Homology between SARS CoV-2 and human proteins

Vladimir Khavinson1,3, Alexander Terekhov1, Dmitry Kormilets2 & Alexander Maryanovich1*

An extremely high contagiousness of SARS CoV-2 indicates that the virus developed the ability to deceive the innate immune system. The virus could have included in its outer protein domains some motifs that are structurally similar to those that the potential victim’s immune system has learned to ignore. The similarity of the primary structures of the viral and human proteins can provoke an autoimmune process. Using an open-access protein database Uniprot, we have compared the SARS CoV-2 proteome with those of other organisms. In the SARS CoV-2 spike (S) protein molecule, we have localized more than two dozen hepta- and octamers homologous to human proteins. They are scattered along the entire length of the S protein molecule, while some of them fuse into sequences of considerable length. Except for one, all these n-mers project from the virus particle and therefore can be involved in providing mimicry and misleading the immune system. All hepta- and octamers of the envelope (E) protein, homologous to human proteins, are located in the viral transmembrane domain and form a 28-mer protein E_{14-41} VNSVLLFLAFVVFLVTLAILTALRLCA. The involvement of the protein E in provoking an autoimmune response (after the destruction of the virus particle) seems to be highly likely. Some SARS CoV-2 nonstructural proteins may also be involved in this process, namely ORF3a, ORF7a, ORF7b, ORF8, and ORF9b. It is possible that ORF7b is involved in the dysfunction of olfactory receptors, and the S protein in the dysfunction of taste perception.

The interaction of SARS CoV-2 with the host immune system is largely determined by the structural similarities between viral and host proteins. The studies of SARS CoV-2 are still focused on the S protein.

An extremely high contagiousness of the coronavirus SARS CoV-2 indicates that during its evolution the virus developed the ability to deceive the innate immune system. The simplest way to achieve this ability would be to incorporate into its membrane the proteins that share structural similarity with those which the immune system of the potential victim has learnt to ignore. Probably, the virus borrowed some n-mers from bats or other mammals. Any motif of any mammalian protein was suitable for borrowing, if only the immune system considered it to be of its own.

The knowledge of the homology between the SARS CoV-2 and human proteins would help understand the mechanisms of mimicry at the moment of infection. The SARS CoV-2 proteins may simulate human proteins, mislead the immune system, and slow down its response.

However, mimicry is not the only process that is determined by the protein homology between the virus and host organism. After the inevitable destruction of the virus particle, the proteins or their domains, which were inside the virus until then, come into contact with the immune system. With some structural similarity, a part of the immune response will be directed against the proteins of the host organism, i.e., an autoimmune response will arise.

This study aimed to identify the human proteins which share a significant structural homology with the SARS CoV-2 proteins. We hope this information will be useful to the developers of vaccines against coronavirus.

Joshua Lederberg5 believed that "microbes and their human hosts constitute a superorganism." According to this, we considered the concept of "human proteins" as a combination of human own proteome and the proteomes of gut microbiota. We have paid particular attention to the proteins that are involved in the three functions that are almost necessarily affected in this disease, namely digestion, olfaction and taste.

1Mechinkov North-Western State Medical University, 47 Piskaryovsky Prosp., 195067 St. Petersburg, Russia. 2Kirov Military Medical Academy, St. Petersburg, Russia. 3Saint Petersburg Institute of Bioregulation and Gerontology, St. Petersburg, Russia. *email: atm52atm52@gmail.com
Methods

Using an open-access protein database Uniprot and our original computer program Ouroboros\(^1\), we compared the SARS CoV-2 proteome\(^2\) with those of other organisms. We also searched for a separate database of 75,777 human proteins\(^5\). The algorithm we used compares primary sequences of SARS CoV-2 and human proteins, presented in the form of a one-letter code. We performed a comparison of proteins by a consecutive search for regions of one protein in the others, which is essentially a standard task of finding a substring in a string. This algorithm is implemented in standard methods of many programming languages, including Python, in which the main program was coded. The URL to the source code is provided above\(^3\).

When assessing the homology between the viral and human proteins, we took into account the presence of the common 7-/8-mers and especially their fusion into longer sequences. For example, 7-dimensional viruses, one of which is homologous to the human protein A, and the other to the protein B, can "overlap" at the ends, forming regions of 8 to 14 amino acid residues in length.

Results and discussion

Structural proteins. Spike glycoprotein. S protein, 1273 aa.

Hereinafter, regions homologous to human proteins are highlighted in red. Transmembrane tail TM\(^{1214-1237}\) is underlined.

In the S protein molecule, we localized more than two dozen of 7-/8-mers homologous to human proteins (Table 1).

Fragments homologous to human proteins are scattered along the entire length of the S protein molecule, and some of them fuse in sequences of considerable length, namely 10-mers

$$\text{SPRRARSVAS} \quad 680-689$$

11-mers

$$\text{GLTVLP-PLLTD} \quad 857-867$$

and two closely spaced 7-mers

$$\text{NASVVNI} \quad 1173-1179$$

and

$$\text{EIDRLNE} \quad 1182-1188.$$  

Octamer

$$\text{RRARSVAS} \quad 682-689$$

is located at the junction of the S1 and S2 subunits. All these n-mers stand out from the virus particles and may be involved in the effect of mimicry.

SARS CoV-2 can cause smell and taste dysfunction, as well as muscle injury\(^6\). The 8-mer

$$\text{DEDDSEPV} \quad 1257-1264,$$

located in the cytoplasmic tail, can be released during the destruction of the virus particle and get involved in orchestrating the immune system’s response, directing a part of it to the homologous 8-mer in human unconventional myosin-XVI\(^1404-1421\). The role of this mechanism in muscle dysfunction in coronavirus infection deserves a special investigation.

The 8-mer

$$\text{RRARSVAS} \quad 682-689$$

is homologous to the amiloride-sensitive sodium channel subunit alpha\(^201-208\), which is involved in salt taste perception\(^7\).

With a high degree of probability, it can be argued that the S protein is involved in the process of mimicry. It may also take some part in provoking an autoimmune response.

We have checked the S protein homology across 10 species, specifically primates, bats and some other mammals. The results are presented in Table entitled Similarity of SARS CoV-2 spike glycoprotein structure with some mammalian proteins in the electronic attachment. Probably, attention should be paid to the homologous regions common to SARS CoV-2, humans, and bats. The data presented so far do not allow us to derive a more general rule.

Envelope small membrane protein. E protein, 75 aa (transmembrane domain\(^8-38\) is underlined).

In the E protein molecule, we localized seven 7-mers and one 8-mer homologous to human proteins (Table 2).

A fragment of the E\(^{8-38}\) protein transmembrane domain can be represented as follows:

$$\text{EGTTLIVNSVLLFAFVVFLVTLTALRlca}$$

Envelope small membrane protein. E protein, 75 aa (transmembrane domain\(^8-38\) is underlined).

In the E protein molecule, we localized seven 7-mers and one 8-mer homologous to human proteins (Table 2).

A fragment of the E\(^{8-38}\) Protein transmembrane domain can be represented as follows:

$$\text{EGTTLIVNSVLLFAFVVFLVTLTALRlca}$$
The simulation targets may have been the proteins synthesized by a macroorganism itself or by its normal gut microbiota.

All protein E 7-/8-mers, homologous to proteins of humans, gut bacteria and cereals, are located in the transmembrane domain of the virus and form the 28-mer protein E14-41. A random selection of 28 amino acid residues in a row would require an astronomical number of iterations: $20^{28} = 2.7 \cdot 10^{36}$.

Table 1. Localization of homologous 7-/8-mers in the S protein and human proteins.

| Subunit | SARS-CoV-2 S protein domain | In S protein | In human proteins |
|---------|-----------------------------|-------------|-------------------|
| S1      | Signal peptide (N-terminus) | None        | –                 |
|         | N-terminus domain NTD 41-335 | DKVFRSS 56-66 | Zinc finger protein 52B 157-161 |
|         |                             | PLPPPFSN 57-61 | OTU domain-containing protein 6A 149-154 |
|         |                             | VSGNNGT 58-64 | Lysosome-associated membrane glycoprotein 1 175-187 |
|         |                             | ELLIVYN 59-65 | ATP-binding cassette sub-family A member 10 205-211 |
|         |                             | FKNLREF 60-66 | Isovaleryl-CoA dehydrogenase, mitochondrial 7 87-93 |
|         |                             | TRFQTL 62-68 | Disheveled-associated activator of morphogenesis 2 213-217 |
|         |                             | KTMSRFH 69-75 | Uncharacterized protein C1orf105 13 |
|         |                             | SSSGWT 76-82 | Uncharacterized protein KIAA1109 (Fragment) 603-616 |
|         | Uncharacterized fragment 56-314 | None | – |
|         | Receptor-binding domain RBD 19-541 | KLNDLC 183-190 | Interleukin-7 149-155 |
|         |                             | DEVRQTA 191-201 | Histone-lysine N-methyltransferase 2C 202-210 |
|         | Uncharacterized fragment 424-787 | VSYSG 250-256 | Neural cell adhesion molecule 1 283-294 |
|         |                             | IGAC 257-262 | Hepatitis A virus cellular receptor 2 285-292 |
|         |                             | SFRAR 263-275 | Hermansky-Pudlak syndrome 1 protein 286-297 |
|         |                             | ARS 276-287 | Amlodipine-sensitive sodium channel subunit alpha 1 288-294 |
| S2      | Fusion peptide FP 377-400 | None | – |
|         | Uncharacterized fragment 367-451 | VTLADA 382-390 | Non-receptor tyrosine-protein kinase TK1 398-404 |
|         |                             | GLTVLP 391-401 | FHI/FH2 domain-containing protein 3 392-400 |
|         | Heptapeptide repeat sequence 1 HR1 404-464 | SSTAS 405-412 | 40S ribosomal protein S13 448-455 |
|         | Uncharacterized fragment 396-1142 | VKEAEV 437-446 | Emilin-3 525-531 |
|         |                             | TGRQL 447-455 | Neuron navigator 3 546-554 |
|         | Heptapeptide repeat sequence 2 HR2 1163-1213 | NASVNU 1209-1215 | Unconventional myosin-XVIIa 1325-1333 |
|         | Transmembrane tail TM 1214-1237 | None | – |
|         | Cytoplasm tail CT 1238-1273 | DEDD 1274-1281 | Unconventional myosin-XVI 1405-1411 |

Table 2. Localization of homologous 7-/8-mers in the E protein and human proteins. *Domain boundaries see in8. b Heptamer TALRLCA 35-41 is located at the junction of the transmembrane domain 35-38 and internal domain 39-75.

| E protein domains | In E protein | In human proteins |
|-------------------|-------------|-------------------|
| Signal peptide (N-terminus domain) 7- | None | – |
| Transmembrane domain 35-38 | VNSVLLF 14-20 | Heterogeneous nuclear ribonucleoprotein L191-197 |
|         | VNSVLLF 14-20 | Ran-binding protein 6 198-204 |
|         | NSSVLLF 14-28 | Lysosomal amino acid transporter 1 homolog 128-139 |
|         | VSSLPLF 19-25 | Cytochrome P450 2B6 34-40; Cytochrome P450 2B7 41-51 |
|         | LAFVVFL 21-27 | Solute carrier family 15 member 5 230-241 |
|         | VFLLVTL 25-31 | Alpha-(1,3)-fucosyltransferase 10 36-42 |
|         | LAILTLAL 31-37 | Transient receptor potential cation channel subfamily M member 6 394-408; Transient receptor potential cation channel subfamily M member 3 409-415 |
|         | TALRLCA 35-41 | Protein disulfide-isomerase TMX3 34-41 |
| Internal domain 39-75 | None | – |

The simulation targets may have been the proteins synthesized by a macroorganism itself or by its normal gut microbiota.

All protein E 7-/8-mers, homologous to proteins of humans, gut bacteria and cereals, are located in the transmembrane domain of the virus and form the 28-mer protein E 35-41. A random selection of 28 amino acid residues in a row would require an astronomical number of iterations: $20^{28} = 2.7 \cdot 10^{36}$. 

---

Scientific Reports | (2021) 11:17199 | https://doi.org/10.1038/s41598-021-96233-7
The involvement of the E protein in mimicry is hardly possible, but its implication in provoking an autoimmune response (after the destruction of the virus particle) seems very likely. As a major target, the viral E protein has usually been used for the development of vaccines, specifically against HIV-1, Dengue virus, hepatitis B virus, SARS-CoV-2 and many other viruses. A deletion of the SARS-CoV E protein reduces pathogenicity and mortality in laboratory animals. In the transmembrane domain of the SARS-CoV E protein, specific critical virulence-determining features have been identified.

Membrane protein. Membrane protein, 222 aa.
In the M protein molecule, we localized six 7-mers homologous to human proteins (Table 4).
A N-terminus fragment 1-19 of the M protein can be represented as follows:

```
MADSNTITVEELKKLEQWNLVIGFLF
```

In the protein M, four 7-dimensional homologues of human proteins are fused into 10-mer VEELKKLLEQ10-19, the hydrophilic composition of which indicates a possible contact with the external environment, i.e., with the host's immune system, and the involvement in mimicry.

Outside of the 10-mer, we found only two homologous 7-mers. It is unlikely that the M protein is involved in provoking an autoimmune response.

Nonstructural proteins. All non-structural proteins of SARS-CoV-2 are located completely inside the virus particle and, by definition, cannot be involved in the process of mimicry. It remains to consider the possibility of their implication in provoking an autoimmune process.

Table 3. Localization of some of homologous 7-/8-mers in the E protein and human gut proteome.

| In E protein | In bacterial and plant proteins |
|--------------|--------------------------------|
| FVVFLLV     | Lpp126 large-conductance mechanosensitive channel: Lactobacillus casei, L. paracasei, L. flororum |
| TLAILTA     | Uncharacterized proteins: Zea mays, Sorghum bicolor, Triticum aestivum, Hordeum vulgare |

Table 4. Localization of homologous 7-mers in the M protein and human proteins.

| In M protein | In human proteins |
|--------------|------------------|
| VEELKKL     | Glutaredoxin-related protein, mitochondrial |
| EELKKLL     | GDP-fucose protein O-fucosyltransferase 2 |
| LKELLEQ     | Calcinin |
| LLESELV     | Filamin-A-interacting protein |
| AGDSGFA     | Myosin |

Table 5. Localization of homologous 7-mers in the N protein and human proteins.

| In N protein | In human proteins |
|--------------|------------------|
| SKQLQQSMSSADS | Myosin |
| AEGSRGGSQA   | Filamin-A-interacting protein |

Membrane protein. Membrane protein, 222 aa.
In the M protein molecule, we localized six 7-mers homologous to human proteins (Table 4).
A N-terminus fragment 1-19 of the M protein can be represented as follows:

```
MADSNTITVEELKKLEQWNLVIGFLF
```

In the protein M, four 7-dimensional homologues of human proteins are fused into 10-mer VEELKKLEQ10-19, the hydrophilic composition of which indicates a possible contact with the external environment, i.e., with the host's immune system, and the involvement in mimicry.

Outside of the 10-mer, we found only two homologous 7-mers. It is unlikely that the M protein is involved in provoking an autoimmune response (after the destruction of the virus particle).

Nonstructural proteins. All non-structural proteins of SARS-CoV-2 are located completely inside the virus particle and, by definition, cannot be involved in the process of mimicry. It remains to consider the possibility of their implication in provoking an autoimmune process.
ORF3a protein. ORF3a protein, 275 aa.
In the ORF3a protein molecule, we localized five 7-mers homologous to human proteins (Table 6).

Table 5. Localization of homologous 7-mers in the N protein and human proteins.

| In N protein | In human proteins |
|--------------|-------------------|
| RPQGLPN| GATOR complex protein WDR5976,763 |
| GGQVPI| Putative uncharacterized protein encoded by LINCO0345,140,160 |
| NSSPDQ| NEDD4-binding protein 2,144,160 |
| GRMKLS| Chromodomain-helicase-DNA-binding protein 1-like175,756 |
| VILPQG| Prestin15,56 |
| AEGRGGG| RNA-activating protein complex subunit33,4 |
| SKGQSO| Ras-associated and dilute domain-containing protein 968,992 |
| KADETQA| Myopalladin95,96 |
| LLPAADL| Probable E3 ubiquitin-protein ligase HERC11588,158 |
| SKGLOQ| Codanin-11,208,209 |
| SMSSADDL| Protein PRRC2B,146,147 |

Table 6. Localization of homologous 7-mers in the ORF3a protein and human proteins.

| In ORF3a protein | In human proteins |
|------------------|-------------------|
| VGVALLA| Manganese-transporting ATPase 13A1576,802 |
| LLVAAL| Glycerophosphoinositol inositolphosphodiesterase GDPD2,229,229 |
| KCRSRKP| Vacular protein sorting-associated protein 13A15808,2072 |
| SVTSSIV| Protein piccolo777,2783 |
| TTSSDST| Septin-14,400,404 |

Table 7. Localization of homologous 7-mers in the ORF7a protein and human proteins.

| In ORF7a protein | In human proteins |
|------------------|-------------------|
| VAAIVFI| Transmembrane protein 25S8,92 |
| FTLKRT| Cytosolic 5'-nucleotidase 3A11,42 |

ORF7a protein. ORF7a protein, 121 aa.
In the ORF7a protein molecule, we found two 7-mers homologous to human proteins and located in close proximity to each other (Table 7).

It is possible that ORF7a is involved in provoking an autoimmune response.

ORF7b protein. ORF7b protein, 43 aa.
In this polypeptide, we found only one 7-mer homologous to the human protein (Table 8).

ORF7b may be involved in provoking an autoimmune response, contributing to olfactory dysfunction.
ORF8 protein. ORF8 protein, 121 aa.

The primary structure of SARS-CoV-2 ORF8 is close to that of bat RaTG13-CoV. In this polypeptide, there are three 7-mers homologous to human proteins (Table 9).

Due to the fusion of two 7-mers into 10-mer LVFLGIIITTV4-13, the ORF8 protein can be involved in provoking an autoimmune response.

Table 8. Localization of the homologous 7-mer in ORF7b and a human protein.

| In ORF7b protein | In human protein |
|------------------|-----------------|
| IIFWFSL26-32     | Olfactory receptor 7D4151-157 |

Table 9. Localization of homologous 7-mers in the ORF8 protein and human proteins.

| In ORF8 protein | In human proteins |
|-----------------|-------------------|
| LVFLGII4-10     | Zinc finger protein 48649–55 |
| LGIITTV7-13     | D-2-hydroxyglutarate dehydrogenase, mitochondrial262-268 |
| KLGSLVV94-100   | Sodium leak channel non-selective protein |

Table 10. Localization some of homologous 7-/8-mers in ORF9b protein and human proteins.

| In ORF9b protein | In human proteins |
|------------------|-------------------|
| LVDPQIQ14-21     | Valine—tRNA ligase, mitochondrial1002 |
| MENAVGR18-32     | Neprilysin19-43 |
| LGSPSL14-54      | Stress-responsive DNAJB4-interacting membrane protein 116-43 |
| GSPLSLN4-55      | E3 ubiquitin-protein ligase HERC2415-459 |
| TEELPDEFVV86-93  | KH homology domain-containing protein 446-471 |
| LGSPLSLN48-55    | E3 ubiquitin-protein ligase HERC2415-459 |

MKFLVFLGIITTVAAFHQ8CSLQSCTQHQPYVDDPCPIHFYSKYIRVAGKSAPLIELCVDEAGSKSPIQYIDIGNYTSVCLPFTINCQEPKLGSLVRC5SFEDELYHDVRVVLDFI

ORF8 protein. ORF8 protein, 121 aa.

The primary structure of SARS-CoV-2 ORF8 is close to that of bat RaTG13-CoV. In this polypeptide, there are three 7-mers homologous to human proteins (Table 9).

Due to the fusion of two 7-mers into 10-mer LVFLGIIITTV4-13, the ORF8 protein can be involved in provoking an autoimmune response.

Table 8. Localization of the homologous 7-mer in ORF7b and a human protein.

| In ORF7b protein | In human protein |
|------------------|-----------------|
| IIFWFSL26-32     | Olfactory receptor 7D4151-157 |

Table 9. Localization of homologous 7-mers in the ORF8 protein and human proteins.

| In ORF8 protein | In human proteins |
|-----------------|-------------------|
| LVFLGII4-10     | Zinc finger protein 48649–55 |
| LGIITTV7-13     | D-2-hydroxyglutarate dehydrogenase, mitochondrial262-268 |
| KLGSLVV94-100   | Sodium leak channel non-selective protein |

Table 10. Localization some of homologous 7-/8-mers in ORF9b protein and human proteins.

| In ORF9b protein | In human proteins |
|------------------|-------------------|
| LVDPQIQ14-21     | Valine—tRNA ligase, mitochondrial1002 |
| MENAVGR18-32     | Neprilysin19-43 |
| LGSPSL14-54      | Stress-responsive DNAJB4-interacting membrane protein 116-43 |
| GSPLSLN4-55      | E3 ubiquitin-protein ligase HERC2415-459 |
| TEELPDEFVV86-93  | KH homology domain-containing protein 446-471 |
| LGSPLSLN48-55    | E3 ubiquitin-protein ligase HERC2415-459 |

MDPKISEMHPALRLVDQIQ1LAVTRMENAVGRQNNVGPKVYPIILRLGSPLSLNMARKTLNSLEDKAFQLTPIAVQMTKLATEELPDEFVVVVTKV

ORF9b protein. ORF9b protein, 97 aa.

In the ORF9b protein molecule, we localized six 7-/8-mers, homologous to human proteins (Table 10).

Some of these 7-/8-mers merge into larger n-mers TEELPDEFVV86-93 and LGSPLSLN48-55. Octamer ELPDEFVV86-93 is homologous to the Maestro heat-like repeat-containing protein family member 2B (Fig. 1), which may play a role in the sperm capacitation. Male reproductive dysfunction was proposed as a likely consequence of COVID-19.
Figure 1. The SARS CoV-2 S, E and ORF9b protein molecules contain hepta/octamers that are homologous to proteins in the human body, including some nutrients and intestinal commensal bacteria.

| In Replicase polyprotein 1a | In human proteins                                                                                     |
|-----------------------------|------------------------------------------------------------------------------------------------------|
| EVEKGVLF         56–62      | Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1214–221                                |
| ESQLKTLDD      156–163     | Annexin A7, 404–411                                                                                   |
| REETGLLM       218–231     | Estrogen-related receptor gamma30–57                                                                  |
| GGSCVLSG       311–318     | Sorting nexin-2, 2712–2719                                                                            |
| D1QLLKSA       1127–1134   | Echinoderm microtubule-associated protein-like 138–145                                                |
| RRSFVYV        2242–2249   | Transmembrane protein adipocyte-associated 1225–232                                                   |
| ARRNLLPFF      2733–2740   | Acyl-CoA:lysophosphatidylglycerol acyltransferase 1199–206                                           |
| YNYEPTQT      2008–2007    | DNA helicase 199–206                                                                                 |
| SLKELQHN       3538–3557   | Centromere protein 146–153                                                                            |
| DTLSGFP        3627–3678   | Solute carrier family 12 member 799–1002                                                              |
| PEANQCFR       4312–4319   | Arachidonate 5-lipoxygenase-activating protein 54–61                                                  |

Table 11. Localization of homologous 8-mers in RPP 1a and human proteins.
Replicase polyprotein RPP 1a. Replicase polyprotein RPP 1a, 4405 aa.

The longest n-mers are underlined. In the RPP 1a molecule, we localized eleven 8-mers (Table 11) and more than a hundred 7-mers homologous to human proteins. Some of the 8-mers are found in more than one human protein, some fold into long n-mers, for example EDIQKLNSAYENFQH

1126-1141, EVEKGVLPQLEQPY

55-68 and SVEEVLSEARQHL34-46.

In the RPP 1a molecule, 7-mers SCGNFKV

505-511 and AIFYLIT

2785-2791 are homologous to human olfactory receptor proteins 52N2190-196 and 2W132-38, respectively. A heptamer LKTLQVA

1556-1562 is homologous to the human bitter taste receptor T2R55181-187 (Fig. 2).

Replicase polyprotein RPP 1ab. This huge (7096 aa; the primary structure see in18) molecule contains 210 hepta- and octamers homologous to human proteins. Some of them fold into long (more than 15 aa) n-mers.

The possibility of the involvement of replicases in provoking an autoimmune response is debatable. Enzymes in general, and cell cycle enzymes in particular, are evolutionarily highly conserved. Fragments homologous to human proteins must be thrown in huge quantities into the gut lumen during the decay of any microorganism that dies there. It is possible that the interaction of replicases with the host's immune system obeys the laws other than for shorter proteins.

ORF6, ORF10, and ORF14. In these polypeptides (61, 38, and 73 aa, respectively), we did not find 7-/8-mers homologous to human proteins. When assessing the role of SARS CoV-2 proteins in mimicry and provoking an autoimmune process in humans, we considered the following parameters: (i) the number of homologous n-mers; (ii) the compactness of their arrangement in the SARS CoV-2 protein molecules; (iii) intradomain localization (external, transmembrane, internal) of the SARS CoV-2 proteins, and (iv) physiological functions that involve the homologous human proteins (Table 12).

Conclusions

Analysis of homology between the SARS CoV-2 and human proteins led us to the following conclusions. Some of the SARS CoV-2 proteins can be implicated in mimicry that can delay the response of innate immunity to the invasion of virus particles into a macroorganism, and in provoking an autoimmune process that directs a part
of the immune response to the proteins of a macroorganism (after the destruction of virus particles). Mimicry is probably more characteristic of the spike (S) protein, and the provocation of an autoimmune response seems to be a distinctive feature of the envelope (E) protein. The ORF7b protein may be involved in the impairment of olfactory receptors, and the S protein may be involved in taste perception dysfunction.

Drugs aimed at destructing or blocking these and alike regions in proteins of SARS CoV-2 and other viruses can enable the human immune system not to succumb to viral deception and destroy the invader shortly after its penetration into a macroorganism. It should also be borne in mind that drugs affecting such imitation regions can damage native proteins present of the human body. Destroying or blocking such regions can weaken the autoimmune response.

**Table 12.** Qualitative assessment of the possibility for the SARS CoV-2 proteins to be involved in the processes of mimicry and provoking an autoimmune response.

| Group of proteins | Protein | Mimicry | Autoimmune response | Comment |
|-------------------|---------|---------|---------------------|---------|
| Structural        | S       | +++     | +                   | Taste?—Amiloride-sensitive sodium channel subunit alpha_201–208
                                                Muscle contraction?—Unconventional myosin-XVI_1404–1421 |
|                   | E       | −       | +++                 | Gut microbiota?—Lactobacillus paracasei
                                                Digestion?—Cereals’ proteins |
|                   | M       | ++      | −                   | |
|                   | N       | −       | +                   | |
| Nonstructural     | ORF3a   | −       | +                   | No homology |
|                   | ORF6    | −       | −                   | |
|                   | ORF7a   | −       | +                   | |
|                   | ORF7b   | −       | +                   | Smell?—Olfactory receptor 7D4
                                                Gut microbiota?—Lactobacillus curvatus |
|                   | ORF8    | −       | ++                  | |
|                   | ORF9b   | −       | ++                  | Sperm capacitation?—Maestro heat-like repeat-containing protein family member 2B_103–110 |
|                   | ORF10   | −       | −                   | No homology |
|                   | ORF14   | −       | −                   | No homology |
|                   | RPP1a   | −       | ?                   | Taste?—T2R55 receptor
                                                Smell?—Olfactory receptors 2W1 and 52N2
                                                Gut microbiota?—Eubacterium sp. |
|                   | RPP1ab  | −       | ?                   | |
Data availability
The highest.

Code availability
Source code of Ouroboros (v. 0.5) is fully available at github. URL: https://github.com/liquidbrainisstrain/ouroboros. Artwork: We used GIMP (Version 2.10.22) to create our artwork. The figures are completely original and have not been published anywhere.

Received: 12 April 2021; Accepted: 2 August 2021
Published online: 25 August 2021

References
1. Sanami, S. et al. Design of a multi-epitope vaccine against SARS-CoV-2 using immunoinformatics approach. Int. J. Biol. Macromol. 164, 871–883. https://doi.org/10.1016/j.ijbiomac.2020.07.117 (2020).
2. Lederberg, J. Infectious history. Science 288(5464), 287–293 (2000).
3. Terekhov, A. Ouroboros (Version 0.5) [Source code]. https://github.com/liquidbrainisstrain/ouroboros.
4. Proteomes: Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2). https://www.uniprot.org/proteomes/UP000464624 SARS-COV-2, accessed 20 Aug 2020.
5. Proteomes: Homo sapiens (Human). https://www.uniprot.org/proteomes/UP00005640 Homo sapiens, accessed 03 Sept 2020.
6. Koralnik, I. J. & Tyler, K. L. COVID-19: A global threat to the nervous system. Ann. Neurol. 88(1), 1–11. https://doi.org/10.1002/ana.25807 (2020).
7. Huang, T. & Stähler, F. Effects of dietary Na+ deprivation on epithelial Na+ channel (ENaC), BDNF, and TrkB mRNA expression in the rat tongue. BMC Neurosci. 10, 19. https://doi.org/10.1186/1471-2202-10-19 (2009).
8. Mandalia, V. S. et al. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. Nat. Struct. Mol. Biol. 27(12), 1202–1208. https://doi.org/10.1038/s41594-020-00536-8 (2020).
9. Li, S. W. et al. Gene editing in CHO cells to prevent proteolysis and enhance glycosylation: Production of HIV envelope proteins as vaccine immunogens. PLoS ONE 15, e0233866. https://doi.org/10.1371/journal.pone.0233866 (2020).
10. Rathore, A. S., Sarker, A. & Gupta, R. D. Production and immunogenicity of Fubc subunit protein redesigned from DENV envelope protein. Appl. Microbiol. Biotechnol. 104, 4333. https://doi.org/10.1007/s00253-020-10541-y (2020).
11. Ho, J. K., Jeevan-Raj, B. & Netter, H. J. Hepatitis B Virus (HBV) subviral particles as protective vaccines and vaccine platforms. Viruses 12, 126. https://doi.org/10.3390/v12020126 (2020).
12. Abdelmaged, M. I. et al. Design of a multiepitope-based peptide vaccine against the E protein of human COVID-19: An immunoinformatics approach. Biomed. Res. Int. 2020, 2633286. https://doi.org/10.1155/2020/2633286 (2020).
13. DeDiego, M. L., et al. Inhibition of NF-κB-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival. J. Virol. 88, 913–924 (2014).
14. Regla-Nava, J. A. et al. Severe acute respiratory coronavirus syndromes with mutations in the E protein are attenuated and promising vaccine candidates. J. Virol. 89, 3870–3887 (2015).
15. Hassan, S. S. et al. A unique view of SARS-CoV-2 through the lens of ORF8 protein. Comput. Biol. Med. 133, 104380. https://doi.org/10.1016/j.compbiomed.2021.104380 (2021).
16. MROH2B: Function. https://www.nextprot.org/entry/NX_Q7Z745.
17. Sansone, A. et al. Addressing male sexual and reproductive health in the wake of COVID-19 outbreak. J. Endocrinol. Invest. 44(2), 223–231. https://doi.org/10.1007/s40618-020-01350-1 (2021).
18. Replicase polyprotein 1ab [Severe acute respiratory syndrome coronavirus 2], https://www.ncbi.nlm.nih.gov/protein/P0DTD1.17

Author contributions
A.M. and V.K. wrote the main manuscript text. A.T. and D.K. prepared data analysis. All authors reviewed the manuscript.

Funding
This research is an authors’ initiative project funded exclusively from their personal sources.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2021