Trp in gWNVp and eWNVp, 6F-Trp and 7F-Trp in gDENV2p, and 5F-Trp in gMVEVp.

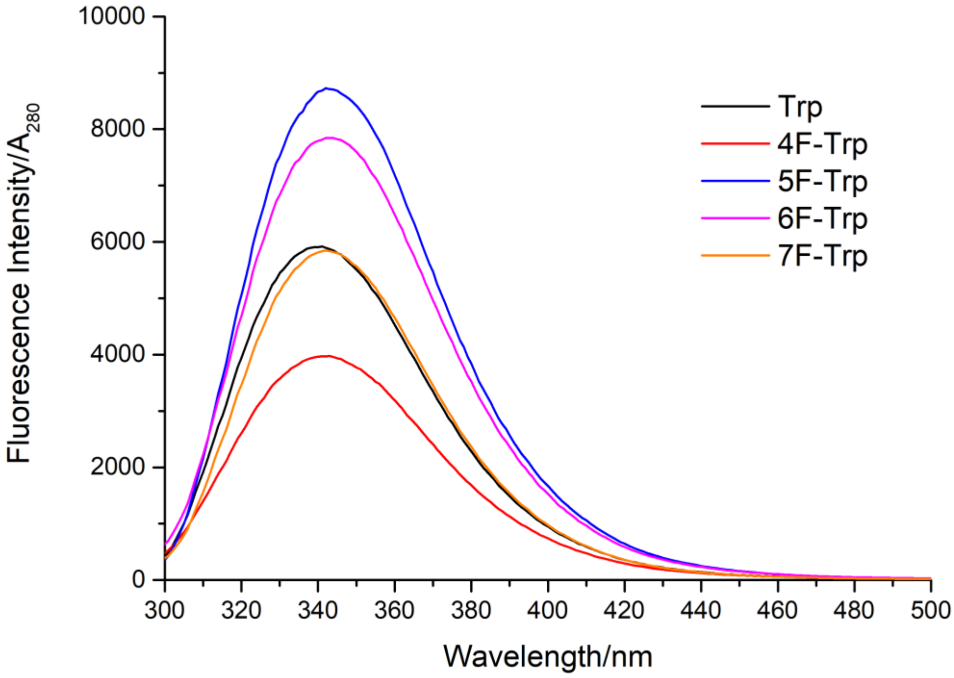

Figure S10. Fluorescence spectra (excitation at 280 nm) of wild-type (Trp) and F-Trp AncCDT-1 mutants. The wild-type protein contains four tryptophans of which only Trp71 is selectively replaced by F-Trp in the mutants. 4F-Trp can be used as a "knock-out" fluorescence analog of Trp due to its low fluorescence quantum yield. The fluorescence was measured for 2 mM protein in PBS buffer at room temperature and normalized by the UV absorbance value at 280 nm of each mutant.

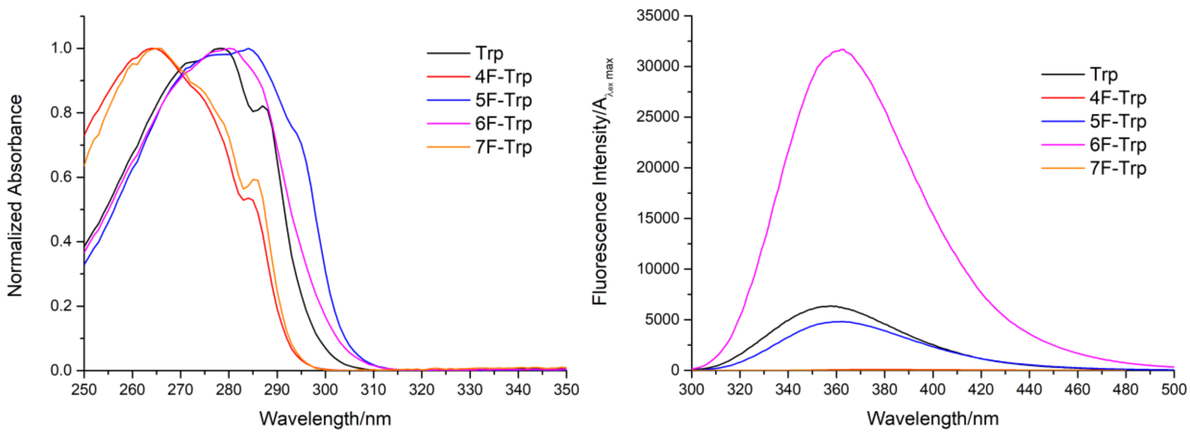

Figure S11. Absorption and fluorescence spectra of 2 mM solutions of pure F-Trp or Trp in PBS buffer at room temperature. Left panel: UV absorption spectra normalized by the value at maximum absorbance wavelength. Right panel: Fluorescence spectra normalized by the UV absorbance value at maximum excitation wavelength (260 nm for 4F-Trp and 7F-Trp; 280 nm for Trp, 5F-Trp, and 6F-Trp).

Supplementary tables

Table S1. Mutations found in *G1*PylRS variants selected for activity with 4F-Trp, 5F-Trp, 6F-Trp, or 7F-Trp.[a]

RS Variants	*Randomized Sites*						
*Mm*PylRS-wt	L305	Y306	N346	V348	Y384	V401	W417
*G1*PylRS-wt	L124	Y125	N165	V167	Y204	A221	W237
F4W11	F	Y	A	A	Y	N	W
F4W27[b]	L	Y	A	F	W	T	N
F4W32[c]	L	Y	A	A	W	N	W
F4W57	L	Y	A	A	Y	G	H
F5W23	L	Y	A	V	W	S	V
F5W28	L	W	A	A	Y	T	S
F5W34	L	Y	G	L	F	S	N
F5W45	I	F	G	A	W	A	H
F5W46	F	V	A	A	W	S	W
F5W48	L	Y	G	V	F	S	I
F6W15	S	V	V	L	W	C	W
F6W23	T	Y	A	G	W	G	W
F6W26	Q	F	S	A	W	G	W
F6W37	L	Y	A	S	Y	G	T
F6W53	Q	F	A	A	W	G	Y
F7W02	G	F	G	F	W	G	Y
F7W15	S	F	A	F	W	G	Y
F7W29	H	F	A	A	W	G	Y
F7W36[b]	Q	F	A	F	W	G	Y
F7W53	L	Y	A	F	W	G	Y

[a] The five mutants identified from selection for 7F-Trp were reported in our previous study.[1]

[b] The mutants *G1*-F4W27 and *G1*-F7W36 were used for all following applications.

[c] An extra mutation P171T was found in the mutant *G1*-F4W32.

Table S2. Crystal structures of flaviviral NS2B-NS3 proteases in the protein data bank.

Protein	NS2Bc conformation[a]	Tryptophan side chain conformations[b]	PDB code	Resolution (Å)	Reference
DENV4 with full-length NS3	absent	canonical	2VBC	3.15	[3]
DENV4 with full-length NS3	absent	canonical	2WHX 2VBC	2.2 3.15	[4]
DENV4 with full-length NS3	absent	W61* undefined	2WZQ	2.8	[4]
DENV4 with full-length NS3	open	canonical	5YVJ	2.5	[2]
DENV4 with full-length NS3	closed	canonical	5YVY	3.2	[2]
DENV4 with full-length NS3	closed	W50 and W89 flipped, W83 undefined	5YVV	3.1	[2]
DENV4 with full-length NS3	closed	W69 and W89 flipped, W83 undefined	5YVW	3.1	[2]
DENV4 with full-length NS3 + aprotinin	closed	W83 flipped	5YVU	2.5	[2]
DENV4 with full-length NS3 + aprotinin	closed	W69 and W83 flipped	5YW1	2.6	[2]
DENV4	half-open	canonical	7VMV	3.35	[5]
DENV3 + inhibitor	closed	canonical	3U1I	2.3	[6]
DENV3 + aprotinin	partly disordered	W50 flipped	3U1J	1.8	[6]
DENV2	open	canonical	2FOM	1.5	[7]
DENV2	open	canonical	4M9F 4M9I 4M9K 4M9M 4M9T	1.5 2.4 1.5 1.5 1.7	[8]
DENV2 + allosteric inhibitor	open	canonical	6MO0 6MO1 6MO2	2.7 3.0 2.8	[9]
DENV1	open	canonical	3LKW 3L6P	2.0 2.2	[10]
WNV + inhibitor	closed	canonical	2FP7	1.7	[7]
WNV + aprotinin	closed	canonical	2IJO	2.3	[11]
WNV	open	canonical	2GGV	1.8	[11]
WNV	closed	canonical	3E90	2.45	[12]

WNV + inhibitor	closed	canonical	2YOL	3.2	[13]
WNV + inhibitor	closed	W50 in two conformations	5IDK	1.5	[14]
WNV	open	canonical	8CO8	1.9	Fairhead et al., to be published
Zika + inhibitor	closed	W89 flipped in one of two conformations	5LC0	2.7	[15]
Zika + inhibitor	closed	canonical	5GJ4	1.8	[16]
Zika	closed	canonical	5GPI 5H4I	1.6 1.2	[17]
Zika	open	canonical	5GXJ	2.6	[18]
Zika	open	canonical	5T1V	3.1	[19]
Zika	open	canonical	5TFN 5TFO	3 2.5	Aleshin et al., to be published
Zika + inhibitor	closed	W50 flipped	5H6V	2.4	[20]
Zika + inhibitor	closed	W50 in two conformations	5YOF	1.5	[21]
Zika + inhibitor	closed	canonical	5YOD 5ZMQ 5ZMS	1.9 2.0 1.8	[22]
Zika + inhibitor	closed	W50 in two conformations	5ZOB	2.0	[22]
Zika + inhibitor	closed	W50 in two conformations	6JPW	1.95	[23]
Zika + inhibitor	closed	canonical	7DOC	1.9	[24]
Zika + inhibitor	closed	canonical	6KK2 6KK5	2.0 2.0	[25]
Zika + inhibitor	closed	W50 flipped	6KK3 6KK4 6KK6 6KPQ 6Y3B	2.05 1.74 1.74 2.6 1.6	[25]
Zika + inhibitor	closed	canonical	6L50 6L4Z	1.95 1.9	[26]
Zika + inhibitor	open	canonical	6UM3	2.5 Å	Aleshin et al., to be published
Zika + inhibitor	open	W50 flipped	7M1V	1.6 Å	Aleshin et al., to be published
Zika + inhibitor	closed	canonical	7O55 7OBV 7PFQ 7PFY 7PFZ	1.95 1.3 1.45 1.4 1.45	[27]

			7PG1 7VLI 7O2M 7ZNO	1.95 2.4 1.9 1.7	
Zika + inhibitor	closed	W50 flipped	7OC2 7VLG 7PGC 7VLH	1.7 1.7 1.55 2.6	[27]
Zika + inhibitor	closed	canonical	7VXX 7VXY	1.9 1.9	[28]
Zika + inhibitor	closed	canonical	7ZLC 7ZLD	1.75 1.6	[29]
Zika + inhibitor	closed	W50 in two conformations	7ZMI	2.15	[29]
Zika	open	canonical	7VMV	3.35	[5]
Zika	closed	W50 in two conformations	8PN6	1.6	Ni et al., to be published
Zika + inhibitor	closed	canonical	7ZQ1 7ZWK	1.5 2.0	Huber and Steinmetzer, to be published
Zika + inhibitor	closed	W50 in two conformations	7ZTM 8A15	1.45 1.23	Huber and Steinmetzer, to be published
Zika + inhibitor	closed	W50 flipped	7ZPD 7ZQF 7ZUM 7ZV4 7ZVV 7ZW5 7ZYS 8AQA 8AQB 8AQK	1.4 1.7 1.75 1.7 1.75 1.4 1.26 1.35 1.3 1.2	Huber and Steinmetzer, to be published
Yellow fever	closed	canonical	6URV	2.9	[30]
Murray Valley encephalitis	open	canonical	2WV9	2.75	[31]
Japanese encephalitis	absent	W50 flipped	4R8T	2.1	[32]

[a] NS2Bc refers to the polypeptide segment forming a β-hairpin, which in the closed conformation contributes to the substrate binding site. Some constructs are devoid of NS2Bc.

[b] The canonical side chain conformations are defined by the high-resolution structure 2FOM. A star marks sequence numbers in NS2B.

Table S3. DNA and corresponding amino acid sequences of the proteins used in the current study.

Protein	DNA sequence	Amino acid sequence[a]
*G1*PylRS	ATGGTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAATA TGGTGATAATGATTGGAGCGAAGCCGAATTTGAAGATGCAGCAGCAC GTGATAAAGAATTTAGCAGCCAGTTTAGCAAACTGAAAAGCGCCAAT GATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAATGATCTGAC CGATCTGGAAAACAAAATTCGCGAAAAACTGGCAGCCCGTGGTTTTA TTGAAGTTCATACCCCGATTTTTGTGAGCAAAAGCGCACTGGCAAAA ATGACCATTACCGAAGATCATCCGCTGTTCAAACAGGTGTTTTGGAT TGATGATAAACGTGCACTGCGTCCGATGCATGCAATGAATCTGTATA AAGTTATGCGTGAACTGCGCGATCATACCAAAGGTCCGGTTAAAATC TTTGAAATTGGTAGCTGCTTTCGCAAAGAAAGCAAAAGCAGTACCCA TCTGGAAGAATTTACCATGCTGAACCTGGTTGAAATGGGTCCTGATG GTGATCCGATGGAACATCTGAAAATGTATATTGGCGATATCATGGAT GCCGTTGGTGTTGAATATACCACCAGTCGTGAAGAATCAGATGTTTA TGTTGAAACCCTGGACGTGGAAATTAATGGCACCGAAGTTGCAAGCG GTGCCGTTGGTCCGCATAAACTGGATCCGGCACATGATGTGCATGAA CCGTGGGCAGGTATTGGTTTTGGTCTGGAACGTCTGCTGATGCTGAA AAATGGTAAAAGCAATGCACGCAAAACCGGCAAAAGTATTACCTATC TGAATGGCTACAAACTGGATTAA	MVVKFTDSQIQHLMEYGDNDWSE AEFEDAAARDKEFSSQFSKLKSA NDKGLKDVIANPRNDLTDLENKI REKLAARGFIEVHTPIFVSKSAL AKMTITEDHPLFKQVFWIDDKRA LRPMHAMNLYKVMRELRDHTKGP VKIFEIGSCFRKESKSSTHLEEF TMLNLVEMGPDGDPMEHLKMYIG DIMDAVGVEYTTSREESDVYVET LDVEINGTEVASGAVGPHKLDPA HDVHEPWAGIGFGLERLLMLKNG KSNARKTGKSITYLNGYKLD
G1F4W27	ATGGTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAATA TGGTGATAATGATTGGAGCGAAGCCGAATTTGAAGATGCAGCAGCAC GTGATAAAGAATTTAGCAGCCAGTTTAGCAAACTGAAAAGCGCCAAT GATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAATGATCTGAC CGATCTGGAAAACAAAATTCGCGAAAAACTGGCAGCCCGTGGTTTTA TTGAAGTTCATACCCCGATTTTTGTGAGCAAAAGCGCACTGGCAAAA ATGACCATTACCGAAGATCATCCGCTGTTCAAACAGGTGTTTTGGAT TGATGATAAACGTGCACTGCGTCCGATGCATGCAATGAATCTGTATA AAGTTATGCGTGAACTGCGCGATCATACCAAAGGTCCGGTTAAAATC TTTGAAATTGGTAGCTGCTTTCGCAAAGAAAGCAAAAGCAGTACCCA TCTGGAAGAATTTACCATGCTGGCCCTGTTCGAAATGGGTCCTGATG GTGATCCGATGGAACATCTGAAAATGTATATTGGCGATATCATGGAT GCCGTTGGTGTTGAATATACCACCAGTCGTGAAGAATCAGATGTTTG GGTTGAAACCCTGGACGTGGAAATTAATGGCACCGAAGTTGCAAGCG GTACGGTTGGTCCGCATAAACTGGATCCGGCACATGATGTGCATGAA CCGAATGCAGGTATTGGTTTTGGTCTGGAACGTCTGCTGATGCTGAA AAATGGTAAAAGCAATGCACGCAAAACCGGCAAAAGTATTACCTATC TGAATGGCTACAAACTGGATTAA	MVVKFTDSQIQHLMEYGDNDWSE AEFEDAAARDKEFSSQFSKLKSA NDKGLKDVIANPRNDLTDLENKI REKLAARGFIEVHTPIFVSKSAL AKMTITEDHPLFKQVFWIDDKRA LRPMHAMNLYKVMRELRDHTKGP VKIFEIGSCFRKESKSSTHLEEF TMLALFEMGPDGDPMEHLKMYIG DIMDAVGVEYTTSREESDVWVET LDVEINGTEVASGTVGPHKLDPA HDVHEPNAGIGFGLERLLMLKNG KSNARKTGKSITYLNGYKLD*
G1F7W36	ATGGTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAATA TGGTGATAATGATTGGAGCGAAGCCGAATTTGAAGATGCAGCAGCAC GTGATAAAGAATTTAGCAGCCAGTTTAGCAAACTGAAAAGCGCCAAT GATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAATGATCTGAC CGATCTGGAAAACAAAATTCGCGAAAAACTGGCAGCCCGTGGTTTTA TTGAAGTTCATACCCCGATTTTTGTGAGCAAAAGCGCACTGGCAAAA ATGACCATTACCGAAGATCATCCGCTGTTCAAACAGGTGTTTTGGAT TGATGATAAACGTGCACTGCGTCCGATGCATGCAATGAATCAGTTTA AAGTTATGCGTGAACTGCGCGATCATACCAAAGGTCCGGTTAAAATC TTTGAAATTGGTAGCTGCTTTCGCAAAGAAAGCAAAAGCAGTACCCA TCTGGAAGAATTTACCATGCTGGCCCTGTTCGAAATGGGTCCTGATG GTGATCCGATGGAACATCTGAAAATGTATATTGGCGATATCATGGAT GCCGTTGGTGTTGAATATACCACCAGTCGTGAAGAATCAGATGTTTG GGTTGAAACCCTGGACGTGGAAATTAATGGCACCGAAGTTGCAAGCG GTGGGGTTGGTCCGCATAAACTGGATCCGGCACATGATGTGCATGAA CCGTATGCAGGTATTGGTTTTGGTCTGGAACGTCTGCTGATGCTGAA AAATGGTAAAAGCAATGCACGCAAAACCGGCAAAAGTATTACCTATC TGAATGGCTACAAACTGGATTAA	MVVKFTDSQIQHLMEYGDNDWSE AEFEDAAARDKEFSSQFSKLKSA NDKGLKDVIANPRNDLTDLENKI REKLAARGFIEVHTPIFVSKSAL AKMTITEDHPLFKQVFWIDDKRA LRPMHAMNQFKVMRELRDHTKGP VKIFEIGSCFRKESKSSTHLEEF TMLALFEMGPDGDPMEHLKMYIG DIMDAVGVEYTTSREESDVWVET LDVEINGTEVASGGVGPHKLDPA HDVHEPYAGIGFGLERLLMLKNG KSNARKTGKSITYLNGYKLD

AncCDT-1-W71TAG	ATGATCGCAGCAAGCACCCTGGATGAAATTATGAAACGTGGCACCCT GCGTGTTGGCACCGATGCAGATTATAAACCGTTTAGCTTCAAAGACA AAAACGGTCAGTATACCGGCTTTGATATTGATCTGGCAAAAGCACTG GCCAAAGAACTGGGTGTTAAAGTTGAATTTGTTCCGACCACCTAGGA TGGTATTATTCCGGCACTGCAGACCGGTAAATTTGATATCGTTATGA GCGGTATGACCATTACACCGGAACGTAAAAAAAAAGTGGATTTCAGC GATCCGTATATGACCGCAGGTCAGACCATTCTGGTTAAAAAAGATAA CGCCGATAAAATCAAAAGCTTTGAAGATCTGAACAAACCGGATGTTA AAGTGGCAGTTCAGCTGGGTACAACCAGCGAACAGGCAGCAAAAGAA TTTCTGCCGAAAGCAAAAATTCGCACCTTTGAAAATAATGCCGAAGC CTTTCAAGAGGTTGTTAGCGGTCGTGCAGATGCAATGGTTACCGATA GTCCGGTTGCAGCATATTACGCCAAAAAAAACCCTGGTCTGGCAGTT GTTGTTGTGGATGAACCGTTTACCCATGAACCGCTGGGTTTTGCAAT TCGTAAAGGTGATCCGGAACTGCTGAATTGGGTGAATAATTGGCTGA AACAAATGAAAAAAGACGGCACCTATGACAAACTGTACGAAAAATGG TTTAAACTCCATCATCATCACCATCATTAA	MIAASTLDEIMKRGTLRVGTDAD YKPFSFKDKNGQYTGFDIDLAKA LAKELGVKVEFVPTTXDGIIPAL QTGKFDIVMSGMTITPERKKKVD FSDPYMTAGQTILVKKDNADKIK SFEDLNKPDVKVAVQLGTTSEQA AKEFLPKAKIRTFENNAEAFQEV VSGRADAMVTDSPVAAYYAKKNP GLAVVVVDEPFTHEPLGFAIRKG DPELLNWVNNWLKQMKKDGTYDK LYEKWFKLHHHHHH
gZIKVp-W83TAG	ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCTGGGA GAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGACGTTGCTC TGGATGAATCGGGTGACTTTAGTCTTGTAGAAGACGATGGTCCTCCA ATGGCGGGTGGAGGTGGCTCAGGAGGGGGTGGGTCTGGTGCATTATG GGACGTACCAGCACCTAAGGAAGTAAAAAACGGCGAAACTACAGACG GTGTGTATCGCGTGATGACACGTGGTTTGTTGGGTAGCACACAAGTG GGCGTCGGAGTAATGCAAGAGGGCGTATTTCACACGATGTGGCATGT AACCAAAGGTTCTGCGTTACGCTCCGGCGAAGGACGCCTTGACCCGT ATTGGGGCGACGTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTAG AAACTGGACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGC CGTTCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGTA TTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAGATTAT CCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAGTGGTCGCGT GATTGGGTTATATGGGAATGGGGTTGTAATCAAGAATGGATCTTACG TCTCGGCGATCACACAGGGTCGCCGCCATCACCACCATCATCACTAA	MSVDMYIERAGDITWEKDAEVTG NSPRLDVALDESGDFSLVEDDGP PMAGGGGSGGGGSGALWDVPAPK EVKNGETTDGVYRVMTRGLLGST QVGVGVMQEGVFHTMWHVTKGSA LRSGEGRLDPYWGDVKQDLVSYS GPXKLDAAWDGHSEVQLLAVPPG ERARNIQTLPGIFKTKDGDIGAV ALDYPAGTSGSPILDKSGRVIGL YGNGVVIKNGSYVSAITQGRRHH HHHH
eZIKVp-W83TAG	ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCTGGGA GAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGACGTTGCTC TGGATGAATCGGGTGACTTTAGTCTTGTAGAAGACGATGGTCCTCCA ATGGCGGTTAAAACCGGTCGTAAATCTGGTGCATTATGGGACGTACC AGCACCTAAGGAAGTAAAAAACGGCGAAACTACAGACGGTGTGTATC GCGTGATGACACGTGGTTTGTTGGGTAGCACACAAGTGGGCGTCGGA GTAATGCAAGAGGGCGTATTTCACACGATGTGGCATGTAACCAAAGG TTCTGCGTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCG ACGTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTAGAAACTGGAC GCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGTTCCGCC AGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGTATTTTCAAAA CTAAGGATGGTGATATCGGGGCGGTGGCCTTAGATTATCCGGCAGGA ACATCCGGGTCTCCCATTTTAGACAAGAGTGGTCGCGTGATTGGGTT ATATGGGAATGGGGTTGTAATCAAGAATGGATCTTACGTCTCGGCGA TCACACAGGGTCGCCGCCATCACCACCATCATCACTAA	MSVDMYIERAGDITWEKDAEVTG NSPRLDVALDESGDFSLVEDDGP PMAVKTGRKSGALWDVPAPKEVK NGETTDGVYRVMTRGLLGSTQVG VGVMQEGVFHTMWHVTKGSALRS GEGRLDPYWGDVKQDLVSYSGPX KLDAAWDGHSEVQLLAVPPGERA RNIQTLPGIFKTKDGDIGAVALD YPAGTSGSPILDKSGRVIGLYGN GVVIKNGSYVSAITQGRRHHHHH H
gWNVp-W83TAG	ATGACCGATATGTGGATTGAACGTACCGCAGATATTACCTGGGAAAG TGATGCAGAAATTACCGGTAGCAGCGAACGTGTTGATGTTCGTCTGG ATGATGATGGTAATTTTCAGCTGATGAATGATCCGGGTGCACCGTGG GCAGGTGGAGGTGGCTCAGGAGGGGGTGGGGGTGGTGTTCTGTGGGA TACCCCGAGTCCGAAAGAATATAAGAAAGGTGATACCACCACCGGTG TGTATCGTATTATGACCCGTGGTCTGCTGGGTAGCTATCAGGCTGGT GCCGGTGTTATGGTTGAAGGTGTTTTTCATACCCTGTGGCATACCAC CAAAGGTGCAGCACTGATGAGCGGTGAAGGTCGCCTGGATCCGTATT GGGGTAGCGTTAAAGAAGATCGTCTTTGTTACGGTGGTCCGTAGAAA CTGCAGCATAAATGGAATGGTCATGATGAAGTTCAGATGATTGTTGT GGAACCGGGTAAAAATGTTAAAAACGTTCAGACCAAACCGGGTGTGT TTAAAACTCCGGAAGGTGAAATTGGTGCAGTTACCCTGGATTATCCG ACCGGTACAAGCGGTAGCCCGATTGTTGATAAAAATGGTGATGTGAT TGGCCTGTATGGTAATGGTGTTATTATGCCGAATGGCAGCTATATTT CAGCAATTGTTCAGGGTGAACGTCATCACCACCATCATCACTAA	MTDMWIERTADITWESDAEITGS SERVDVRLDDDGNFQLMNDPGAP WAGGGGSGGGGGGVLWDTPSPKE YKKGDTTTGVYRIMTRGLLGSYQ AGAGVMVEGVFHTLWHTTKGAAL MSGEGRLDPYWGSVKEDRLCYGG PXKLQHKWNGHDEVQMIVVEPGK NVKNVQTKPGVFKTPEGEIGAVT LDYPTGTSGSPIVDKNGDVIGLY GNGVIMPNGSYISAIVQGERHHH HHH

eWNVp- W83TAG	ATGACCGATATGTGGATTGAACGTACCGCAGATATTACCTGGGAAAG TGATGCAGAAATTACCGGTAGCAGCGAACGTGTTGATGTTCGTCTGG ATGATGATGGTAATTTTCAGCTGATGAATGATCCGGGTGCACCGTGG GCACTGCAGTATACCAAACGTGGTGGTGTTCTGTGGGATACCCCGAG TCCGAAAGAATATAAGAAAGGTGATACCACCACCGGTGTGTATCGTA TTATGACCCGTGGTCTGCTGGGTAGCTATCAGGCTGGTGCCGGTGTT ATGGTTGAAGGTGTTTTTCATACCCTGTGGCATACCACCAAAGGTGC AGCACTGATGAGCGGTGAAGGTCGCCTGGATCCGTATTGGGGTAGCG TTAAAGAAGATCGTCTTTGTTACGGTGGTCCGTAGAAACTGCAGCAT AAATGGAATGGTCATGATGAAGTTCAGATGATTGTTGTGGAACCGGG TAAAAATGTTAAAAACGTTCAGACCAAACCGGGTGTGTTTAAAACTC CGGAAGGTGAAATTGGTGCAGTTACCCTGGATTATCCGACCGGTACA AGCGGTAGCCCGATTGTTGATAAAAATGGTGATGTGATTGGCCTGTA TGGTAATGGTGTTATTATGCCGAATGGCAGCTATATTTCAGCAATTG TTCAGGGTGAACGTCATCACCACCATCATCACTAA	MTDMWIERTADITWESDAEITGS SERVDVRLDDDGNFQLMNDPGAP WALQYTKRGGVLWDTPSPKEYKK GDTTTGVYRIMTRGLLGSYQAGA GVMVEGVFHTLWHTTKGAALMSG EGRLDPYWGSVKEDRLCYGGPXK LQHKWNGHDEVQMIVVEPGKNVK NVQTKPGVFKTPEGEIGAVTLDY PTGTSGSPIVDKNGDVIGLYGNG VIMPNGSYISAIVQGERHHHHHH
gDENV2p- W83TAG	ATGGCAGATCTGGAACTGGAACGTGCAGCAGATGTTCGTTGGGAAGA ACAGGCAGAAATTAGCGGTAGCAGCCCGATTCTGAGCATTACCATTA GCGAAGATGGTAGCATGAGCATCAAAAACGAAGAAGAAGAACAGACC CTGGGTGGAGGTGGCTCAGGAGGGGGTGGGGCCGGTGTTCTGTGGGA TGTTCCGAGTCCGCCTCCGGTTGGTAAAGCAGAACTGGAAGATGGTG CCTATCGTATTAAACAGAAAGGTATTCTGGGTTACAGCCAGATTGGT GCGGGTGTTTATAAAGAAGGCACCTTTCATACCATGTGGCATGTTAC CCGTGGTGCAGTTCTGATGCATAAAGGTAAACGTATTGAACCGAGCT GGGCAGATGTTAAAAAGGATCTGATTAGCTATGGTGGTGGTTAGAAA CTGGAAGGTGAATGGAAAGAAGGTGAAGAAGTTCAGGTTCTGGCCCT GGAACCGGGTAAAAATCCGCGTGCAGTTCAGACCAAACCGGGTCTGT TTAAAACCAATACCGGCACCATTGGTGCAGTGAGCCTGGATTTTAGT CCGGGTACAAGCGGTAGCCCGATTGTTGATAAAAAGGGTAAAGTTGT TGGCCTGTATGGTAATGGTGTTGTGACCCGTAGCGGTGCATATGTTA GCGCAATTGCAAATACCGAAAAACATCACCACCATCATCACTAA	MADLELERAADVRWEEQAEISGS SPILSITISEDGSMSIKNEEEEQ TLGGGGSGGGGAGVLWDVPSPPP VGKAELEDGAYRIKQKGILGYSQ IGAGVYKEGTFHTMWHVTRGAVL MHKGKRIEPSWADVKKDLISYGG GXKLEGEWKEGEEVQVLALEPGK NPRAVQTKPGLFKTNTGTIGAVS LDFSPGTSGSPIVDKKGKVVGLY GNGVVTRSGAYVSAIANTEKHHH HHH
gMVEVp- W83TAG	ATGGCAACCGATATGTGGCTGGAACGTGCAGCAGATGTTAGCTGGGA AGCCGGTGCAGCAATTACCGGCACCAGCGAACGTCTGGATGTTCAGC TGGATGACGATGGTGATTTTCATCTGCTGAATGATCCGGGTGTTCCG TGGAAAGGTGGAGGTGGCTCAGGAGGGGGTGGGGGTGGTGTGTTTTG GGATACCCCGAGTCCGAAAGTTTATCCGAAAGGTGATACCACACCGG GTGTTTATCGTATTATGGCACGTGGTATTCTGGGTCGTTATCAGGCA GGCGTTGGTGTTATGCATGAAGGTGTTTTTCATACCCTGTGGCATAC AACCCGTGGTGCAGCCATTATGAGCGGTGAAGGTCGTCTGACCCCGT ATTGGGGTAATGTTAAAGAAGATCGCGTTACCTACGGTGGCCCTTAG AAACTGGATCAGAAATGGAATGGTGTTGATGATGTTCAGATGATTGT TGTGGAACCGGGTAAACCGGCAATTAATGTTCAGACCAAACCGGGTA TCTTTAAAACCGCACATGGTGAAATTGGTGCAGTTAGCCTGGATTAT CCGATTGGTACAAGCGGTAGCCCGATTGTTAATAGCAATGGCGAAAT TATTGGCCTGTATGGTAATGGTGTGATTCTTGGTAATGGCGCATATG TTAGCGCAATTGTTCAGGGTGAACGTCATCACCACCATCATCACTAA	MATDMWLERAADVSWEAGAAITG TSERLDVQLDDDGDFHLLNDPGV PWKGGGGSGGGGGGVFWDTPSPK VYPKGDTTPGVYRIMARGILGRY QAGVGVMHEGVFHTLWHTTRGAA IMSGEGRLTPYWGNVKEDRVTYG GPXKLDQKWNGVDDVQMIVVEPG KPAINVQTKPGIFKTAHGEIGAV SLDYPIGTSGSPIVNSNGEIIGL YGNGVILGNGAYVSAIVQGERHH HHHH
eMVEVp- W83TAG	ATGGCAACCGATATGTGGCTGGAACGTGCAGCAGATGTTAGCTGGGA AGCCGGTGCAGCAATTACCGGCACCAGCGAACGTCTGGATGTTCAGC TGGATGACGATGGTGATTTTCATCTGCTGAATGATCCGGGTGTTCCG TGGAAACTGAAATATACCAAACGTGGTGGTGTGTTTTGGGATACCCC GAGTCCGAAAGTTTATCCGAAAGGTGATACCACACCGGGTGTTTATC GTATTATGGCACGTGGTATTCTGGGTCGTTATCAGGCAGGCGTTGGT GTTATGCATGAAGGTGTTTTTCATACCCTGTGGCATACAACCCGTGG TGCAGCCATTATGAGCGGTGAAGGTCGTCTGACCCCGTATTGGGGTA ATGTTAAAGAAGATCGCGTTACCTACGGTGGCCCTTAGAAACTGGAT CAGAAATGGAATGGTGTTGATGATGTTCAGATGATTGTTGTGGAACC GGGTAAACCGGCAATTAATGTTCAGACCAAACCGGGTATCTTTAAAA CCGCACATGGTGAAATTGGTGCAGTTAGCCTGGATTATCCGATTGGT ACAAGCGGTAGCCCGATTGTTAATAGCAATGGCGAAATTATTGGCCT GTATGGTAATGGTGTGATTCTTGGTAATGGCGCATATGTTAGCGCAA TTGTTCAGGGTGAACGTCATCACCACCATCATCACTAA	MATDMWLERAADVSWEAGAAITG TSERLDVQLDDDGDFHLLNDPGV PWKLKYTKRGGVFWDTPSPKVYP KGDTTPGVYRIMARGILGRYQAG VGVMHEGVFHTLWHTTRGAAIMS GEGRLTPYWGNVKEDRVTYGGPX KLDQKWNGVDDVQMIVVEPGKPA INVQTKPGIFKTAHGEIGAVSLD YPIGTSGSPIVNSNGEIIGLYGN GVILGNGAYVSAIVQGERHHHHH H

gJEVp-W83TAG	ATGGCAACCGATATGTGGCTGGATCGTGCAGCAGATATTAGCTGGGA AATGGAAGCAGCAATTACCGGTAGCAGCCGTCGTCTGGATGTTAAAC TGGATGACGATGGTGATTTTCACCTGATTGATGATCCGGGTGTTCCG TGGAAAGGTGGAGGTGGCTCAGGAGGGGGTGGGGGTGGTGTGTTTTG GGATACCCCGAGTCCGAAACCGTGTCTGAAAGGTGATACCACCACCG GTGTTTATCGTATTATGGCACGTGGTGTTCTGGGCACCTATCAGGCA GGCGTTGGTGTTATGTATGAAAATGTTTTTCATACCCTGTGGCATAC CACACGTGGTGCAGCCATTATGAGCGGTGAAGGTAAACTGACCCCGT ATTGGGGTAGCGTTAAAGAAGATCGTATTAGCTACGGTGGTCCGTAG CGTTTTGATCGTAAATGGAATGGCACCGATGATGTTCAGGTTATTGT TGTTGAACCGGGTAAACCGGCAGTTAATATTCAGACCAAACCGGGTG TGTTTCGTACCCCGTTTGGTGAAGTTGGTGCAGTTAGCCTGGATTAT CCGCGTGGCACCAGCGGTAGCCCGATTCTGGATAGCAATGGTGATAT TATTGGTCTGTATGGTAATGGCGTGGAATTAGGTGATGGTAGCTATG TTAGCGCAATTGTTCAGGGTGATCGTCATCACCACCATCATCACTAA	MATDMWLDRAADISWEMEAAITG SSRRLDVKLDDDGDFHLIDDPGV PWKGGGGSGGGGGGVFWDTPSPK PCLKGDTTTGVYRIMARGVLGTY QAGVGVMYENVFHTLWHTTRGAA IMSGEGKLTPYWGSVKEDRISYG GPXRFDRKWNGTDDVQVIVVEPG KPAVNIQTKPGVFRTPFGEVGAV SLDYPRGTSGSPILDSNGDIIGL YGNGVELGDGSYVSAIVQGDRHH HHHH

[a] X indicates the positions of FTrp. The linkers in different flaviviral NS2B-NS3p are colored in blue.

References

1. Qianzhu, H.; Abdelkader, E. H.; Herath, I. D.; Otting, G.; Huber, T., Site-specific incorporation of 7-fluoro-L-tryptophan into proteins by genetic encoding to monitor ligand binding by ^{19}F NMR spectroscopy. *ACS Sens.* **2022,** *7*, 44–49.
2. Phoo, W. W.; El Sahili, A.; Zhang, Z.; Chen, M. W.; Liew, C. W.; Lescar, J.; Vasudevan, S. G.; Luo, D., Crystal structures of full length DENV4 NS2B-NS3 reveal the dynamic interaction between NS2B and NS3. *Antivir. Res.* **2020,** *182*, 104900.
3. Luo, D.; Xu, T.; Hunke, C.; Grüber, G.; Vasudevan, S. G.; Lescar, J., Crystal structure of the NS3 protease-helicase from dengue virus. *J. Virol.* **2008,** *82*, 173–183.
4. Luo, D.; Wei, N.; Doan, D. N.; Paradkar, P. N.; Chong, Y.; Davidson, A. D.; Kotaka, M.; Lescar, J.; Vasudevan, S. G., Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. *J. Biol. Chem.* **2010,** *285*, 18817–18827.
5. Quek, J.-P.; Ser, Z.; Chew, B. L. A.; Li, X.; Wang, L.; Sobota, R. M.; Luo, D.; Phoo, W. W., Dynamic interactions of post cleaved NS2B cofactor and NS3 protease identified by integrative structural approaches. *Viruses* **2022,** *14*, 1440.
6. Noble, C. G.; Seh, C. C.; Chao, A. T.; Shi, P. Y., Ligand-bound structures of the dengue virus protease reveal the active conformation. *J. Virol.* **2012,** *86*, 438–446.
7. Erbel, P.; Schiering, N.; D'Arcy, A.; Renatus, M.; Kroemer, M.; Lim, S. P.; Yin, Z.; Keller, T. H.; Vasudevan, S. G.; Hommel, U., Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat. Struct. Mol. Biol.* **2006,** *13*, 372–373.
8. Yildiz, M.; Ghosh, S.; Bell, J. A.; Sherman, W.; Hardy, J. A., Allosteric inhibition of the NS2B-NS3 protease from dengue virus. *ACS Chem. Biol.* **2013,** *8*, 2744–2752.
9. Yao, Y.; Huo, T.; Lin, Y.-L.; Nie, S.; Wu, F.; Hua, Y.; Wu, J.; Kneubehl, A. R.; Vogt, M. B.; Rico-Hesse, R.; Song, Y., Discovery, X-ray crystallography and antiviral activity of allosteric inhibitors of flavivirus NS2B-NS3 protease. *J. Am. Chem. Soc.* **2019,** *141*, 6832–6836.
10. Chandramouli, S.; Joseph, J. S.; Daudenarde, S.; Gatchalian, J.; Cornillez-Ty, C.; Kuhn, P., Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. *J. Virol.* **2010,** *84*, 3059–3067.

11. Aleshin, A. E.; Shiryaev, S. A.; Strongin, A. Y.; Liddington, R. C., Structural evidence for regulation and specificity of flaviviral proteases and evolution of the *Flaviviridae* fold. *Protein Sci.* **2007,** *16*, 795–806.

12. Robin, G.; Chappell, K.; Stoermer, M. J.; Hu, S.-H.; Young, P. R.; Fairlie, D. P.; Martin, J. L., Structure of West Nile virus NS3 protease: Ligand stabilization of the catalytic conformation. *J. Mol. Biol.* **2009,** *385*, 1568–1577.

13. Hammamy, M. Z.; Haase, C.; Hammami, M.; Hilgenfeld, R.; Steinmetzer, T., Development and characterization of new peptidomimetic inhibitors of the West Nile virus NS2B–NS3 protease. *ChemMedChem* **2013,** *8*, 231–241.

14. Nitsche, C.; Zhang, L.; Weigel, L. F.; Schilz, J.; Graf, D.; Bartenschlager, R.; Hilgenfeld, R.; Klein, C. D., Peptide–boronic acid inhibitors of flaviviral proteases: medicinal chemistry and structural biology. *J. Med. Chem.* **2017,** *60*, 511–516.

15. Lei, J.; Hansen, G.; Nitsche, C.; Klein, C. D.; Zhang, L.; Hilgenfeld, R., Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. *Science* **2016,** *353*, 503–505.

16. Phoo, W. W.; Li, Y.; Zhang, Z.; Lee, M. Y.; Loh, Y. R.; Tan, Y. B.; Ng, E. Y.; Lescar, J.; Kang, C.; Luo, D., Structure of the NS2B-NS3 protease from Zika virus after self-cleavage. *Nat. Commun.* **2016,** *7*, 13410.

17. Zhang, Z.; Li, Y.; Loh, Y. R.; Phoo, W. W.; Hung, A. W.; Kang, C.; Luo, D., Crystal structure of unlinked NS2B-NS3 protease from Zika virus. *Science* **2016,** *354*, 1597-1600.

18. Chen, X.; Yang, K.; Wu, C.; Chen, C.; Hu, C.; Buzovetsky, O.; Wang, Z.; Ji, X.; Xiong, Y.; Yang, H., Mechanisms of activation and inhibition of Zika virus NS2B-NS3 protease. *Cell Res.* **2016,** *26*, 1260–1263.

19. Lee, H.; Ren, J.; Nocadello, S.; Rice, A. J.; Ojeda, I.; Light, S.; Minasov, G.; Vargas, J.; Nagarathnam, D.; Anderson, W. F.; Johnson, M. E., Identification of novel small molecule inhibitors against NS2B/NS3 serine protease from Zika virus. *Antivir. Res.* **2017,** *139*, 49–58.

20. Li, Y.; Zhang, Z.; Phoo, W. W.; Loh, Y. R.; Wang, W.; Liu, S.; Chen, M. W.; Hung, A. W.; Keller, T. H.; Luo, D.; Kang, C., Structural Dynamics of Zika Virus NS2B-NS3 Protease Binding to Dipeptide Inhibitors. *Structure* **2017,** *25*, 1242-1250.e3.

21. Li, Y.; Zhang, Z.; Phoo, W. W.; Loh, Y. R.; Li, R.; Yang, H. Y.; Jansson, A. E.; Hill, J.; Keller, T. H.; Nacro, K.; Luo, D.; Kang, C., Structural insights into the inhibition of Zika virus NS2B-NS3 protease by a small-molecule inhibitor. *Struct.* **2018,** *26*, 555–564.e3.

22. Phoo, W. W.; Zhang, Z.; Wirawan, M.; Chew, E. J. C.; Chew, A. B. L.; Kouretova, J.; Steinmetzer, T.; Luo, D., Structures of Zika virus NS2B-NS3 protease in complex with peptidomimetic inhibitors. *Antivir. Res.* **2018,** *160*, 17–24.

23. Nitsche, C.; Onagi, H.; Quek, J.-P.; Otting, G.; Luo, D.; Huber, T., Biocompatible macrocyclization between cysteine and 2-cyanopyridine generates stable peptide inhibitors. *Org. Lett.* **2019,** *21*, 4709–4712.

24. Patil, N. A.; Quek, J.-P.; Schroeder, B.; Morewood, R.; Rademann, J.; Luo, D.; Nitsche, C., 2-Cyanoisonicotinamide conjugation: A facile approach to generate potent peptide inhibitors of the Zika virus protease. *ACS Med. Chem. Lett.* **2021,** *12*, 732–737.

25. Braun, N. J.; Quek, J. P.; Huber, S.; Kouretova, J.; Rogge, D.; Lang-Henkel, H.; Cheong, E. Z. K.; Chew, B. L. A.; Heine, A.; Luo, D.; Steinmetzer, T., Structure-based macrocyclization of substrate analogue NS2B-NS3 protease inhibitors of Zika, West Nile and dengue viruses. *ChemMedChem* **2020,** *15*, 1439–1452.

26. Quek, J. P.; Liu, S.; Zhang, Z.; Li, Y.; Ng, E. Y.; Loh, Y. R.; Hung, A. W.; Luo, D.; Kang, C., Identification and structural characterization of small molecule fragments targeting Zika virus NS2B-NS3 protease. *Antivir. Res.* **2020,** *175*, 104707.

27. Huber, S.; Braun, N. J.; Schmacke, L. C.; Quek, J. P.; Murra, R.; Bender, D.; Hildt, E.; Luo, D.; Heine, A.; Steinmetzer, T., Structure-based optimization and characterization of macrocyclic Zika virus NS2B-NS3 protease inhibitors. *J. Med. Chem.* **2022,** *65*, 6555–6572.

28. Xiong, Y.; Cheng, F.; Zhang, J.; Su, H.; Hu, H.; Zou, Y.; Li, M.; Xu, Y., Structure-based design of a novel inhibitor of the ZIKA virus NS2B/NS3 protease. *Bioorg. Chem.* **2022,** *128*, 106109.

29. Hammerschmidt, S. J.; Huber, S.; Braun, N. J.; Lander, M.; Steinmetzer, T.; Kersten, C., Thermodynamic characterization of a macrocyclic Zika virus NS2B/NS3 protease inhibitor and its acyclic analogs. *Arch. Pharm.* **2023,** *356*, 2200518.

30. Noske, G. D.; Gawriljuk, V. O.; Fernandes, R. S.; Furtado, N. D.; Bonaldo, M. C.; Oliva, G.; Godoy, A. S., Structural characterization and polymorphism analysis of the NS2B-NS3 protease from the 2017 Brazilian circulating strain of Yellow Fever virus. *Biochim. Biophys. Acta, Gen. Subj.* **2020,** *1864*, 129521.

31. Assenberg, R.; Mastrangelo, E.; Walter, T. S.; Verma, A.; Milani, M.; Owens, R. J.; Stuart, D. I.; Grimes, J. M.; Mancini, E. J., Crystal structure of a novel conformational state of the flavivirus NS3 protein: Implications for polyprotein processing and viral replication. *J. Virol.* **2009,** *83*, 12895–12906.

32. Weinert, T.; Olieric, V.; Waltersperger, S.; Panepucci, E.; Chen, L.; Zhang, H.; Zhou, D.; Rose, J.; Ebihara, A.; Kuramitsu, S.; Li, D.; Howe, N.; Schnapp, G.; Pautsch, A.; Bargsten, K.; Prota, A. E.; Surana, P.; Kottur, J.; Nair, D. T.; Basilico, F.; Cecatiello, V.; Pasqualato, S.; Boland, A.; Weichenrieder, O.; Wang, B.-C.; Steinmetz, M. O.; Caffrey, M.; Wang, M., Fast native-SAD phasing for routine macromolecular structure determination. *Nat. Methods* **2015,** *12*, 131–133.