Electronic Supplementary Material (ESI) for RSC Chemical Biology.

Helix-based screening with structure prediction using artificial intelligence has potential for the rapid development of peptide inhibitors targeting class I viral fusion

Satoshi Suzuki,[a] Mio Kuroda,[b] Keisuke Aoki,[b,c] Kumi Kawaji,[d] Yoshiki Hiramatsu,[a] Mina Sasano,[d] Akie Nishiyama,[a] Kazutaka Murayama,[e] Eiichi N Kodama,[a,d,f] Shinya Oishi,[b] and Hironori Hayashi[d]*

[a.] *Department of Infectious Diseases, Tohoku University Graduate School of Medicine, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan. E-mail: satoshi-suzuki@med.tohoku.ac.jp*

[b.] *Laboratory of Medicinal Chemistry, Kyoto Pharmaceutical University, 1, Misasagi-Shichono-cho, Yamashina-ku, Kyoto 607-8412, Japan*

[c.] *Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan*

[d.] *Division of Infectious Diseases, International Research Institute of Disaster Science, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan*

[e.] *Division of Biomedical Measurements and Diagnostics, Graduate School of Biomedical Engineering, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan*

[f.] *Department of Infectious Diseases, Graduate School of Medicine and Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan*

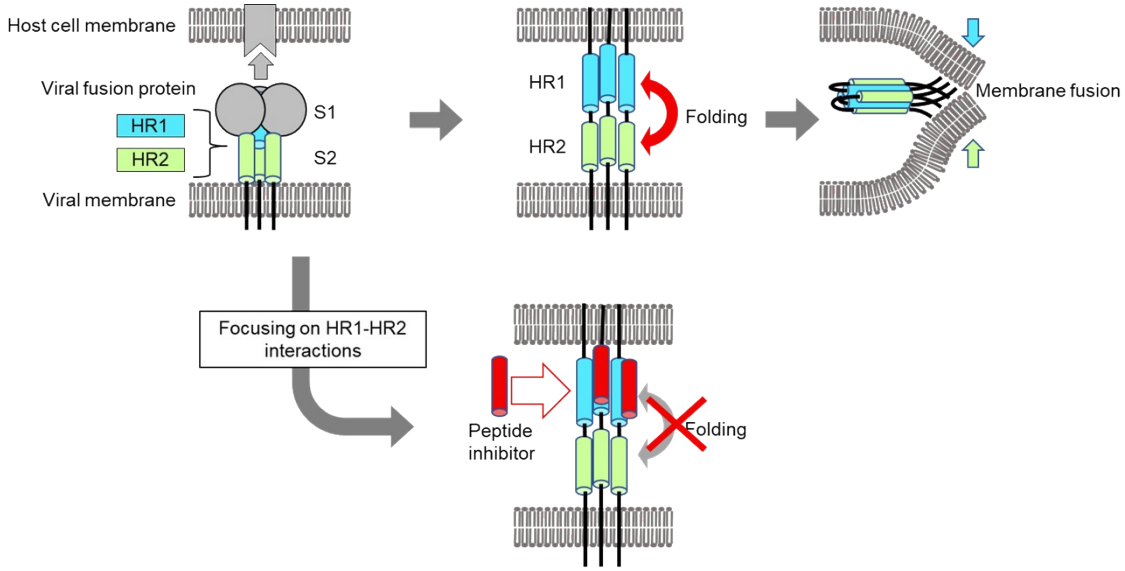

Fig. S1. Mechanism of virus-cell fusion and the inhibition by peptides.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) uses two spike (S) proteins to enter the host cells. The S1 protein binds to receptors of the host cells, such as angiotensin-converting enzyme 2. This binding activates the S2 protein and leads to penetration of the N-terminal fusion domain into the host cell membrane. After anchoring the cell and viral membranes with the S2 protein, the heptad repeat 1 (HR1) and heptad repeat 2 (HR2) interact with each other and form a six helix bundle (6HB) through intramolecular folding. The 6HB formation induces the fusion of the membranes. Agents preventing the intramolecular folding of the HR1-HR2 complex can inhibit viral entry.

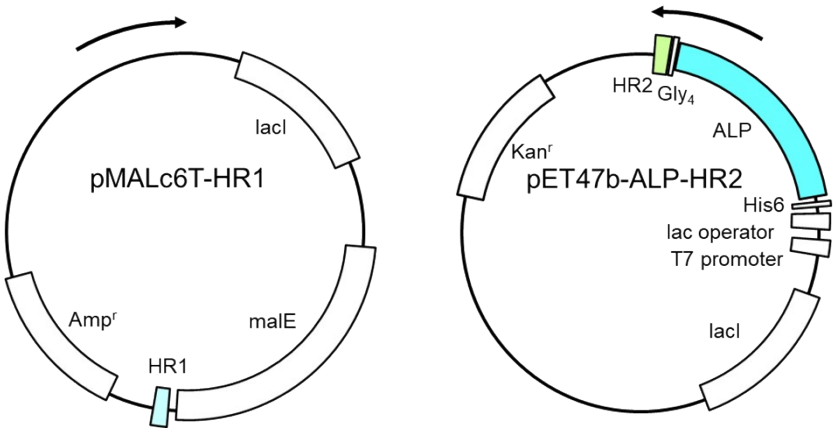

Fig. S2. Construction of plasmids.

The amplified DNA fragment coding HR1 was inserted into the pMAL-c6T vector to generate maltose binding protein (MBP)-fused HR1 (MBP-HR1). The DNA fragment of the alkaline phosphatase (ALP) coding region (amino acids 22−471) from *Escherichia coli*, JM109 genome cDNA for a tetra-glycine (Gly_4) linker, and cDNA for HR2 were sequentially inserted downstream of the original a hexa-histidine tag (His6) coding frame of a pET47b vector. The different lengths of DNA fragments derived from the HR2 of SARS-CoV-2 were inserted into the pET47b-ALP vector. ALP-HR2 was designed to be expressed with a His6 fused to the N-terminal amino acids of ALP.

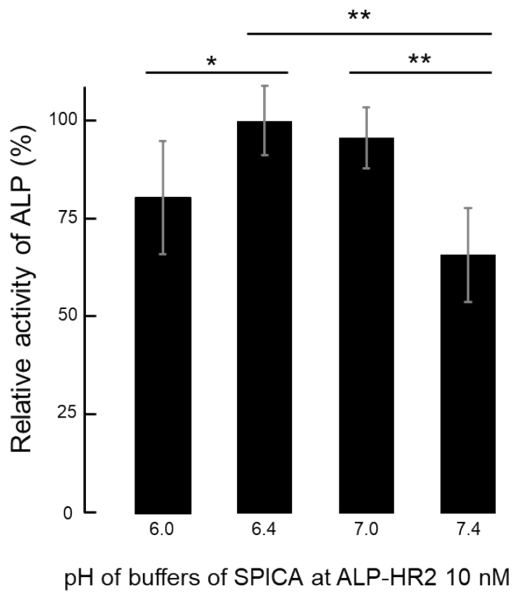

Fig. S3. Optimization of the pH value for the interaction of ALP-HR2 with MBP-HR1 in SPICA.
The pH value of the SPICA buffer (to match the optimal buffer range for each buffer, 100 mM MES at pH 6.0 and 6.4, 100 mM HEPES at pH 7.0, PBS at pH 7.4, with 100 mM NaCl and 2 mM $MgCl_2$, were used) influenced the activity of ALP. The activity of ALP at different pH values was calculated relative to the ALP activity at pH 6.4. Sigle asterisk means p value < 0.05, double asterisks mean p value < 0.01.

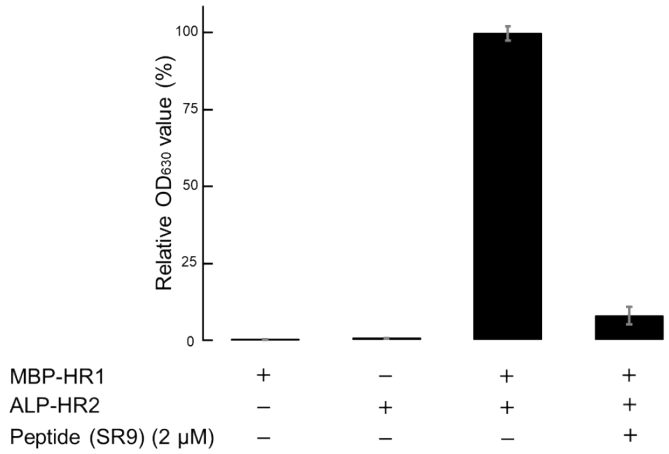

Fig. S4. SPICA with control well.

Little ALP activity was observed as indicated by the absorbance at 630 nm (OD_{630}) with MBP-HR1 or ALP-HR2 alone, but ALP activity was detected with MBP-HR1 and ALP-HR2. The activity of ALP was suppressed with SR9.

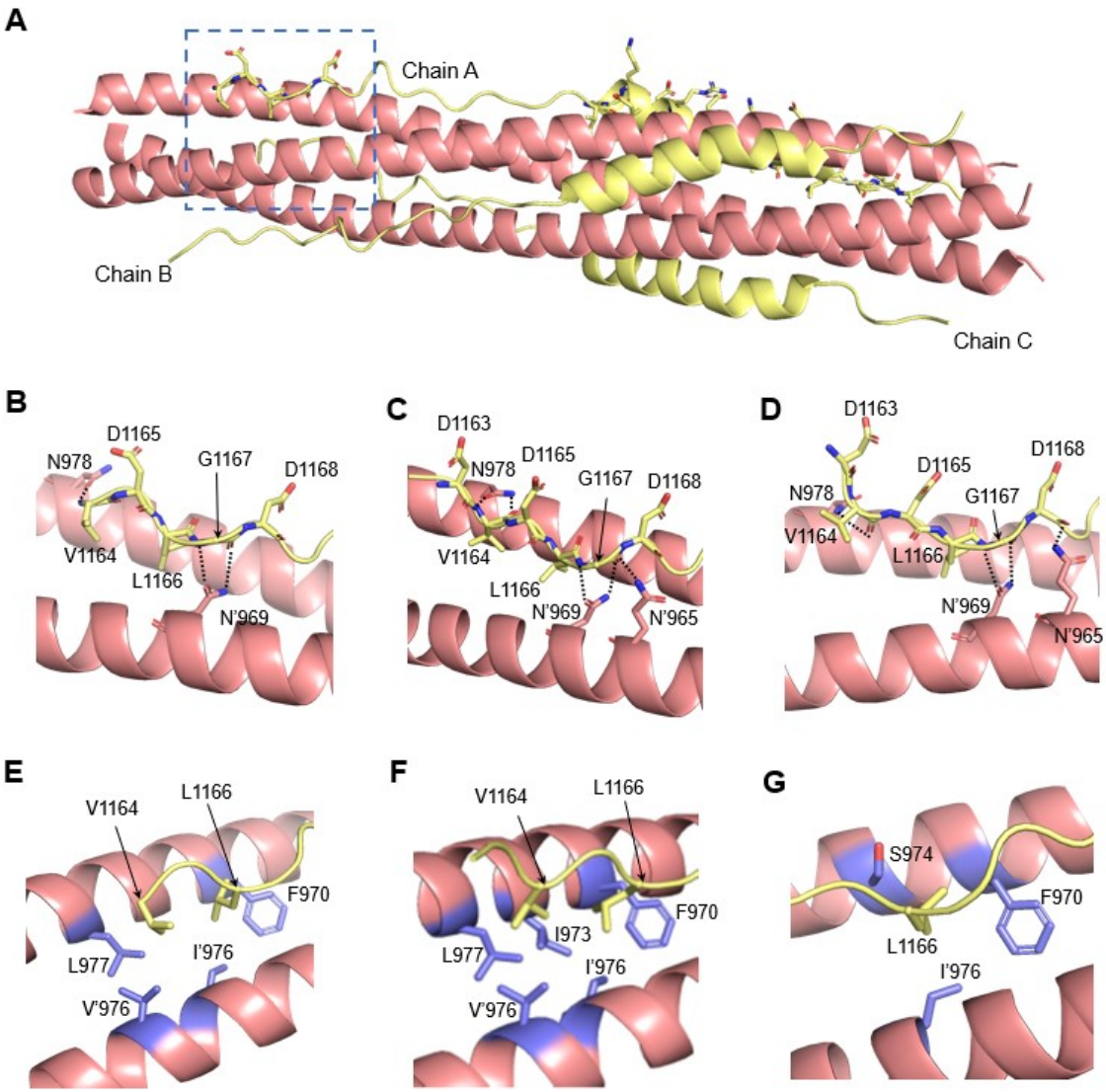

Fig. S5. Binding mode between DVDLGD and a HR1 trimer in previous reported crystal structure. (A) showed a crystal structure of the six-helix bundle formed by a HR1 trimer and three HR2s (Chain A, B and C). HR1 and HR2 in 6HB (6LXT) are shown in ribbon model, and colored pink and light yellow, respectively. DVDLGD was surrounded by a dotted line. (B) to (D) showed hydrogen bonds (H-Bonds) formed between DVDLGD and HR1 trimer. DVDLGD was displayed in stick model. Amino acids of HR1 forming H-bonds with DVDLGD are shown in stick model. Black dotted lines mean H-bonds. (E) to (G) exhibited amino acids forming hydrophobic interactions. Amino acids forming hydrophobic interactions with V1164 and/or L1166 of HR2 were displayed in stick model and colored light blue. Additionally, (B) and (E) show interactions around chain A in the crystal structure. (C) and (F) are chain B. (D) and (G) are chain C.

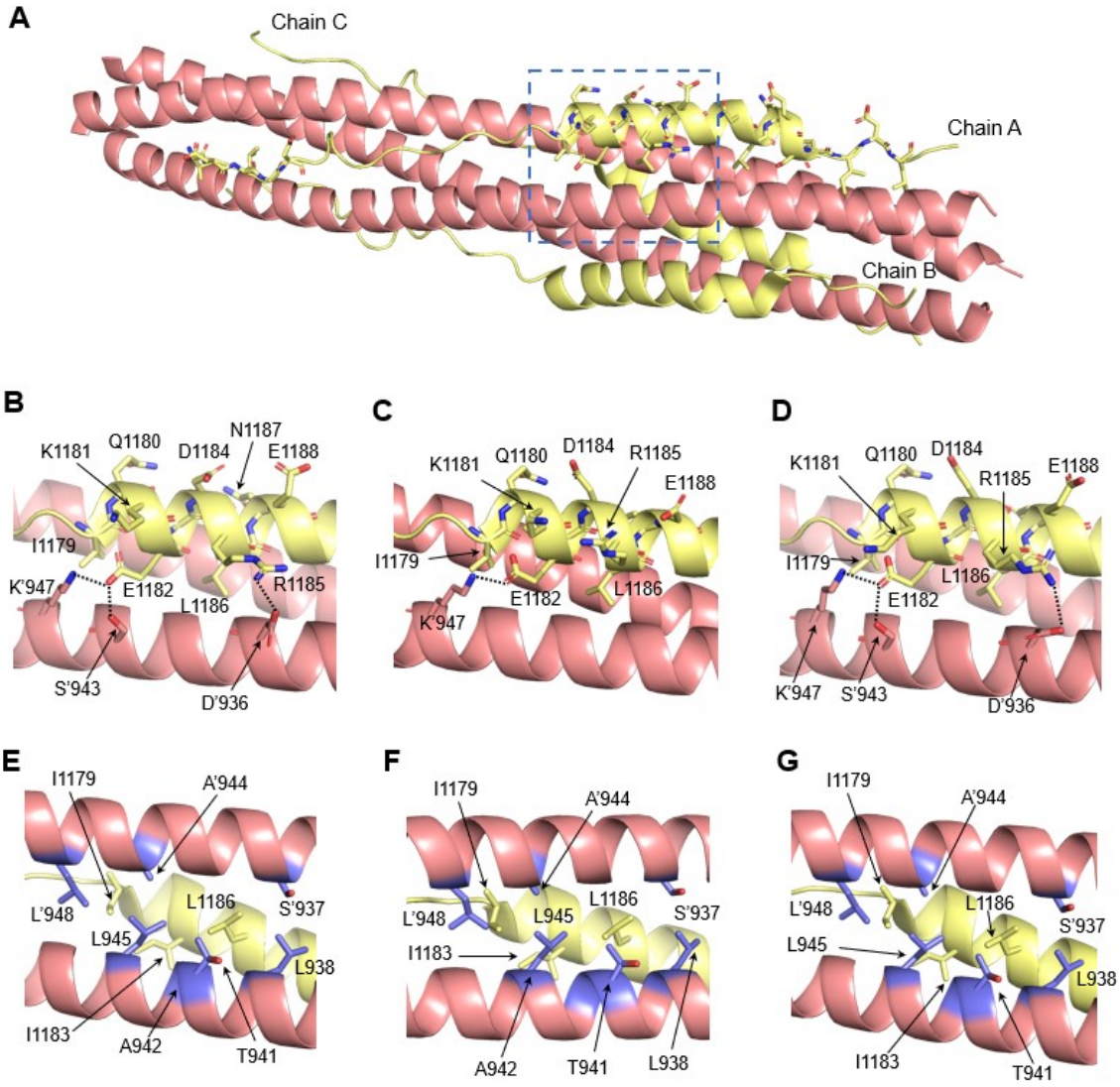

Fig. S6. Binding mode between IQKEIDRLNE and a HR1 trimer in previous reported crystal structure.

(A) showed a crystal structure of the six-helix bundle formed by a HR1 trimer and three HR2s (Chain A, B and C). HR1 and HR2 in 6HB (6LXT) are shown in ribbon model, and colored pink and light yellow, respectively. IQKEIDRLNE was surrounded by a dotted line. (B) to (D) showed hydrogen bonds (H-Bonds) formed between IQKEIDRLNE and HR1 trimer. IQKEIDRLNE is displayed in stick model. Amino acids of HR1 forming H-bonds with IQKEIDRLNE are shown in stick model. Black dotted lines mean H-bonds. (E) to (G) exhibited amino acids forming hydrophobic interactions. Amino acids forming hydrophobic interactions with I1179, I1183 and/or L1186 of HR2 were displayed in stick model and colored light blue. Additionally, (B) and (E) show interactions around chain A in the crystal structure. (C) and (F) are chain B. (D) and (G) are chain C.

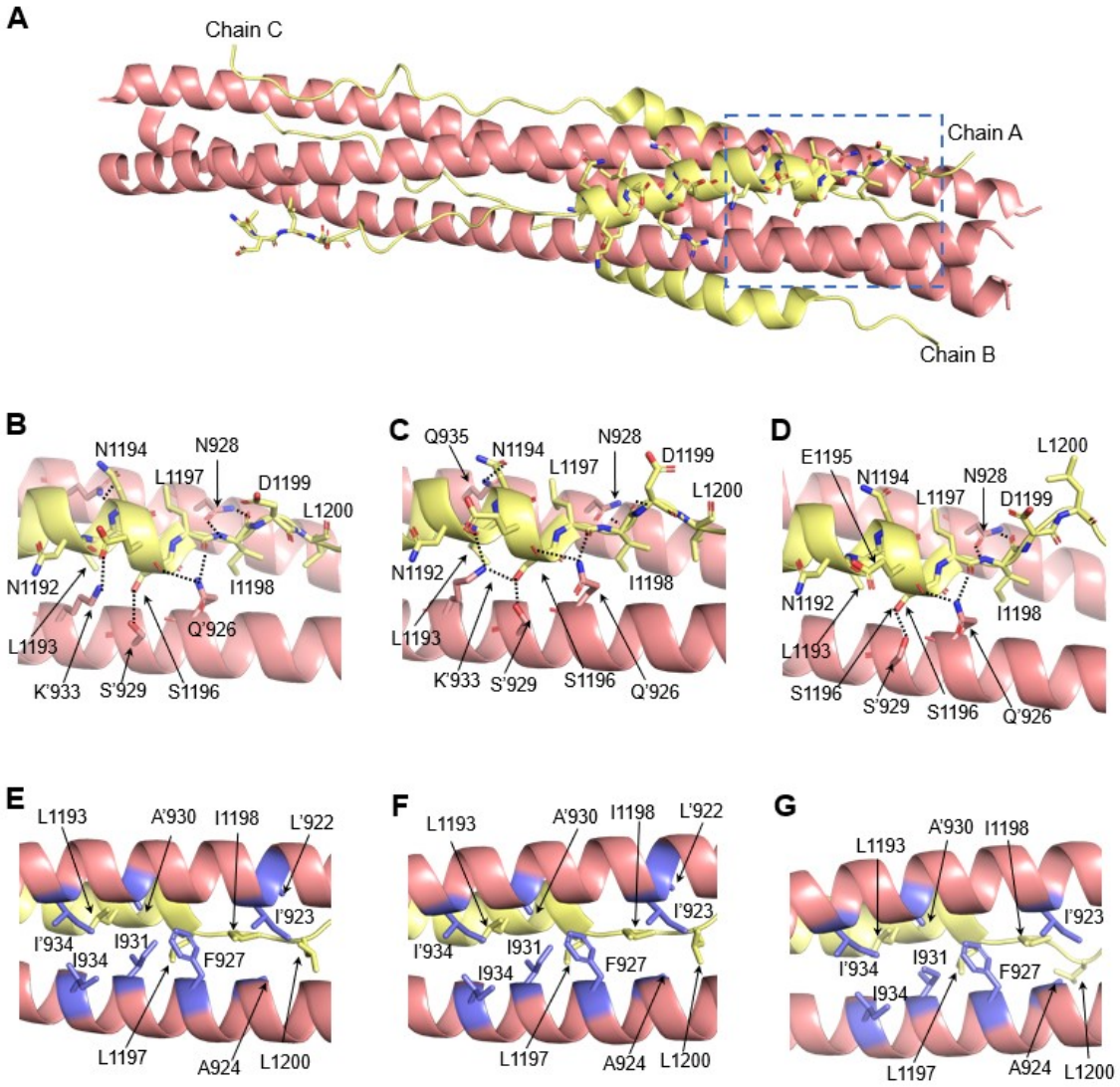

Fig. S7. Binding mode between NLNESLIDL and a HR1 trimer in previous reported crystal structure. (A) showed a crystal structure of the six-helix bundle formed by a HR1 trimer and three HR2s (Chain A, B and C). HR1 and HR2 in 6HB (6LXT) are shown in ribbon model, and colored pink and light yellow, respectively. NLNESLIDL was surrounded by a dotted line. (B) to (D) showed hydrogen bonds (H-Bonds) formed between NLNESLIDL and HR1 trimer. NLNESLIDL was displayed in stick model. Amino acids of HR1 forming H-bonds with NLNRSLIDL are shown in stick model. Black dotted lines mean H-bonds. (E) to (G) exhibited amino acids forming hydrophobic interactions. Amino acids of HR1 forming hydrophobic interactions with L1193, L1197, I1198 and/or L1200of HR2 were displayed in stick model and colored light blue. Additionally, (B) and (E) show interactions around chain A in the crystal structure. (C) and (F) are chain B. (D) and (G) are chain C.

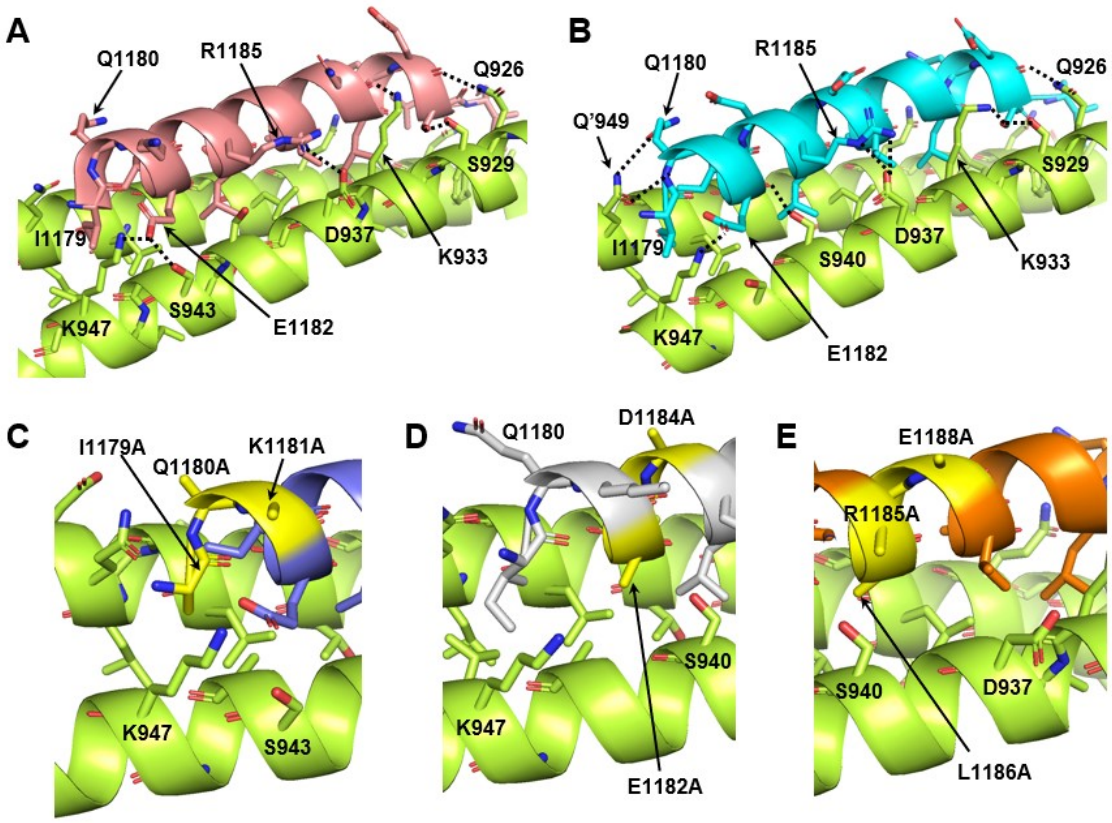

Fig. S8. Predicted 3D structures of Vir^{SC2}4 and derivatives.

(A) Crystal structure (PDB code: 6LXT) of the 6HB from SARS-CoV-2. Light yellow and light pink indicate the α-helices in the HR2 and HR1 regions, respectively. The dotted lines indicate hydrogen bonds (H-bonds). **(B)** Predicted 3D structure of the 6HB formed by Vir^{SC2}4 and HR1. Cyan represents the α-helix in Vir^{SC2}4. Light green represents HR1. The dotted lines indicate predicted H-bonds. **(C)** to **(E)** N-terminal region of the predicted structures formed by the Vir^{SC2}4 derivatives and HR1. Blue, white, and orange represents Vir^{SC2}4$_{IQK}$, Vir^{SC2}4$_{EID}$, and Vir^{SC2}4$_{RLNE}$, respectively. Light green represents HR1. Substitutions with alanine in Vir^{SC2}4 are indicated in yellow. The 3D structures suggested that alanine substitutions in Vir^{SC2}4 resulted in the loss of H-bonds and/or hydrophobic interactions.

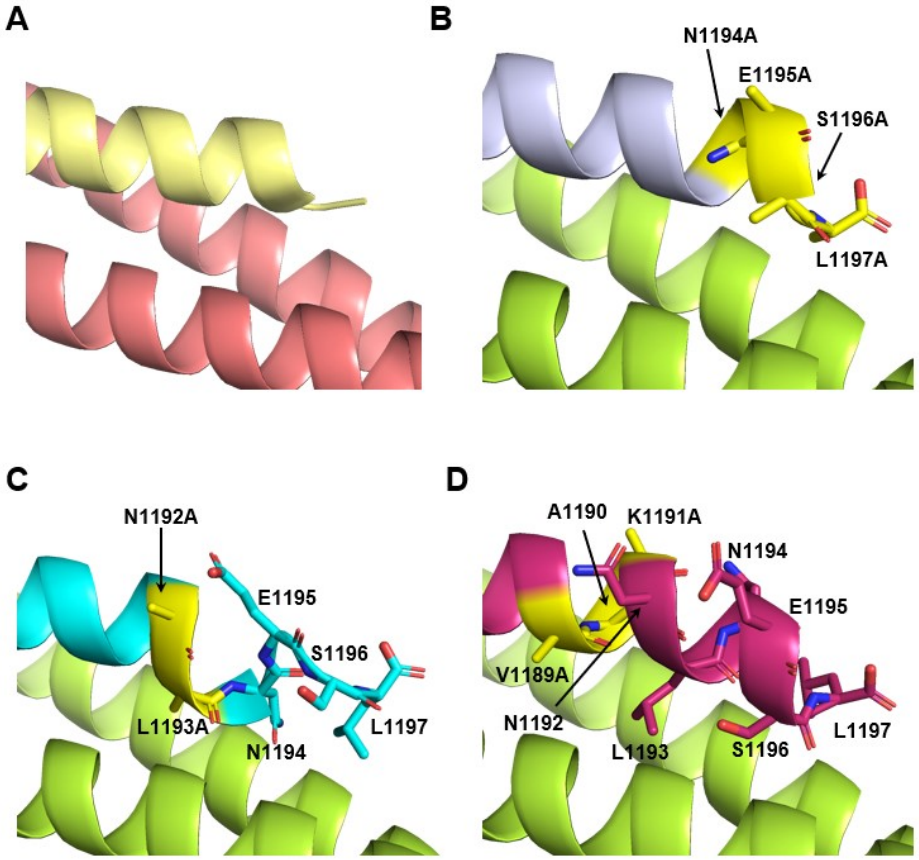

Fig. S9. Predicted 3D structures of Vir^{SC2}8 and derivatives.

(A) Light yellow and light pink represents C-terminal HR2 region of the α-helix and HR1 derived from a previously reported crystal structure, respectively (PDB code: 6LXT). **(B)** to **(D)** C-terminal region in the predicted structures of the complexes of Vir^{SC2}8 derivatives with HR1. **(B)** Vir^{SC2}8$_{NESL}$ (light blue). **(C)** Vir^{SC2}8$_{NL}$ (cyan). **(D)** Vir^{SC2}8$_{VAK}$ (wine red). HR1 is shown in light green. Alanine substitutions in Vir^{SC2}8 are indicated in yellow.

Table S1. Peptide inhibitors reported in previous studies.

Peptide	Amino-acid sequence	Reference
EK1	SLDQINVTFLDLEYEMKKLEEAIKKLEESYIDLKEL	[1]
IPB1 (SR9)	ISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL	[2]
#2	DISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL	[3]
IPB19	SVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIK	[4]
EKL1	NVTFLDLEYEMKKLEEAIKKLEESYIDLKELGTYEY	[5]
AntiSCV2P1	AAVEQIQKEIERLQEAAKNLLESLK	[6]
AntiSCV2P7	ASVRNIQEEIKRLNEVAKKLAESLI	[6]
longHR2_42	PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL	[7]
42G	PDVDLGGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL	[7]
HR2-P40	VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL	[8]
3a	ISGINASVVNIQEEIKRLNEVAKKLNESLIDLQE	[9]

Table S2. Interactions between amino acids in HR1 and 2 measured by protein crystallography.

Structure of HR2	Amino acids of HR2 (chain A)	Amino acids of HR1 (chain A)	Amino acids of HR1 (chain B)	Binding	Consistency with the results of SPICA
Disordered	Asp1163	Asn978		Hydrogen	+
	Val1164	Asn978		Hydrogen**	+
		Leu977		Hydrophobic	+
			Val976	Hydrophobic	+
	Leu1166		Ile973	Hydrophobic	+
		Phe970		Hydrophobic	+
	Gly1167		Asn969	Hydrogen**	+
	Ile1169	Leu966		Hydrophobic	-
			Leu962*	Hydrophobic	-
		Val963		Hydrophobic	-
	Ile1172	Val963		Hydrophobic	-
			Leu962	Hydrophobic	-
			Ala958	Hydrophobic	-
	Asn1173	Asn960		Hydrogen	-
	Ala1174	Asn960		Hydrogen**	-
		Leu959		Hydrophobic	-
		Ala956		Hydrophobic	-
	Ser1175		Asn955	Hydrogen**	-
			Gln954	Hydrogen	-
	Val1176	Ala956		Hydrophobic	-
	Val1177	Asn953		Hydrogen**	-
		Val952		Hydrophobic	-
			Val951	Hydrophobic	-
		Gln949		Hydrogen**	-
			Leu948	Hydrophobic	-
Helix	Ile1179		Leu948	Hydrophobic	+
		Leu945		Hydrophobic	+
			Ala944	Hydrophobic	+
	Glu1182		Lys947	Hydrogen	+
			Ser943	Hydrogen	+
	Ile1183	Leu945		Hydrophobic	+
		Ala942		Hydrophobic	+
	Arg1185		Asp936	Hydrogen	+
	Leu1186	Leu938*		Hydrophobic	+
		Leu945		Hydrophobic	+
		Ala942		Hydrophobic	+
		Thr941		Hydrophobic	+
			Ser937	Hydrophobic	+
	Val1189	Leu938		Hydrophobic	-
	Ala1190	Leu938		Hydrophobic	-
		Gln935		Hydrogen**	-
	Asn1192		Lys933	Hydrogen**	+
	Leu1193	Leu938		Hydrophobic	+
		Ile934		Hydrophobic	+
		Ile931		Hydrophobic	+
			Ala930	Hydrophobic	+
	Asn1194	Gln935		Hydrogen	+
	Glu1195		Gln926	Hydrogen**	+
	Ser1196		Ser929	Hydrogen	+
	Leu1197	Ile931		Hydrophobic	+
			Gln926	Hydrogen**	+
Disordered	Ile1198	Asn928		Hydrogen**	+
		Phe927		Hydrophobic	+
		Ala924		Hydrophobic	+
			Ile923	Hydrophobic	+
			Leu922	Hydrophobic	+
	Leu1200	Ala924		Hydrophobic	+
			Ile923	Hydrophobic	+
	Gln1201		Asn919	Hydrogen**	-
		Tyr917*		Hydrogen**	-

The crystal structure from the Protein Data Bank data, 6LXT[10] (Asp1163 from 6XRA[11]), was analyzed and compared with the results of SPICA. * indicates hydrogen bonds or hydrophobic interactions observed only on some chains; ** indicates hydrogen bonds to peptide bonds. The gray columns show hydrophobic interactions. The structure of Leu1203 to Pro1213 was not able to be observed in the crystal structure and cryo-electron microscopy (cryo-EM) structure.[11]

Table S3. Nucleotide sequence of the primers for RT-PCR, the sequence reaction, and cloning.

	Primer name	Sequences
RT-PCR	RT-F	5’- CTTGGTGATATTGCTGCTAGAGA -3’
	RT-R	5'- CGTCTTCATCAAATTTGCAGCAGGA -3'
For Sequencing	HR2-F1	5'- CACGATGCGTCCGGCGTAGAGGATCGA -3'
	HR2-F2	5'- GGACTACGTCACCGACTCGGCTGCA -3'
	HR2-F3	5'- AGAGGATTCACAAGAACATACCGGCA -3'
	HR2-R1	5'- GCGCGTCCCATTCGCCAATCCGGATA -3'
	HR2-R2	5'- CTTTGGTATTTAGCGCCTGGGTGA -3'
	T7 pro	5'- TAATACGACTCACTATAGGG -3'
	T7 term	5'- ATGCTAGTTATTGCTCAGCGG-3'
	HR1-F1	5'- GACAATTAATCATCGGCTCGTATA -3'
	HR1-F2	5'- CCCGCCAAAAACCTGGGAAGAGA -3'
	HR1-R1	5'- TCCGCTCCCGGCGGATTTGTCCTA -3'
	HR1-R2	5'- TCGTAAGACTTCAGCGCTACGGCA -3'
For Cloning	ALP-F	5'- GACGGATCCCCGGACACCAGAAATGCCTGTTCTGGAA -3'
	ALP-R	5'- CCCAAGCTTACCACCACCACCTTTCAGCCCCAGAGCGGCTTTCATGGTG -3'
	HR1-F	5'- CCCGTCGACACACAGAATGTTCTCTATGAGAACCA -3'
	HR1-R	5'- CCCGGATCCTTAAAGACGTGAAAGGATATCATTTAAAA -3'
	HR2-F	5'- CCCAAGCTTGATGTTGATTTAGGTGACATCTCTGGCA -3'
	HR2-F-Vir^{SC2}1	5'- CCCAAGCTTATCTCTGGCATTAATGCTTCAGTTGTAAA -3'
	HR2-F-Vir^{SC2}2	5'- CCCAAGCTTAATGCTTCAGTTGTAAACATTCAAA -3'
	HR2-F-Vir^{SC2}3	5'- CCCAAGCTTGTTGTAAACATTCAAAAAGAAA -3'
	HR2-F-Vir^{SC2}4	5'- CCCAAGCTTATTCAAAAAGAAATTGACCGCCTCA -3'
	HR2-F-Vir$^{SC2}4_{IQK}$	5'- CCCAAGCTTGCAGCAGCAGAAATTGACCGCCTCAATGAGGTTGCCA -3'
	HR2-F-Vir$^{SC2}4_{EID}$-1	5'- GCAGCAGCACGCCTCAATGAGGTTGCCAAGAATTTA -3'
	HR2-F-Vir$^{SC2}4_{EID}$-2	5'- CCCAAGCTTATTCAAAAAGCAGCAGCACGCCTCAA -3'
	HR2-F-Vir$^{SC2}4_{RLNE}$-1	5'- GCAGCAGCAGCAGTTGCCAAGAATTTAAATGAATCTCTCA -3'
	HR2-F-Vir$^{SC2}4_{RLNE}$-2	5'- CCCAAGCTTATTCAAAAAGAAATTGACGCAGCAGCAGCAGTTG -3'
	HR2-R	5'- CCCCTCGAGTTATGGCCATTTTATATACTGCTCATA -3'
	HR2-R-Vir^{SC2}5	5'- CCCCTCGAGTTACTGCTCATACTTTCCAAGTTCTTGGAGA -3'
	HR2-R-Vir^{SC2}6	5'- CCCCTCGAGTTAAAGTTCTTGGAGATCGATGAGAGATTCA -3'
	HR2-R-Vir^{SC2}7	5'- CCCCTCGAGTTAGAGATCGATGAGAGATTCATTTA -3'
	HR2-R-Vir^{SC2}8	5'- CCCCTCGAGTTAGAGAGATTCATTTAAATTCTTGGCA -3'
	HR2-R-Vir^{SC2}9	5'- CCCCTCGAGTTATAAATTCTTGGCAACCTCATTGA -3'
	HR2-R-Vir$^{SC2}8_{NESL}$-1	5'- GGCGGCGGCGGCTAAATTCTTGGCAACCTCATTGA -3'

HR2-R-Vir$^{SC2}8_{NESL}$-2	5'- CCCCTCGAGTTAGGCGGCGGCGGCTAAATTCTTGGCA -3'
HR2-R-Vir$^{SC2}8_{NL}$-1	5'- GAGAGATTCATTTGCTGCCTTGGCAACCTCATTGAGGCGGTCA -3'
HR2-R-Vir$^{SC2}8_{NL}$-2	5'- CCCCTCGAGTTAGAGAGATTCATTTGCTGCCTTGG -3'
HR2-R-Vir$^{SC2}8_{VAK}$-1	5'- TAAATTTGCTGCTGCCTCATTGAGGCGGTCAATTTCTTTTTGAA -3'
HR2-R-Vir$^{SC2}8_{VAK}$-2	5'- CCCCTCGAGTTAGAGAGATTCATTTAAATTTGCTGCTGCCTCA -3'

Underlining indicates restriction enzyme sites. Orange letters indicate a tetra-glycine linker (Gly_4).

References

1. Xia S, Zhu Y, Liu M, Lan Q, Xu W, Wu Y, Ying T, Liu S, Shi Z, Jiang S, Lu L, *Cell Mol Immunol*, 2020, **17**, 765–767.
2. Zhu Y, Yu D, Yan H, Chong H, He Y, *J Virol*, 2020, **94**, e00635-20.
3. Kandeel M, Yamamoto M, Tani H, Kobayashi A, Gohda J, Kawaguchi Y, Park BK, Kwon HJ, Inoue JI, Alkattan A, *Biomol Ther (Seoul)*, 2021, **29**, 282–289.
4. Yu D, Zhu Y, Jiao T, Wu T, Xiao X, Qin B, Chong H, Lei X, Ren L, Cui S, Wang J, He Y, *Emerg Microbes Infect*, 2021, **10**, 1227-1240.
5. Zhou J, Xu W, Liu Z, Wang C, Xia S, Lan Q, Cai Y, Su S, Pu J, Xing L, Xie Y, Lu L, Jiang S, Wang Q, *Acta Pharm Sin B*, 2022, **12**, 1652-1661.
6. Behzadipour Y, Hemmati S, *Int J Pept Res Ther*, 2022, **28**, 42.
7. Yang K, Wang C, Kreutzberger AJB, White KI, Pfuetzner RA, Esquivies L, Kirchhausen T, Brunger AT, *Proc Natl Acad Sci USA*, 2023, **120**, e2300360120.
8. Hu Y, Zhu Y, Yu Y, Liu N, Ju X, Ding Q, He Y, *Antiviral Res*, 2023, **212**, 105571.
9. Tsuji K, Baffour-Awuah Owusu K, Miura Y, Ishii T, Shinohara K, Kobayakawa T, Emi A, Nakano T, Suzuki Y, Tamamura H, *RSC Adv*, 2023, **13**, 8779-8793.
10. Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, Qi F, Bao L, Du L, Liu S, Qin C, Sun F, Shi Z, Zhu Y, Jiang S, Lu L, *Cell Res*, 2020, **30**, 343-355.
11. Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM Jr, Rawson S, Rits-Volloch S, Chen B, *Science*, 2020, **369**, 1586-1592.