

Supplementary Materials

ARTICLE

## Self and Nonself Short Constituent Sequences of Amino Acids in the SARS-CoV-2 Proteome for Vaccine Development

Joji M. Otaki <sup>1,\*</sup>, Wataru Nakasone <sup>2</sup> and Morikazu Nakamura <sup>2</sup>

<sup>1</sup> The BCPH Unit of Molecular Physiology, Department of Chemistry, Biology and Marine Science, University of the Ryukyus, Okinawa 903-0213, Japan

<sup>2</sup> Computer Science and Intelligent Systems Unit, Department of Information and Engineering, Faculty of Engineering, University of the Ryukyus, Okinawa 903-0213, Japan

\*e-mail: [otaki@sci.u-ryukyu.ac.jp](mailto:otaki@sci.u-ryukyu.ac.jp)

Supplementary Materials and Methods

Supplementary Tables S1, S2, S3, and S4

## Supplementary Materials and Methods

**Variant proteomes.** We first downloaded 93 variant proteomes in addition to the reference proteome (ASM985889v3). Among them, the 67 variant proteomes that were analyzed for proteome-wide mutations were as follows: ASM993789v1, ASM993791v1, ASM993792v1, ASM993793v1, ASM993794v1, ASM1153697v1, ASM1153700v1, ASM1153701v1, ASM1153706v1, ASM1153707v1, ASM1153708v1, ASM1153713v1, ASM1153714v1, ASM1153722v1, ASM1153723v1, ASM1153726v1, ASM1153735v1, ASM1153736v1, ASM1153742v1, ASM1153745v1, ASM1153747v1, ASM1153750v1, ASM1154508v2, ASM1154509v2, ASM1154512v2, ASM1154513v1, ASM1154528v2, ASM1154529v1, ASM1154532v2, ASM1154533v2, ASM1154534v1, ASM1154535v1, ASM1154542v1, ASM1154545v1, ASM1154548v1, ASM1154550v1, ASM1154553v1, ASM1154555v1, ASM1174199v1, ASM1174200v2, ASM1174201v1, ASM1174 202v1, ASM1174203v1, ASM1153751v1, ASM1153752v2, ASM1153755v1, ASM1153756v1, ASM1153761v1, ASM1153762v1, ASM1153769v1, ASM1153770v1, ASM1153771v1, ASM1153781v1, ASM1153782v2, ASM1153783v2, ASM1153786v2, ASM1153789v2, ASM1153794v2, ASM1153797v2, ASM1153798v2, ASM1153802v2, ASM1153805v2, ASM1153810v2, ASM1154497v2, ASM1154502v2, ASM1154506v2, ASM1154507v2.

Other 26 proteomes that were not analyzed further due to insertions and deletions were as follows: ASM993788v1, ASM993790v1, ASM993805v1, ASM993806v1, ASM994842v1, ASM994846v1, ASM994849v1, ASM994852v1, ASM994855v1, ASM1153693v1, ASM1153715v1, ASM1153729v1, ASM1153732v1, ASM1153739v1, ASM1153746v1, ASM1153778v1, ASM1153801v2, ASM1154503v2, ASM1154516v1, ASM1154523v1, ASM1154524v1, ASM1154527v1, ASM1154549v1, ASM1154554v1, ASM1308859v1, ViralProj15500.

**Point mutations in the spike protein.** The point mutations that were analyzed for the spike mutations in this study were as follows [46–49]: L5F, L18F, D80A, N122Q, N165Q, D215G, N234Q, Q239K, R246I, N331Q, V341I, N343Q, D364Y, V367F, D405V, Q409E, Q414E/P, K417N, I434K, A435S, S438F, N439K, L452R, K458R, D467V, I468F/T, I472V, A475V, V483A, E484K, F490L, P491R, N501Y, V503F, Y508H, R509K, V510L, P521S, A522V, A570D, D614G, Q675H, P681H, A701V, N709Q, T716I, N717Q, T719A, N801Q, A831V, D839Y, D936Y, S939F, S943T, S982A, N1074Q, D1118H, P1263L.

**Table S1. Number of self and nonself SCSs in the SARS-CoV-2 proteins.**

	<b>Number of Self SCS</b>	<b>Number of Nonself SCS</b>	<b>Number of Total SCS</b>	<b>Nonself SCS [%]</b>
<b>ORF1ab</b>	6450	642	7092	9.05
<b>Spike (S)</b>	1172	97	1269	7.64
<b>ORF3a</b>	240	31	271	11.44
<b>Envelope (E)</b>	64	7	71	9.86
<b>Membrane (M)</b>	203	15	218	6.88
<b>ORF6</b>	51	6	57	10.53
<b>ORF7a</b>	111	6	117	5.13
<b>ORF8</b>	104	13	117	11.11
<b>Nucleocapsid (N)</b>	386	29	415	6.99
<b>ORF10</b>	28	6	34	17.65
<b>Total</b>	8809	852	9661	8.82

**Table S2. Distribution of nonself SCS clusters in the SARS-CoV-2 proteins.**

aa	ORF1ab	S	ORF3a	E	M	OR6	OR7a	ORF8	N	ORF10	Total
5aa	152	23	6	0	4	2	1	5	4	0	198
6aa	31	6	1	0	0	1	1	1	1	0	42
7aa	24	1	1	0	0	0	0	0	4	0	30
8aa	21	5	1	0	1	1	1	0	1	1	32
9aa	27	3	1	1	0	0	0	0	1	1	34
10aa	13	1	0	1	1	0	0	0	0	0	16
11aa	8	2	0	0	0	0	0	0	1	0	11
12aa	6	1	0	0	1	0	0	0	1	0	9
13aa	1	0	0	0	0	0	0	1	0	0	2
14aa	6	0	0	0	0	0	0	0	0	0	6
15aa	6	2	0	0	0	0	0	0	0	0	8
16aa	2	1	0	0	0	0	0	0	0	0	3
17aa	1	0	1	0	0	0	0	0	0	0	2
18aa	1	0	1	0	0	0	0	0	0	0	2
19aa	0	0	0	0	0	0	0	0	0	0	0
20aa	1	0	0	0	0	0	0	0	0	0	1
21aa	1	0	0	0	0	0	0	0	0	0	1
22aa	0	0	0	0	0	0	0	0	0	0	0
23aa	0	0	0	0	0	0	0	0	0	0	0
24aa	0	0	0	0	0	0	0	0	0	0	0
25aa	0	0	0	0	0	0	0	0	0	0	0
26aa	0	0	0	0	0	0	0	0	0	0	0
27aa	1	0	0	0	0	0	0	0	0	0	1
≥6aa	150	22	6	2	3	2	2	2	9	2	200

5aa SCSs here are solitary SCSs, which do not form any nonself cluster with other adjacent nonself SCSs.

**Table S3. Self-nonsel status changes of SCSs due to point mutations in 68 variant proteomes of SARS-CoV-2.**

Variant proteome	Serial SCS site number	Protein	Status change site	Original SCS (RefSeq)	Variant SCS	Self/nonsel change
ASM993794v1	3606 (1-3606)	ORF1ab	L3606F	LYENA	FYENA	0 to 1
ASM1153713v1	2705 (1-2705)	ORF1ab	K2705K	KSHNI	KSHSI	1 to 0
ASM1153723v1	901 (1-901)	ORF1ab	S901S	SMATY	SIATY	0 to 1
ASM1153723v1	902 (1-902)	ORF1ab	M902I	MATYY	IATYY	1 to 0
ASM1153723v1	8485 (3-124)	ORF3a	I124I	IMRLW	IMRLL	1 to 0
ASM1153723v1	8489 (3-128)	ORF3a	W128L	WLCWK	LLCWK	1 to 0
ASM1153723v1	8608 (3-247)	ORF3a	H247H	HTIDG	HTIDV	1 to 0
ASM1153736v1	3073 (1-3073)	ORF1ab	E3073E	EYSHV	EYSYV	1 to 0
ASM1153745v1	7888 (2-796)	Spike (S)	D796D	DFGGF	DCGGF	0 to 1
ASM1153762v1	3606 (1-3606)	ORF1ab	L3606F	LYENA	FYENA	0 to 1
ASM1153771v1	1836 (1-1836)	ORF1ab	Y1836Y	YKHIT	YKHII	0 to 1
ASM1153810v2	1856 (1-1856)	ORF1ab	S1856S	SSEYK	SSEYN	0 to 1
ASM1153810v2	1858 (1-1858)	ORF1ab	E1858E	EYKGP	EYNGP	0 to 1
ASM1153810v2	1859 (1-1859)	ORF1ab	Y1859Y	YKGPI	YNGPI	0 to 1
ASM1154548v1	5826 (1-5826)	ORF1ab	R5826R	RNPAW	RNLAW	0 to 1
ASM1154548v1	5827 (1-5827)	ORF1ab	N5827N	NPAWR	NLAWR	1 to 0
ASM1154548v1	5863 (1-5863)	ORF1ab	S5863S	SEYDY	SECDY	0 to 1
ASM1174201v1	9552 (9-340)	Nucleocapsid (N)	D340D	DDKDP	DDKVP	0 to 1
ASM1174202v1	1607 (1-1607)	ORF1ab	I1607V	IKPHN	VKPHN	1 to 0

Redundant sites in plural variant proteomes are not included.

Status change site shows the first aa of SCS. For example, K2705K indicates that KSHNI changed to KSHSI.

Self and nonself SCSs are indicated by 0 and 1, respectively.

**Table S4. Self-nonsel status changes of SCSs due to point mutations in the spike protein.**

Status change site	Mutation site	Original SCS (RefSeq)	Variant SCS	Self/nonsel change
<b>C15C</b>	L18F	CVNLT	CVNFT	0 to 1
<b>R78R</b>	D80A	RFDNP	RFANP	0 to 1
<b>A163A</b>	N165Q	ANNCT	ANQCT	0 to 1
<b>N164N</b>	N165Q	NNCTF	NQCTF	0 to 1
<b>P230P</b>	N234Q	PIGIN	PIGIQ	1 to 0
<b>F338F</b>	V341I	FGEVF	FGEIF	0 to 1
<b>C361C</b>	D364Y	CVADY	CVAYY	1 to 0
<b>V367F</b>	V367F	VLYNS	FLYNS	0 to 1
<b>G431G</b>	I434K	GCVIA	GCVKA	0 to 1
<b>C432C</b>	A435S	CVIAW	CVISW	0 to 1
<b>C432C</b>	I434K	CVIAW	CVKAW	0 to 1
<b>V433V</b>	A435S	VIAWN	VISWN	1 to 0
<b>V433V</b>	I434K	VIAWN	VKAWN	1 to 0
<b>I434I</b>	S438F	IAWNS	IAWNF	0 to 1
<b>A435S</b>	A435S	AWNSN	SWNSN	1 to 0
<b>A435A</b>	N439K	AWNSN	AWNSK	1 to 0
<b>Y449Y</b>	L452R	YNYLY	YNYRY	0 to 1
<b>C484C</b>	V483A	CNGVE	CNGAG	0 to 1
<b>E484K</b>	E484K	EGFNC	KGFNC	0 to 1
<b>F486F</b>	F490L	FNCYF	FNCYL	1 to 0
<b>N487N</b>	F490L	NCYFP	NCYLP	0 to 1
<b>Q498Q</b>	N501Y	QPTNG	QPTYG	0 to 1
<b>N501Y</b>	N501Y	NGVGY	YGVGY	0 to 1
<b>C671C</b>	Q675H	CASYQ	CASYH	0 to 1
<b>Y674Y</b>	Q675H	YQTQT	YHTQT	1 to 0
<b>Q675H</b>	Q675H	QTQTN	HTQTN	1 to 0
<b>Q677Q</b>	P681H	QTNSP	QTNSH	0 to 1
<b>T678T</b>	P681H	TNSPR	TNSHR	0 to 1
<b>A706A</b>	N709Q	AYSNN	AYSQN	0 to 1
<b>I714I</b>	N717Q	IPTNF	IPTQF	0 to 1
<b>A831V</b>	A831V	AGFIK	VGFIK	0 to 1
<b>K1073K</b>	N1074Q	KNFTT	KQFTT	1 to 0
<b>I1115I</b>	D1118H	ITTDN	ITTHN	0 to 1
<b>T1116T</b>	D1118H	TTDNT	TTHNT	0 to 1
<b>T1117T</b>	D1118H	TDNTF	THNTF	0 to 1
<b>D1118H</b>	D1118H	DNTFV	HNTFV	0 to 1

Self and nonself SCSs are indicated by 0 and 1, respectively.