Supplementary Information

ARTICLE

Self and nonself short constituent sequences of amino acids in the SARS-CoV-2 proteome for vaccine development

Joji M. Otaki [1]*, Wataru Nakasone [2], Morikazu Nakamura [2]

[1] The BCPH Unit of Molecular Physiology, Department of Chemistry, Biology and Marine Science, University of the Ryukyus, Okinawa 903-0213, Japan
[2] Computer Science and Intelligent Systems Unit, Department of Information and Engineering, Faculty of Engineering, University of the Ryukyus, Okinawa 903-0213, Japan

*email: otaki@sci.u-ryukyu.ac.jp

Supplementary Results

Self-nonself status changes in the SARS-CoV-2 proteome

Mutations can be categorized into two distinct groups: one causes a self-to-nonself (0-to-1) status change, and the other causes a nonself-to-self (1-to-0) status change. The former may increase the chance of being recognized and eliminated by the immune system, and the latter may decrease this chance. In the 68 SARS-CoV-2 variant proteome sequences examined, we found 19 SCSs that changed self-nonself status due to their own or surrounding mutations (Supplementary Table S3; Additional Data 3 at GitHub). Among them, 11 were self-to-nonself (0-to-1) changes, and 8 were nonself-to-self (1-to-0) changes. Since there were 8,809 self SCSs and 852 nonself SCSs in the SARS-CoV-2 proteome, 0.125% of self SCSs changed to nonself SCSs (status change rate), whereas 0.939% of nonself SCSs changed to self SCSs. The percentage ratio of nonself-to-self changes to self-to-nonself changes was 7.51, and these two frequencies were significantly different (χ^2 test: $df = 1$, $t = 26.0$, $p < 0.0001$; $df = 1$, $t = 22.0$, $p < 0.0001$ after Yates's adjustment). This means that nonself-to-self changes and their associated mutations (escaping mutations or mimicry mutations) were more frequent than self-to-nonself changes and their associated mutations (exposing mutations).

Because self SCSs and nonself SCSs are not equally common in the 3.2 million SCS repertoire, probabilistically, a mutated residue may produce nonself SCSs less frequently than self SCSs. Conversion factors for this weight consideration in the human proteome were defined as ×(self/nonself) = ×(2,401,598/798,402) = ×3.0 for self-to-nonself changes and ×1.0 for nonself-to-self changes. Using these conversion factors based on the human proteome, the number of self-to-nonself changes was calculated as 33, and the number of nonself-to-self changes was calculated as 8. The self-to-nonself status change rate was 0.375%, and the nonself-to-self status change rate was 0.939%. The percentage ratio of nonself-to-self changes to self-to-nonself changes was thus 2.50. When the converted mutation frequencies between the two were subjected to the χ^2 test, a significant difference was still obtained at the level of $p < 0.05$ ($df = 1$, $t = 5.6$, $p = 0.016$; $df = 1$, $t = 4.5$, $p = 0.033$ after Yates's adjustment).

From an immunological point of view, nonself-to-self changes would confer a survival advantage to the coronavirus, and this result is in accordance with this interpretation, if not just a coincidence. Among the 19 SCSs that changed self-nonself status as discussed above, only a single SCS outside the RBD was found in the spike protein: a self SCS (796-DFGGF-800) that changed to a nonself SCS (796-DCGGF-800).

Self-nonself status changes in the spike protein

A similar analysis was performed to focus on the spike protein using a set of functional mutations (48 point mutations, three of which were at redundant positions)[40,41], a recent England variant (B.1.1.7) (6 additional point mutations)[42] and a South African variant (501Y.V2) (7 additional point mutations)[43]. Due to these point mutations, we discovered 26 self-to-nonself status changes out of 1172 self SCSs (status change rate: 2.22%) and 10 nonself-to-self status changes out of 97 nonself SCSs (status change rate: 10.31%) (Supplementary Table S4; Additional Data 4 at GitHub). The percentage ratio of nonself-to-self changes to self-to-nonself changes was 4.65. When the status change frequencies were subjected to the χ^2 test, a significant difference was obtained ($df = 1$, $t = 18.9$, $p < 0.0001$; $df = 1$, $t = 16.3$, $p < 0.0001$ after Yates's adjustment).

Using conversion factors for weight adjustment based on the human proteome (×3.0 for self-to-nonself changes and ×1.0 for nonself-to-self changes), the number of self-to-nonself changes was now calculated as 78 (status change rate: 6.66%), and the number of nonself-self changes was calculated as 10 (status change rate: 10.31%). The percentage ratio of nonself-to-self changes to self-to-nonself changes was 1.55. When the converted number of changes was subjected to the χ^2 test, no significant difference was obtained ($df = 1$, $t = 1.57$, $p = 0.21$; $df = 1$, $t = 1.10$, $p = 0.29$ after Yates's adjustment). This result suggests that the ongoing spike protein evolution in humans may not be driven by mimicry to escape immunological recognition and elimination.

Among the mutations that caused the nonself-to-self status changes detected above, 4 mutation sites were directly located within the nonself SCS clusters in the RBD discussed above (Fig. 2b; Supplementary Table 4). Three of them (I434K, A435S, and N439K) were located in the VIAWNSNN cluster of the 17-aa supercluster. An additional mutation site that does not change the self-nonself status was located in the same cluster, suggesting that this is an unstable nonself cluster. In contrast, only one status-changing site (F490L) was located in the 19-aa supercluster. This supercluster contained 3 additional mutation sites, but they did not change the self-nonself status and were at the self-nonself boundaries.

Supplementary Discussion

Consistent with the discussion in the main text, throughout the proteome of SARS-CoV-2, nonself-to-self status changes were significantly greater than self-to-nonself status changes even after weighting, supporting the sequence mimicry hypothesis of host-parasite interactions; however, the sample size was small, and the conclusion here may thus be considered tentative. Similarly, the self-nonself status changes in SARS-CoV-2 spike protein also showed significant differences but only without a weight consideration. Spike protein evolution may not be driven much by sequence mimicry at present. Nonetheless, it is also true that nonself-to-self changes indeed occur in the spike protein at a probabilistically reasonable rate.

These results describe the real-time evolution of the virus, which may provide a hint at viral dynamics. The accumulation of nonself-to-self mutations (hereafter called mimicry mutations) may be an unavoidable evolutionary route for any pathogen after its host change as a consequence of immunological escape. These mutations would occur independently of those that increase the virulence of the virus. However, for the sake of discussion, we assume here that mutations occur exclusively for sequence mimicry and that viral virulence is determined by the mimicry level. First, the number of mimicry mutations is considered a function of time after a host change. Mimicry mutations are advantageous for any pathogen and will accumulate rapidly until a saturation point at which further accumulation of such mutations harms the molecular functions of viral proteins (route *A* to *B* in Supplementary Fig. S1a). Alternatively, harmful mutations may gradually accumulate to increase the mimicry level (routes *A* to *C* in Supplementary Fig. S1a). These harmful mutations will be eliminated by natural selection when they reduce the survival of the virus.

Second, possible contributions of mimicry mutations to survival and virulence are considered. At a relatively low level, an increase in mimicry mutations simply means that the virus can escape the host immune system better, directly contributing to viral virulence and simultaneously contributing to viral survival (route *A* for both survival and virulence in Supplementary Fig. S1b). Here, survival and virulence may show similar dynamics, and mimicry mutations may reach the maximum level of virulence without harming molecular functions. The level of mimicry mutations may become stable at the maximum point. A further increase in mimicry mutations, if it occurs, may negatively impact the molecular functions of viral proteins, reducing virulence, and this may result in a simultaneous decrease in the survival of the virus (routes *A* to *B* for both survival and virulence in Supplementary Fig. S1b). This excessive mimicry is not favorable for the

viral existence. Alternatively, such an increase in mimicry mutations may enhance the survival of the virus at the expense of virulence or molecular functionality (routes *A* to *C* for survival and routes *A* to *B* for virulence in Supplementary Fig. S1b). In this way, survival and virulence may show two different dynamics in response to mimicry mutations. This scenario may be more likely to sustain viral existence due to a higher level of survival, and such mutations may open the possibility of a benign life cycle in response to a decrease in immunological selection pressure. Over time, the virus may find an equilibrium between functional compromise (low virulence) and immunological escape (high invisibility) for coexistence. The present study suggests that the SARS-CoV-2 proteome (although not spike protein) may currently be under such an evolutionary process, and we predict that the current pandemic of SARS-CoV-2 may cease over time due to the accumulation of mimicry mutations. Unfortunately, we are unable to predict when this might occur. Large-scale bioinformatics studies may make such predictions possible.

Additionally, many further studies remain unexplored. For example, the nonself SCSs in the SARS-CoV-2 proteome (other than spike proteins) identified in this study have not been examined thoroughly in terms of the possibilities of their functioning as epitopes. An analysis of human proteome variants may reveal differences in infectivity among human individuals and could predict prognosis. An analysis of other pathogens in relation to the human proteome may further test the sequence mimicry hypothesis. This hypothesis may be widely applicable not only to a host-pathogen relationship but also to a host-parasite relationship widely seen in biological mutualism. The nonself definition in the present study (i.e., zero-count SCSs) may vary under different genetic backgrounds and environmental conditions, which may trigger or prevent autoimmune diseases. Reflecting the worldwide pandemic, numerous mutations in SARS-CoV-2 have been detected[54,55], and these mutations should be analyzed more comprehensively *in silico* in the future. Such studies will be able to contribute to improved vaccines. Other lines of preventive and treatment measures, such as nutritional balance that could induce nitric oxide (NO) production[55] and traditional complementary medicine[57,58], should also be encouraged.

The present study can be considered an application of the SCS concept to physiological problems. The SCS concept is simple, and its applications are diverse[10]. This field of study has expanded *in silico*[32,33,59-67], but the SCS concept has not yet been sufficiently explored to understand physiological systems. To our knowledge, the first

physiological application in this field is the use of a group of peptides as immunological adjuvants[68,69]. The search for human-specific proteins is also an example of SCS-based physiological studies[70]. We anticipate that the present study will expand important frontiers of physiological studies from the viewpoint of the SCS concept.

Supplementary Methods

Characterization of the human and SARS-CoV-2 proteomes. Source Codes for Human SCS Analysis and for SARS-CoV-2 SCS Analysis were used to calculate parameters of SCS distributions in the human and SARS-CoV-2 proteomes, which were placed at GitHub, https://adslab-uryukyu.github.io/scs-sars-cov-2/. Calculations were performed not only for SCS frequencies shown in Fig. 1a-c but also for other tasks such as exact SCSs for a particular rank range and availability scores[24,25]. SCS length can be adjusted as 3-aa, 4-aa, or 5-aa residues.

Variant proteomes. We first downloated 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.

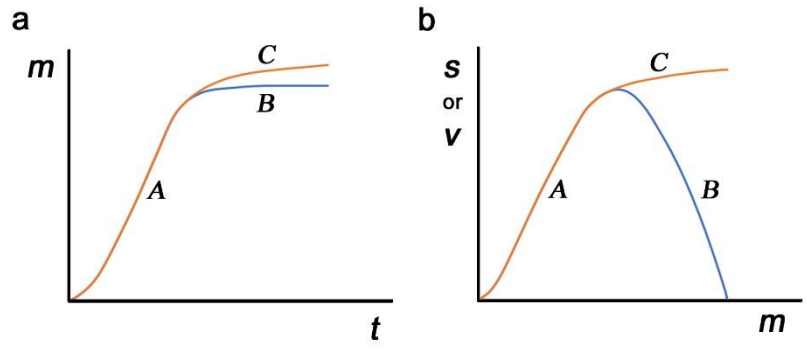
Other 26 proteomes that were not analyzed further due to insertions and deletions were as follows: ASM993788v1, ASM993790v1, ASM993805v1, ASM993806v1, ASM994842v1, ASM994846v1, ASM994849v1, ASM994852v1, ASM994855v1, ASM1153693v1, ASM1153715v1, ASM1153729v1, ASM1153732v1, ASM1153739v1, ASM1153746v1, ASM1153778v1, 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[39-42]: 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.

Identification of nonself SCSs in a 3D model of the spike protein. Nonself SCSs identified as potential candidates for vaccine targets (**Fig. 2b**) were further examined in their locations in a 3D model of the spike protein (**Fig. 3**). To do so, a 3D structural data of the spike protein (PDB ID: 6VYB)[36] in the Protein Data Bank (PDB) managed by the Research Collaboratory for Structural Bioinformatics (RCSB) were accessed[71]. This structure has been determined by cryo-EM at 3.20-Å resolution[36]. The structure was viewed and the nonself SCSs were highlighted by Mol*, a built-in viewer of the RCSB-PDB[72].

Supplementary References

54. Toyoshima, Y. et al. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J. Hum. Genet.* **65**, 1075-1082 (2020).
55. Laha, S. et al. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect. Genet. Evol.* **85**, 104445 (2020).
56. Yamasaki, H. Blood nitrate and nitrite modulating nitric oxide bioavailability: Potential therapeutic functions in COVID-19. *Nitric Oxide* **103**, 29-30 (2020).
57. Yang, Y. et al. Traditional Chinese Medicine in the treatment of patients infected with 2019-new coronavirus (SARS-CoV-2): A review and perspective. *Int. J. Biol. Sci.* **16**, 1708-1717 (2020).
58. Liu, M. et al. Efficacy and safety of Integrated Traditional Chinese and Western Medicine for corona virus disease 2019 (COVID-19): a systematic review and meta-analysis. *Pharmacol. Res.* **158**, 104896 (2020).
59. Yu, L. et al. Grammar of protein domain architectures. *Proc. Natl. Acad. Sci. USA* **116**, 3636-3645 (2019).
60. Figureau, A., Soto, M. A. & Tohá, J. A pentapeptide-based method for protein secondary structure prediction. *Protein Engineering* **16**, 103-107 (2003).
61. Pe'er, I. et al. Proteomic signatures: Amino acid and oligopeptide compositions differentiates among phyla. *Proteins* **54**, 20-40 (2004).
62. Vries, J.K., Liu, X. & Bahar, I. The relationship between n-gram patterns and protein secondary structure. *Proteins* **68**, 830-838 (2007).
63. Daeyaert, F., Moereels, H. & Lewi, P. J. Classification and identification of proteins by means of common and specific amino acid n-tuples in unaligned sequences. *Comput. Methods Programs Biomed.* **56**, 221-233 (1998).
64. Imai, K. & Nakai, K. Tools for the recognition of sorting signals and the prediction of subcellular localization of proteins from their amino acid sequences. *Front. Genet.* **11**, 607812 (2020).
65. Chou, K. C. Prediction of protein cellular attributes using pseudo amino acid composition. *Proteins* **43**, 246-255 (2001).
66. Tsutsumi, M. & Otaki, J. M. Parallel and antiparallel β-strands differ in amino acid composition and availability of short constituent sequences. *J. Chem. Inf. Model.* **51**: 1457-1464 (2011).
67. Otaki, J. M., Tsutsumi, M., Gotoh, T. & Yamamoto, H. Secondary structure characterization based on amino acid composition and availability in proteins. *J. Chem. Inf. Model.* **50**: 690-700 (2010).
68. Patel, A. et al. Pentamers not found in the universal proteome can enhance antigen specific immune responses and adjuvant vaccines. *PLoS One* **7**, e3802 (2012).
69. Le, H.-T. et al. Enhancing the immune response of a nicotine vaccine with synthetic small "non-natural" peptides. *Molecules* **25**, 1290 (2020).
70. Endo, S. et al. (2020) Search for human-specific proteins based on availability scores of short constituent sequences: Identification of a WRWSH protein in human testis. In: Behzadi, P. & Bernabò, N. (Eds) *Computational Biology and Chemistry.* pp. 11-33. IntechOpen, London.
71. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H.; Shindyalov, I. N.& Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242 (2000).
72. Sehnal, D., Rose, A. S., Koča, J., Burley, S. K. & Velankar, S. Mol*: toward a common library and tools for web molecular graphics. *MolVA '18: Proceedings of the Workshop on Molecular Graphics and Visual Analysis of Molecular Data.* 29-33 (2018).



Supplementary Fig. S1. Possible dynamics of mimicry (*m*), survival (*s*), and virulence (*v*) of a parasite. (**a**) Mimicry mutations (*m*) as a function of time (*t*). Routes *A*-*B* and routes *A*-*C* may be possible. (**b**) Survival (*s*) and virulence (*v*) as a function of mimicry (*m*). At the relatively low level of *m*, both *s* and *v* may follow *A*, but after the critical point, *v* may follow *B*, and *s* may separately follow *C*.

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

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

Supplementary 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	24	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.

Supplementary Table S3. Self-nonself 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	Self/nonself change
ASM993794v1	3606 (1-3606)	ORF1ab	L3606F	0 to 1
ASM1153713	2705 (1-2705)	ORF1ab	K2705K	1 to 0
ASM1153723	901 (1-901)	ORF1ab	S901S	0 to 1
ASM1153723	902 (1-902)	ORF1ab	M902I	1 to 0
ASM1153723	8485 (3-124)	ORF3a	I124I	1 to 0
ASM1153723	8489 (3-128)	ORF3a	W128L	1 to 0
ASM1153723	8608 (3-247)	ORF3a	H247H	1 to 0
ASM1153736	3073 (1-3073)	ORF1ab	E3073E	1 to 0
ASM1153745	7888 (2-796)	Spike (S)	D796D	0 to 1
ASM1153762	3606 (1-3606)	ORF1ab	L3606F	0 to 1
ASM1153771	1836 (1-1836)	ORF1ab	Y1836Y	0 to 1
ASM1153810	1856 (1-1856)	ORF1ab	S1856S	0 to 1
ASM1153810	1858 (1-1858)	ORF1ab	E1858E	0 to 1
ASM1153810	1859 (1-1859)	ORF1ab	Y1859Y	0 to 1
ASM1154548	5826 (1-5826)	ORF1ab	R5826R	0 to 1
ASM1154548	5827 (1-5827)	ORF1ab	N5827N	1 to 0
ASM1154548	5863 (1-5863)	ORF1ab	S5863S	0 to 1
ASM1174201	9552 (9-340)	Nucleocapsid (N)	D340D	0 to 1
ASM1174202	1607 (1-1607)	ORF1ab	I1607V	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.

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

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

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