SUPPLEMENT

Fomina A.D., Palyulin V.A., Osolodkin D.I. (2024) Homology modeling of the orthoflavivirus NS1 protein for virtual screening of potential ligands. Biomeditsinskaya Khimiya, **70**(6), 456-468.

DOI: 10.18097/PBMC20247006456

Supplementary materials

Table of Contents

SUPPLEMENTARY DISCUSSION

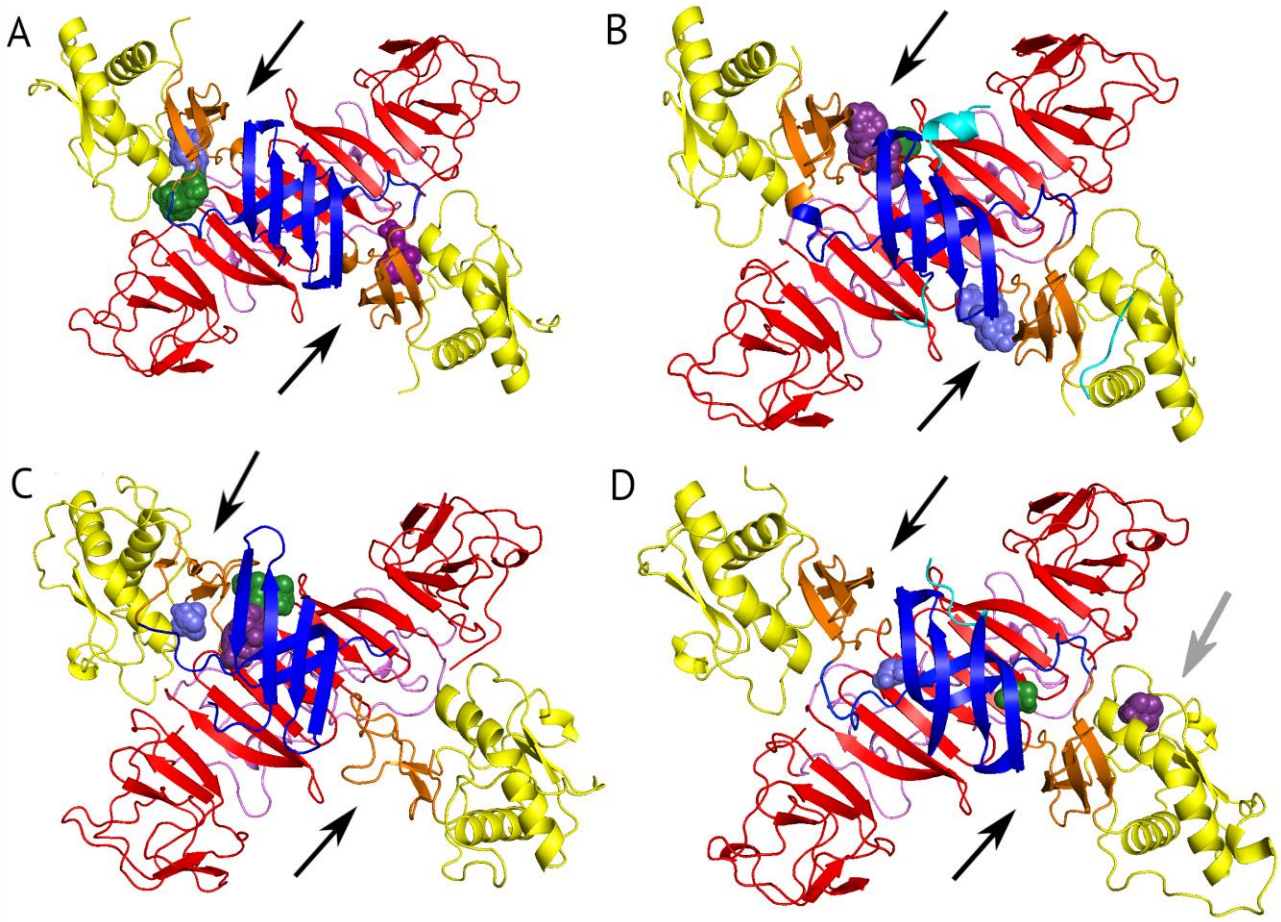
Search for potential binding sites for small molecules

Pocket Search with FTSite

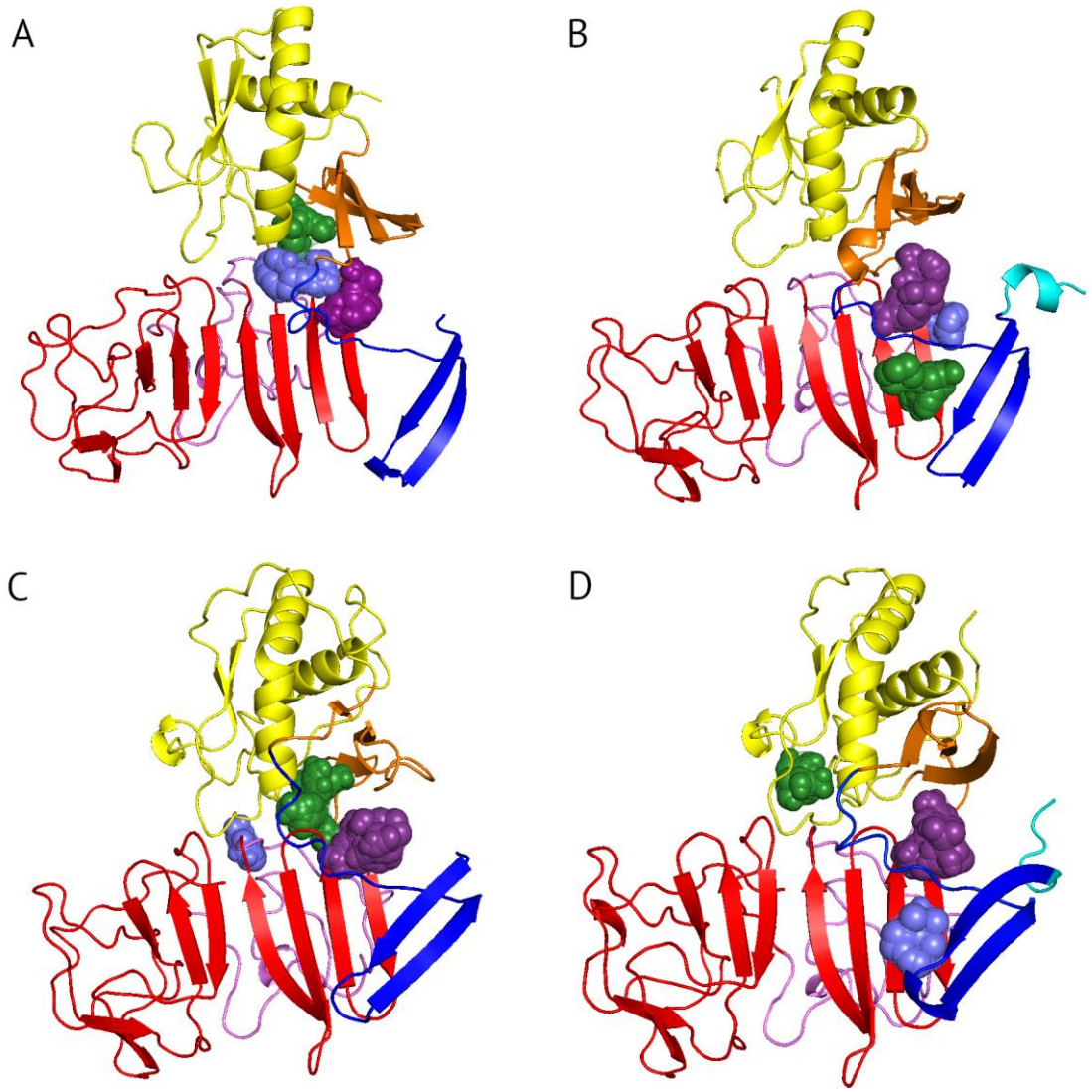
FTSite outputs only the three highest-scoring pockets. For each pocket, the total surface area and solvent-accessible surface area were calculated using UCSF Chimera (Supplementary Discussion Figure SDF1). The calculations were performed separately for dimers and monomers (Supplementary Discussion Figure SDF1, SDF2), but the results were similar. For all structures, both dimeric and monomeric, pockets were found in the dimerization site region between the β-barrel, connector domain, α-helical wing, and β-ladder (Supplementary Discussion Figure SDF1, indicated by black arrows), as well as pockets on the outer surface of the α-helical wing (Supplementary Discussion Figure SDF1 D, indicated by gray arrow).

Supplementary Discussion Table SDT1 – FTSite pockets. 1 - Surface area, $Å^2$, 2 - Surface area accessible to solvent $Å^2$

Dimers								
Pocket	4O6C		4O6D		5K6K		5GS6	
	1	2	1	2	1	2	1	2
FT1	254.523	496.587	378.484	684.588	317.113	559.968	241.590	472.430
FT2	278.533	521.433	66.372	202.736	286.198	543.746	215.907	435.082
FT3	207.685	428.336	309.400	565.472	124.496	293.716	124.955	296.750
Monomers								
Pocket	4O6C		4O6D		5K6K		5GS6	
	1	2	1	2	1	2	1	2
FT1	229.094	453.904	317.113	559.968	304.575	550.396	324.625	563.421
FT2	232.496	454.591	286.198	543.746	191.382	392.054	324.261	588.642
FT3	272.384	511.209	124.496	293.716	189.763	391.433	147.632	333.571



Supplementary Discussion Figure SDF1 – FTSite pockets for dimeric structures. The pockets are colored in descending order of score: purple, dark green, blue. (A) – 4O6C (WNV), (B) – 4O6D (WNV), (C) – 5GS6 (ZIKV), (D) – 5K6K (ZIKV). Domain coloring: blue – β-barrel, orange – connector domain, yellow – α-helical wing, red – β-ladder), pink – "spaghetti loop". Arrows are interpreted in the main text.



Supplementary Discussion Figure SDF2 – FTSite pockets for monomeric structures. (A) — 4O6C (WNV), (B) — 4O6D (WNV), (C) — 5GS6 (ZIKV), (D) — 5K6K (ZIKV). Domains and pockets are coloured similarly to Figure 3, the monomer structures of 4O6C and 5K6K contain a polyhistidine tag, highlighted in bright turquoise.

Results of DoGSiteScorer calculations

The DoGSiteScorer program returns 20 pockets ranked by volume. From these, 3 to 5 highest-ranked pockets were selected (they received index a), as well as symmetrical ones (indicated with index b; if two pockets with a lower rating were in the region symmetrical to pocket a, they were assigned indices b1 and b2).

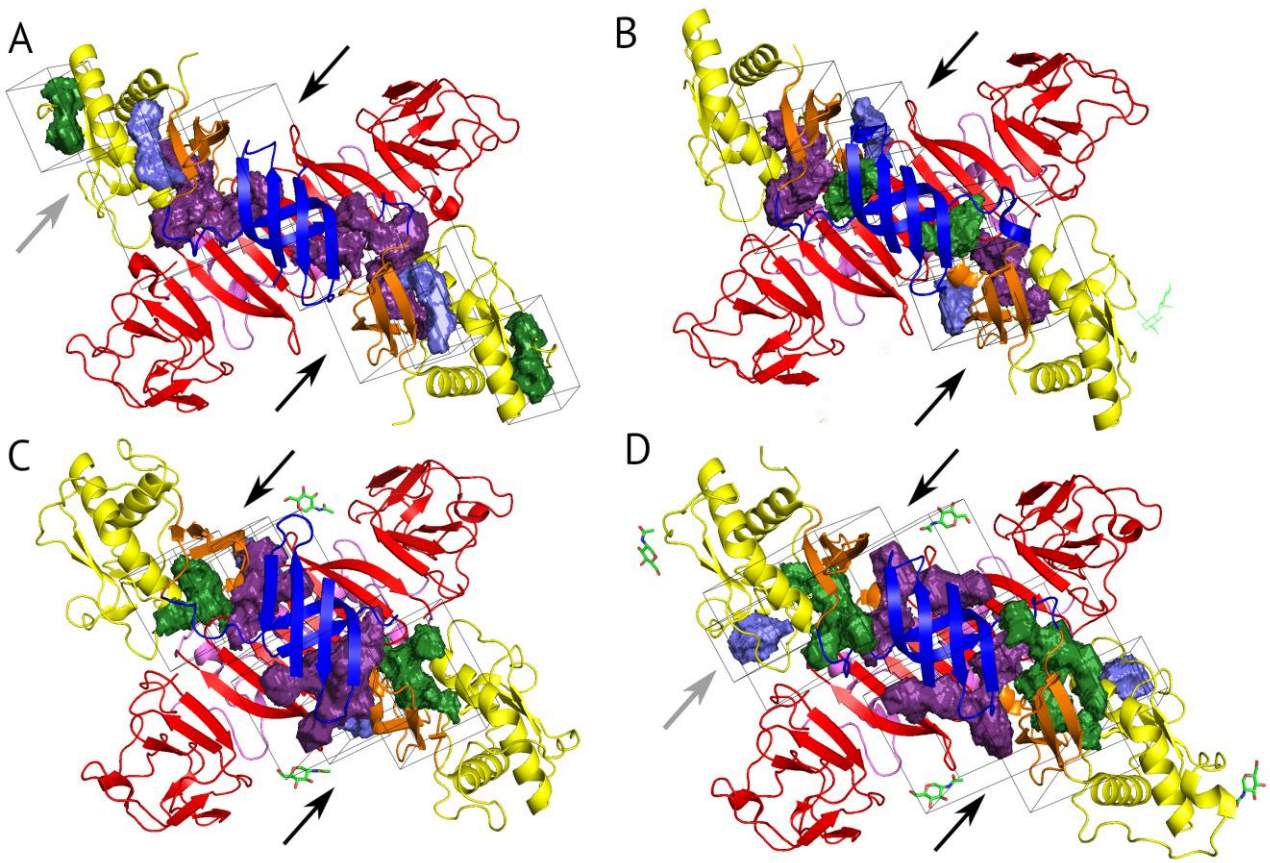
The volume of the pockets identified by DoGSiteScorer was larger compared to the FTSite pockets (Supplementary Discussion Table SDT2 , SDT3). However, the pockets were located in a similar way between the β-barrel, connector domain, α-helical wing and β-ladder, as well as on the outer surface of the α-helical wing (Supplementary Discussion Figure SDF3, SDT4).

Supplementary Discussion Table SDT2 – Characteristics of DoGSiteScorer pockets for monomers.

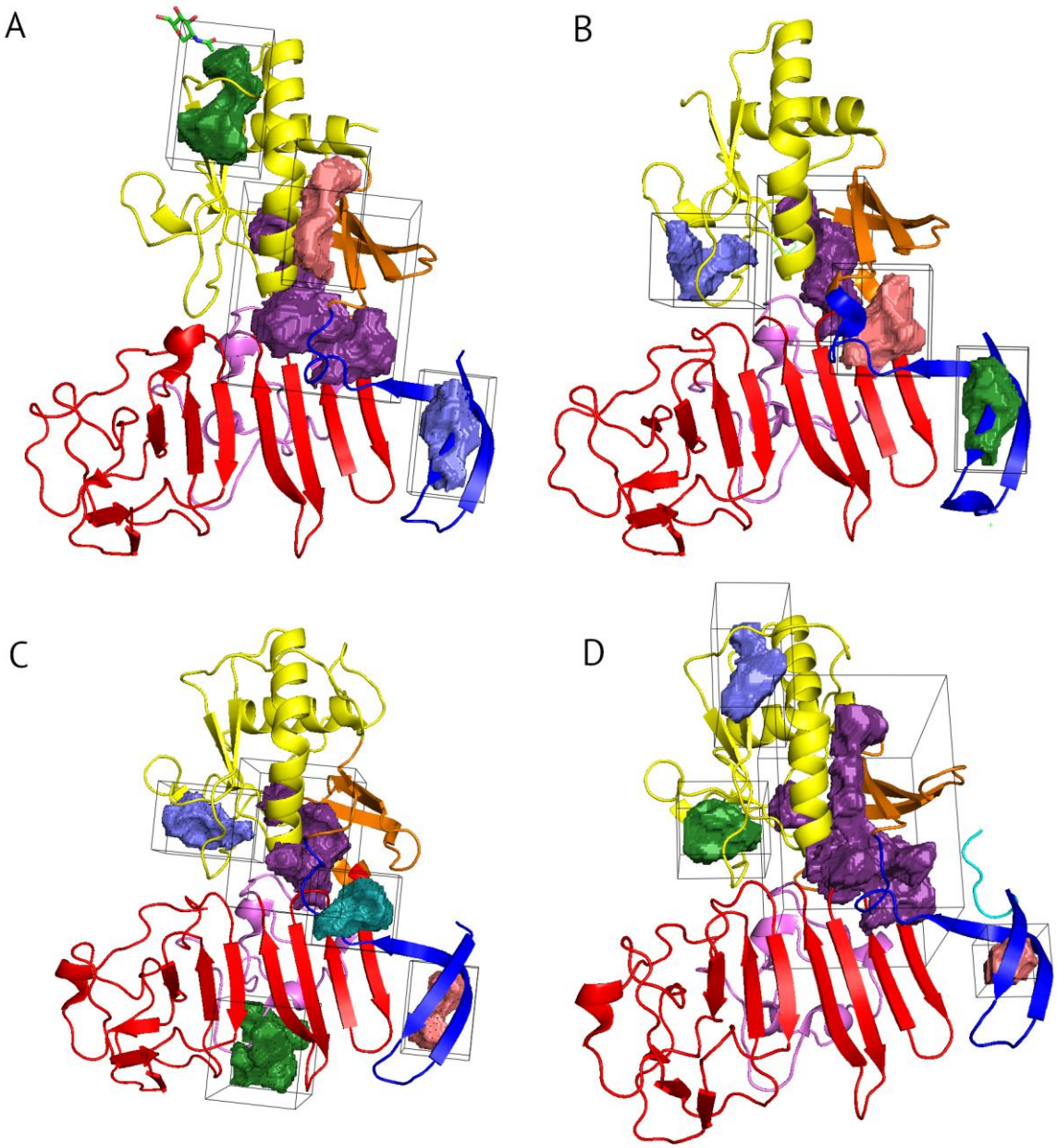
№		Volume, Å^3	Surface area, Å^2	Depth, Å	Hydrophobisity	simpleScore	drugScore
4O6C	1	999.17	1044.15	26.18	0.21	0.57	0.812306
	2	345.98	673.26	10	0.4	0.19	0.493328
	3	294.08	620.58	12.46	0.47	0.17	0.513667
	4	265.02	534.23	15.23	0.52	0.15	0.624596
4O6D	1	441.41	303.44	19.88	0.14	0.17	0.801797
	2	300.61	522.24	12.14	0.48	0.17	0.5
	3	275.26	570.65	14.32	0.43	0.13	0.539993
	4	270.02	298.72	12.76	0.3	0.07	0.544148
5K6K	1	1222.46	1542.01	27.04	0.34	0.62	0.812363
	2	268.29	378.66	10.8	0.42	0.12	0.473985
	3	244.93	640.01	10.2	0.44	0.1	0.41238
	4	143.36	353.74	7.82	0.26	0	0.186699
5GS6	1	569.73	519.83	20.29	0.23	0.31	0.828564
	2	500.86	890.33	12.27	0.36	0.3	0.621492
	3	270.08	427.18	13.08	0.38	0.11	0.540729
	4	228.99	555.95	10.75	0.41	0.07	0.360466
	5	221.7	288.07	12.2	0.34	0.04	0.491086

Supplementary Discussion Table SDT3 – Characteristics of DoGSiteScorer pockets for dimers.

№		Volume, Å^3	Surface area, Å^2	Depth, Å	Hydrophobisity	simpleScore	drugScore
4O6C	1.a	1034.5	965.84	23.79	0.1	0.53	0.81
	1.b1	811.61	891.58	31.44	0.15	0.44	0.86
	2.a	306.42	568.33	10.97	0.44	0.16	0.5
	2.b	288.97	574.57	9.76	0.39	0.13	0.43
	1.b2	259.92	255.98	17.32	0.09	0	0.68
	3.a	247.27	512.7	15.13	0.53	0.14	0.61
	3.b	212.36	500	14.69	0.55	0.11	0.57
4O6D	1.b	683.14	652.83	24.76	0.11	0.34	0.86
	1.a	420.66	316.57	18.96	0.12	0.15	0.78
	2.a	286.69	222.41	16.38	0.05	0	0.66
	2.b	261.47	203.41	16.63	0.04	0	0.66
	3	213.44	286.62	12.36	0.19	0	0.42
	4	206.21	453.81	13.73	0.17	0	0.52
5K6K	1.b	902.45	1008.15	25.22	0.26	0.53	0.83
	2.b	889.25	1135.66	33.77	0.31	0.55	0.84
	1.a	778.21	852.62	23.79	0.21	0.44	0.85
	2.a	752.7	843.88	31.09	0.27	0.45	0.88
	3.a	260.16	386.31	11.14	0.44	0.12	0.47
	3.b	258.64	414.12	9.4	0.46	0.13	0.41
5GS6	1.b	1160.06	1336.87	26.63	0.28	0.6	0.81
	1.a	577.05	505.01	20.96	0.23	0.31	0.84
	2.b	550.05	644.1	19.31	0.32	0.33	0.81
	2.a	511.39	501.47	20.42	0.2	0.25	0.82
	3.a	260.26	393.83	10.89	0.47	0.13	0.46
	3.b	188.89	292.77	8.94	0.43	0.03	0.32



Supplementary Discussion Figure SDF3 – DoGSiteScorer pockets for dimers. (A) – 4O6C (WNV), (B) – 4O6D (WNV), (C) – 5GS6 (ZIKV), (D) – 5K6K (ZIKV). The coloring of the domains and pockets and the interpretation of the arrows are similar to Figure 3.



Supplementary Discussion Figure SDF4 – DoGSiteScorer pockets for monomeric structures. FTSite pockets for monomeric structures. (A) — 4O6C (WNV), (B) — 4O6D (WNV), (C) — 5GS6 (ZIKV), (D) — 5K6K (ZIKV). Domains and pockets are coloured similarly to Figure 4

For the monomeric structures 4O6C and 4O6D, the pockets in the β-barrel region clearly have calculation defects, since in both cases the pocket surface intersects with the protein structure. In the other cases, no such errors were observed. However, this suggests that the program may overestimate the volumes of potential binding sites.

SUPPLEMENTARY TABLES

Supplementary table ST1 – Basic information about the described NS1 structures available in the PDB.

PDB ID	Virus	Year	Method	Completeness of the structure (resolved areas are indicated)	Oligomeric state	Resolution, Å	Source (DOI)
4O6B	DENV2	2014	X-ray diffraction	A: 0-7, 11-107, 129-158, 166-349 B: 1-106, 131-161, 164-349	Dimer	3.0005	science.1247749
4OIG	DENV1	2014	X-ray diffraction	C-domain: 178-352	Dimer	2.6900	10.1073/pnas.1322036111
5IY3	ZIKV	2016	X-ray diffraction	C-domain: 176-352	Dimer	2.2000	10.1038/nsmb.3213
5K6K	ZIKV	2016	X-ray diffraction	A:-5 - 112, 121-352 B: -1 - 352	Dimer	1.8900	10.1038/nsmb.3268
5GS6	ZIKV	2016	X-ray diffraction	A: 1-352 B: -4 - 25, 34-352	Dimer	2.8520	10.15252/embj.201695290
5X8Y	ZIKV	2017	X-ray diffraction	C-domain: A,B: 176-351	Dimer	2.8170	10.1038/srep42580
5YXA	YFV	2018	X-ray diffraction	C-domain: 176-352	Monomer	2.10	10.1007/s11427-017-9238-8
4O6D	WNV	2014	X-ray diffraction	A:-6 - 107, 129-352 B: -4 - 107, 130-352	Dimer	2.5936	10.1126/science.1247749
4O6C	WNV	2014	X-ray diffraction	A,B: 0-109, 125-352	Dimer	2.7508	10.1126/science.124774]
4OII	WNV	2014	X-ray diffraction	C-domain A,B: 176-352	Complex with antibodies	3.0	10.1073/pnas.1322036111
4OIE	WNV	2014	X-ray diffraction	C-domain 176-352	Monomer	1.85	10.1073/pnas.1322036111

PDB ID	Virus	Year	Method	Completeness of the structure (resolved areas are indicated)	Oligomeric state	Resolution, Å	Source (DOI)
4TPL	WNV	2014	X-ray diffraction	A: -10-107, 129-352 B: -20 - 108,130-352	Dimer	2.90	10.1107/S1399004714017556
5O19	JEV	2018	X-ray diffraction	C-domain 177-352	Monomer	2.10	10.1128/JVI.01868-17
5O36	JEV	2018	X-ray diffraction	C-domain 177-354	Monomer	2.60	10.1128/JVI.01868-17
7WUS	DENV2	2022	Cryo Electron microscopy	Full structure	Dimer	3.40	10.1038/s41467-022-34415-1
7WUT	DENV2	2022	Cryo Electron microscopy	Full structure	Tetramer	3.5	10.1038/s41467-022-34415-1
7WUU	DENV2	2022	Cryo Electron microscopy	Full structure	Tetramer	8.0	10.1038/s41467-022-34415-1
7WUV	DENV2	2022	Cryo Electron microscopy	Full structure	Tetramer	8.3	10.1038/s41467-022-34415-1
7WUR	DENV2	2022	Cryo Electron microscopy	Full structure	Complex with antibodies	3.5	10.1038/s41467-022-34415-1

Supplementary table ST2 – Results of stereochemical validation of NS1 structures. The numbers and percentages of residues belonging to the corresponding region of the Ramachandran plots are indicated.

Structure	4O6B	4O6D	4TPL	4O6C	5K6K	5GS6
favoured plot areas	475 (85.7%)	510 (86.6%)	525 (86.9%)	536 (91.5%)	524 (89.0%)	492 (81.6%)
allowed areas of the plot	76 (13.7%)	76 (12.9%)	76 (12.6%)	46 (7.8%)	64 (10.5%)	104 (17.2%)
"generously allowed" areas of the plot	1 (0.2%)	1 (0.2%)	3 (0.5%)	2 (0.3%)	3 (0.5%)	6 (1.0%)
unfavourable areas of the plot	2 (0.4%)	2 (0.3%)	0 (0.0%)	2 (0.3%)	0 (0.0%)	1 (0.2%)

Supplementary table ST4 — Epidemiologically significant orthoflaviviruses

Species	Virus name	Abbreviation (Eng)	GenBank ID
Tick-borne viruses			
Orthoflavivirus kyasanurense	Kyasanur Forest disease virus	KFDV	AY323490
Orthoflavivirus langatense	Langat virus	LGTV	AF253419
Orthoflavivirus loupingi	Louping ill virus	LIV	Y07863
Orthoflavivirus omskense	Omsk hemorrhagic fever virus	OHFV	AY193805
Orthoflavivirus powassanense	Powassan virus	POWV	L06436
Orthoflavivirus encephalitidis	Tick-borne Encephalitis Virus - European subtype	TBEV-Eur	U27495
	Tick-borne Encephalitis Virus - Far-eastern subtype	TBEV-FE	X07755
	Tick-borne Encephalitis Virus - Siberian subtype	TBEV-Sib	L40361
Mosquito-borne viruses			
Orthoflavivirus denguei	Dengue virus 1	DENV-1	U88536
	Dengue virus 2	DENV-2	U87411
	Dengue virus 3	DENV-3	M93130
	Dengue virus 4	DENV-4	AF326573
Orthoflavivirus japonicum	Japanese encephalitis virus	JEV	M18370
Orthoflavivirus murrayense	Murray Valley encephalitis virus	MVEV	AF161266
Orthoflavivirus louisense	St. Louis encephalitis virus	SLEV	DQ525916
Orthoflavivirus nilense	West Nile virus	WNV	M12294
Orthoflavivirus zikaense	Zika virus	ZIKV	AY632535
Orthoflavivirus flavi	yellow fever virus	YFV	X03700

Supplementary table ST4. Kolmogorov-Smirnov test statistics for ensembles of NS1 protein models combined by reference structure

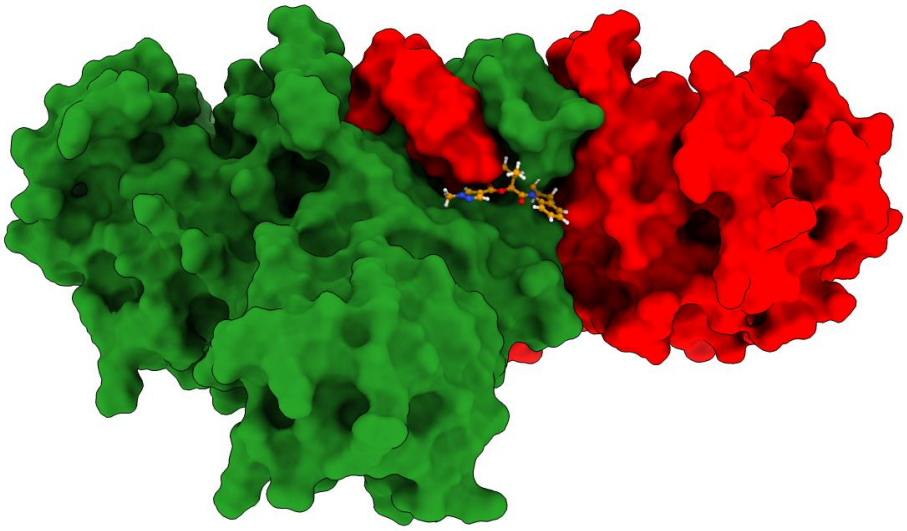
Reference structure	Mosquito-borne VS All		Tick-borne VS All		Mosquito-borne VS Tick-borne	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
4O6B	0.4239	0	0.6406	0	0.2784	$4.29*10^{-170}$
4O6C	0.4215	0	0.6671	0	0.3051	$8.35*10^{-205}$
4O6D	0.4207	0	0.6418	0	0.2766	$7.25*10^{-168}$
5GS6	0.5747	0	0.6744	0	0.1363	$1.12*10^{-40}$
5K6K	0.4623	0	0.5688	0	0.1316	$5.54*10^{-38}$
7WUS	0.4587	0	0.6113	0	0.2036	$1.26*10^{-90}$

Supplementary table ST5. Number of molecules for which the normalised sum of ranks is less than or equal to 500 for ensembles of mode

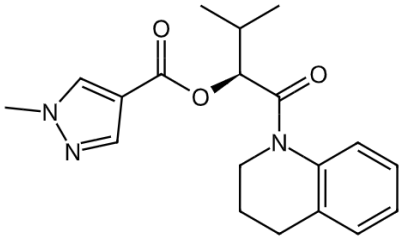
PDB ID	All viruses	Mosquito-borne	Tick-borne
4O6B	81	95	170
4O6C	67	117	100
4O6D	79	91	129
5K6K	40	54	63
5GS6	121	193	145
7WUS	124	173	167

SUPPLEMENTARY FIGURES

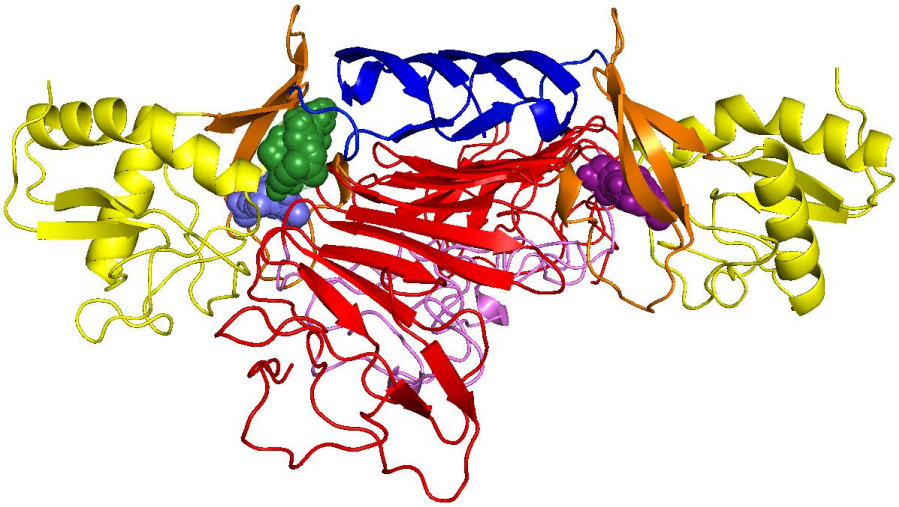
A



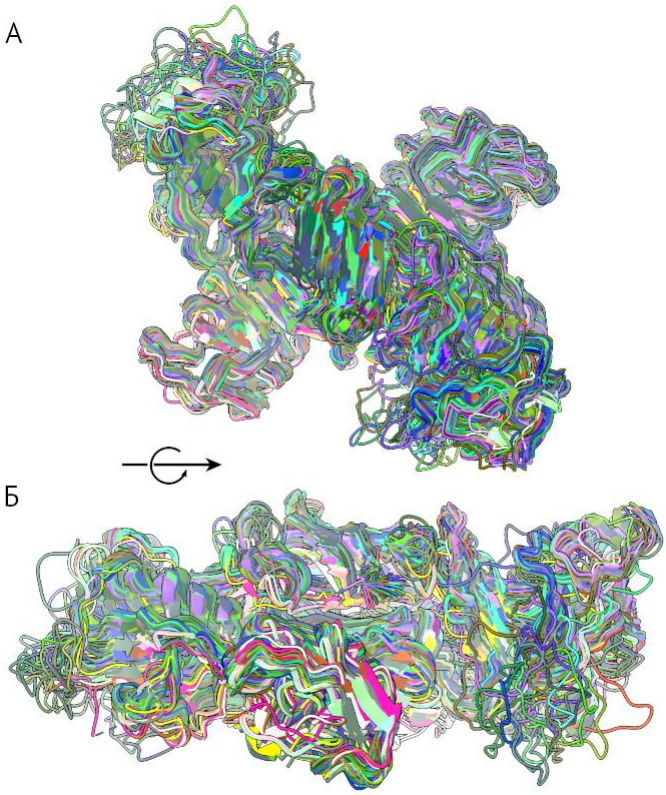
Б



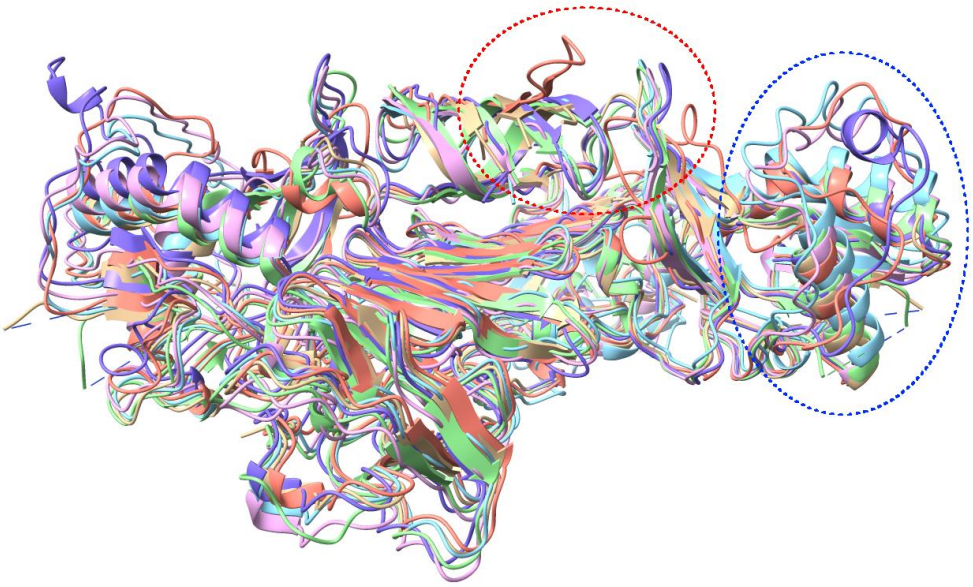
В



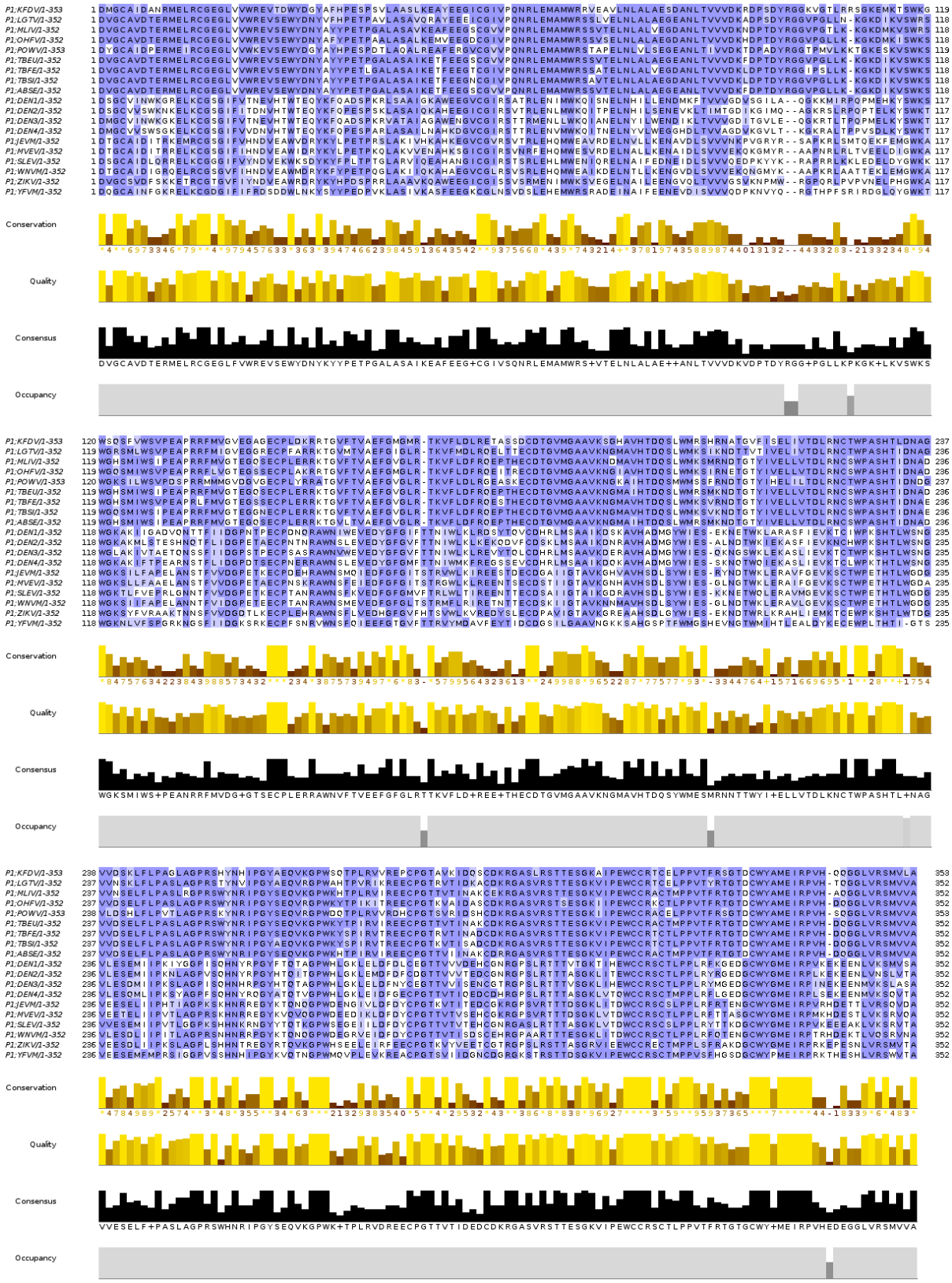
Supplementary figure SF1. Docking site location. (A) Position of the reference ligand ZINC000734046780 in the surface pocket of NS1 protein (PDB ID: 4O6D). (B) Structure of ZINC000734046780. (C) Pockets predicted by FTSite, which were used to label the docking site.



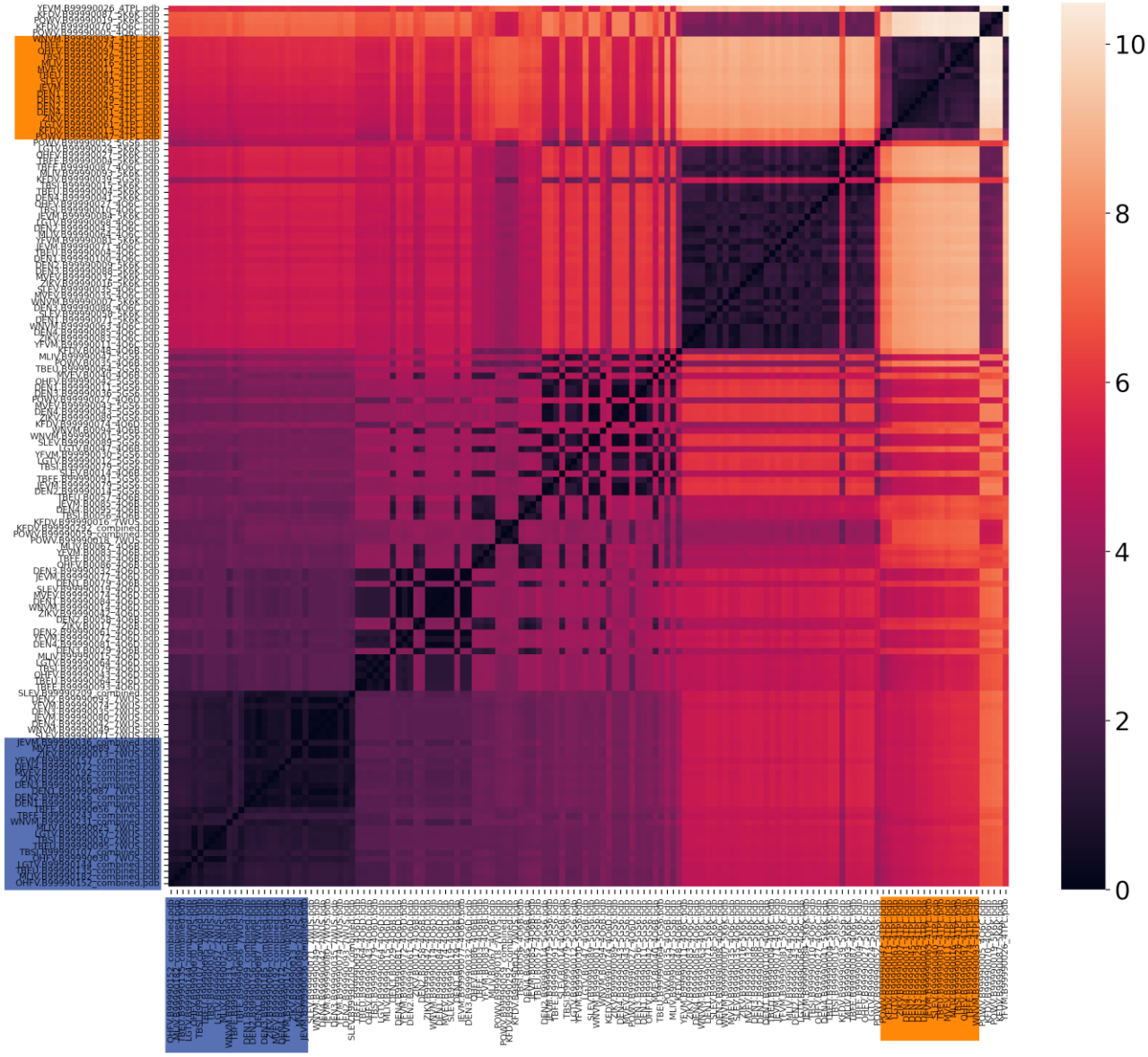
Supplementary figure SF2. Spatial alignment of NS1 protein models of epidemiologically significant orthoflaviviruses.



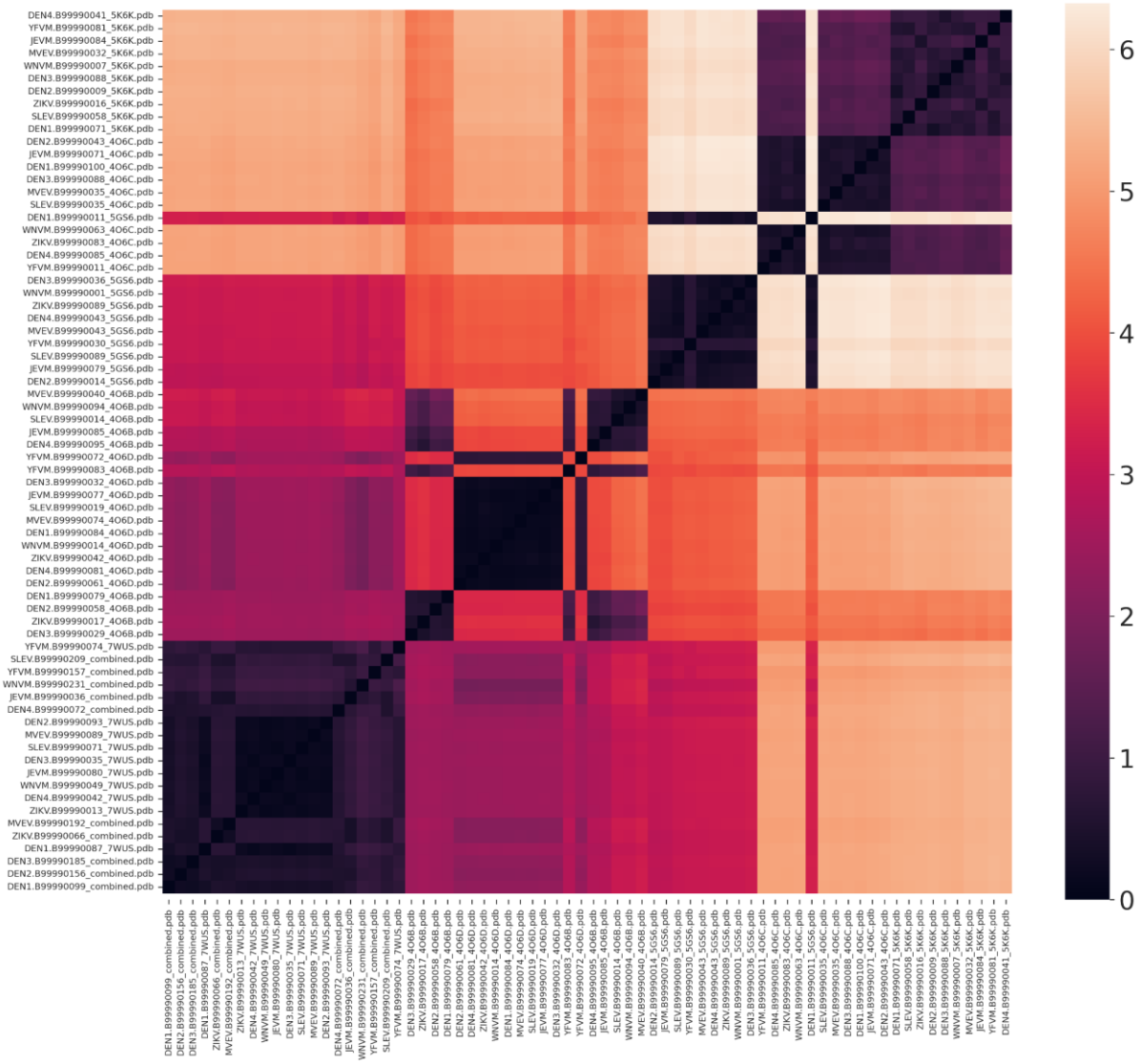
Supplementary figure SF3. Spatial alignment of complete dimeric crystal structures of the NS1 protein of orthoflaviviruses. The regions of the protein that contribute the most to the increase in pairwise RMSD are highlighted: red - β-barrel, residues 6-14; blue - α-wing, residues 62-82.



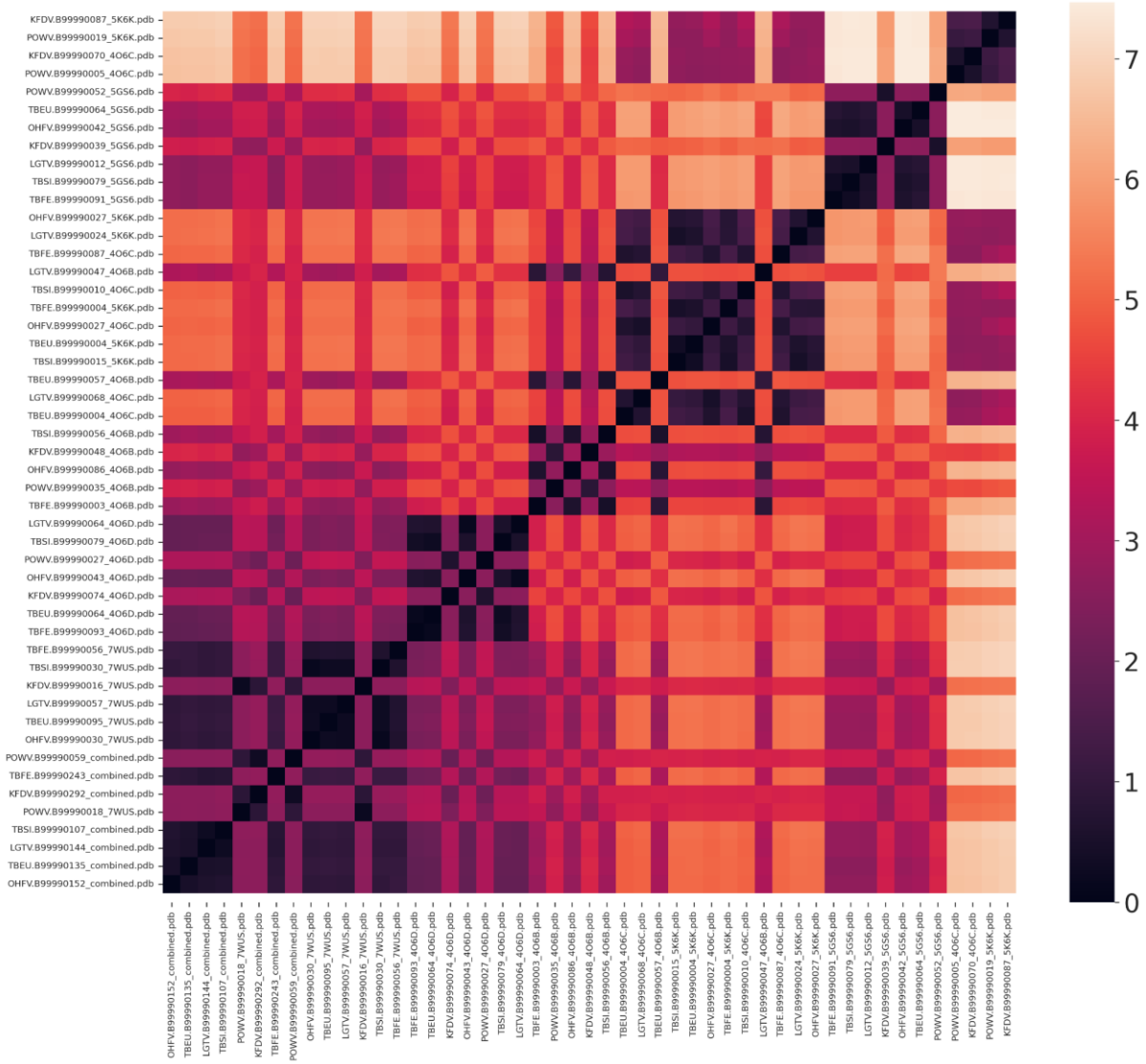
Supplementary figure SF4. Amino acid sequence alignment of NS1 proteins of orthoflaviviruses, stained with BLOSUM62 [10.1038/nbt0308-274].



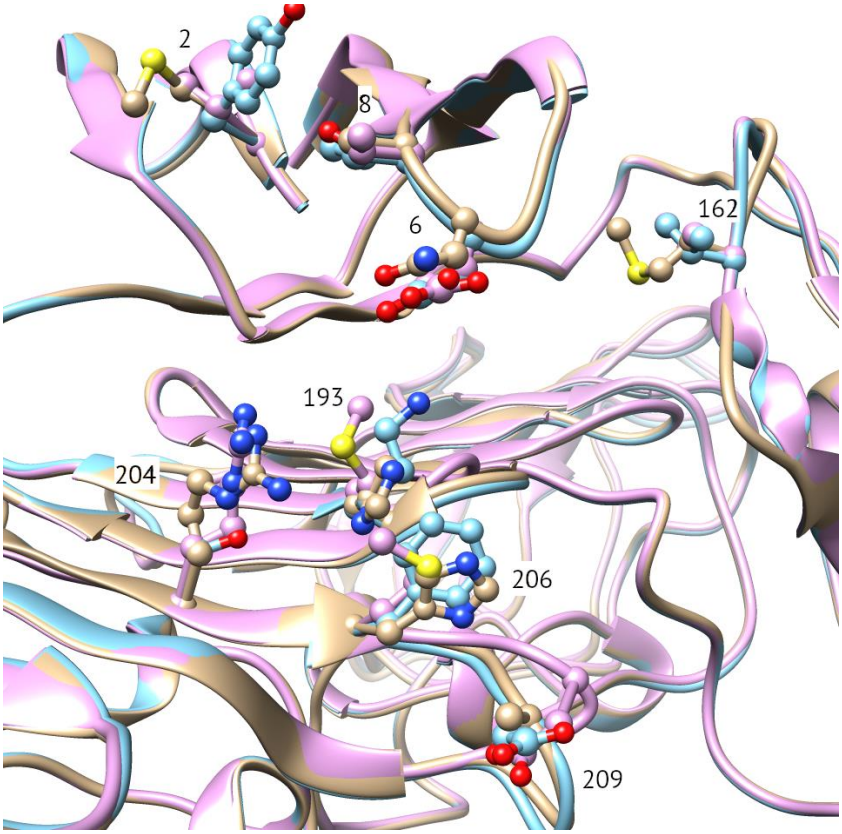
Supplementary Figure SF5. Heat map of pairwise RMSDs between Cα-atoms of residues forming the pocket in the dimerization site region. Models are ordered by average pairwise RMSD. Models built using the 4TPL template are highlighted in orange, and models built using the combined template are highlighted in blue.



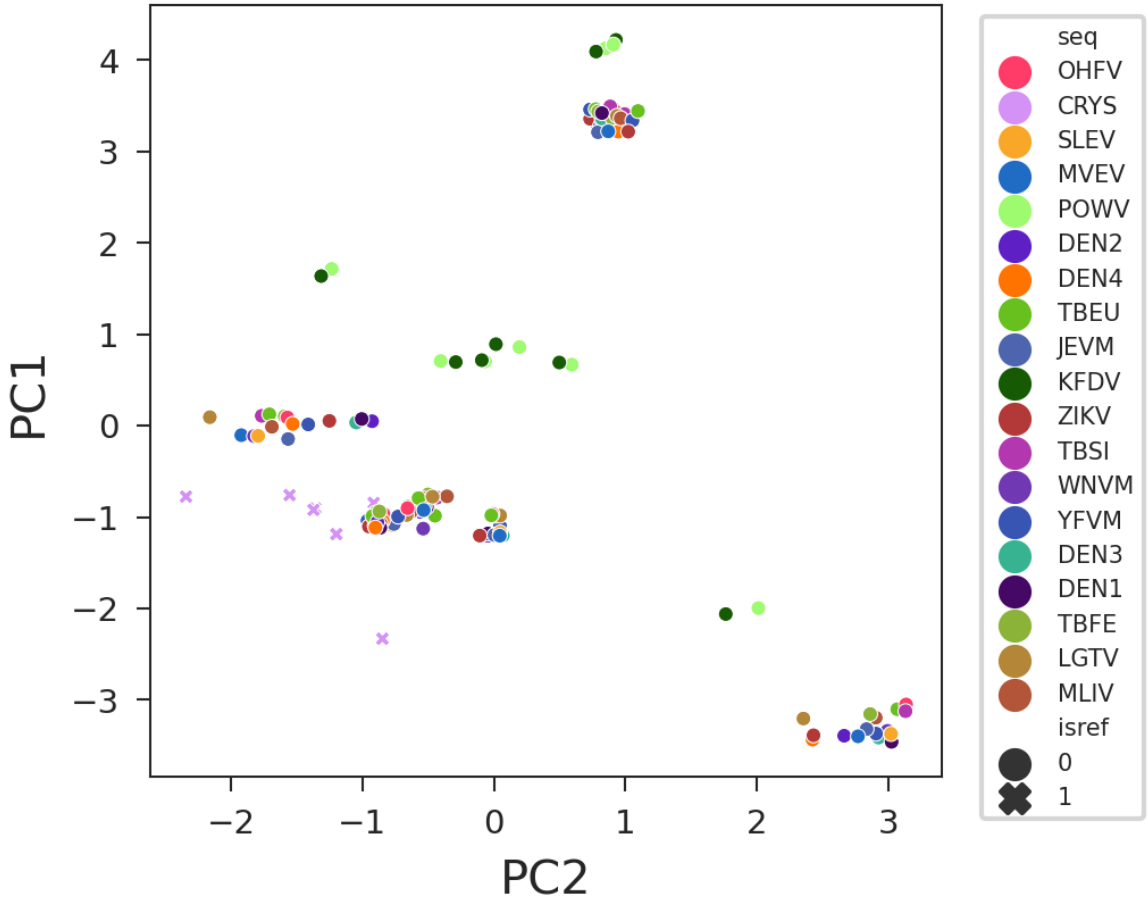
Supplementary figure SF6. Heat map of pairwise RMSDs between Cα atoms forming the target pocket for NS1 protein models of mosquito-borne viruses



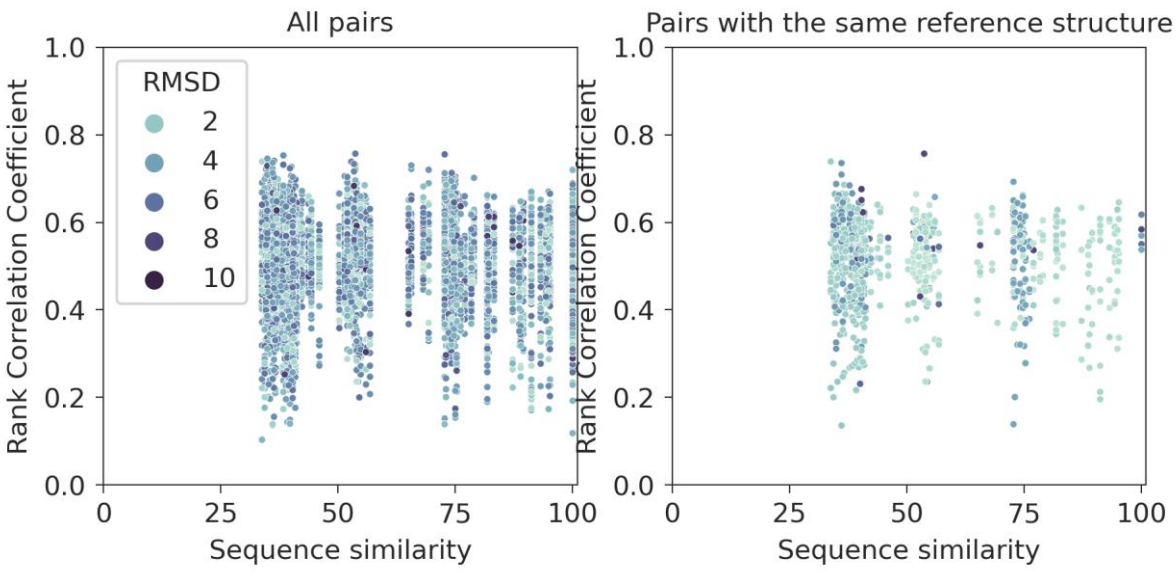
Supplementary figure SF7. Heat map of pairwise RMSDs between Cα atoms forming the target pocket for NS1 protein models of tick-borne viruses



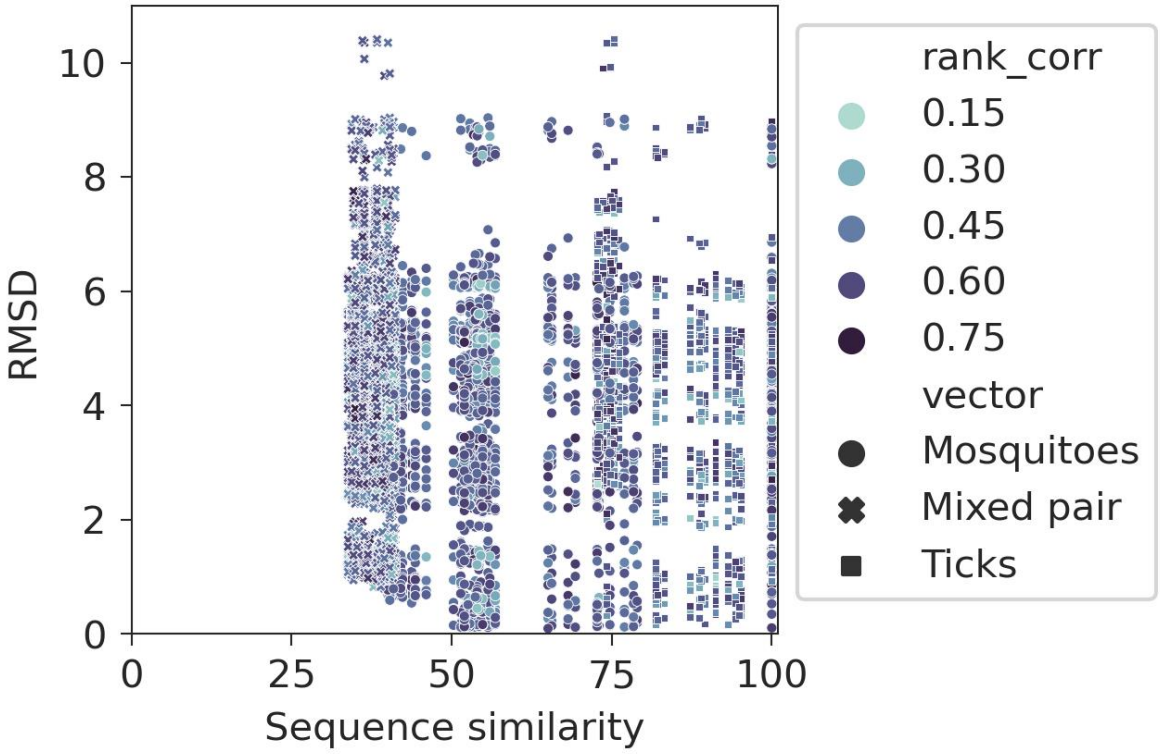
Supplementary figure SF8. Models of POWV NS1 (beige), KFDV (blue) and TBEV-Eu (pink) proteins, showing amino acid residues with non-conservative substitutions occur in POWV and KFDV NS1, their positions are given according to the POWV NS1 protein sequence.



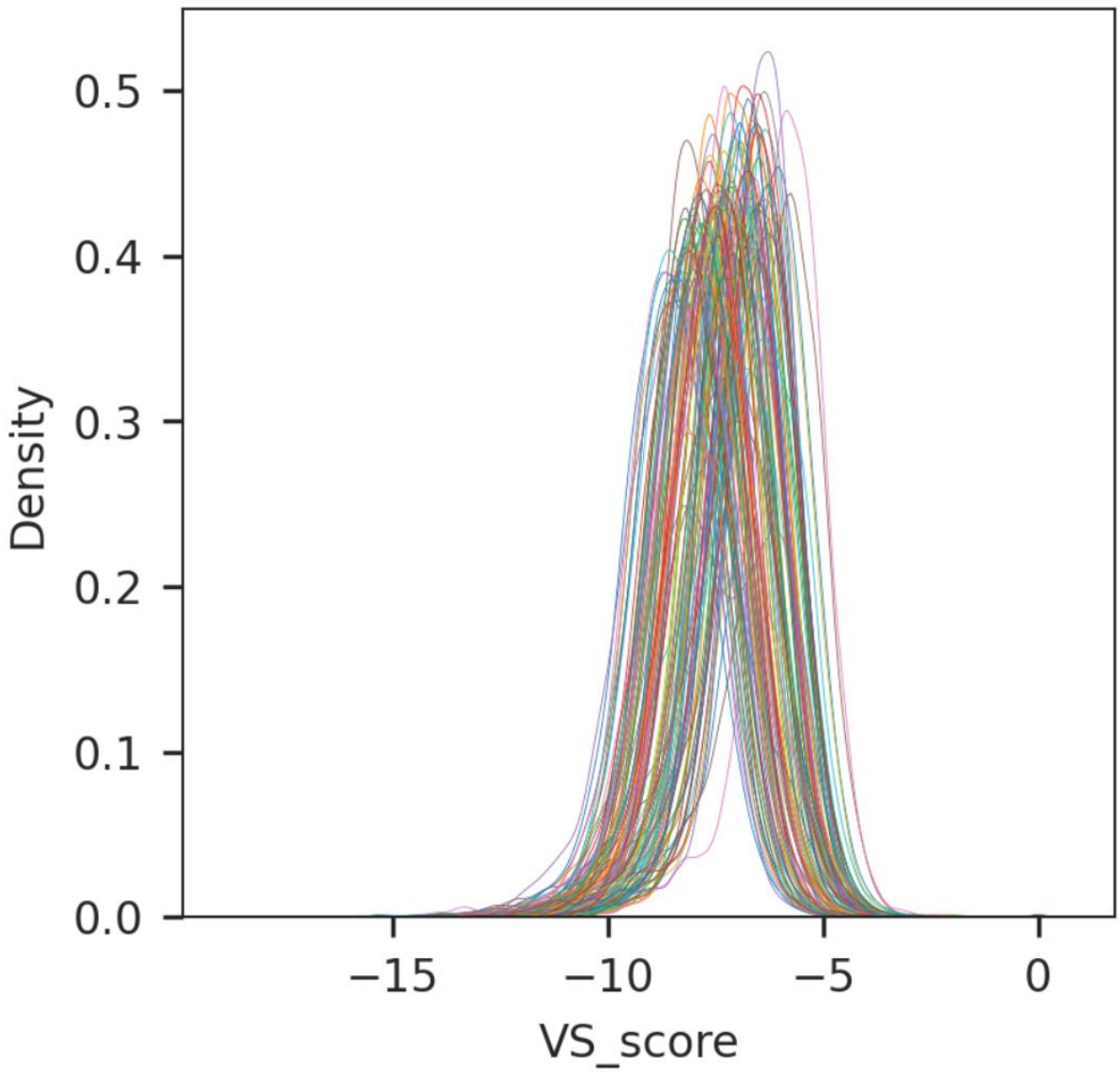
Supplementary figure SF9. Principal component analysis of the coordinates of Ca atoms of residues forming the orthoflaviviruses NS1 protein target pocket of models and reference crystal structures. Models are coloured by viruses. Models are shown as circles, reference structures - as crosses.



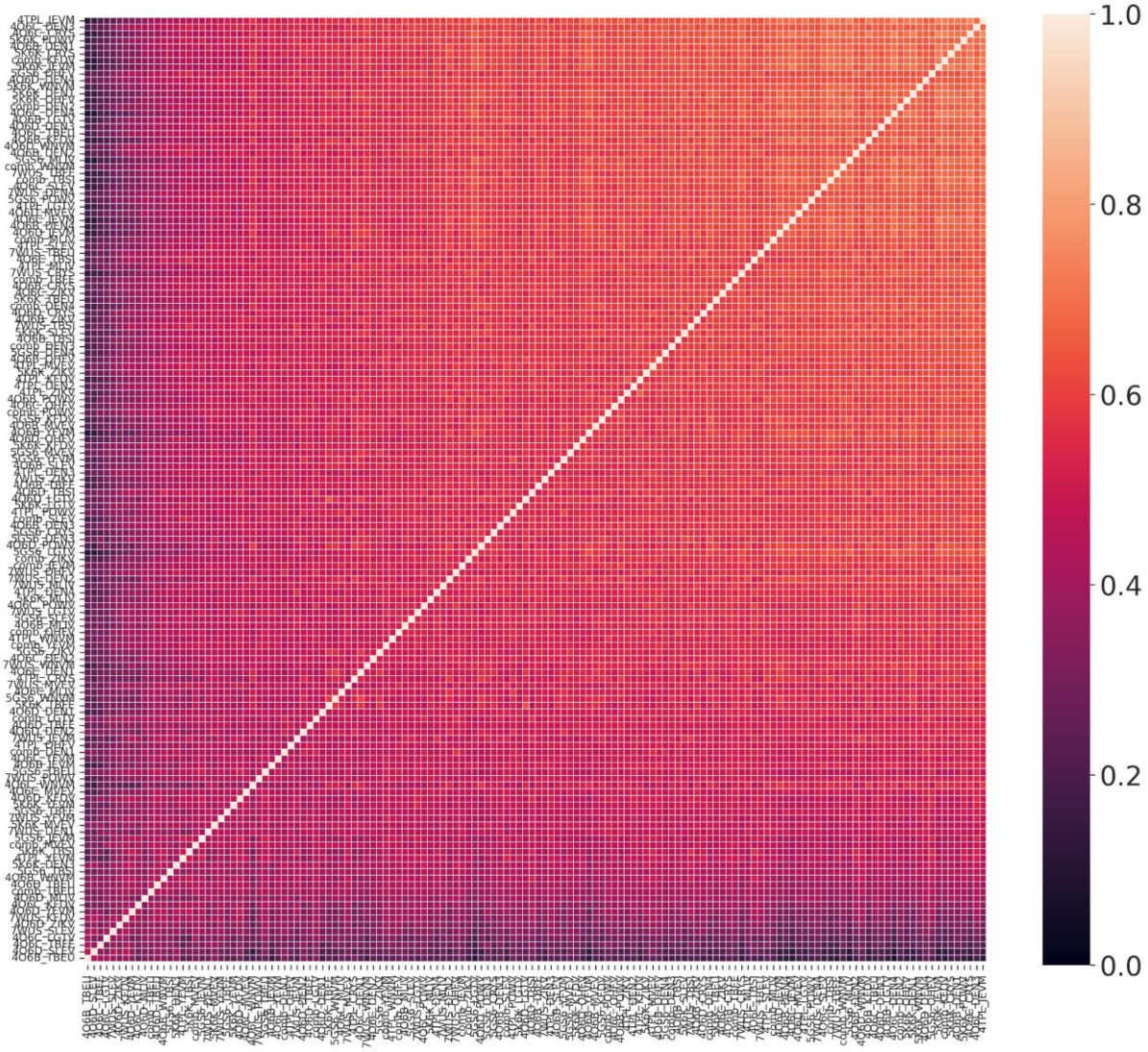
Supplementary figure SF10. Dependence of the pairwise correlation coefficient of the docking results ranks in the structure of the NS1 protein models of flaviviruses on the percentage of similarity between the corresponding sequences, coloring by pairwise RMSD between the corresponding structures.



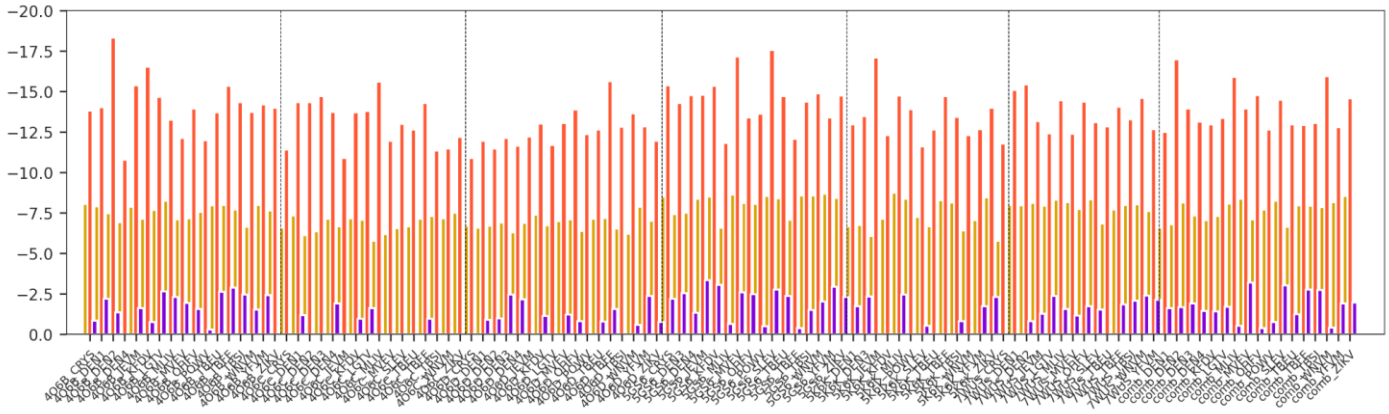
Supplementary figure SF11. Pairwise RMSD between Ca atoms of residues forming a potential binding pocket in NS1 protein models of different orthoflaviviruses versus the percentage of similarity of their amino acid sequences. Coloring of dots by the degree of correlation of ranks of docking results in the corresponding models.



Supplementary figure SF12. Docking scores distribution for a library of 5000 diverse compounds from the ZINC15 database



Supplementary figure SF13. Rank-order correlation coefficients between docking results of diverse drug-like compounds into model structures and crystal structures of the orthoflaviviruses NS1 protein.



Supplementary figure SF14. Minimum (red) and maximum (purple) docking scores of a diverse compound library in the orthoflavivirus NS1 protein model and modes (yellow) of the docking scores distributions.