Coronavirus-associated molecular mimicry through homology to a SARS-CoV-2 peptide could be leading to susceptibility in patients with HLA-A*02:01 and HLA-A*24:02 serotypes

Y Adiguzel*

*Department of Biophysics, School of Medicine, Altinbas University, Kartaltepe Mah. Incirli Cad. No11, 34147 Bakirkoy, Istanbul, Turkey. E-mail: yekbun.adiguzel@altinbas.edu.tr

Abstract

Aim-This study aims to predict autoimmunity-related pathological mechanisms that posses risk for individuals with certain HLA serotypes and are common to the pathogenicity of certain coronaviruses including SARS-CoV-2, based on homology to a SARS-CoV-2 peptide.

Methods-Coronavirus-associated sequences, which are homologous to the SARS-CoV-2 peptide CFLGYFCTCYFGLFC, are obtained. Human peptides that have at least 7 residue matches with those coronavirus sequences, and the SARS-CoV-2 peptide, are obtained. Then, epitope pairs, which are sourced by those coronavirus and human sequences' alignments, are identified. There, epitope pairs that are predicted to bind strongly not only to the same HLA allele with each other but also to the same HLA allele with the respective alignment of the SARS-CoV-2 peptide are selected.

Results-Following is the list of proteins (or regions) with predicted HLA-A*02:01 or HLA-A*24:02 epitopes, which are not only common but also homologous to the epitopes that are sourced by the SARS-CoV-2 peptide and its homologous coronavirus peptides: Immunoglobulin heavy chain junction regions, CRB1 isoform I precursor, slit homolog 2 protein, hCG1995581, hCG2028737, phospholipid phosphatase-related protein type 2. Among those, CRB1 isoform I precursor sequence with the predicted HLA-A*24:02 epitope aligns with the highest number of different coronavirus sequences.

Conclusion-Results imply autoimmunity risk in COVID-19 patients with HLA-A*02:01 and HLA-A*24:02 serotypes, through molecular mimicry, as a common pathogenicity risk that can be prevalent upon getting infected with certain other coronaviruses. These results can pave the way to improved risk groups' assessment and autoimmunity treatment options, to be used in COVID-19 and its associated diseases.

Keywords: SARS-CoV-2, coronavirus, autoimmunity, molecular mimicry, HLA class I, MHC class I, disease susceptibility, HLA-A*24:02, HLA-A*02:01

1. Introduction

Upon infection, presence of similar sequences to the pathogens' proteins in the human proteome can be a potential risk of causing autoimmune response. This molecular mimicry is among the preliminary conditions. Kerkar and Vergani [1] mentioned molecular mimicry, sequence similarities of pathogenic organisms with autoantigens, as a possible mechanism of autoimmunity, and is conceptually reinforced through the discoveries of de novo autoimmune hepatitis associated with certain viral infections. In 90s, a sequence similarity search revealed 70 % overlap of a 10mer within the V3-loop of the envelope glycoprotein gp120 of HIV-1 isolates, with the collagen-like region of the human complement component C1q-A [2]. Follow-up studies by the same group addressed the complications that would be caused by vaccines, which are based on gp120 of HIV-1 [3]. Later, their findings on the presence of antibodies that are reactive for the peptide in the V3-loop of HIV-1 in the healthy individuals [4], and the complementarity of the antibodies for V3-loop of HIV-1 and IgG of human [5] were reported. Presence of homologies between human proteins and virus proteins, and the examination of the pathways of those homologous peptides, drew attention to the pathologies of viral infections that are related to the immune response [6]. Even therapies targeting such complexifications were suggested. In correlation with these, Kanduc and Shoenfeld [7] considered the intrinsic hazards in the vaccines that are based on such pathogen sequences. These authors [8] later added onto their earlier findings by looking for pentapeptide sharing between the viral, bacterial, protozoan pathogens, and human. There, it was suggested that the respective results could elucidate the decades-old "original antigenic sin" phenomenon. Further, "negative selection of self-reactive lymphocytes" was concluded to be unlikely, in a later study of the researchers [9]. Similarities among the SARS-CoV-spike glycoprotein sequences and those of the human surfactant protein and the related proteins were revealed by similar methodology [10]. Specifically, 13 of 24 shared pentapeptides were found to be present in 52 SARS-CoV-derived immunoreactive epitopes. In addition, heptapeptide-sharing among pathogens, which are also containing SARS-CoV-2, and the non-human primates, was observed and it revealed that high level of peptide-sharing is somewhat unique to human [11]. Among the other animals that are used for preclinical tests, mice would more expectedly reveal autoimmune reactions. Eventually, "aged mice" was suggested to be appropriate for testing of vaccines, which are based on spike glycoproteins of SARS-CoV-2 [12].

In case of rheumatoid arthritis, patients expressing HLA-DRB1*04:01 are more frequently experiencing severe form of the disease, and there is a 5mer peptide of HLA-DRB1*04:01 that is also present in the heat shock protein of *E. coli* [13]. It is suggested to be a risk factor upon being exposed to Enterobacteriaceae, for patients with rheumatoid arthritis, who are expressing HLA-DRB1*04:01. There are other examples, e.g.: In systemic lupus erythematosus, a 7mer peptide region of Epstein–Barr virus nuclear antigen 1 has cross-reactivity with an 8mer region of Sm [14]. In systemic sclerosis, a 12mer peptide has homology with the human cytomegalovirus late

protein UL94, and it reacts with the IgG of patients [15]. In primary biliary cholangitis, E2 component of the pyruvate dehydrogenase complexes of human and *E. coli* have similar regions at a peptide of the human protein that is 31 amino acids (aa) in length [16].

Relationship of COVID-19 and immunity is complex [17–22] and it can involve autoimmune reactions through molecular mimicry. Woodruff *et al*. [23] recently reported that critically ill patients of COVID-19 "displayed hallmarks of extrafollicular B cell activation as previously described in autoimmune settings." Rodríguez and co-workers [24] recently reviewed autoinflammatory and autoimmune conditions in COVID-19. Accordingly, antiphospholipid syndrome, autoimmune cytopenia, Guillain-Barré syndrome and Kawasaki disease are known to be reported in COVID-19 patients [24]. Regarding molecular mimicry, Cappello and co-workers [25] hypothesized that SARS-CoV-2 may be generating autoimmunity through molecular mimicry, induced by stress. Rodríguez and co-workers [24] mentioned the molecular mimicry and bystander activation as the mechanisms that can link COVID-19 to autoimmunity. In relation, Lucchese and Flöel [26] reported three 6mers in the human brainstem respiratory pacemaker proteins that are present in the SARS-CoV-2 proteome. These authors also reported molecular mimicry with SARS-CoV-2 and heat shock proteins 90 and 60 of human, which are known to be associated with Guillain-Barré syndrome and other autoimmune diseases [27]. Importantly, the shared peptides are part of those epitopes that were experimentally shown to be immunoreactive. Also, one of the works that is mentioned above [10] reported 5mers of human surfactant protein to be present in the SARS-CoV-2 proteome. Angileri and co-workers [28] reported a 7mer of human Odorant Receptor 7D4, a 6mer of human Poly ADP-Ribose Polymerase Family Member 9, and a 7mer of Solute Carrier Family 12 Member 6, which are present in the putative epitopes of SARS-CoV-2. There are also human proteins that have strong immune cross-reactions with the spike protein antibody of SARS-CoV-2 [29], which can be suggestive of autoimmunity by means of molecular mimicry in susceptible individuals [24]. In relation, aim of the current study is to look for autoimmunity related pathological mechanisms that are common to certain coronaviruses, including SARS-CoV-2, by means of a selected sequence (CFLGYFCTCYFGLFC), which is obtained through our ongoing study [30] involving tblastx search of SARS-CoV-2 and plasmodium species that cause malaria in human [31,32].

2. Methods

The CFLGYFCTCYFGLFC sequence was obtained by performing blastx [33] at NCBI [34], between the reference genome of query input SARS-CoV-2 (NC_045512.2) and *P. vivax* (taxid:5855) [30]. It was the aligned query sequence in the tblastx output that revealed the top identity between the query and subject, which was afterwards utilized as input for NCBI blastp search, by limiting the search to SARS-CoV-2 (taxid:2697049), to ensure that the sequence is

expressed [30]. Here, in this study, blast of the sequence is performed at Uniprot with threshold 10 (*performed on 5 August 2020*). The associated coronaviruses with homologous sequences are selected from the results. Then, blastp of the SARS-CoV-2 sequence and the coronavirus sequences that are homologous, are performed separately, by limiting the searches to *H. sapiens* (taxid:9606). The query-subject sequence pairs that have at least 7 residue matches are found in those results. Within those results, the ones with the same names and sequence IDs as those in the respective results of SARS-CoV-2 blastp search, are identified. This is followed by major histocompatibility complex (MHC) class I binding predictions for those identified query-subject sequence pairs. This is done to find the coronavirus and human sequences that are homologous, and which are predicted to bind strongly to the same HLA alleles as that with the SARS-CoV-2 sequence. To do that, binding affinities to the MHC class I (*MHC class I genes are human leukocyte antigen A [HLA-A], -B, and -C genes* [35]) proteins are predicted with the use of a tool that integrates NetMHC 4.0 [36,37], NetMHCpan 4.1 [38], and PickPocket 1.1 [39]. That tool is NetMHCcons 1.1 [40]. Predictions are performed for 8-15mers, with default parameters, and by performing the predictions for 12 MHC supertype representatives. The 12 MHC supertype representatives are HLA-A*01:01 (A1), HLA-A*02:01 (A2), HLA-A*03:01 (A3), HLA-A*24:02 (A24), HLA-A*26:01 (A26), HLA-B*07:02 (B7), HLA-B*08:01 (B8), HLA-B*27:05 (B27), HLA-B*39:01 (B39), HLA-B*40:01 (B44), HLA-B*58:01 (B58), HLA-B*15:01 (B62). NetCTLpan 1.1 [41] is utilized as well, similarly, for the prediction of the epitopes of cytotoxic T lymphocyte (CTL), as 8-11mers. Within the prediction results, epitopes with at least 5 residue matches and strong binding affinities to the same MHC supertype representative are considered for possible risk of autoimmunity-related pathological mechanisms that are common to SARS-CoV-2 and associated coronaviruses, based on similarity to the selected short SARS-CoV-2 sequence. Epitope-pairs with the highest number of residue-matches are displayed.

Summary information about the identified proteins (or peptides in case of immunoglobulin heavy chain junction regions) is collected by searching the sequence ID of the aligned subject sequence in the blastp results at Entrez (NCBI), and then searching the encoding gene ID, which is indicated at the UniProt (www.uniprot.org) [42], to retrieve the UniProtKB number. Information on the associated diseases is obtained from the human gene database GeneCards (www.genecards.org) [43], wherever readily available.

3. Results

The query peptide with the sequence CFLGYFCTCYFGLFC in the single letter code representation is the outcome of initial tblastx search (see s2 file of [30]). Peptide with that sequence is present in the SARS-CoV-2 proteome (see s6 file of [30]), as part of non-structural protein 6 (nsp6) that is cleaved from the replicase polyprotein 1a. Blast search of the sequence is

performed at Uniprot with threshold 10 (s1, *performed on 5 August 2020*). The associated coronaviruses are retrieved from that search and they are shown in Table 1, with their related sequences.

Table 1

Associated coronaviruses with their sequences that are obtained from the blast search results of the CFLGYFCTCYFGLFC, performed at Uniprot with threshold 10.

Sequence	Protein ID	Protein Abbreviation	Protein name	Organism
CFLGYFCTCYFGLFC	P0DTD1	R1AB_SARS2	Replicase polyprotein 1ab	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)
	P0DTC1	R1A_SARS2	Replicase polyprotein 1a	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)
CFLGYCCCCYFGLFC	P0C6V9	R1AB_BC279	Replicase polyprotein 1ab	Bat coronavirus 279/2005 (BtCoV) (BtCoV/279/2005)
	P0C6X7	R1AB_CVHSA	Replicase polyprotein 1ab	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)
	P0C6W6	R1AB_BCRP3	Replicase polyprotein 1ab	Bat coronavirus Rp3/2004 (BtCoV/Rp3/2004) (SARS-like coronavirus Rp3)
	P0C6W2	R1AB_BCHK3	Replicase polyprotein 1ab	Bat coronavirus HKU3 (BtCoV) (SARS-like coronavirus HKU3)
	P0C6F5	R1A_BC279	Replicase polyprotein 1a	Bat coronavirus 279/2005 (BtCoV) (BtCoV/279/2005)
	P0C6U8	R1A_CVHSA	Replicase polyprotein 1a	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)
	P0C6T7	R1A_BCRP3	Replicase polyprotein 1a	Bat coronavirus Rp3/2004 (BtCoV/Rp3/2004) (SARS-like coronavirus Rp3)
	P0C6F8	R1A_BCHK3	Replicase polyprotein 1a	Bat coronavirus HKU3 (BtCoV) (SARS-like coronavirus HKU3)
	A0A0K1YZY7	A0A0K1YZY7_CVHSA	Replicase polyprotein 1ab	Bat SARS-like coronavirus YNLF_31C
	F2YCN6	F2YCN6_CVHSA	Polyprotein orf1ab	SARS coronavirus MA15 ExoN1
	D2E1D0	D2E1D0_CVHSA	Orf1ab polyprotein	SARS coronavirus ExoN1
	A0A0U1WHG0	A0A0U1WHG0_CVHSA	Orf1ab polyprotein	BtRf-BetaCoV/JL2012

	D2DJW2	D2DJW2_CVHSA	Non-structural polyprotein pp1ab	SARS coronavirus Rs_672/2006
	R9QTH2	R9QTH2_CVHSA	Orf1a polyprotein	Bat coronavirus Cp/Yunnan2011
	A0A0K1Z0N1	A0A0K1Z0N1_CVHSA	Polyprotein 1a	Bat SARS-like coronavirus YNLF_31C
	F2YCN7	F2YCN7_CVHSA	Polyprotein orf1a	SARS coronavirus MA15 ExoN1
	D2E1D1	D2E1D1_CVHSA	Orf1a polyprotein	SARS coronavirus ExoN1
	R9QTB2	R9QTB2_CVHSA	Orf1a protein	Bat coronavirus Rp/Shaanxi2011
	D2DJW3	D2DJW3_CVHSA	Non-structural polyprotein pp1a	SARS coronavirus Rs_672/2006
CYLIVGYFCTCYFGVF	E0ZN35	E0ZN35_BCHK9	Orf1ab polyprotein	Bat coronavirus HKU9-5-1
CYLVVGYFCTCYFGVF	E0ZN59	E0ZN59_BCHK9	Orf1ab polyprotein	Bat coronavirus HKU9-10-2
	E0ZN43	E0ZN43_BCHK9	Orf1ab polyprotein	Bat coronavirus HKU9-5-2
CYLLVGYLCTCYFGVF	P0C6W5	R1AB_BCHK9	Replicase polyprotein 1ab	Bat coronavirus HKU9 (BtCoV) (BtCoV/HKU9)
	P0C6T6	R1A_BCHK9	Replicase polyprotein 1a	Bat coronavirus HKU9 (BtCoV) (BtCoV/HKU9)
	A3EXI9	A3EXI9_BCHK9	Orf1ab polyprotein	Bat coronavirus HKU9-4
FIGYVCTCYFGVF	A0A023YA54	A0A023YA54_MERS	Orf1ab polyprotein	BtVs-BetaCoV/SC2013
GYFCTVYFGVF	P0C6W4	R1AB_BCHK5	Replicase polyprotein 1ab	Bat coronavirus HKU5 (BtCoV) (BtCoV/HKU5/2004)
	P0C6T5	R1A_BCHK5	Replicase polyprotein 1a	Bat coronavirus HKU5 (BtCoV) (BtCoV/HKU5/2004)
	A0A0U1WHL2	A0A0U1WHL2_BCHK5	Orf1ab polyprotein	BtPa-BetaCoV/GD2013
LGFMCTCYFGVF	K9N7C7	R1AB_MERS1	Replicase polyprotein 1ab	Middle East respiratory syndrome-related coronavirus (isolate United Kingdom/H123990006/2012)
	K9N638	R1A_MERS1	Replicase polyprotein 1a	Middle East respiratory syndrome-related coronavirus (isolate United Kingdom/H123990006/2012)

	T2B9U0	T2B9U0_MERS	Orf1ab	Middle East respiratory syndrome-related coronavirus (MERS-CoV)
	T2B9I2	T2B9I2_MERS	Orf1a	Middle East respiratory syndrome-related coronavirus (MERS-CoV)
GWFCTCYFGLY	A0A0S2ZX33	A0A0S2ZX33_9GAMC	Polyprotein 1a	European turkey coronavirus 080385d

Blastp search results of the sequences that are displayed in the first column of Table 1 are obtained by limiting the search to *H. sapiens* (s2-10). The query-subject sequence pairs in those results, which have at least 7 residue matches, are found (s11, *file displays the respective subject sequences and the proteins containing them*). Among those, the ones with the same protein name and sequence IDs as those in the respective results of SARS-CoV-2 blastp, are identified (Table 2). The LGFMCTCYFGVF sequence that is displayed at Table 1 do not have any alignment with the same protein name and sequence ID as those that are aligned with the SARS-CoV-2 sequence. The GWFCTCYFGLY sequence that is displayed at Table 1 do not have any alignment with the same protein name and sequence ID as neither those that are aligned with the SARS-CoV-2 sequence nor the others. Yet, it should be reminded that we looked for at least 7 residue matches, which indicates that the outcome could be different with 6 residue matches, for instance.

Table 2

Human proteins and their subject sequences, which are aligned with the coronavirus sequences that are homologous with the 15mer SARS-CoV-2 sequence (query number 1, seq-1). Aligned subject sequences that are shown in the second column contain at least 7 residues that are matching with the query sequences, which are indicated by a designating-number in the first column. Descriptions of those numbers are at the end of the table. Aligned sequences in the second column are displayed together with the gaps that are introduced to the sequence for the alignment.

[a]Query	[b]Subject sequence	[c]Protein name	[d]Sequence ID
1	**CFFNY----YFGL**	immunoglobulin heavy chain junction region	MOR79299.1
2	**CFFNY----YFGL**		
1	**CFLHY----YYGL**	immunoglobulin heavy chain junction region	MOQ87140.1
2	**CFLHY----YYGL**		
1	**CFSSYF----FLLFC**	hCG1995581, partial	EAW57092.1
2	**CFSSY----FFLLFC**		
1	**CFVG---SC-FGLF**	immunoglobulin heavy chain junction region	MON95290.1
2	**CFVGSC----FGLF**		
1	**FIGY-CSSTSCYTGGFC**	immunoglobulin heavy chain variable region, partial	CEF94348.1

2	**FIGYCSSTSCYTGGFC**		
6	**FIGY-CSSTSCYTGGF**		
1	**FLG----VYSFGLF**	phospholipid phosphatase-related protein type 2 isoform	XP_024307423.1
2	**FLGVYS----FGLF**		
1	**FLGY----YYGL**	immunoglobulin heavy chain junction region	MOP50498.1
2	**FLGY----YYGL**		
1	**GYFCTNYF**	hCG2028737	EAW73174.1
4	**GYFCTNYF**		
7	**GYFCTNYF**		
1	**GYSCLC-FGNF**	CRB1 isoform I precursor	AAL10681.1
3	**GYSCLC-FGNF**		
4	**GYSCLC-FGNF**		
5	**GYSCLC-FGNF**		
6	**GYSCLC-FGNF**		
1	**GYTCICPEGYSGLFC**	slit homolog 2 protein isoform	XP_011512212.2
2	**GYTCICPEGYSGLFC**		
1	**LGYCCTNSCNYNGL**	immunoglobulin heavy chain variable region, partial	ACT68971.1
2	**LGYCCTNSCNYNGL**		
1	**LGY-CSSTSCYFGFF**	immunoglobulin heavy chain junction region	MCG41834.1
2	**LGYCSSTSCYFGFF**		
1	**LGY---CYGGL**	immunoglobulin heavy chain variable region, partial	AAK13839.1
2	**LGY---CYGGL**		
1	**LGYLCT-----FC**	glutamate receptor, metabotropic 5, isoform CRA_b, partial	EAW59359.1
5	**LLVA-L-TCYF**		
1	**LGYY---YFGL**	immunoglobulin heavy chain junction region	MOR77883.1
2	**LGYY---YFGL**		
1	**MGY---CYFGL**	immunoglobulin heavy chain junction region	MON20268.1
2	**MGYC---YFGL**		

[a]Query sequence (seq)-1: CFLGYFCTCYFGLFC, seq-2: CFLGYCCCCYFGLFC, seq-3: CYLIVGYFCTCYFGVF, seq-4: CYLVVGYFCTCYFGVF, seq-5: CYLLVGYLCTCYFGVF, seq-6: FIGYVCTCYFGVF, seq-7: GYFCTVYFGVF.

[b]Aligned sequence is displayed as it is shown in the alignment, i.e., together with the gaps that are introduced for the alignment.

[c]Aligned proteins with the same name and sequence IDs as those aligned with the SARS-CoV-2 sequence (query seq-1).

[d]Single sequence ID is indicated here, but there can be different sequence IDs with the same alignment.

It is observed at Table 2 that there are 24 different alignments of 16 human protein sequences, with the coronavirus sequences, which contains the CFLGYFCTCYFGLFC sequence (query seq-1, the SARS-CoV-2 sequence) or those homologous to that sequence. So, all the listed proteins in Table 2 are aligning with the query seq-1, which is a SARS-CoV-2 sequence, and with at least one other coronavirus sequence that is homologous to the query seq-1. Briefly, Table 2 presents the alignments of CFLGYFCTCYFGLFC as query seq-1, with the human proteins, and those alignments are common to the respective sequences of other coronaviruses, which are homologous to the 15mer SARS-CoV-2 sequence (query seq-1). Finally, Table 3 presents the MHC supertype representative-binding predictions for the coronavirus (query) and human (subject) sequence pairs, which are predicted to bind strongly to the same HLA alleles, not only with each other but also with those of the SARS-CoV-2. Namely, MHC supertype representative-binding predictions are performed for (both the query and subject) sequences that are informed at Table 2. Those that are predicted to bind strongly to the same HLA alleles with each other and at the same time, with the query and subject sequences of SARS-CoV-2, are displayed in Table 3. Immunoglobulin heavy chain junction regions with sequence IDs MOR79299.1, MOR77883.1, and MON20268.1; immunoglobulin heavy chain variable regions with sequence IDs ACT68971.1 and AAK13839.1; and metabotropic glutamate receptor 5 isoform with sequence ID EAW59359.1 are present in Table 2, but not displayed in Table 3. The query-subject sequence pairs of those protein sequences are not predicted to bind strongly to the same HLA alleles. On the other hand, immunoglobulin heavy chain variable region with the sequence ID CEF94348.1 is also absent in Table 3 because query-subject sequence pairs of those other than those of the SARS-CoV-2 are not predicted to bind strongly to the same HLA alleles there. Here, it is reminded that the query-subject sequence pairs that are predicted to bind strongly to the same HLA alleles with each other but not with that of the SARS-CoV-2 sequence are not displayed in Table 3 and prediction results with weak binding affinities are also not considered, which may have eliminated some potentially significant results.

Table 3

Human proteins and their (subject) sequences, which are aligned with the coronavirus (query) sequences that are not only homologous to the SARS-CoV-2 15mer (query seq-1) but also predicted to bind strongly to the same HLA alleles as that of the SARS-CoV-2 (query) and human (subject) epitope-pairs. These epitope-pairs are shown in consecutive rows and each epitope-pair have at least 5 matching-residues. Residues in these epitope-pairs are written bold if they are among the matching-residues in the original alignments. Those residues are further underlined if they are still present as matching-residues in the predicted epitope-pairs. HLA allele is indicated in the first column. Accordingly, the same HLA allele is displayed in case of the coronavirus (query) and human (subject) epitope-pairs. Epitopes predicted by NetMHCcons are displayed here. Epitope-pairs that are sourced by the alignments of the SARS-CoV-2 15mer (query seq-1) are shown in the third column. In the second column, 1 indicates the SARS-CoV-2 (query) sequence-number and it is written at the rows, which correspond to the same rows as the SARS-CoV-2 (query) epitopes of the epitope-pairs at the third column. Epitope-pairs that are sourced by the

alignments of the other coronavirus sequences are shown in the fifth column. The numbers in the fourth column indicates the other coronaviruses' (queries') sequence-numbers and they are written at the rows, which correspond to the same rows as the respective coronaviruses' (queries') epitopes of the epitope-pairs at the fifth column. Names of the human proteins with the aligned subject sequences are displayed at the last column.

[a]HLA	[b]Seq	Top: query epitope Bottom: subject epitope	[b]Seq	Top: query epitope Bottom: subject epitope	[c]Protein name
A2	1	FLGYFCTCYFGL	2	FLGYCCCCYFGL	
A2		FLHYYYYGL		FLHYYYYGL	Immunoglobulin heavy chain junction region
A24	1	CFLGYFCTCYFGLF	2	CFLGYCCCCYFGLF	
A24		CFSSYFFLLF		CFSSYFFLLF	hCG1995581, partial
A24	1	CFLGYFCTCYFGLF	2	CFLGYCCCCYFGLF	
A24		CFVGSCFGLF		CFVGSCFGLF	Immunoglobulin heavy chain junction region
A2	1	FLGYFCTCYFGL	2	FLGYCCCCYFGL	
A2		FLGVYSFGL		FLGVYSFGL	Phospholipid phosphatase-related protein type 2
A2	1	FLGYFCTCYFGL	2	FLGYCCCCYFGL	
A2		FLGYYYYGL		FLGYYYYGL	Immunoglobulin heavy chain junction region
A24	1	GYFCTCYFGLF	4	GYFCTCYFGVF	
A24		GYFCTNYF		GYFCTNYF	hCG2028737
A24			7	GYFCTVYF	
A24				GYFCTNYF	
A24	1	GYFCTCYFGLF	3,4	GYFCTCYFGVF	
A24		GYSCLCFGNF		GYSCLCFGNF	CRB1 isoform I precursor
A24			5	GYLCTCYFGVF	
A24				GYSCLCFGNF	
A24			6	GYVCTCYFGVF	
A24				GYSCLCFGNF	
A24	1	GYFCTCYFGLF	2	GYCCCCYFGLF	
A24		GYTCICPEGYSGLF		GYTCICPEGYSGLF	Slit homolog 2 protein isoform
A24	1	GYFCTCYFGLF	2	GYCCCCYFGLF	
A24		GYCSSTSCYFGFF		GYCSSTSCYFGFF	Immunoglobulin heavy chain junction region

[a]A2: HLA-A*02:01, A24: HLA-A*24:02.

[b]Query sequence (seq)-1: CFLGYFCTCYFGLFC, seq-2: CFLGYCCCCYFGLFC, seq-3: CYLIVGYFCTCYFGVF, seq-4: CYLVVGYFCTCYFGVF, seq-5: CYLLVGYLCTCYFGVF, seq-6: FIGYVCTCYFGVF, seq-7: GYFCTVYFGVF.

[c]Name of the proteins with the sequences in the second (bottom) lines of the query and subject epitope-pairs, which are shown side by side, as the query seq-1 (on top) and its homologous human sequence (at the bottom), and the other coronavirus sequences that are homologous to the query seq-1 (on top) and their homologous human sequences (at the bottom). The latter query and subject epitope-pairs are shaded in grey, together with the number(s) designating the respective query sequences in the front.

The predictions that are presented in Table 3 belong to NetMHCcons, but predictions are performed by using NetCTLpan as well. NetCTLpan made similar predictions for those that have presumably high affinity to the HLA-A*24:02 allele, according to NetMHCcons. The exceptions are the predictions for the alignments of query seq-2 (CFLGYCCCCYFGLFC) with hCG1995581 (sequence ID EAW57092.1), predictions for the alignments of both query seq-1 and seq-2 with the slit homolog 2 protein (sequence ID XP_011512212.2), and the alignments of query seq-2 with one of the immunoglobulin heavy chain junction region (sequence ID MCG41834.1).

It is observed in Table 3 that the predicted epitope-pairs, which include peptides that are sourced by two immunoglobulin heavy chain junction regions (sequence IDs MOQ87140.1 and MOP50498.1) and phospholipid phosphatase-related protein type 2 isoform (sequence ID XP_024307423.1), bind strongly to the HLA-A*02:01 (A2) allele. On the other hand, those that include peptides that are sourced by hCG1995581 (sequence ID EAW57092.1), two other immunoglobulin heavy chain junction regions (sequence IDs MON95290.1 and MCG41834.1), hCG2028737 (sequence ID EAW73174.1), CRB1 isoform I precursor (sequence ID AAL10681.1) and slit homolog 2 protein isoform (sequence ID XP_011512212.2), bind strongly to the HLA-A*24:02 (A24) allele. Peptides of most of these proteins that overlap with the SARS-CoV-2 peptide's sequence (seq-1) overlap with the CFLGYCCCCYFGLFC sequence (seq-2) as well. This is expected since CFLGYCCCCYFGLFC (seq-2) is closer to the SARS-CoV-2 peptide (seq-1), compared to the other sequences (Figure 1). Yet, the CRB1 isoform I precursor is the protein that is containing an antigenic, similar peptide to the highest number of different coronavirus sequences (query seq-3 to -6) that are homologous to the SARS-CoV-2 15mer (query seq-1) (Table 3). This makes the epitope, which is part of the CRB1 isoform I precursor, a strong candidate of a common mechanism of autoimmune reaction risk in the individuals with the HLA-A*24:02 serotype, sourced by SARS-CoV-2 and other coronaviruses. CRB1 related coronavirus pathogenicity is already suggested [45,46]. Here, it is suggested that possible CRB1 related coronavirus pathogenicity is additionally furthered by autoimmunity risk in certain individuals.

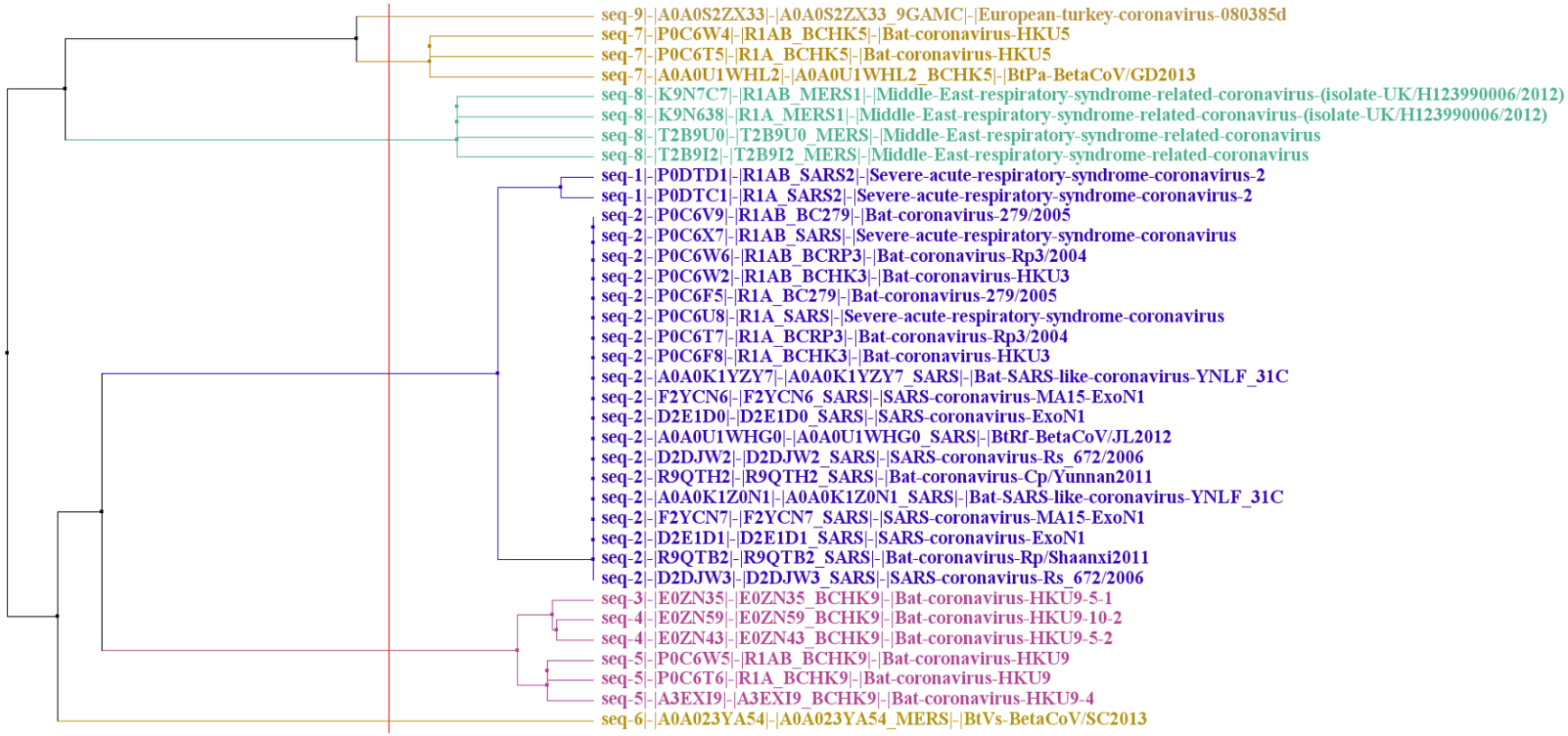

Fig. 1. Tree calculated with the alignment of 8 coronavirus sequences homologous to the SARS-CoV-2 sequence CFLGYFCTCYFGLFC (seq-1), and the SARS-CoV-2 sequence (seq-1) itself.[1] Sequences are obtained by performing Blast of the SARS-CoV-2 sequence at Uniprot, with threshold 10 (s1, *performed on 5 August 2020*). The tree is calculated with Jalview (version 2.11.1.3) [44], Blosum 62, average distance.

[1]Sequence (seq)-1: CFLGYFCTCYFGLFC, seq-2: CFLGYCCCCYFGLFC, seq-3: CYLIVGYFCTCYFGVF, seq-4: CYLVVGYFCTCYFGVF, seq-5: CYLLVGYLCTCYFGVF, seq-6: FIGYVCTCYFGVF, seq-7: GYFCTVYFGVF, seq-8: LGFMCTCYFGVF, seq-9: GWFCTCYFGLY.

Summary information of all the proteins (or peptides) in Table 3 are as follows:

Immunoglobulin heavy chain junction region

- There is 1 alignment in case of SARS-CoV-2 query sequence
- NCBI sequence ID MOQ87140.1
- 22 amino acids (*NCBI information*)

hCG1995581

- There is 1 alignment in case of SARS-CoV-2 query sequence
- NCBI sequence ID EAW57092.1

Immunoglobulin heavy chain junction region

- There are 2 alignments in case of SARS-CoV-2 query sequence, which are the same, except that they are at 2 different subject-sequence regions
- NCBI sequence IDs MON77051.1 and MON95290.1
- 19 amino acids (*NCBI information*)

Phospholipid phosphatase-related protein type 2

- There are 12 alignments in case of SARS-CoV-2 query sequence, which are the same, except that they are at 7 different subject-sequence regions
- UniProtKB - Q96GM1 (PLPR2_HUMAN)
- Gene PLPPR2
- It has binary interactions with 21 proteins. (*UniProt information*)

Immunoglobulin heavy chain junction region

- There is 1 alignment in case of SARS-CoV-2 query sequence
- NCBI sequence ID MOP50498.1
- 26 amino acids (*NCBI information*)

hCG2028737

- There is 1 alignment in case of SARS-CoV-2 query sequence
- NCBI sequence ID EAW73174.1

Protein crumbs homolog 1

- There are 17 alignments in case of SARS-CoV-2 query sequence, which are the same, except that they are at 13 different subject-sequence regions
- UniProtKB - P82279 (CRUM1_HUMAN)
- Gene CRB1
- It takes role in photoreceptor morphogenesis and may maintain cellular adhesion and polarization. (*UniProt information*)
- It has binary interactions with PATJ [Q8NI35]. (*UniProt information*)
- The associated diseases involve Leber Congenital Amaurosis 8 and Retinitis Pigmentosa 12. (*GeneCards information*)

Slit homolog 2 protein

- There are 16 alignments in case of SARS-CoV-2 query sequence, which are the same, except that they are at 12 different subject-sequence regions
- *Neurogenic extracellular slit protein Slit2 is also the name that is given at the blastp results*
- UniProtKB - O94813 (SLIT2_HUMAN)
- Gene SLIT2
- It is believed to function as a molecular guidance cue in cellular migration and it takes roles in axonal guidance and development of the parts of the neural system. (*UniProt information*)
- It has binary interactions with itself and ROBO1 [Q9Y6N7], which is its receptor. (*UniProt information*)
- The associated diseases involve Cakut and Crohn's Colitis. (*GeneCards information*)

Immunoglobulin heavy chain junction region

- There is 1 alignment in case of SARS-CoV-2 query sequence
- NCBI sequence ID MCG41834.1
- 21 amino acids (*NCBI information*)

4. Discussion

Bianchi and co-workers [47] indicated in their paper that studies involving peptide elution confirmed the presentation of transmembrane helices by the MHC class I molecules. In accordance, CFLGYFCTCYFGLFC is also likely a transmembrane region, and is predicted as a strong binder to certain HLA alleles, in the same manner as its homologous sequences in the human proteome [48]. In this study, its subset that is common to other coronaviruses, which have a peptide sequence that is homologous to CFLGYFCTCYFGLFC, is found (Table 3). Some associated diseases of the proteins with the aligned peptides as potential epitopes are already mentioned in the "summary information." In relation to this study, Lyons-Weiler [49] identified immunogenic epitopes that are present in the proteome of SARS-CoV-2. They then compared those epitopes with the proteome of human. More than 1/3 of peptides that are immunogenic were found to be homologous to the proteins that are significant in the adaptive immune system. Kanduc [50] suggested that there is an extensive range of health disorders, related to probable autoimmunity reactions against peptides of the human proteome, which have homology with the immunogenic peptides of SARS-CoV-2. Also, it was indicated in another study [51] that cerebrospinal fluids from the patients of COVID-19 were suggestive of autoimmunity. In line with those studies, the results presented here suggest autoimmunity risk in COVID-19 patients with HLA-A*02:01 and HLA-A*24:02 serotypes, by means of molecular mimicry, as a common risk to their exposure to the other coronaviruses. This is because immune responses to the peptides and proteins with similar sequences and strong binding affinities to the same HLA allele

can lead to autoimmune responses [52–58]. Yet, Trost and co-workers [55] pointed at dramatically high number of common heptapeptides (7mers) between bacterial and human proteomes. In relation, Amela and co-workers [59] demonstrated that most of the pathogen proteins that can cause antibody generation by the host immune system are not homologous to the human proteins and *vice versa*. Distinguishing self from non-self was suggested to be one contributing mechanism [60]. In addition, there need to be genetic, physiological, and environmental [13] variations in action. Besides, studies are generally performed or initiated by analyzing the reference genomes, but individual genetic variations in the identified sequences could well influence the outcomes in real, including those for the vaccines [61]. For evaluation, it is offered that the detection of cross-reacting T-cells or antibodies, epidemiological connection between exposures to or possessing the risk factors and following development of autoimmune disease, and duplicability of the conditions in animal model(s), are the needed proofs, in addition to the presence of similarity. Autoimmune reactions were shown to be developed in the animal models under such conditions (e.g. [62,63]). Yet, the last two criteria of those are challenging, and besides, not free from concerns [13].

Involvement of evolutionary processes [64] is a related concern, not only to these discussions but also to this study and its results. Vaccine targets are already studied for COVID-19 [65], but cross-reactivity risk in the adjuvant-vaccine for the individuals with genetic susceptibility should also be considered [61], together with the discussed considerations in terms of the autoimmunity risk, in general. Similar studies as those mentioned here [7–12,49,50] need to be performed, maybe even by taking possible variations in the genetic makeup into account. We eventually would like to remind once more the useful suggestion of implementing HLA typing into COVID-19 tests and clinical trials [35], and the considerations about tests on the animals [11].

It is observed here that the epitope, which is part of the CRB1 isoform I precursor, is the candidate of a common mechanism of coronavirus-sourced risk of autoimmune reaction in the individuals with the HLA-A*24:02 serotype. Interestingly, Warren and Birol [66] identified the same allele in predictions from the transcriptome sequences of bronchoalveolar lavage fluid of 4 among 5 COVID-19 patients from China. Differently, this allele was reported by the authors as not (recognized as) a SARS-risk-factor. On the other hand, only the query-subject sequence pairs that are predicted to bind strongly to the same HLA alleles with each other and with that of the SARS-CoV-2 sequence are considered here and prediction results with weak binding affinities are eliminated. This may have diminished the number of results here. Also, current study involves one SARS-CoV-2 peptide and performing a similar study with the whole proteome of the virus will be providing a much comprehensive view.

5. Conclusion

We investigated autoimmunity related common pathological mechanisms of coronaviruses, through a 15mer peptide of SARs-CoV-2 with the sequence CFLGYFCTCYFGLFC, in one letter code. For that, coronavirus sequences homologous to the SARS-CoV-2 peptide are initially obtained. Afterwards, we identified those homologous coronavirus sequences that are aligning with the human proteins with at least 7 residue matches, which are common to the alignments of the SARS-CoV-2 15mer with the human protein sequences, in a similar fashion. So, coronavirus (query) and human (subject) epitope-pairs are identified. Those epitope-pairs are the ones that are predicted to bind strongly to the same HLA alleles, not only with each other but also with that of the SARS-CoV-2 15mer and its aligned sequence in the same human protein. Immunoglobulin heavy chain junction regions, phospholipid phosphatase-related protein type 2, CRB1 isoform I precursor, and slit homolog 2 protein are among the proteins with those predicted epitopes. They can be related to pathological conditions in susceptible patients. Particularly the probable coronavirus pathogenicity related to CRB1 is suggested to be promoted by autoimmunity risk in the individuals with the HLA-A*24:02 serotypes. Overall, the results infer autoimmunity risk in COVID-19 individuals with HLA-A*02:01 and HLA-A*24:02 serotypes, in general, by means of molecular mimicry. Further, it is common to other coronaviruses. This is based on their homology to the SARS-CoV-2 15mer, and their aligning with the same human proteins as those that are aligning with the SARS-CoV-2 15mer, and their binding affinity predictions, which are revealed to be strong for the same HLA alleles. These findings can pave the way to clinical studies for autoimmunity treatment options to be used in COVID-19 and associated diseases; identification of individuals with specific HLA alleles as disease risk groups; and consideration of genetic susceptibilities in vaccine studies. Such studies can be further nourished by identification of novel alleles [67–69].

Acknowledgments - Ecology and Evolutionary Biology Society of Turkey is acknowledged.

Supporting information - Supplementary files available at shorturl.at/rswS6

Conflicts of interests - There are no conflicts of interests to declare.

References

1. Kerkar N, Vergani D. De novo autoimmune hepatitis –is this different in adults compared to children. J Autoimmun. 2018;95:26–33.

2. Metlas R, Skerl V, Veljkovic V, et al. Immunoglobulin-like domain of HIV-1 envelope glycoprotein gpl20 encodes putative internal image of some common human proteins. Viral Immunol. 1994;7(4):215–219.

3. Veljkovic V, Johnson E, Metlaš R. Molecular basis of the inefficacy and possible harmful effects of AIDS vaccine candidates based on HIV-1 envelope glycoprotein gp120. Vaccine. 1997;15(2):473–474.

4. Metlas R, Trajkovic D, Srdic T, et al. Human immunodeficiency virus V3 peptide-reactive antibodies are present in normal HIV-negative sera. AIDS Res Hum Retrovir. 1999;15(7):671–677.

5. Metlas R, Trajkovic D, Srdic T, et al. Anti-V3 and anti-IgG antibodies of healthy individuals share complementarity structures. J Acquir Immune Defic Syndr. 1999;21(4):266–270.

6. Carter CJ. Extensive viral mimicry of 22 AIDS-related autoantigens by HIV-1 proteins and pathway analysis of 561 viral/human homologues suggest an initial treatable autoimmune component of AIDS, FEMS Immunol. Med Microbiol. 2011;63:254–268.

7. Kanduc D, Shoenfeld Y. From HBV to HPV: designing vaccines for extensive and intensive vaccination campaigns worldwide. Autoimmun Rev. 2016;15:1054–1061.

8. Kanduc D, Shoenfeld Y. Inter-Pathogen Peptide Sharing and the original antigenic sin: solving a paradox. The Open Immunology Journal. 2018;8:16–27.

9. Kanduc D, Shoenfeld Y. Human papillomavirus epitope mimicry and autoimmunity: the molecular truth of peptide sharing. Pathobiology. 2019;86:285–295.

10. Kanduc D, Shoenfeld Y. On the molecular determinants of the SARS-CoV-2 attack. Clin Immunol. 2020;215:108426.

11. Kanduc D, Shoenfeld Y. Medical, genomic, and evolutionary aspects of the peptide sharing between pathogens, primates, and humans. Global Med Genet. 2020;7:64–67.

12. Kanduc D, Shoenfeld Y. Molecular mimicry between SARS-CoV-2 spike glycoprotein and mammalian proteomes: implications for the vaccine. Immunol Res. 2020;68:310–313.

13. Rojas M, Restrepo-Jiménez P, Monsalve DM, et al. Molecular mimicry and autoimmunity. J Autoimmun. 2018;95:100–123.

14. James JA, Harley JB. Linear epitope mapping of an Sm B/B′ polypeptide. J Immunol. 1992;148:2074–2079.

15. Lunardi C, Bason C, Navone R, et al. Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells. Nat Med. 2000;6:1183–1186.

16. Fussey SP, Ali ST, Guest JR, et al. Reactivity of primary biliary cirrhosis sera with Escherichia coli dihydrolipoamide acetyltransferase (E2p): characterization of the main immunogenic region. Proc Natl Acad Sci Unit States Am. 1990;87(10):3987–3991.

17. Atyeo C, Fischinger S, Zohar T, et al. Distinct early serological signatures track with SARS-CoV-2 survival. Immunity. 2020;53:524–532.

18. Kaneko N, Kuo H-H, Boucau J, et al. Loss of Bcl-6-expressing T follicular helper cells and germinal centers in COVID-19. Cell. 2020;183:1–15.

19. Kuri-Cervantes L, Pampena MB, Meng V, et al. Comprehensive mapping of immune perturbations associated with severe COVID-19. Sci Immunol. 2020;5:eabd7114.

20. Laing AG, Lorenc A, del Molino del Barrio I, et al. A dynamic COVID-19 immune signature includes associations with poor prognosis. Nat Med. 2020;26:1623–1635.

21. Lucas C, Wong P, Klein J, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. Nature. 2020;584:463–469.

22. Mathew D, Giles JR, Baxter AE, et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. Science. 2020;369:eabc8511.

23. Woodruff MC, Ramonell RP, Cashman KS, et al. Dominant extrafollicular B cell responses in severe COVID-19 disease correlate with robust viral-specific antibody production but poor clinical outcomes. MedRxiv. 2020. DOI: 10.1101/2020.04.29.20083717

24. Rodríguez Y, Novelli L, Rojas M, et al. Autoinflammatory and autoimmune conditions at the crossroad of COVID-19. J Autoimmun. 2020;114:102506.

25. Cappello F, Gammazza AM, Dieli F, et al. Does SARS-CoV-2 trigger stress-induced autoimmunity by molecular mimicry? A Hypothesis. J Clin Med. 2020;9:2038.

26. Lucchese G, Flöel A. 2020. Molecular mimicry between SARS-CoV-2 and respiratory pacemaker neurons. Autoimmun Rev. 2020;19:102556.

27. Lucchese G, Flöel A. SARS-CoV-2 and Guillain-Barré syndrome: molecular mimicry with human heat shock proteins as potential pathogenic mechanism. Cell Stress Chaperones. 2020;25:731–735.

28. Angileri F, Legare S, Gammazza AM, et al. Molecular mimicry may explain multi-organ damage in COVID-19. Autoimmun Rev. 2020;19:102591.

29. Vojdani A, Kharrazian D. Potential antigenic cross-reactivity between SARS-CoV-2 and human tissue with a possible link to an increase in autoimmune diseases. Clin Immunol. 2020;217:108480.

30. Adiguzel Y. Possible molecular mimicry through homology to a SARS-CoV-2 peptide in Plasmodium species and Human. 2021. 28 p. Accompanied by: Supplementary files available at shorturl.at/glzJ8. Located at: https://arxiv.org/abs/2101.07356

31. Blanquart S, Gascuel O. Mitochondrial genes support a common origin of rodent malaria parasites and Plasmodium falciparum's relatives infecting great apes. BMC Evol Biol. 2011;11:70.

32. Déchamps S, Maynadier M, Wein S, et al. Rodent and nonrodent malaria parasites differ in their phospholipid metabolic pathways. J Lipid Res. 2010;51:81–96.

33. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–3402.

34. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2017;46:D8–D13.

35. Nguyen A, David JK, Maden SK, et al. Human leukocyte antigen susceptibility map for Severe Acute Respiratory Syndrome Coronavirus 2. J Virol. 2020;94(13):e00510-20.

36. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2016;32(4):511–517.

37. Nielsen M, Lundegaard C, Worning P, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003;12:1007–1017.

38. Reynisson B, Alvarez B, Paul S, et al. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. 2020;48(W1):W449–W454.

39. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics. 2009;25(10):1293–1299.

40. Karosiene E, Lundegaard C, Lund O, et al. NetMHCcons: a consensus method for the Major Histocompatibility Complex Class I predictions. Immunogenetics. 2012;64(3):177–186.

41. Stranzl T, Larsen MV, Lundegaard C, et al. NetCTLpan. Pan-specific MHC class I pathway epitope predictions. Immunogenetics. 2010;62(6):357–368.

42. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47(D1):D506–D515.

43. Stelzer G, Rosen R, Plaschkes I, et al. The GeneCards suite: from gene data mining to disease genome sequence analysis. Curr Protoc Bioinformatics. 2016;54:1.30.1–1.30.33.

44. Waterhouse AM, Procter JB, Martin DMA, et al. Jalview version 2 – a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–1191.

45. De Maio F, Lo Cascio E, Babini G, et al. Improved binding of SARS-CoV-2 envelope protein to tight junction-associated PALS1 could play a key role in COVID-19 pathogenesis. Microb Infect. 2020;22(10):592–597.

46. Teoh K-T, Siu Y-L, Chan W-L, et al. The SARS Coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. Mol Biol Cell. 2010;21(22):3838–3852.

47. Bianchi F, Textor J, van den Bogaart G. Transmembrane helices are an overlooked source of Major Histocompatibility Complex Class I epitopes. Front Immunol. 2017;8:1118.

48. Adiguzel Y. Molecular mimicry between SARS-CoV-2 and human proteins [letter to the editor]. Autoimmun Rev. 2020.

49. Lyons-Weiler J. Pathogenic priming likely contributes to serious and critical illness and mortality in COVID-19 via autoimmunity. Journal of Translational Autoimmunity. 2020;3:100051.

50. Kanduc D. From Anti-SARS-CoV-2 immune responses to COVID-19 via molecular mimicry. Antibodies. 2020;9:33.

51. Lucchese G. Cerebrospinal fluid findings in COVID-19 indicate autoimmunity. Lancet Microbe. 2020;1:e242.

52. Kohm AK, Fuller KG, Miller SD. Mimicking the way to autoimmunity: an evolving theory of sequence and structural homology. Trends Microbiol. 2003;11:101–105.

53. Lule S, Colpak AI, Balci-Peynircioglu B, et al. Behcet Disease serum is immunoreactive to neurofilament medium which share common epitopes to bacterial HSP-65, a putative trigger. J Autoimmun. 2017;84:87–96.

54. Negi S, Singh H, Mukhopadhyay A. Gut bacterial peptides with autoimmunity potential as environmental trigger for late onset complex diseases: in-silico study. PloS One. 2017;12:e0180518.

55. Trost B, Lucchese G, Stufano A, et al. No human protein is exempt from bacterial motifs, not even one. Self Nonself. 2010;1:328–334.

56. Vellozzi C, Iqbal S, Broder K, Guillain-Barre syndrome, influenza, and influenza vaccination: the epidemiologic evidence. Clin Infect Dis. 2014;58:1149–1155.

57. Yuki N. Ganglioside mimicry and peripheral nerve disease. Muscle Nerve. 2007;35:691–711.

58. Zabriskie JB, Freimer EH, An immunological relationship between the group. A streptococcus and mammalian muscle. J Exp Med. 1966;124:661–678.

59. Amela I, Cedano J, Querol E. Pathogen proteins eliciting antibodies do not share epitopes with host proteins: A Bioinformatics Approach. PLoS One. 2007;2(6):e512.

60. Matzinger P. The danger model: a renewed sense of self. Science. 2002;296:301–305.

61. Kanduc D. Peptide cross-reactivity: the original sin of vaccines. Front Biosci. 2012;4:1393–1401.

62. Fujinami RS, Oldstone MB, Wroblewska Z, et al. Molecular mimicry in virus infection: crossreaction of measles virus phosphoprotein or of herpes simplex virus protein with human intermediate filaments. Proc Natl Acad Sci Unit States Am. 1983;80:2346–2350.

63. Fujinami RS, Oldstone MB. Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity. Science. 1985;230(4729):1043–1045.

64. Kanduc, D. The comparative biochemistry of viruses and humans: An evolutionary path towards autoimmunity. Biol Chem. 2019;400:629–638.

65. Ahmed SF, Quadeer AA, McKay MR. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. Viruses. 2020;12:254.

66. Warren RL, Birol I. HLA predictions from the bronchoalveolar lavage fluid samples of five patients at the early stage of the Wuhan seafood market COVID-19 outbreak. 2020. 4p. Located at: https://arxiv.org/abs/2004.07108v3

67. Cheranev V, Loginova M, Jankevic T, Kutyavina S, Rebrikov D. A novel allele, HLA-C*15:227, identified when typing COVID-19 patients. HLA. 2020;doi: 10.1111/tan.14199.

68. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. Nucleic Acids Research. 2020;48:D948-D955.

69. Marsh SGE, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. 2010;75:291–455.