Electronic Supplementary Information

The SARS-CoV-2 Spike Protein Structure: A Symmetry Tale on Distortion Trail

Inbal Tuvi-Arad[*] and Yaffa Shalit

Department of Natural Sciences, The Open University of Israel, Raanana, Israel

Table S1. Domains of the SARS-CoV-2 spike protein.

| Subunit	Domain	Residues Number
S1	NTD	≤ 293
	N2R	294-330
	RBD*	331-527
	SD1	528-590
	SD2	591-685
S2		≥ 686

The boundaries of the domains, and especially the RBD, vary by few residues in different papers. We followed Gobeil et al. *(Molecular Cell, 82*(11), 2050-2068.e6, 2022) and defined the RBD between P330 and P527. It should be noted that a mismatch of several residues in the boundaries of each domain has a minor effect on the CSM analysis.

Table S2. Number of proteins in our dataset of spike proteins, by state and variant.
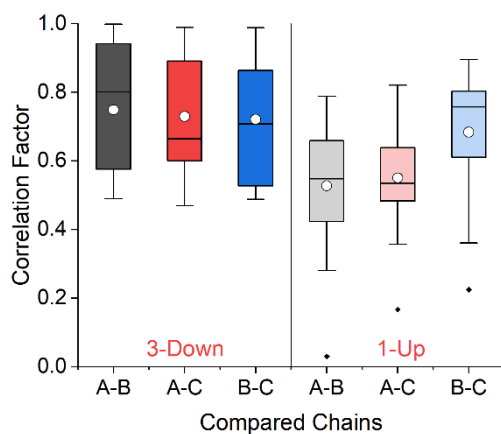
	State		
Variant	3-Down	1-Up	Total
Alpha	3	3	6
Beta	1	5	6
Delta	7	5	12
Omicron	2	4	6
WT	2	1	4
Total	**15**	**18**	**33**

Table S3. PDB-Ids of proteins in our dataset.

| Variant	3-Down	1-Up	Total
Alpha	7LWS, 7OD3, 7ODL	7EDF, 7FET , 7LWV	6
Beta	7LYM	7LYN, 7LYO, 7LYP, 7V76, 7VX1	6
Delta	7SBK, 7TOU, 7TOX, 7TOY, 7TP0, 7TP1, 7TP2	7V7O, 7V7P, 7V7Q, 7V7R, 7W92	12
Omicron	7TF8, 7WK2	7QO7, 7TGW, 7TO4, 7WVN	6
WT	6ZB4, 7FCE	6Z97	4
Total	**15**	**18**	**33**
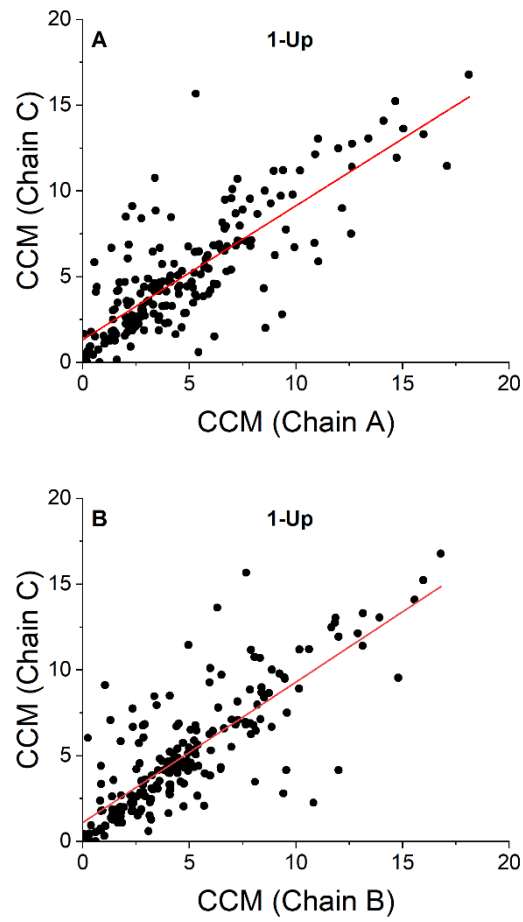
Figure S1. Correlation plots for the CCM per residue of the 1-Up conformer of the Omicron variant along the RBD. **A.** Correlation between chains A and C. **B.** Correlation between chains B and C. Red lines represent the regression lines. PDB-ID = 7TO4.

Figure S2. Box and whisker plots of Pearson correlation factors for residues CCM between chains of the RBD, calculated at the backbone level. The 3-Down and 1-Up states are represented by 15 and 18 proteins respectively. Box boundaries represent 25-75% of the data, horizontal line within the box is the median, and the white circle is the mean value in each box.
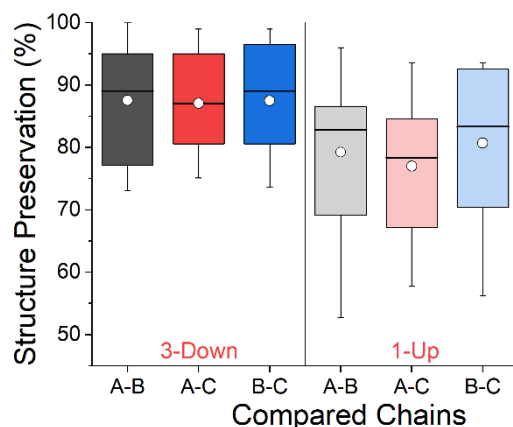
Figure S3. Box and whisker plots of secondary structure preservation between chains of the RBD. The 3-Down and 1-Up states are represented by 15 and 18 proteins respectively. Structure preservation for each two chains was based on comparing the secondary structure of each pair of residues, as determined by DSSP (https://swift.cmbi.umcn.nl/gv/dssp/). Numbers on the vertical axes represent the percentages of pairs with the same secondary structure. Box boundaries represent 25-75% of the data, horizontal line within the box is the median, and the white circle is the mean value in each box.
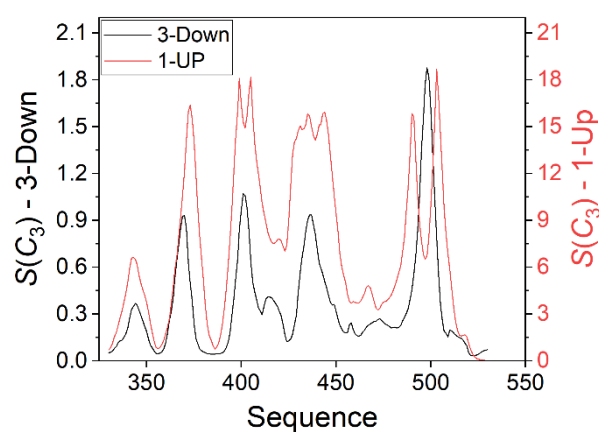
Figure S4. Distortion with respect to C_3 symmetry of 10-residues fragments (only backbone atoms included) along the sequence of the RBD domain of the Omicron BA.1 variant of the SARS-CoV-2 spike protein. Black: 3-Down state, left Y scale (PDB-ID: 7TF8). Red: 1-Up state, right Y scale (PDB-ID: 7TO4).
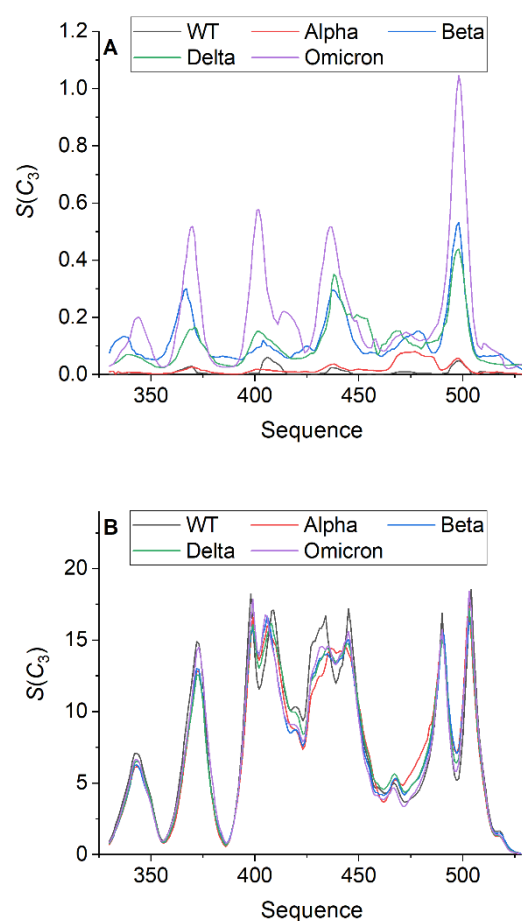
Figure S5. CSM spectra of 10-residues fragments along the sequence of the RBD for 33 SARS-CoV-2 spike protein at the backbone level, averaged by variant. **A.** 3-Down state. **B.** 1-Up state. PDB-Ids are listed in Table S2.

S5

Table S4. Mutations of the SARS-CoV-2 spike protein at the RBD.

Source: https://www.who.int/activities/tracking-SARS-CoV-2-variants.

Alpha	Beta	Delta	Omicron BA.1
			G339D
			S371F
			S373P
			S375F
	K417N		K417N
			N440K
			G446S
		L452R	
			S477N
		T478K	T478K
	E484K		E484A
			Q493R
			Q496S
			Q498R
N501Y	N501Y		N501Y
			Y505H