Supplemental Information

Identification of SARS-CoV-2 Vaccine

Epitopes Predicted to Induce

Long-Term Population-Scale Immunity

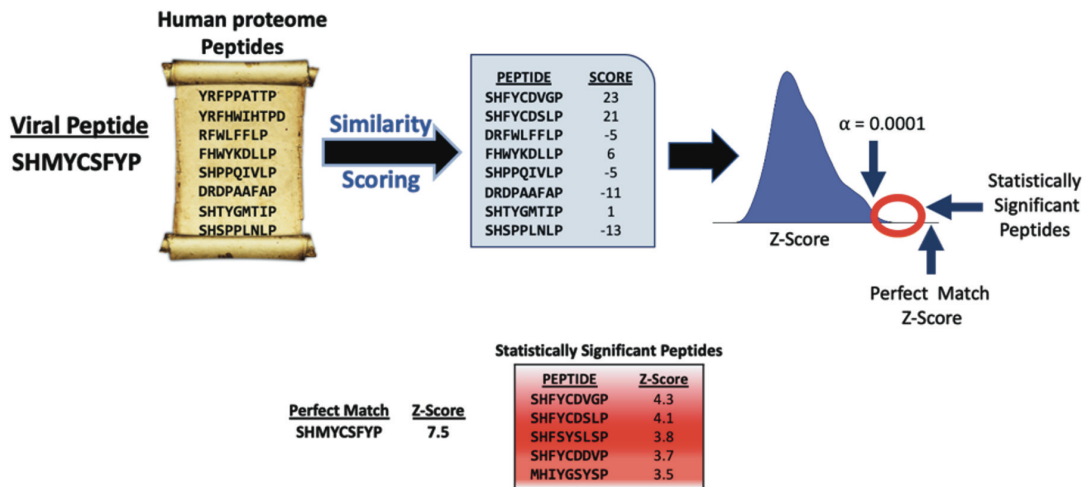**Mark Yarmarkovich, John M. Warrington, Alvin Farrel, and John M. Maris**

A

Metric	Score
Perfect Match	5
Matched Group	2
Unmatched Polarity	-2

Groups	Amino Acids
Short Chains	A, G
Acidic	D, E
Basic	K, R, H
Amines	N, Q
Sulfides	C, M
Alcohols	S, T, Y
Aliphatic	I, L, V, M, A
Aromatic	F, Y, W, H
Proline	P
Polar	D, E, N, Q, R, K, H, Y, C, S, T
Hydrophobic	G, A, F, W, P, I, L, V, M

Viral Peptide S H M Y C S F Y P
Human Peptide S H F Y C D V G P
AA Match Scores 5 5 0 5 5 0 0 -2 5
Weights 1 0 1 2 2 2 1 1 0
∑(5 0 0 10 10 0 0 -2 0)

Score 23

B

Human proteome Peptides

Viral Peptide SHMYCSFYP

Similarity Scoring

PEPTIDE	SCORE
SHFYCDVGP	23
SHFYCDSLP	21
DRFWLFFLP	-5
FHWYKDLLP	6
SHPPQIVLP	-5
DRDPAAFAP	-11
SHTYGMTIP	1
SHSPPLNLP	-13

α = 0.0001

Z-Score

Statistically Significant Peptides

Perfect Match Z-Score

Statistically Significant Peptides

PEPTIDE	Z-Score
SHFYCDVGP	4.3
SHFYCDSLP	4.1
SHFSYSLSP	3.8
SHFYCDDVP	3.7
MHIYGSYSP	3.5

Perfect Match	Z-Score
SHMYCSFYP	7.5

$$Disimilarity\ Score = Z_{Max} - \left(Z_{Top} + \frac{N_{Sig}^{\frac{\overline{Z_{Sig}}}{Z_{max}}}}{1000} \right)$$

$$Disimilarity\ Score = 7.5 - \left(4.3 + \frac{6^{\frac{3.9}{7.5}}}{1000} \right)$$

$$Disimilarity\ Score = 3.13$$

Figure S1. Dissimilarity Scoring, related to STAR Methods Dissimilarity Scoring. A) 3,524 viral epitopes (12,383 total peptide/MHC pairs) were compared against the normal human proteome. Non-anchor residues were used to calculate similarity scores based on amino acid classifications as described in methods. Residues in the same position of the viral and human peptides with a perfect match, similar amino acid classification, or different polarity, were assigned scores of five, two, or negative two, respectively. B) Each viral peptide/HLA pair was compared against the set of normal peptides presented on the same MHC. Dissimilarity score for each viral peptide was calculated by comparing against the most similar group of peptides with p < 0.0001 and reported as the difference in Z-scores between the viral peptide and closest-scoring peptides.