Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Peptides of *H. sapiens* and *P. falciparum* that are predicted to bind strongly to HLA-A*24:02 and homologous to a SARS-CoV-2 peptide

Yekbun Adiguzel

Department of Biophysics, School of Medicine, Altinbas University, Kartaltepe Mah. Incirli Cad. No11, 34147 Bakirkoy, Istanbul, Turkey

**ARTICLE INFO**

**Keywords:**
SARS-CoV-2
Plasmodium
human leukocyte antigen
molecular mimicry
autoimmunity
disease susceptibility

**ABSTRACT**

**Aim:** This study is looking for a common pathogenicity between SARS-CoV-2 and Plasmodium species, in individuals with certain HLA serotypes.

**Methods:** 1. Tblastx searches of SARS-CoV-2 are performed by limiting searches to five Plasmodium species that infect humans. 2. Aligned sequences in the respective organisms’ proteomes are searched with blastp. 3. Binding predictions of the identified SARS-CoV-2 peptide to HLA supertype representatives are performed. 4. Blastp searches of predicted epitopes that bind strongly to the identified HLA allele are performed by limiting searches to H. sapiens and Plasmodium species, separately. 5. Peptides with minimum 60% identity to the predicted epitopes are found in results. 6. Peptides among those, which bind strongly to the same HLA allele, are predicted. 7. Step-4 is repeated by limiting searches to H. sapiens, followed by the remaining steps until step-7, for peptides sourced by Plasmodium species after step-6.

**Results:** SARS-CoV-2 peptide with single letter amino acid code CFLGYFCTCYFGLFC has the highest identity to *P. vivax*. Its YFCTCYFGLF part is predicted to bind strongly to HLA-A*24:02. Peptides in the human proteome both homologous to YFCTCYFGLF and with a strong binding affinity to HLA-A*24:02 are YYCARRFGLF, YYCHCPFGVF, and YYCQQYFFILF. Such peptides in the Plasmodium species’ proteomes are FFYTFYFFELF, YPVACLFFILF, and YFPITTFHLF. The first one belonging to *P. falciparum* has a homologous peptide (YFYLFSLELF) in the human proteome, which also has a strong binding affinity to the same HLA allele.

**Conclusion:** Immune responses to the identified-peptides with similar sequences and strong binding affinities to HLA-A*24:02 can be related to autoimmune response risk in individuals with HLA-A*24:02 serotypes, upon getting infected with SARS-CoV-2 or *P. falciparum*.

1. Introduction

COVID-19 pandemic is an ongoing global crisis with devastating effects. Chloroquine derivatives and hydroxychloroquine were among the treatment options (Perricone et al., 2020). However, they were not deemed significantly effective in decreasing death rates or mechanical ventilation risk (Magagnoli et al., 2020), in addition to serious side effects (Touret and De Lamballerie, 2020). Chloroquine and hydroxychloroquine are registered compounds for the treatment and prevention of malaria and for treatment of autoimmune diseases (Charon et al., 2020). Accordingly, this study is initiated by searching for a relationship between SARS-CoV-2 and five Plasmodium species that cause malaria in humans (Blanquart and Gascuel, 2011; Décamps et al., 2010), followed by investigating potential of autoimmune reactions through molecular mimicry. In relation, SARS-CoV-2 and Plasmodium falciparum common immunodominant regions were suggested to be explaining low COVID-19 incidence in the “malaria-endemic belt” (Iesa et al., 2020). That study was initiated mainly due to the small number of confirmed cases of COVID-19 at the African region with high malaria-prevalence (references 4–6 of (Iesa et al., 2020)). This lower incidence was highlighted in an earlier study as well (references 1,2 of Mehta et al., 2020). The presence of zoonotic viral genome in *P. vivax* (Charon et al., 2019), which was pointed also by others (De Souza, 2020), was mentioned as a basis of related work (Mehta et al., 2020). Also, one case study reported co-infection with COVID-19 and *P. vivax* in a 10-year-old boy (Kishore et al., 2020). The boy received incomplete primaquine therapy six months before, for the *P. vivax* infection. Authors concluded that it was a relapse of earlier infection, triggered by SARS-CoV-2. Yet, re-infection with *P. vivax* simultaneously with COVID-19 or a relapse in its natural course was not ruled out. However, incidences of malaria and SARS-CoV-2 were found to have a negative correlation (Mehta et al., 2020). The genomic similarity between various strains of SARS-CoV-2...
and the matryoshka RNA virus 1 (MaRNAV-1), associated with P. vivax, was also investigated (Charon et al., 2019). Accordingly, 15-base pair (bp) alignment with 93.75% identity (e-value 3.1) between SARS-CoV-2 (NC_045512.2) 22487–22502 bp region and 2796–2781 bp region of MaRNAV-1 was observed. The aligned region, which is encoding surface glycoprotein, was suggested to be small to draw any conclusion. Yet, authors stated that a shared immunogenicity cannot be ruled out. Further, their observation was valid in all different strains’ data that was used. Cross-species’ peptide sharing and high number of matches of amino acids (aa) are expected. Ultimate claim of this work is that such similarities (homologies) carry the potential of leading to inherently related autoimmunity pathologies.

The immune response to pathogen proteins is a defence mechanism of the host. However, it can cause adverse effects when pathogen protein has similar sequences to that of humans. That is molecular mimicry, which can lead to autoimmune response. Researchers Kerkar and Vergani (2018) supported this notion by findings of de novo autoimmune hepatitis associated with certain viral infections. Earlier, related work on HIV supported the concept of immune network manipulation and autoimmunity induced by HIV, through mechanisms involving molecular mimicry (Metlas et al., 1994; Metlas et al., 1999a; Metlas et al., 1999b). Associated work pointed at related problems in HIV-1 gp120-based vaccines (Veljkovic et al., 1997). Similarities and pathway analysis of homologous peptides implicated viral infection associated immune response pathologies (Carter, 2011). Kanduc and Shoefield contributed to the field with several related works (Kanduc and Shoefield, 2016; Kanduc and Shoefield, 2018; Kanduc and Shoefield, 2019) and they considered risks inherent in associated vaccines, investigated vaccines for papilloma virus L1 (HPV L1) proteins of 4 strains, and surface antigen of hepatitis B virus (Kanduc and Shoefield, 2016). Cross-reactions of immune responses against HPV L1 proteins with 20 human proteins were deemed feasible, revealing a probable connection with pathophysiological conditions, implying that peptide-based vaccines need to be pathogen-specific. Regarding SARS-CoV-2, researchers found homologous sequences between spike glycoprotein of the virus, and human surfactant and related proteins, indicative of associated pathologies (Kanduc and Shoefield, 2020a). In relation, autoimmune reaction risk of vaccines based on whole SARS-CoV-2 antigens was told to have less likelihood of being revealed through tests on primates. This expectation is due to lower peptide sharing between primates and human pathogens, including SARS-CoV-2 (Kanduc and Shoefield, 2020b). Researchers suggested ‘aged mice’ as suitable animal model for testing SARS-CoV-2 spike glycoproteins-based vaccines (Kanduc and Shoefield, 2020c).

The relationship of COVID-19 with immunity is multifaceted (Atyeo et al., 2020; Kaneko et al., 2020; Kuri-Cervantes et al., 2020; Laing et al., 2020; Lucas et al., 2020; Mathew et al., 2020). COVID-19 patients experiencing serious health threats displayed hallmarks of extrafollicular B cell activation like in autoimmune settings (Woodruff et al., 2020). Autoinflammatory and autoimmune conditions in COVID-19 are also reviewed (Rodriguez et al., 2020). COVID-19 was suggested to be linked to autoimmunity through mechanisms such as molecular mimicry and bystander activation. SARS-CoV-2 was also hypothesised to be triggering stress-induced autoimmunity through molecular mimicry (Cappello et al., 2020). Also, studies on cerebrospinal fluids of COVID-19 patients were indicative of autoimmunity Lucshee (2020). Peptide sharing between SARS-CoV-2 proteome and human brainstem respiratory pathway proteins (Lucshee and Fioel, 2020a); human heat shock proteins 90 and 60 (Lucshee and Fioel, 2020b); odorant receptor 7D4, Poly ADP-Ribose Polymerase Family Member 9, and Solute Carrier Family 12 Member 6 proteins of human (Angileri et al., 2020) were reported. Strong immune cross-reactions with SARS-CoV-2 spike protein antibody (Vojdani and Kharrziaz, 2020) can infer autoimmunity risk in susceptible individuals, through molecular mimicry (Rodrigue et al., 2020). More than one-third of the immunogenic peptides of SARS-CoV-2 have homology with proteins that are of importance for adaptive immune system (Lyons-Weiler, 2020). Also, various disorders associated with possible autoimmunity against human peptides homologous to immunogenic SARS-CoV-2 peptides were pointed at, and examination of patients’ sera for autoantibodies against those human peptides were suggested Kanduc (2020).

Certain major histocompatibility complex (MHC) allele types can cause risk in certain diseases. For instance, patients expressing human leukocyte antigen (HLA)-DRB1*04:01 are suffering from severe form of rheumatoid arthritis more, and in relation, HLA-DRB1*04:01 has a 5mer, which is shared with E. coli heat shock protein (Rojas et al., 2018). This similarity can bear risk in rheumatoid arthritis patients with HLA-DRB1*04:01, when exposed to Enterobacteriaceae. Besides, pathogen and human proteins’ peptide similarities are present in systemic lupus erythematosus (James and Harley, 1992), systemic sclerosis (Lunardi et al., 2000), and primary biliary cholangitis (Fussey et al., 1990). Regarding SARS-CoV-2, conserved viral proteome regions with high MHC allele-binding affinities are not found under positive or negative selective pressure (Nguyen et al., 2020). Nevertheless, different studies (e.g., (Cappello, 2020a; Cappello, 2020b; Sedaghat and Karimi, 2020), in addition to those mentioned) suggested related autoimmune reaction risk. Also, association of several alleles through exclusive routes can prevent sorting the risk posed by individual allele types in population-based studies. Additionally, HLA-B*46:01 (Lin et al., 2003; Alicia, 2020) and HLA-B*07:03 (Ng et al., 2004; Alicia, 2020) were found to be leading to susceptibility to the associated disease, SARS.

In Plasmodium infections, pathogen-infected cell-detection by cytotoxic CD8(+) T cells of adaptive immune system was thought to be protective through MHC class I molecules in the liver-stage malaria, but not in the blood stage malaria because erythrocytes do not express MHC class I molecules (Rivera-Correa and Rodriguez, 2020). Finding that P. vivax-infected reticulocytes keep expressing MHC class I molecules changed this concept (Junqueira et al., 2018). Regarding molecular mimicry, that between histamine-releasing factor of human and translationally controlled tumour protein of P. falciparum (MacDonald et al., 2001); vitronectin and erythrocyte membrane protein 1 of P. falciparum (Ludin et al., 2011); and ankynir, spectrin, and actin proteins of erythrocytes and P. vivax proteins (Mourao et al., 2018), were mentioned (Mourao et al., 2020).

2. Methods

2.1. Blast search

Tblastx (Altschul et al., 1997) at NCBI (NCBI Resource Coordinators, 2017) is used to compare SARS-CoV-2 (NC_045512.2) and five Plasmodium species that infect humans (P. falciparum, taxid:5833; P. malariae, taxid:5858; P. vivax, taxid:5855; P. ovale, taxid:36330; P. knowlesi, taxid:5850). Separate searches are performed by limiting searches to individual Plasmodium species. Default algorithm parameters are used, except for the number of maximum target sequences. Accordingly, the highest number of allowed target sequences was used. It was 20000 in the first searches and diminished to 5000 later, by the server, for all blast searches.

Aligned sequences with the highest identity in the tblastx outputs are used as inputs of blastp search at NCBI, performed to check the presence of those peptides in the respective organisms’ proteomes. Blastp searches are also performed using default parameters, except for increasing the number of maximum target sequences, as mentioned. Sequences are displayed as single letter amino acid (aa) codes.

2.2. MHC class I-binding

SARS-CoV-2 peptide identified in the previous step is used as input for predicting its binding affinities to MHC class I (MHC class I genes are HLA-A, -B, and -C genes) (Nguyen et al., 2020) proteins. NetMHC 4.0 (Andreatta and Nielsen, 2016; Nielsen et al., 2003), NetMHCpan 4.1
Acta Tropica 221 (2021) 106013

Y. Adiguzel

1. Introduction

1.1. Methodology

1.1.1. Blastp search of SB-peptide and specific HLA allele-binding of SB-peptide homologues

1.1.2. Specific HLA allele-binding predictions of SB-peptide homologues

1.2. Results

1.2.1. Blast search

1.2.2. Visualisation

2. Discussion

2.3. Conclusion

References

1. Introduction

1.1. Methodology

1.1.1. Blastp search of SB-peptide and specific HLA allele-binding of SB-peptide homologues

1.1.2. Specific HLA allele-binding predictions of SB-peptide homologues

1.2. Results

1.2.1. Blast search

1.2.2. Visualisation

2. Discussion

2.3. Conclusion

References
organism in tblastx search of SARS-CoV-2, other alignment results than that with the highest identity contain whole or part of CFLGYFCT-CYFGLFC in several queries (s14, s18). YFCT repeat is abundant in the alignments among alignment-hits with P. malariae genome assembly chromosome 10 (sequence ID LT594498.1) (s14). So, subject sequences in the alignments contain repeats, i.e., low complexity regions. It was mentioned that there are 12 alignments with the highest (73%) identity to CFLGYFCTCYFGLFC, within alignment results of tblastx search of SARS-CoV-2, limited to P. vivax (s15). That tblastx search result has several other alignments with the same sequence, as part or whole of aligned query (i.e., 73 alignments of a total number of 143 alignments, see Table 2) (s15, s18). In case of P. ovale and P. knowlesi, situation is like

![Diagram of tblastx and Plasmodium species' sequence similarities](image)

**Fig. 1.** The outline of the process that is followed.

| Query (COVID-19) sequence | Subject (Plasmodium) sequence | Subject name |
|---------------------------|------------------------------|--------------|
| seq.1 CNFFNCHYFR          | seq.1' CHYFRCHYFR            | *P. falciparum* |
| seq.2 CFLGYFCTCYFGLFC     | seq.2' CFVGFCTCFVGYFC        | *P. malariae*  |
| seq.3 FCFHKCFC            | seq.3' FCCFSCFC              | *P. vivax*    |
| seq.4 LVLWLLCNCFLC        | seq.4' VVLWLLCCLCWLW        | *P. ovale*    |
| seq.4' VVLWLLCCLCWLW      | -                            | *P. knowlesi* |

**Table 1**

Aligned SARS-CoV-2 query sequences and Plasmodium subject sequences with the highest identities.

![Bar chart of alignments by species](image)

**Fig. 2.** The number of alignments of whole or part of query sequences with the highest percent identities, within alignment results of tblastx searches of SARS-CoV-2, limited to Plasmodium species.
Table 3

| MHC class I type | Length | NetMHC (SB) | PickPocket (≥0.5) | NetMHCons (SB) | NetCTLpan (E) |
|------------------|--------|-------------|-------------------|----------------|---------------|
| HLA-A*24:02      | 8mer   |             |                   |                | GYFCTCYFGL    |
| HLA-A*01:01      | 9mer   |             |                   |                | FLGYFCTCYF    |
| HLA-A*02:01      | 9mer   |             |                   |                | YFCTCYFGL     |
| HLA-B*15:01      | 10mer  |             |                   |                | GYFCTCYFGL    |
| HLA-A*24:02      | 11mer  |             |                   |                | GYFCTCYFGL    |
| HLA-A*24:02      | 12mer  |             |                   |                | GYFCTCYFGL    |
| HLA-A*02:01      | 12mer  |             |                   |                | GYFCTCYFGL    |
| HLA-A*24:02      | 14mer  |             |                   |                | GYFCTCYFGL    |
| HLA-A*24:02      | 16mer  |             |                   |                | GYFCTCYFGL    |

* NetMHCons results are not displayed since there was no SB prediction there. (SB: strong binder; ≥0.5 stands for 1-log50k(aff)≥0.5; E: epitope.)
Fig. 3. Three-dimensional structure model of TCR alpha, TCR beta, YFCTCYFGLF, and MHC complex, from different perspectives. The model is obtained through TCRpMHCmodels (Jensen et al., 2019).

Table 4

| Species  | 10mer sequence | NetMHC (SB / WB / none)* | PickPocket (≥0.5 / <0.5)* | NetMHCpan (SB / WB / none)* | NetMHCcons (SB / WB / none)* | NetCTLpan (E)* |
|----------|----------------|--------------------------|---------------------------|-------------------------------|-------------------------------|----------------|
| H. sapiens | YFCARNFGPF      | SB                        | <0.5 (0.436)              | WB                            | SB                            | E              |
|          | YFCASSFGSF      | SB                        | <0.5 (0.456)              | WB                            | SB                            | E              |
|          | YYCARRFGLF      | SB                        | ≥0.5 (0.536)              | SB                            | SB                            | E              |
|          | YYCHCPFGVF      | SB                        | ≥0.5 (0.554)              | SB                            | SB                            | E              |
|          | YYQQYFILF       | SB                        | ≥0.5 (0.677)              | SB                            | SB                            | E              |

* WB: weak binder; SB: strong binder; ≥0.5 and <0.5 stands for 1-log50k(aff)≥0.5 and 1-log50k(aff)<0.5, respectively; and E: epitope.

Table 5

| Species  | 10mer sequence | NetMHC (SB / WB / none)* | PickPocket (≥0.5 / <0.5)* | NetMHCpan (SB / WB / none)* | NetMHCcons (SB / WB / none)* | NetCTLpan (E)* |
|----------|----------------|--------------------------|---------------------------|-------------------------------|-------------------------------|----------------|
| P. falciparum | FFCTPPFILF     | SB                        | <0.5 (0.446)              | WB                            | SB                            | E              |
|          | YICFYFILF      | SB                        | <0.5 (0.345)              | none                          | none                          | none           |
|          | YITCISGLF      | SB                        | <0.5 (0.288)              | none                          | none                          | none           |
| *         | FFYTFYFELF     | SB                        | ≥0.5 (0.500)              | WB                            | SB                            | E              |
|          | YICFGPSLFLF    | WB                        | <0.5 (0.316)              | none                          | none                          | none           |
| P. malariae | FFCTKYFAHF     | WB                        | <0.5 (0.642)              | WB                            | SB                            | E              |
|          | YFYACLFLF      | SB                        | ≥0.5 (0.563)              | WB                            | SB                            | E              |
| P. vivax  | YFATFYFILY     | SB                        | <0.5 (0.334)              | none                          | none                          | none           |
| P. ovale  | YFPITTFHFLF    | SB                        | ≥0.5 (0.589)              | SB                            | SB                            | E              |
| P. knowlesi | FFCAACLFLFLF   | SB                        | ≥0.5 (0.472)              | none                          | SB                            | E              |
| * H. sapiens | YYYLPSLELF    | SB                        | <0.5 (0.455)              | SB                            | SB                            | E              |

* HLA-A*24:02 binding prediction results until the last row present predictions for Plasmodium species’ sequences homologous to SB-peptide YFCTCYFGLF. Among those, the ones that are identified as SBs are written bold and termed as SB ‘-peptide. Results in the last row are the human peptide homologous to YFCTCYFGLF, the only peptide identified at the end of the search for homologous peptides of SB ‘-peptides in human proteome. (WB: weak binder; SB: strong binder; ≥0.5 and <0.5 stands for 1-log50k(aff)≥0.5 and 1-log50k(aff)<0.5, respectively; and E: epitope.)
YFCARNFGPF: Ig heavy chain variable region (Sequence ID (seq.ID) AAQ05358.1).

YFCASSFGSF: T cell receptor beta chain variable region (seq.ID ANO56516.1).

YYCARRFGLF: Immunoglobulin heavy chain variable region (seq. ID QGT38216.1).

YYCHCPFGVF: Protocadherin FAT4 (seq.ID AHN13824.1), unnamed protein product (seq.IDs BAF84150.1, BAG53346.1), protocadherin Fat 4 isoform X2 (seq.ID XP_011530539.1), Fat tumor suppressor homolog 1 (seq.ID EAX05203.1), and protocadherin Fat 4 isoforms 3, 2, and 1 (seq.IDs NP_078858.4 and Q6V0I7.2, NP_001278214.1, and NP_001278232.1, respectively).

YYCQQYFFLF: Anti-HIV immunoglobulin kappa chain variable region (seq.IDs AVQ94610.1 and AVQ94611.1).

As mentioned, there are 12 homologous sequences of YFCTCYFGLF, within Plasmodium species’ proteomes (Table 5, until the last row, s36–40). Among those, 3 sequences are predicted as SBs to HLA-A*24:02 (Table 5, bold sequences until the last row, s36–40). Those are termed SB'-peptides. To determine if any of those SB'-peptides are homologous to human proteins while binding to HLA-A*24:02 with high affinity: First, blastp searches of SB'-peptides are performed separately, by limiting searches to H. sapiens (s46–48). One homologous sequence (YFYLFSLELF) is identified through analysis of blastp search results (Table 5, the last row). That is homologous to SB'-peptide FFYTFYFELF. It belongs to LOC441426 protein in human proteome (seq.IDs AAI36787.1 and AAI36791.1), and is also SB to HLA-A*24:02 (Table 5, the last row, and s49–53).

Peptide in Plasmodium species’ proteomes homologous to the second SB-peptide GYFCTCYFGLF is the HYFPTITFHLF sequence of P. ovale. It is also predicted as SB to HLA-A*24:02 (s54). Analysis of blastp search results of that sequence in human proteome (s55) did not reveal homologous sequences.

3.4. Visualization

Fig. 4 displays sequence profiles of human- and Plasmodium-sourced 10mers that are both SBs to HLA-A*24:02 and have 6 identical residues with YFCTCYFGLF, compared to the 9mer motif of HLA-A*24:02. Three-dimensional structure model of TCR alpha, TCR beta, YFCTCYFGLF, and MHC complex is already presented (Fig. 3).

4. Discussion

CFLGYFCTCYFGLFC sequence is identified here as query sequence with the highest identity to a to P. vivax sequence. This 15mer sequence is prevalently observed within tblastx alignment result and is common within respective results of the other Plasmodium species as well (Fig. 2, Table 2). It is suggested to be a transmembrane region (UniProtKB: P0DTC1). In relation, it is mentioned in other studies (Bianchi et al., 2017) that MHC class I molecules present transmembrane helices. YFCTCYFGLF part of this 15mer peptide is predicted as SB to HLA-A*24:02 of HLA supertype representatives. Yet, the other alleles can also be SB to parts of this 15mer and they are eliminated here because we looked for agreement of prediction results of different tools. YVCARRFGLF, YYCHCPFGVF, and YYCQQYFFLF sequences in human proteome, which are among homologous sequences of YFCTCYFGLF, are predicted as SB to the same HLA allele (Table 4, bold sequences). Upon

![Image of Sequence logos](https://services.healthtech.dtu.dk/services/MHCMotifViewer/HLA-A_files/Media/HLA-A2429/HLA-A2429.jpg?disposition=download)
being infected with SARS-CoV-2, individuals with HLA-A*24:02 can develop autoimmune responses, accordingly. Further, among sequences in Plasmodium species’ proteomes homologous to YFCCTCYFGLF, 3 sequences are predicted as SBs to the same HLA allele (Table 5, bold sequences until the last row). Those pathogen sequences can trigger the immune response in HLA-A*24:02 serotypes, upon getting infected with respective Plasmodium species. Yet, the question here is whether it can lead to an autoimmune response as well, through homology to SARS-CoV-2 peptide. Therefore, we looked if any of those are homologous to human proteins that are also SB to the same HLA allele. YFYLSLELF is identified accordingly (Table 5, the last row).

Aligned SARS-CoV-2 query sequences with gaps are eliminated in this study. Additionally, aligned subject sequences with gaps in regions other than the positions that align with the investigated sequence regions were also eliminated. This process diminished the number of sequences that were obtained. In support, relevant work Adiguzel, 2021a reported a higher number of results by presenting alignments of the same 15mer with human proteome, unrestricted to alignments’ gaps. Yet, results outlined here are obtained although restrictions implemented.

Other than the discussions above, this study is conducted using reference genomes, meaning that individual variations in indicated sequences can influence outcomes, but they were not evaluated. Further genetic, physiological, and environmental variations can also contribute. All can influence possible immune responses and autoimmune reactions, but they cannot be evaluated in studies with single, representative reference genomes.

Peptides with homologous regions and strong binding affinities to the same HLA allele possess autoimmune reaction risk (Zabriskie and Freimer, 1966; Kohn et al., 2003; Yuki, 2007; Lule et al., 2017; Negi et al., 2017; Trost et al., 2010; Vellozzi et al., 2014). Intense rise in the number of shared heptapeptides among bacteria and human proteomes is reported (Trost et al., 2010). Yet, antibody generating pathogen proteins are not homologous to human proteins, and vice versa (Amela et al., 2015). These indicate that mere resemblance is generally not the only reason for developing autoimmunity, which should involve more parameters, including environmental factors (Rojas et al., 2018).

Distinguishing self from non-self is a probable contributor (Matzinger, 2002). To account for molecular mimicry leading to an autoimmune disease, in addition to homologous host and pathogen proteins, the presence of cross-reacting T-cells or antibodies, epidemiological link and autoimmune disease generation, and reproducibility in an animal model, are required. In accordance, autoimmune reactions were observed in animal models (e.g., Fujimami et al., 1983; Fujimami and Oldstone, 1985)). However, the latter two criteria are still challenging and prone to concerns (Rojas et al., 2018). In any case, findings of in silico studies as this one need to be supported by further evidence, in terms of autoimmune susceptibility risk. Considering these criticisms, this work presents a condition of autoimmunity susceptibility risk of certain HLA serotypes, upon being infected with SARS-CoV-2 and P. falciparum, through homology to the same SARS-CoV-2 peptide.

Implications of the current study also highlight the importance of accounting for genetic variations in vaccine development. Far more genetic variations than HLA serotypes exist. Cross-reactivity between adjuvant-vaccine and patient proteome can possess autoimmune reaction risk in genetically susceptible individuals (Kanduc, 2012). Therefore, homologies between adjuvant vaccines and human proteome need to be investigated (Kanduc and Shoenfeld, 2016; Kanduc and Shoenfeld, 2018; Kanduc and Shoenfeld, 2019; Kanduc and Shoenfeld, 2020a; Kanduc and Shoenfeld, 2020b; Kanduc and Shoenfeld, 2020c) by taking genetic variations into account. Individuals to be included in the respective clinical trials should be representative of allelic variation of HLA types. As suggested (Nguyen et al., 2020), the implementation of HLA typing into ongoing and upcoming clinical trials and COVID-19 tests would be extremely useful. Such considerations are of importance for tests on animals as well since autoimmune responses would not necessarily be indicative of the same situation in humans, and vice versa (Kanduc and Shoenfeld, 2020b).

In case of COVID-19, HLA class I and II alleles’ prediction from transcriptome sequencing data of bronchoalveolar lavage fluid samples of 5 COVID-19 patients from China also led to the identification of HLA-A*24:02, and it was deemed significant (Warren and Birol, 2020), compared to the frequency of the allele in South Han Chinese population (Middleton et al., 2003). Authors mentioned that HLA-A*24:02 (A24 allele group) was unknown as a risk factor for SARS, but it is associated with diabetes (Adamashvili et al., 1997; Noble et al., 2002; Nakashima and Inoko, 2006; Kronenberg et al., 2012), which is known as a risk factor in COVID-19 (Guan et al., 2020). HLA-A*24 is reported as an independent predictor of 5-year progression to diabetes in autoantibody-positive first-degree relatives of patients with Type 1 diabetes (Mbenue et al., 2013). Malaria is also associated with diabetes (Danquah et al., 2010; Acquah, 2019; Carrillo-Larco et al., 2019). In the mentioned COVID-19 study (Warren and Birol, 2020), HLA class II alleles DPA1*02:02 and DPB1*05:01 were found to be frequent as well. Those alleles are common in Han Chinese and associated with autoimmune diseases Graves’ (Chu et al., 2018) and narcolepsy (Ollila et al., 2015). The latter one, DPB1*05:01, is associated with chronic hepatitis B in Asians, in addition to being a risk factor effecting viral infection clearance ability ((Kamataki et al., 2009; Ollila et al., 2015) cited in (Middleton et al., 2003)). Accordingly, class II allele-binding predictions of SARS-CoV-2 15mer identified here is planned as a future work.

4.1. Evolutionary perspective

Peptide sharing events or high number of matches in peptides potentially carry autoimmune response risk in individuals with genetic susceptibilities. Observation of matches in peptides of different species that possess such risks is expected. The probability of a hexapeptide to occur is 1/64000000 (Kanduc, 2020), which is calculated through equal likelihood of occurrence of 20 aa at any of the 6 positions in a hexapeptide. Proteome size of the organism, to which blastp search is limited, is the site to realise that possibility: 1/64000000. An unbiased proteome with 64 million aa would be sufficient for observing any 6mer for once, considering that such a proteome would be comprised of 64 million hexamers (ignoring the dissection of proteome into proteins). 122962 proteins with the total number of 80769298 aa are present in human proteome (retrieved from the protein-list at NCBI, genome assembly id 582967). This size of a proteome can let every possible 6mer to be observed once, at least. In this study, we looked for a minimum of 6 aa-matches in a 10mer and did not observe more than 6 aa-matches. The expected occurrence of 6 aa-matches in a 10mer is 210 and that of observing 6mers is 5. 210 is the number of 6-membered subsets in a 10-membered set, calculated by 10!/6!(4!) (each member is the aa at each position of the 10mer). However, we observed only few sequences with 6 aa-matches (Table 4), namely, few of the 6-membered subsets. So, the observation is less than expected. This lessening is because the calculated number is expected number of observations for unbiased conditions, but human proteome is biased, and most proteins have several isomers. These results somehow support views about involvement of evolutionary processes between human hosts and SARS-CoV-2 (Kanduc, 2019; Kanduc, 2020), in relation to a different part of the concept. The evolution of viral proteome can be providing a selection advantage based on infectivity, or elimination by the host. For instance, proteome of P. falciparum strain 3D7 reference genome has 5387 proteins with total number of 4103467 aa (data retrieved from the protein-list at NCBI, genome assembly id 895506). It is about 20 times smaller than that of the human proteome. Yet, that proteome also revealed 5 peptides with 6 aa-matches to the 10mer query peptide, the first SB-peptide (Table 5). This amount can be interpreted as less bias in P. falciparum proteome than in human proteome. P. falciparum proteins do not have several isomers like in case of human proteins, which supports this interpretation. Still, this observation should be validated with
a larger and varied query set. Currently, it is interesting that only one of the three SB'-peptides of Plasmodium species in Table 5 has a 6 aa-matching peptide in human proteome. This observation can be related to host-pathogen evolution or co-evolution of pathogens, or both. Also, 32 Plasmodium sequences are present among 17657 sequences in descriptions table of blastn search of SARS-CoV-2 with somewhat similar sequences (s56), which contributed to this view. In relation, 10mer part of SARS-CoV-2 peptide investigated here have homologous peptides with 6 aa-matches both in human proteome and in Plasmodium species’ proteomes. Yet only one of those Plasmodium species’ peptides has one such peptide in human proteome. In relation, shared immunodominant regions between P. falciparum and SARS-CoV-2 were suggested to be explaining low incidence of COVID-19 within the malaria-endemic belt (Iesa et al., 2020). Here, autoimmune reaction risk is studied through such a similarity. Accordingly, low incidence in malaria-endemic belt can be due to the higher number of homologous peptides of Plasmodium species and SARS-CoV-2. It can cause more protection by triggering immune response than an autoimmune reaction. This claim is supported here by that only one of the three SB’-peptides of Plasmodium species have a homologous peptide in human proteome that is SB to the same allele (Table 5). Accordingly, probability of a protective effect due to shared peptides between pathogen is higher than their risk of causing autoimmune response. This difference is because only a subset, i.e., a smaller number, of those shared peptides are SB to the same HLA allele and have a homologous peptide in human proteome, which is also SB to the same HLA allele. It can be interpreted also as follows: Host-pathogen evolution and pathogen co-evolution at the “malaria-endemic belt” worked towards diminished autoimmune responses that involve molecular mimicry. It can be protective against severe cases of COVID-19 with autoimmune reactions involving associated mechanisms. Yet, this scenario does not rule out the potential of inherently related autoimmune pathologies through peptide homologies. Besides, such similarities are expected, although physiological outcomes are complex. Still, results of in silico studies like this one require experimental validation or clinical support, or both.

Eventually, studies like this one can pave grounds to certain HLA serotypes’ recognition as risk groups and identification of novel alleles (Marsh et al., 2010; Cherevanev et al., 2021; Robinson et al., 2020). It also highlights the importance of relevant considerations in vaccine studies and can contribute to studies involving commonalities between pathogens of human host, driven by host-pathogen evolution. As outlined nicely (Córdoba-Aguilar et al., 2021), what is as much important is the possibility of lessons that will be learned from the current pandemic to trigger an understanding leading to the development of disease prevention and environmental protection strategies for the well-being of not only humanity but also nature itself.

5. Conclusion

In this study, 10mer peptide at positions 5–14 of CFLGYFCTCYFGLFC sequence of SARS-CoV-2 is identified as a possible relationship between SARS-CoV-2 and P. falciparum. Homologous 10mers of that peptide in human proteome are predicted to have strong binding affinities to the HLA-A*24:02 allele. Plasmodium species other than P. falciparum have homologous 10mers of that peptide with high affinities to the same HLA allele. However, only the 10mer of P. falciparum has a homologous 10mer in human proteome, with strong affinity to that HLA allele. These results imply autoimmunity susceptibility of HLA-A*24:02 serotypes through molecular mimicry in SARS-CoV-2 and P. falciparum infections, due to homology to the same SARS-CoV-2 peptide. This result is yet to be supported by clinical and experimental findings.

Supporting information

Supplementary files available at shorrturl.at/bmzP6 and deposited at Mendeley Data under the title Dataset1_ADIGUZEL (Adiguzel, 2021b)

Availability of data and materials

Data as supplementary files deposited at Database: Mendeley Data (Dataset1_ADIGUZEL) (Adiguzel, 2021b) and available as supplementary files under the shortened URL (URL shortened at https://www.shorrturl.at/ ) of the Google Drive link.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

CRediT authorship contribution statement

Yekbun Adiguzel: Conceptualization, Funding acquisition, Formal analysis, Data curation, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Ecology and Evolutionary Biology Society of Turkey is acknowledged.

References

Acquah, S., 2019. Linking malaria to type 2 diabetes mellitus: a review. J. Ghana Sci. Assoc. 18, 56–70.
Adamashvili, I., McCabe, R., Gelder, F., Gautreaux, M., Jaramillo, J., Roggero, T., McDonald, J., 1997. Soluble HLA class I antigens in patients with type I diabetes and their family members. Human Immunol. 55, 176–183.
Adiguzel, Y., 2021a. Molecular mimicry between SARS-CoV-2 and human proteins. Autoimmun. Rev. 20, 102791.
Adiguzel, Y., 2021b. Dataset1_ADIGUZEL. Mendeley Data, v1. https://doi.org/10.17632/ccwy9yp9v1.1.
Alicia, S.M., 2020. HLA studies in the context of coronavirus outbreaks. Swiss Med. Wkly. 150, w20248.
Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402.
Amelio, I., Cedano, J., Querol, E., 2007. Pathogen proteins eliciting antibodies do not share epitopes with host proteins: a bioinformatics approach. PLoS One 2, e512.
Andreatta, M., Nielsen, M., 2016. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 32, 511–517.
Angileri, F., Legare, S., Gammazza, A.M., de Macario, E.C., Macario, A.J., Cappello, F., 2020. Molecular mimicry may explain multi-organ damage in COVID-19. Autoimmun. Rev. 19, 102591.
Ayten, C., Fischinger, S., Zohar, T., Slein, M.D., Burke, J., Loos, C., McCulloch, D.J., Newman, K.L., Wolf, C., Yu, J., Shury, K., Feldman, J., Hauser, B.M., Caradonna, T., Schmidt, A.G., Surochov, T.J., Linde, C., Cai, Y., Baroch, D., Ryan, E.T., et al., 2020. Distinct early serological signatures track with SARS-CoV-2 survival. Autoimmun. Rev. 19, 102591.
Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P., 2000. The protein data bank. Nucleic Acids Res 28, 235–242.
Bianchi, F., Textor, J., van den Bogaart, G., 2017. Transmembrane helices are an overlooked source of Major Histocompatibility Complex Class I epitopes. Front. Immunol. 8, 1118.
Cappello, F., 2020a. Is COVID-19 a proteiform disease inducing also molecular mimicry?

Chu, X., Yang, M., Song, Z.J., Dong, Y., Li, C., Shen, M., Zhu, Y.Q., Song, H.D., Chen, S.J., Carter, C.J., 2011. Extensive viral mimicry of 22 AIDS-related autoantigens by HIV-1

Carrillo-Larco, R., Altez-Fernandez, C., Ugarte-Gil, C., 2019. Is diabetes associated with molecular mimicry? A systematic review of observational studies.

Wellcome Open Res 4, 136.

Carter, C.J., 2011. Extensive viral mimicry of 22 AIDS-related autoantigens by HIV-1

Fujinami, R.S., Oldstone, M.B., 1985. Amino acid homology between the matrix protein of the human cytomegalovirus and a virus with poor prognosis. Nat. Med. 26, 1623–1625.

Gouma, S., Weirick, M.E., et al., 2020. Comprehensive mapping of immune perturbations associated with severe COVID-19. Sci. Immunol. 5, eabd7114.

Vantourout Laing, A.G., Lorenc, A., del Molino del Barrio, I., Das, A., Isho, Y., Sulaiman, M., de Ru, A., Kronenberg, D., Knight, R.R., Estorninho, M., Ellis, R.J., Kester, M.G., de Ru, A., Kanduc, D., Shoenfeld, Y., 2020a. Molecular mimicry between SARS-CoV-2 spike glycoprotein and mammalian proteomes: implications for the vaccine. Immunol. Lett. 210, 10832.

Kanduc, D., Kuo, H.-L., Boucau, J., Farmer, J.R., Allard-Chamard, H., Mahajan, V.S., Piechocka-Trocha, A., Lefteri, K., Osborn, M., Bals, J., Bartsch, Y.C., Bonhur, N., Caradonna, T.M., Chevalier, V., Chowdhury, F., Dienfench, T.J., Einkauf, K., Fallon, J., Feldman, J., Finn, K.K., et al., 2020. Loss of Bcl-6-expressing T follicular helper cells and germinal centers in COVID-19. Cell 183, 1–15.

Kurose, E., Lundegaard, C., Lund, O., Nielsen, M. 2012. NetMHCcons: a consensus method for the best histocompatibility class complex I predictions.

Immunogenetics 64, 177–184.

Kanduc, D., Shoenfeld, Y., 2020. Molecular mimicry between SARS-CoV-2 spike glycoprotein and mammalian proteomes: implications for the vaccine. Immunol. Lett. 1604–1606.

Kanduc, D., 2020. From anti-SARS-CoV-2 immune responses to COVID-19 via molecular mimicry. Antibodies 9, 33.

Kanduc, D., 2020. From anti-SARS-CoV-2 immune responses to COVID-19 via molecular mimicry. Antibodies 9, 33.

Kanduc, D., 2020. From anti-SARS-CoV-2 immune responses to COVID-19 via molecular mimicry. Antibodies 9, 33.

Kanduc, D., 2020. From anti-SARS-CoV-2 immune responses to COVID-19 via molecular mimicry. Antibodies 9, 33.
Mathew, D., Giles, J.R., Baxter, A.E., Oldridge, D.A., Greenleaf, A.R., Wu, J.E., Alanis, C., Kuri-Cervantes, L., Pampena, M.B., D’Andrea, K., Manne, S., Chen, Z., Huang, V.J., Beilby, J.P., Weisman, A.R., Ittner, C.A., Kuthuru, O., Dougherty, J., Nzingha, K., Han, N., et al., 2020. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. Science 369, eabc5111.

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

Metlas, R., Trajkovic, D., Srdic, T., Veljkovic, V., Colombatti, A., 1999a. Human immunodeficiency virus V3 peptide-reactive antibodies are present in normal HIV-negative sera. AIDS Res. Hum. Retrovir. 15, 671–677.

Metlas, R., Trajkovic, D., Srdic, T., Veljkovic, V., Colombatti, A., 1999b. Anti-V3 and anti-IgG antibodies of healthy individuals share complementarity structures. J. Acquir. Immune. Defic. Syndr. 21, 266–270.

Middleton, D., Menchaca, L., Rood, H., Vagcrew, R., 2003. New allele frequency database: http://www.allelefrequencies.net. Tissue Antigens 61, 403–406.

Negi, S., Singh, H., Mukhopadhyay, A., 2006. Combination of HLA-A24, -DQA1*03, and -DR9 locus-disparate pHLA recognition by delta /alpha beta T cell receptors (TCRs): insight into recognition of HIV peptides by TCRs. J. Virol. 91, e00717–e00725.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Negi, S., Singh, H., Mukhopadhyay, A., 2006. Combination of HLA-A24, -DQA1*03, and -DR9 locus-disparate pHLA recognition by delta /alpha beta T cell receptors (TCRs): insight into recognition of HIV peptides by TCRs. J. Virol. 91, e00717–e00725.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Negi, S., Singh, H., Mukhopadhyay, A., 2006. Combination of HLA-A24, -DQA1*03, and -DR9 locus-disparate pHLA recognition by delta /alpha beta T cell receptors (TCRs): insight into recognition of HIV peptides by TCRs. J. Virol. 91, e00717–e00725.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Nakamura, H., Koibuchi, T., Carlson, J., Martin, E., Brumme, C.J., Shi, Y., Gao, G.F., Brumme, Z.L., Fukai, S., Iwamoto, A., 2013. Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. Sci. Rep. 3, 2701.

Negi, S., Singh, H., Mukhopadhyay, A., 2006. Combination of HLA-A24, -DQA1*03, and -DR9 locus-disparate pHLA recognition by delta /alpha beta T cell receptors (TCRs): insight into recognition of HIV peptides by TCRs. J. Virol. 91, e00717–e00725.