scientific reports

Check for updates

OPEN

Homology between SARS CoV-2 and human proteins

Vladimir Khavinson[1,3], Alexander Terekhov[1], Dmitry Kormilets[2] & Alexander Maryanovich[1]✉

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 $E_{14\text{-}41}$ VNSVLLFLAFVVFLLVTLAILTALRLCA. 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[1].

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 Lederberg[2] 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.

[1]Mechinkov North-Western State Medical University, 47 Piskaryovsky Prosp., 195067 St. Petersburg, Russia. [2]Kirov Military Medical Academy, St. Petersburg, Russia. [3]Saint 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[3], we compared the SARS CoV-2 proteome[4] 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.

S protein, 1273 aa

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRG
WIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSK
HTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVE
KGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAP
GQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSF
ELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCT
EVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTIS
VTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFN
KVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQF
NSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMS
ECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIV
NNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCC
MTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

Hereinafter, regions homologous to human proteins are highlighted in red. Transmembrane tail $TM_{1214\text{-}1237}$ 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 $SPRRARSVAS_{680\text{-}689}$, 11-mers $GLTVLP\text{–}PLLTD_{857\text{-}867}$ and two closely spaced 7-mers $NASVVNI_{1173\text{-}1179}$ and $EIDRLNE_{1182\text{-}1188}$. Octamer $RRARSVAS_{682\text{-}689}$ 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 $DEDDSEPV_{1257\text{-}1264}$, 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-$XVI_{1404\text{-}1421}$. The role of this mechanism in muscle dysfunction in coronavirus infection deserves a special investigation.

The 8-mer $RRARSVAS_{682\text{-}689}$ is homologous to the amiloride-sensitive sodium channel subunit $alpha_{201\text{-}208}$, 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 across10 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 attachement. 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.

MYSFVSEETGTLIVN**SVLLFLA**FVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV

Envelope small membrane protein. E protein, 75 aa (transmembrane $domain_{8\text{-}38}$ 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 $E_{8\text{-}38}$ protein transmembrane domain can be represented as follows:

ETGTLIVN**SVLLFLA**FVVFLLVTLAILTALRLCA

The size of the letters (point size) corresponds to the frequency of the viral 7-/8-mers in the human proteome.

The protein E transmembrane domain contains 7-/8-mers, homologous to the proteins of some gut bacteria and even cereals, for example, corn, sorghum, wheat, and barley (Table 3).

Subunit	SARS CoV-2 S protein domain	In S protein	In human proteins
S1	Signal peptide (N-terminus)$_{1–13}$	None	–
	N-terminus domain NTD$_{14-305}$	DKVFRSS$_{40-46}$	Zinc finger protein 528$_{275–281}$
		FLPFFSN$_{55-61}$	OTU domain-containing protein 6A$_{185-191}$
		VSGTNGT$_{70-76}$	Lysosome-associated membrane glycoprotein 1$_{171–177}$
		SLLIVNN$_{116-122}$	ATP-binding cassette sub-family A member 10$_{825–831}$
		FKNLREF$_{186-192}$	Isovaleryl-CoA dehydrogenase, mitochondrial$_{77-83}$
		TRFQTLL$_{236-242}$	Disheveled-associated activator of morphogenesis 2$_{251–257}$
		KIYSKHT$_{202-208}$	Uncharacterized protein C1orf105$_{7-13}$
		SSSGWTA$_{254-260}$	Uncharacterized protein KIAA1109 (Fragment)$_{610–616}$
	Uncharacterized fragment$_{306-318}$	None	—
	Receptor-binding domain RBD$_{319-541}$	KLNDLCF$_{386-392}$	Interleukin-7$_{149–155}$
		DEVRQIA$_{405-411}$	Histone-lysine N-methyltransferase 2C$_{4530-4536}$
	Uncharacterized fragment$_{542-787}$	VYSTGSN$_{635-641}$	Neural cell adhesion molecule L1-like protein$_{341-347}$
		IGAGICA$_{666-672}$	Hepatitis A virus cellular receptor 2$_{205–211}$
		SPRRARS$_{680-686}$	Hermansky-Pudlak syndrome 1 protein$_{258-264}$
		RRARSVAS$_{682-689}$	Amiloride-sensitive sodium channel subunit alpha$_{201-208}$
S2	Fusion peptide FP$_{788-806}$	None	—
	Uncharacterized fragment$_{807-911}$	VTLADAG$_{826-832}$	Non-receptor tyrosine-protein kinase TNK1$_{440-446}$
		GLTVLPP$_{857-863}$	FH1/FH2 domain-containing protein 3$_{972–978}$
		LPPLLTD$_{861-867}$	Maestro heat-like repeat-containing protein family member 9$_{250–256}$
	Heptapeptide repeat sequence 1 HR1$_{912-984}$	SSTASAL$_{939-945}$	40S ribosomal protein S13$_{143-149}$
		LVKQLSS$_{962-968}$	E3 SUMO-protein ligase PIAS1$_{284-290}$
	Uncharacterized fragment$_{985-1162}$	KVEAEVQ$_{986-974}$	Emilin-3$_{625–631}$
		TGRLQSL$_{998-1004}$	Neuron navigator 3$_{1610–1616}$
		LIRAAEI$_{1012-1018}$	Unconventional myosin-XVIIIa$_{1352-1358}$
		LDKYFKN$_{1152-1158}$	Follistatin-related protein 1$_{149–155}$
	Heptapeptide repeat sequence 2 HR2$_{1163-1213}$	NASVVNI$_{1173-1179}$	Thyroid adenoma-associated protein$_{1022-1028}$
		EIDRLNE$_{1182-1188}$	Protein SETSIP$_{64-70}$; Protein SET$_{54-60}$
	Transmembrane tail TM$_{1214-1237}$	None	–
	Cytoplasm tail CT$_{1238-1273}$	DEDDSEPV$_{1257-1264}$	Unconventional myosin-XVI$_{1404-1421}$

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

E protein domains[a]	In E protein	In human proteins
Signal peptide (N-terminus domain)$_{1–7}$	None	–
Transmembrane domain$_{8-38}$	VNSVLLF$_{14-20}$	Heterogeneous nuclear ribonucleoprotein L$_{191-197}$
	VNSVLLFL$_{14-21}$	Ran-binding protein 6$_{409–416}$
	NSVLLFL$_{15-21}$	Lysosomal amino acid transporter 1 homolog$_{133-139}$
	SVLLFLA$_{16-22}$	Cytochrome P450 2B6$_{4-10}$; Cytochrome P450 2B7$_{4-10}$; GPI ethanolamine phosphate transferase 3$_{5–11}$
	LAFVVFL$_{21-27}$	Solute carrier family 15 member 4$_{235–241}$
	VFLLVTL$_{25-31}$	Alpha-(1,3)-fucosyltransferase 10$_{20–26}$
	LAILTAL$_{31-37}$	Transient receptor potential cation channel subfamily M member 6$_{394–400}$; Transient receptor potential cation channel subfamily M member 3$_{465–471}$
	TALRLCA$_{35-41}$[b]	Protein disulfide-isomerase TMX3$_{8-14}$
Internal domain$_{39-75}$	None	–

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

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$_{14-41}$. A random selection of 28 amino acid residues in a row would require an astronomical number of iterations: $20^{28} = 2.7 \cdot 10^{36}$.

In E protein	In bacterial and plant proteins
AFVVFLLV$_{22-29}$	Lpp126 large-conductance mechanosensitive channel: Lactobacillus casei$_{80-87}$; L. paracasei$_{80-87}$; L. florum$_{80-87}$
TLAILTA$_{30-36}$	Uncharacterized proteins: Zea mays$_{90-164}$; Sorghum bicolor$_{97-127}$; Triticum aestivum$_{116-190}$; Hordeum vulgare$_{87-161}$

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

In M protein	In human proteins
VEELKKL$_{10-16}$	Glutaredoxin-related protein 5, mitochondrial$_{135-141}$
EELKKLL$_{11-17}$	GDP-fucose protein O-fucosyltransferase 2$_{340-346}$
ELKKLLE$_{12-18}$	Cullin-1$_{335-341}$
LKKLLEQ$_{13-19}$	Filamin-A-interacting protein 1$_{211-217}$
LLESELV$_{133-139}$	Leucine-rich repeat-containing protein 71$_{439-445}$
AGDSGFA$_{188-194}$	Myosin-14$_{359-365}$

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

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[9], Dengue virus[10], hepatitis B virus[11], SARS CoV-2[12] and many other viruses. A deletion of the SARS-CoV E protein reduces pathogenicity and mortality in laboratory animals[13]. In the transmembrane domain of the SARS-CoV E protein, specific critical virulence-determining features have been identified[14].

MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNRFLYIIKLIFLWLLWPVTLACFVLAAVYRINWITGGIAIAMACL
VGLMWLSYFIASFRLFARTRSMWSFNPETNILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSR
TLSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIALLVQ

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:

MADSNGTITVEELKKLLEQWNLVIGFLF

In the protein M, four 7-dimensional homologues of human proteins are fused into 10-mer VEELKKLLEQ$_{10-19}$, 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).

MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHGKEDLKFPRGQGVPINTNSSPDDQ
IGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTT
LPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQQQGQTVTKK
SAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLT
YTGAIKLDDKDPNFKDQVILLNKHIDAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSAD
STQA

Nucleoprotein. Nucleoprotein, 419 aa.
In the N protein molecule, we localized eleven 7-mers homologous to human proteins (Table 5).

The N protein is located completely inside the virus particle and cannot be involved in mimicry. All heptamers homologous to human proteins form several rather long fragments, including the 13-mer SKQLQQSMSSADS$_{404-416}$ and 10-mer AEGSRGGSQA$_{173-182}$, which increases the likelihood of the protein involvement 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.

In N protein	In human proteins
$RPQGLPN_{41-47}$	GATOR complex protein $WDR59_{757-763}$
$RGQGVPI_{68-74}$	Putative uncharacterized protein encoded by $LINC00346_{154-160}$
$NSSPDDQ_{77-83}$	NEDD4-binding protein $2_{154-160}$
$GKMKDLS_{99-105}$	Chromodomain-helicase-DNA-binding protein 1-$like_{770-776}$
$VLQLPQG_{157-163}$	$Prestin_{92-98}$
$AEGSRGG_{173-179}$	snRNA-activating protein complex $subunit3_{2-8}$
$SRGGSQA_{176-182}$	Ras-associating and dilute domain-containing $protein_{886-892}$
$KADETQA_{375-381}$	$Myopalladin_{90-96}$
$LLPAADL_{394-400}$	Probable E3 ubiquitin-protein ligase $HERC1_{1098-1104}$
$SKQLQQS_{404-410}$	$Codanin\text{-}1_{259-265}$
$SMSSADS_{410-416}$	Protein $PRRC2B_{416-422}$

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

In ORF3a protein	In human proteins
$VGVALLA_{48-54}$	Manganese-transporting ATPase $13A1_{876-882}$
$LLVAAGL_{95-101}$	Glycerophosphoinositol inositolphosphodiesterase $GDPD2_{129-135}$
$KCRSKNP_{132-138}$	Vacuolar protein sorting-associated protein $13A_{2066-2972}$
$SVTSSIV_{162-168}$	Protein $piccolo_{2779-2785}$
$TQLSTDT_{217-223}$	$Septin\text{-}14_{418-424}$

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

In ORF7a protein	In human proteins
$VAAIVFI_{104-110}$	Transmembrane protein $255B_{86-92}$
$FTLKRKT_{114-120}$	Cytosolic 5'-nucleotidase $3A_{36-42}$

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

MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFGWLIVGVALLAVFQSASKIITLKKRWQLALSKGVHFVCNLLLLF
VTVYSHLLLVAAGLEAPFLYLYALVYFLQSINFVRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPYNSVTSSIVITSGDG
TTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVME
PIYDEPTTTTSVPL

ORF3a protein. ORF3a protein, 275 aa.
In the ORF3a protein molecule, we localized five 7-mers homologous to human proteins (Table 6).

The 7-mers scattered along the entire length of its molecule do not form long n-mers anywhere else. ORF3a does not appear to be involved in provoking an autoimmune response.

MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNSPFHPLADNKFALTCFSTQFAFACPDGVKHVYQLRARSVSPKLF
IRQEEVQELYSPIFLIVAAIVFITLCFTLKRKTE

ORF7a protein. ORF7a 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.

MIELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNETCHA

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.

In ORF7b protein	In human protein
IIFWFSL$_{26-32}$	Olfactory receptor 7D4$_{151-157}$

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

In ORF8 protein	In human proteins
LVFLGII$_{4-10}$	Zinc finger protein 486$_{49-55}$
LGIITTV$_{7-13}$	D-2-hydroxyglutarate dehydrogenase, mitochondrial$_{262-268}$
KLGSLVV$_{94-100}$	Sodium leak channel non-selective protein$_{505-511}$

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

In ORF9b protein	In human proteins
LVDPQIQL$_{14-21}$	Valine—tRNA ligase, mitochondrial$_{996-1002}$
MENAVGR$_{26-32}$	Neprilysin$_{419-425}$
LGSPLSL$_{48-54}$	Stress-responsive DNAJB4-interacting membrane protein 1$_{37-43}$
GSPLSLN$_{49-55}$	E3 ubiquitin-protein ligase HERC2$_{4533-4539}$
TEELPDE$_{84-90}$	KH homology domain-containing protein 4$_{465-471}$
ELPDEFVV$_{86-93}$	Maestro heat-like repeat-containing protein family member 2B$_{103-110}$

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

MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYTVSCLPFT
INCQEPKLGSLVVRCSFYEDFLEYHDVRVVLDFI

ORF8 protein. ORF8 protein, 121 aa.
The primary structure of SARS-CoV-2 ORF8 is close to that of bat RaTG13-CoV[15]. 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 LVFLGIITTV$_{4-13}$, the ORF8 protein can be involved in provoking an autoimmune response.

MDPKISEMHPALRLVDPQIQLAVTRMENAVGRDQNNVGPKVYPIILRLGSPLSLNMARKTLNSLEDKAFQLTPIAVQMTKLAT
TEELPDEFVVVTVK

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 TEELPDEFVV$_{84-93}$ and LGSPLSLN$_{48-55}$.

Octamer ELPDEFVV$_{86-93}$ is homologous to the Maestro heat-like repeat-containing protein family member 2B (Fig. 1), which may play a role in the sperm capacitation[16]. Male reproductive dysfunction was proposed as a likely consequence of COVID-19[17].

After the destruction of the virus particle, ORF9b can take part in provoking an autoimmune response.

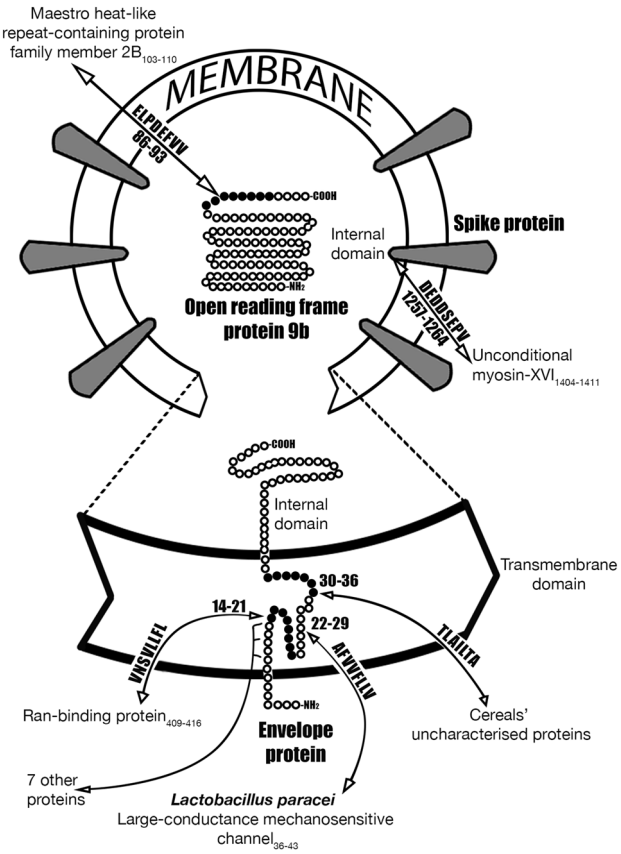

Figure 1. The SARS CoV-2 S, E and ORF9b protein molecules contain hepta/octamers that are homologous to proteins in the human body, including some nutrients and intestinal commensal bacteria.

In Replicase polyprotein 1a	In human proteins
EVEKGVLP$_{55-62}$	Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1$_{214–221}$
ESGLKTIL$_{390-397}$	Annexin A7$_{404-411}$
REETGLLM$_{724-731}$	Estrogen-related receptor gamma$_{30-37}$
GGSCVLSG$_{1100-1107}$	Sorting nexin-27$_{112–119}$
DIQLLKSA$_{1127-1134}$	Echinoderm microtubule-associated protein-like 1$_{38–45}$
RRSFYVYA$_{2431-2438}$	Transmembrane protein adipocyte-associated 1$_{225–232}$
AKKNNLPF$_{2733-2740}$	Acyl-CoA:lysophosphatidylglycerol acyltransferase 1$_{199–206}$
YNYEPLTQ$_{3500-3507}$	DNA helicase$_{199-206}$
SLKELLQN$_{3530-3537}$	Centromere protein I$_{496-503}$
DTSLSGFK$_{3671-3678}$	Solute carrier family 12 member 7$_{995–1002}$
PEANMDQE$_{4312-4319}$	Arachidonate 5-lipoxygenase-activating protein$_{54-61}$

Table 11. Localization of homologous 8-mers in RPP 1a and human proteins.

MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQHLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGETLGVLVPH
VGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQLD
FIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFPLNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETS
WQTGDFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESGLKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVG
EGSEGLNDNLLEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGNFKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSR
TLETAQNSVRVLQKAAITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKLKPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIV
GGQIVTCAKEIKESVQTFFKLVNKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEIIFLEGETLPTEVLTEEVVLKTGDLQPLEQPT
SEAVEAPLVGTPVCINGLMLLEIKDTEKYCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVADAVI
KTLQPVSELLTPLGIDLDEWSMATYYLFDESGEFKLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPEEEQEEDWLDDDSQQTVGQ
QDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSGYLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESDDYIATNGPLKVG
GSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLAPLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIAEIPKEE
VKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGNLHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV
PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREMLAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTT
VASLINTLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSSKTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSN
PTTFHLDGEVITFDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPHNSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSAL
NHTKKWKYPQVNGLTSIKWADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELGDVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQT
TLKGVEAVMYMGTLSYEQFKKGVQIPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSKETLYCIDGALLTKSSEYKGPITDVFYKE
NSYTTTIKPVTYKLDGVVCTEIDPKLDNYYKKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVTFFPDLNGDVVAIDYKHYTPSFK
KGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWSTKPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGDIILKPANNSLKITEEV
GHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNSVPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRIKASMPTTIAK
NTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTAALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLETIQIT
ISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWLMWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATR
VECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRPINPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRAN
NTKGSLPINVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVNTFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQG
FVDSDVETKDVVECLKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALIWNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQV
VNVVTTKIALKGGKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITR
EVGFVVPGLPGTILRTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVAYESLRPDTRYVLMDGSIIQFPNTYLE
GSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHVVAF
NTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMY
LKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLAKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVY
CPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGS
VGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQT
GIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHWLLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVK
HKHAFLCLFLLPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISV
TSNYSGVVTTVMFLARGIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYLVSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGK
PCIKVATVQSKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKLCEEMLDNRATLQAIASEFSSLPSYAAFATA
QEAYEQAVANGDSEVVLKKLKKSLNVAKSEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNIIPLTTAAKL
MVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDADSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTACTDDNALAYYNTTKGGRF
VLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAKAYKDYLASGGQP
ITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSADAQSFL
NGFAV

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 EDIQLLKSAYENFNQH$_{1126\text{-}1141}$, EVEKGVLPQLEQPY$_{55\text{-}68}$ and SVEEVLSEARQHL$_{34\text{-}46}$.

In the RPP 1a molecule, 7-mers SCGNFKV$_{505\text{-}511}$ and AIFYLIT$_{2785\text{-}2791}$ are homologous to human olfactory receptor proteins 52N2$_{190\text{-}196}$ and 2W1$_{32\text{-}38}$, respectively. A heptamer LKTLLSL$_{1556\text{-}1562}$ is homologous to the human bitter taste receptor T2R55$_{181\text{-}187}$ (Fig. 2).

Replicase polyprotein RPP 1ab. This huge (7096 aa; the primary structure see in[18]) 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 response 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

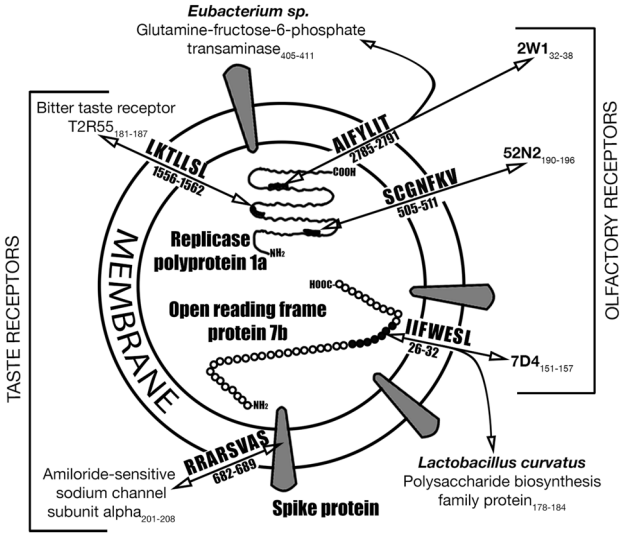

Figure 2. Some SARS CoV-2 hepta/octamers are homologous to human olfactory and taste receptor proteins. Homology to some proteins of commensal gut bacteria is also shown.

Goup 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	–	+	
	ORF6	–	–	No homology
	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	–	?	

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.

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.

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/UP000464024 SARS-COV-2, accessed 20 Aug 2020.
5. Proteomes: Homo sapiens (Human). https://www.uniprot.org/proteomes/UP000005640 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. Mandala, 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. Abdelmageed, 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**, 2653286. https://doi.org/10.1155/2020/2683286 (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 syndrome coronaviruses 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/P0DTD1.1?report=fasta.

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.