

DOI: 10.56871/RBR.2024.11.95.006
УДК 578.834.1+616-036.21+577.217+577.112+577.322.7+547.96

БЕЛКИ SARS CoV-2 И БЕЛКИ ЧЕЛОВЕЧЕСКОГО ОРГАНИЗМА

© Александр Тимурович Марьянович¹, Дмитрий Юрьевич Кормилец²

¹Северо-Западный государственный медицинский университет им. И.И. Мечникова. 195067, Российская Федерация, г. Санкт-Петербург, Пискаревский пр., 47

²Военно-медицинская академия им. С.М. Кирова. 194044, Российская Федерация, г. Санкт-Петербург, ул. Академика Лебедева, 6

Контактная информация: Александр Тимурович Марьянович — д.б.н., профессор, заведующий кафедрой нормальной физиологии.
E-mail: atm52@mail.ru ORCID: <https://orcid.org/0000-0001-7482-3403> SPIN: 5957-2347

Для цитирования: Марьянович А.Т., Кормилец Д.Ю. Белки SARS CoV-2 и белки человеческого организма // Российские биомедицинские исследования. 2024. Т. 9. № 1. С. 48–58. DOI: <https://doi.org/10.56871/RBR.2024.11.95.006>

Поступила: 21.11.2023

Одобрена: 08.01.2024

Принята к печати: 04.03.2024

Резюме. Белки SARS CoV-2 представляют собой молекулы с массой от нескольких десятков до нескольких тысяч аминокислотных остатков. Существуют структурные и неструктурные белки. К первым относятся шиповый гликопротеин, или S-белок (S), малый мембранный оболочечный белок (Е), мембранный белок (М) и нуклеопротеин или нуклеокапсид (N). Вторая группа состоит из 16 неструктурных белков (Nsp1-16, включая полипротеины репликазы RPP 1a и 1ab) и 10 вспомогательных факторов или белков открытой рамки считывания (ORF3a, 3b, 6, 7a, 7b, 8, 9b, 9c, 10 и 14). Белки S, E и M, расположенные снаружи и в мембране вириона, участвуют в контакте вириона с клеткой и проникновении в нее. Другие белки участвуют в захвате внутриклеточных механизмов и их использовании в собственных интересах вируса. Большинство этих белков содержат многочисленные мотивы, гомологичные человеческим белкам, в том числе таким важным, как интерлейкин-7. Возможно, эта гомология является важным фактором, позволяющим «обмануть» иммунную систему на начальных стадиях инфекции и спровоцировать аутоиммунный ответ впоследствии. Гомология белков SARS CoV-2, с одной стороны, и белков вкусовых и обонятельных рецепторов — с другой, возможно, объясняет причины нарушения восприятия вкусовых и обонятельных раздражителей, характерного для COVID-инфекции.

Ключевые слова: COVID-19, SARS CoV-2, гомология белка, рецептор-связывающий домен, интерлейкин-7, рецептор ACE2, врожденный иммунитет, аутоиммунитет, обоняние, вкус

SARS CoV-2 PROTEINS AND HUMAN PROTEINS

© Alexander T. Maryanovich¹, Dmitry Yu. Kormilets²

¹North-Western State Medical University named after I.I. Mechnikov. Piskarevskiy ave., 47, Saint Petersburg, Russian Federation, 195067

²Military Medical Academy named after S.M. Kirov. Akademician Lebedev str., 6, Saint Petersburg, Russian Federation, 194044

Contact information: Alexander T. Maryanovich — PhD, D.Sc. (Biology), Professor, Head of Department of Physiology. E-mail: atm52@mail.ru
ORCID: <https://orcid.org/0000-0001-7482-3403> SPIN: 5957-2347

For citation: Maryanovich AT, Kormilets DY. SARS CoV-2 proteins and human proteins. Russian Biomedical Research. 2024;9(1):48-58. DOI: <https://doi.org/10.56871/RBR.2024.11.95.006>

Received: 21.11.2023

Revised: 08.01.2024

Accepted: 04.03.2024

Abstract. SARS CoV-2 proteins are molecules with a mass of several tens to several thousand amino acid residues. There are structural and nonstructural proteins. The former include Spike glycoprotein (S), small membrane envelope protein (E), membrane protein (M), and nucleoprotein or nucleocapsid (N). The second group consists of 16 nonstructural proteins (Nsp1-16, including replicase polyproteins RPP 1a and 1ab) and 10 accessory factors or open reading frame

proteins (ORF3a, 3b, 6, 7a, 7b, 8, 9b, 9c, 10 and 14). Proteins S, E and M, located outside and in the membrane of a virion, are involved in the contact of the virion with a cell and penetration into it. Other proteins are involved in the hijacking of intracellular mechanisms and their use in the virus's own interests. Most of these proteins contain numerous motifs that are homologous to human proteins including such important ones as Interleukin-7. Perhaps this homology is an important factor in deceiving the immune system at the initial stages of infection and provoking an autoimmune response later. The homology of SARS CoV-2 proteins on the one hand and taste and olfactory receptor proteins on the other hand may possibly explain the causes of the impaired perception of taste and olfactory stimuli characteristic of COVID infection.

Keywords: COVID-19, SARS CoV-2, protein homology, receptor-binding domain, interleukin-7, ACE2 receptor, congenital innuity, autoimmunity, sense of smell, sense of taste

INTRODUCTION

Why is the new SARS CoV-2 coronavirus so infectious? Why many cases of COVID-19 infection are so severe? Why many patients are reported with complete loss of smell and taste? Answers to these questions should be received as soon as possible. The COVID-19 pandemic, which has plagued humanity for almost two years now, does not allow us to follow the usual course of encyclopedia authors: wait, carefully sift through ideas and facts, and wait for new ones to come. In this case (which should not become a precedent), one has to pay attention not only to firmly established experimental data, but also to some hypotheses.

The study of the proteome of the original ("canonical" or Wuhan) variant of SARS CoV-2 proceeds in two directions. Using 3D models, the researchers can determine how the spike protein binds to the ACE2 receptor [1]. This knowledge will help in creating binding blockers.

The alignment method compares the primary structures of millions of proteins [2]. The method allows you to detect the homology of SARS CoV-2 proteins with proteins of humans and other organisms [3]. Probably, the information on homology makes it possible to understand the mechanisms of the virus bypassing the innate immunity system (evasion) at the early stages of the development of the infectious process and the process of provoking an autoimmune response at later stages.

The issue of mutations in SARS CoV-2, and especially of its S protein, is becoming increasingly important. With the help of mutations, the virus "learns" to avoid immune responses [4].

SARS CoV-2 proteins vary significantly in length — from several tens to several thousands amino acid residues. They are traditionally divided into *structural* and *nonstructural* (Table 1).

STRUCTURAL PROTEINS

Spike glycoprotein

Spike (S) protein molecule consists of 1273 amino acid residues:

MFVFLVLLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDK
VFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDN
PVLFPNDGVYFASTEKSNIIRGWIFGTTLDSTQSLILVNNNA
TNVVIKVFCEFCNDPFLGVYHKNKSWMESEFRVYSS
ANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFNIDGYFK
IYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALH
RSYLTGPDSSSGWTAGAAAYVGYLQPRFTLLKYNENGTI
TDAVDCALDPLSEKCTLKSFSTVEKGIYQTSNFRVQPTESI
VRFPNITNLCPFGEVFNATRFASVYAWNKRKISNCVADYSV
LYNSASFSTFKCYGVSPSTKLNDLCFTNVYADSFVIRGDEV
QIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGN
YNYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCY
FPLQSYGFGPTNGVGYQPYRVVLSFELLHAPATVCGPKK
STNLVKNKCVNFNFGTLTGVLTESNKKFLPFQFGRDI
ADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVL
YQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLI
GAEHVNSYECDIPIGAGICASYQTQTNSPRRARSVASQSI
IAYTMSLGAENSVAYSNNIAIPTNFTISVTTEILPVSMTKTS
VDCTMYICGDSTECNLLQYGSFCTQLNRALTGIAVEQD
KNTQEVFAQVKQIYKTPPIKDFGFGNFSQILPDPSKPSKRS
FIEDLLFNKVTADAGFIKQYGDCLGDIARDLCAQKFNG
LTVLPPLLTDEMQYTSALLAGTITSGWTFGAGAALQIPF
AMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDLSL
STASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLND
ILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRAS
ANLAATKMSECVLGQSKRVDFCGKGYHLSFPQSAPHGV
VFLHVTYVPAQEKNTTAPAICHGKAHFPREGVVFVSNGT
HWFVTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDPL
QPELDSFKEELDQYFKNHTSPDVLGDISGINASVNIQKE
IDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIA
LIAIVMVTIMLCMTSCCSCCLKGCCSCGSCCKFDEDDSEP
VLKGVKLHYT

Table 1

SARS CoV-2 proteins: structural and nonstructural

Group	Proteins
Structural	Spike glycoprotein, S Envelope small membrane protein, E Membrane protein, M Nucleoprotein, N
Nonstructural	ORF3a, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF10, ORF14, RPP 1a, RPP 1ab

ORF, open reading frame. RPP, Replicase polyprotein.

Table 2

SARS CoV-2 S protein domains

Subunit	Positions	Domain
S1	1-13	Signal peptide (N-terminus)
	14-305	N-terminus domain (NTD)
	306-318	Uncharacterized fragment
	319-541	Receptor-binding domain (RBD)
	542-787	Uncharacterized fragment
	788-806	Fusion peptide (FP)
S2	807-911	Uncharacterized fragment
	912-984	Heptapeptide repeat sequence 1 (HR1)
	985-1162	Uncharacterized fragment
	1163-1213	Heptapeptide repeat sequence 2 (HR2)
	1214-1237	Transmembrane tail (TM)
	1238-1273	Cytoplasm tail (CT)

See Table 2 for color code. Receptor-binding motif (RBM) $S_{438} - Y_{508}$ underlined. Hereinafter, the primary structures of SARS CoV-2 proteins are given according to Uniprot database [5].

S protein molecule consists of subunits and domains (Table 2).

The Receptor-binding domain (RBD) continues to be of great interest to the researchers. RBD boundaries are estimated differently by different authors, namely: $R_{319} - F_{541}$ (refs. see [3]) or $T_{333} - T_{523}$ [1]. Within this region, the Receptor-binding motif (RBM₄₃₈₋₅₀₈) and amino acid residues N_{439} , L_{452} , T_{47K} , E_{484} , Q_{498} , and N_{501} are considered critical for binding affinity [6]:

SNNLDSKVGGNYNLYRLFRKSNLKPFERDISTEIYQAGS
TPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY.

At the border between subunits S1 and S2, the S protein molecule forms a loop. The authors believe that the loop is a key component in determining virus stability and transmission [7].

There are more than two dozen hepta- and octamers homologous to human proteins in the S protein molecule. Localization of these n-mers is shown in the Table 3.

Of all the regions listed in Table 3 that are homologous to human proteins, only two, KLNDLCF₃₈₆₋₃₉₂ and DEVRQIA₄₀₅₋₄₁₁, are located in RBD and none are located in RBM.

Table 3 shows that heptamer KLNDLCF₃₈₆₋₃₉₂ is homologous to motif in the interleukin 7 (IL-7₁₄₉₋₁₅₅) molecule. In severe cases of COVID-19 patients have increased level of IL-7 [8, 9], which is highly associated with disease severity [10]. IL-7 administered to critically ill COVID-19 patients has been associated with a return of lymphocytes to normal levels [11]. As a vaccine adjuvant, IL-7

Table 3

Localization of homologous hepta- / octamers 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) 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
		SLIVNN 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
S2		SPRRARS ₆₈₀₋₆₈₆	Hermansky-Pudlak syndrome 1 protein ₂₅₈₋₂₆₄
		RRARSVAS ₆₈₂₋₆₈₉	Amiloride-sensitive sodium channel subunit alpha ₂₀₁₋₂₀₈
	Fusion peptide FP ₇₈₈₋₈₀₆	none	—
	Uncharacterized fragment ₈₀₇₋₉₁₁	VTLADAG ₈₂₆₋₈₃₂	Non-receptor tyrosine-protein kinase TNK1 ₄₄₀₋₄₄₆

Ending of the table 3

Subunit	SARS CoV-2 S protein domain	In S protein	In human proteins
		GLTVLPP ₈₅₇₋₈₆₃	FH1/FH2 domain-containing protein 3 ₉₇₂₋₉₇₈
		LPPLTD 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 ; SP-A receptor subunit SP-R210 alphaS ₈₉₄₋₉₀₀
		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

[3, adapted]. Hereinafter, the primary structures of human proteins are given according to Uniprot database — Homo sapiens [5].

Localization of homologous hepta- / octamers in the transmembrane domain₈₋₃₈ E protein and human proteins [3, adapted]

In E protein	In human proteins
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
VLLVTL 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

Heptamer TALRLCA₃₅₋₄₁ is located at the junction of the transmembrane domain₈₋₃₈ and internal domain 39-75.

could enhance the immune responses to vaccines against SARS CoV-2 [12]. IL-7 is beneficial cytokine to the pathophysiology of COVID-19 [13]. At the same time, IL-7 induces SARS CoV-2 receptor ACE2 expression in human vascular endothelial cells [14]. In patients with COVID-19, respiratory failure is associated with

Table 4

an increase in systemic blood pressure, probably due to modulation of the renin-angiotensin-aldosterone system by SARS-CoV-2 infection [15].

Heptamer DEVQRQA₄₀₅₋₄₁₁ is homologous to motif in the Histone-lysine N-methyltransferase 2C (HLNMT 2C₄₅₃₀₋₄₅₃₆). Histone methylation plays an important role in such a critical process as the epigenetic regulation of genes [16].

Presumably, IL-7 is an outpost defense trigger. When cell destruction begins in COVID-19, IL-7 turns on the last reserve of life, activating immunological memory cells. SARS CoV-2 tricks the immune system into presenting a motif homologous to IL-7.

S protein is involved in the organization of virion assembly in the intermediate compartment ER-Golgi [17]. The C-terminal truncation of the protein S molecule results in a variant that easily passes through the Golgi complex to the plasma membrane in a pre-activated conformation, causing increased syncytium formation [18].

Envelope small membrane protein

Envelope small membrane (E) protein is the shortest (75 amino acid residues) of all SARS CoV-2 structural proteins.

MYSFVSEETGTLIVNSVLLFLAFVVLLVTLAILTALRLCA
YCCNIVNVSLVKPSFYVYSRVKNLNSSSRVPDLLV

Transmembrane domain₈₋₃₈ is underlined. Hereinafter, n-mers homologous to human proteins are highlighted in red.

Only a small part of it, namely heptamer M₁ – Y₆, protrudes from the virion outwards.

E protein contains eight hepta- / octamer homologous to human proteins (Table 4).

Homologous n-mers merge into a single 2octamer, which is almost entirely located in the thickness of the envelope of the virion. A random selection of 28 letters in a word would require an astronomical number of iterations: $20^{28} = 2.7 \cdot 10^{36}$. (This number is slightly less than the mass of the Earth, measured in nanograms.)

The degree of homology within this 28-measure can be represented as follows:

VNSVLLFLA_FVVFLVLT_LAILTALRLCA_A

where the size of the letters corresponds to the frequency of the viral hepta- / octamers in the human proteome.

Besides, the protein E transmembrane domain contains an octamer and a heptamer, homologous to the proteins of some gut bacteria *Lactobacillus sp.* and even cereals, including corn *Zea mays*, sorghum *Sorghum bicolor*, wheat *Triticum aestivum*, and barley *Hordeum vulgare* (Table 5).

Table 5

Localization of some of homologous n-mers in the E protein and human gut proteome [3]

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

E protein is integrated into the human cell membrane; later it is transported closer to the endoplasmic reticulum and the Golgi apparatus, where viral replication occurs [19]. E protein can affect the properties of S proteins and contribute to the assembly of viral particles [20].

Membrane protein

Membrane (M) protein consists of 222 amino acid residues, and its structure contains six heptamers homologous to human proteins (Table 6).

MADSNGTIT**VEELKKLLEQ**WNLVIGFLFTWICLLQFAYAN RNRFLYIIKLIFLWLLWPVTLACFVLAAYRINWITGGIAIAMACLV GLMWLSYFIASFRLFARTRSMWSFNPETNILLNVPLHGTILTRP **LLESELV**IGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSY YKLGASQRV**AGDSGFA**AYSRYRIGNYKLNTDHSSSSDNIALLVQ

Four heptamers are located close to the N-terminus of the molecule, merging into a single decamer V₁₀ – Q₁₉. Taking into account the number of homologous amino acid residues, this decamer can be represented as follows:

VEELKKLLEQ.

Table 6

Localization of homologous heptamers in the M protein and human proteins [3]

7bIn 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

Outside of the decamer, there are two homologous heptamers. Protein M is a candidate for participation in mimicry processes.

Like E protein, M protein can affect the properties of S proteins and contribute to the assembly of viral particles [20].

S, E, and M proteins cause Golgi fragmentation; disruption of the Golgi apparatus appears to be a critical component of SARS CoV-2 replication [21].

Nucleoprotein

The nucleoprotein (N-protein) consists of 419 amino acid residues and contains eleven heptamers homologous to human proteins (Table 7).

MSDNGPQNQRNAPRITFGGSPDSTGSNQNNGERSGARSQQR **RPQGLPN**NTASWFTALTQHGKEDLKFP**RGQGV**PINT**NSSPDDQ** IGYRRATRRIRGGD**GKMKDLS**PRWYFYFLGTGPEAGLPYGA NKDGIWVATEGALNTPKDHIGTRNPANNAI**VLQLPQG**TTLPK GFY**AEGRSGGSQA**SSRSSRSRNRSSRNSTPGSSRGTSARM AGNGGDAALALLLDRLNQLSKMSGKGGQQQQQTVTKKSAA EASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQ GTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAI KLDDKDPNFKDQVILLNKHIDAYKTFPPTEPKDKKK**KADETQA** LPQRQKKQQT**LLPAADLDDF****SKQLQSSMSSADSTQA**

Some of the heptamers fuse into several rather long fragments, including the decamer A₁₇₃ – A₁₈₂, and 13-mer S₄₀₄ – S₄₁₆. It increases the likelihood of the protein involvement in provoking an autoimmune response. Protein N is located completely inside the SARS CoV-2 virion and cannot participate in mimicry, but can be involved in provoking an autoimmune response.

In comparison with SARS-CoV, SARS-CoV-2 contains six times more acetyl-lysine residues. This suggests that acetylation of N proteins plays crucial roles in SARS-CoV-2 functions [22].

NONSTRUCTURAL PROTEINS

All nonstructural proteins of SARS CoV-2 (ORF3a, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF10, ORF14, RPP 1a, and

Table 7

Localization of homologous heptamers in the N protein and human proteins [3]

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 subunit ₃₋₈
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-1 ₂₅₉₋₂₆₅
SMSSADS ₄₁₀₋₄₁₆	Protein PRRC2B ₄₁₆₋₄₂₂

RPP 1ab) are located completely inside the SARS CoV-2 virion and, by definition, cannot be involved in the process of mimicry. What remains to consider the possibility of their implication in provoking an autoimmune process [3].

ORF3a protein

ORF3a protein molecule consists of 275 amino acid residues:

MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPQASLPFG
WLI**VGVALLA**VFQSASKIITLKKRWQLALSKGVHFVCNLLLLF
VTVYSHLL**LLVAAGL**EAPFLYLYALVYFLQSINFVRIIMRLWLCW
KCRSKNPLLYDANYFLCWHTNCYDYCIPYN**SVTSSIV**ITSGDG
TTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYS
TQLSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVME
PIYDEPTTTTSVPL

In the ORF3a protein molecule, there are five heptamers homologous to human proteins (Table 8).

Table 8

Localization of homologous heptamers in the ORF3a protein and human proteins [3]

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-14 ₄₁₈₋₄₂₄

The heptamers 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.

ORF6 protein

ORF6 protein molecule consists of 61 amino acid residues:

MFHLVDFQVTIAEILLIMRTFKVSIWNLDYIINLIKNLSKSLTENKY
SQLDEEQPMEID

In the molecule, there is no heptamers homologous to human proteins.

ORF7a protein

ORF7a protein molecule consists of 121 amino acid residues:

MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNSPF
HPLADNKFALTCFSTQFAFACPDGVKHVYQLRARSVSPKLFIRQ
EEVQELYSPIFLI**VAAIVFITLCFTLKRKTE**

In the ORF7a protein molecule, there are only two heptamers homologous to human proteins located in close proximity to each other (Table 9).

Table 9

Localization of homologous heptamers in the ORF7a protein and human proteins [3]

In ORF7a protein	In human proteins
VAAIVFI ₁₀₄₋₁₁₀	Transmembrane protein 255B ₈₆₋₉₂
FTLKRKT ₁₁₄₋₁₂₀	Cytosolic 5'-nucleotidase 3A ₃₆₋₄₂



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

ORF7b protein

ORF7b protein molecule consists of 43 amino acid residues:

MIELSLIDFYLCFLAFLFLVLIMLIIFWFSLELQDHNETCHA

In this polypeptide, there are only one heptamer homologous to the human protein (Table 10).

Table 10

Localization of the homologous heptamer in ORF7b and a human protein [3]

In ORF7b protein	In human protein
IIFWFSL ₂₆₋₃₂	Olfactory receptor 7D4 ₁₅₁₋₁₅₇

The ORF7b protein may be involved in provoking an autoimmune response and, in particular, contribute to olfactory dysfunction.

ORF8 protein

ORF8 protein molecule consists of 121 amino acid residues:

MKF**LVFLGIITV**AAFHQECSLQSQCTQHQPYYVDDPCPIHFYSK
WYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINC
QEP**KLGS**LVVRCSFYEDFLEYHDVRVVLDFI

In this polypeptide, there are three heptamers homologous to human proteins (Table 11).

Table 11

Localization of homologous heptamers in the ORF8 protein and human proteins [3]

In ORF8 protein	In human proteins
LVFLGI ₄₋₁₀	Zinc finger protein 486 ₄₉₋₅₅
LGIITV ₇₋₁₃	D-2-hydroxyglutarate dehydrogenase, mitochondrial ₂₆₂₋₂₆₈
KLGS ₉₄₋₁₀₀	Sodium leak channel non-selective protein ₅₀₅₋₅₁₁

In this case, two heptamers merge into a decamer $L_4 - V_{13}$. Due to the fusion of two heptamers into a decamer $L_4 - V_{13}$, the ORF8 can be involved in provoking an autoimmune response.

ORF9b protein

ORF9b protein molecule consists of 97 amino acid residues:

MDPKISEMHPALR**LVDPQIQL**AVTR**MENAVGR**DQNNVGP
KVYPILRL**LGSP**LSLNMARKTLNSLEDKAFQLTPIAVQMTKLAT
TEELPDEFVVTVK

Table 12

Localization some of homologous hepta- / octamers in ORF9b protein and human proteins [3]

In ORF9b protein	In human proteins
LVDPQIQL ₁₄₋₂₁	Valine-tRNA ligase, mitochondrial ₉₉₆₋₁₀₀₂
MENAVGR ₂₆₋₃₂	Neprilysin ₄₁₉₋₄₂₅
LGSPSL 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

In the ORF9b protein molecule, there are six hepta- / octamers, homologous to human proteins (Table 12).

Some of these hepta- / octamers merge into octamer $L_{48} - N_{55}$ and decamer $T_{84} - V_{93}$.

Octamer ELPDEFVV₈₆₋₉₃ is homologous to the Maestro heat-like repeat-containing protein family member 2B, which may play a role in the sperm capacitation [23]. Male reproductive dysfunction has been proposed as a likely consequence of COVID-19 [24].

After the destruction of the SARS CoV-2 virion, ORF9b can take part in provoking an autoimmune response. This protein plays a special role in hijacking mitochondrial metabolic processes in COVID-19 infection [25].

ORF10 protein

ORF10 protein (traditional name, but more correctly: polypeptide) molecule consists of 38 amino acid residues:

MGYINVFAIPFTIYSLLLCRMNSRSYTAQVGIVNFNLT

In the molecule, there is no heptamers homologous to human proteins.

ORF14 protein

ORF14 protein (synonym: ORF9c) molecule consists of 73 amino acid residues:

MLQSCYNFLKEQHCQKASTQKGAEAAVKPLLVPHHVATVQEIQ
LQAAVGEILLLEWLAMAVMLLLCCCLTD

In the molecule, there is no heptamers homologous to human proteins.

Replicase polyprotein RPP 1a

Replicase polyprotein 1a (RPP 1a) consists of 4405 amino acid residues.

MESLVPGFNEKTHVQLSLPVLQVRDLVRGFGD**SVEEVLSEARQHL**KDGTGCL
VEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVEL**LVAELEGI**QYGRSGETLGLV
 VPHVGEIPVAYRKVLLRKNGKNGAGGHSYGADLKSFDLGDELGTDPYEDFQEN
 WNTKHSSGVTRMRELNGGAYTRYVDNFCGPDGYPLECI**KDLLARAGKASCT**
 LSEQLDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFN
 GECPNFVPLNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLM
 KCDHCGETSWQTGDFVKATCEFCG**TENLTKE**GATTGGLPQNAVVKIYCPACH
 NSEVGPEHSLAEYHNE**ESGLKTIL**RKGGRTIAFGGCVFSYVGCHNKCAYWVPRA
 SANIGCNHTGVVGESEGLNDNL**LEILQKE**KVNINIVGDFKLN**EEIAIL**ASFSAS
 TSAFVETVKGLDYKAFKQIVE**SCGNFKV**TGKAKKGAWNIGEKSILSPYAF
 SEAAVRVSIFSRLETAQNSVRVLQKAAITLDGISQSLRLIDAMMFTSDLATN
 NLVVMAYITGGVVQLTSQWLTNIFGTVEYKLPVLDWLEEFKEGVEFLRDGW
 EIVKFISTCAEIVGGQIVTCAKEIKESVQTFKLVN**KFLALCA**DSIIIG**AKLKAL**
 NLGETFVTHSKGLYRKCVKS**REETG**LLMPLKAPKEIFLE**GETLPTEV**L**TEEVVLT**K
 GDLPLEQPTSEAVEAPLVGTPVCINGLM**LEIKDTE**KYCALAPNMVNTNFT
 LKGGAPTQVTFGDDTVIEVQGYKSNITFELDERIDKVLNEKCSAYTVELGTEVN
 EFACVAVDAVIKTLQ**VSELLT**PLGIDLDEWSMATYYLFDESGEFKLASHMYCS
 FYPPDEDEEGDCEEEFEPSTQYEGTEDDYQGKPLEFGAT**SAALQPEE**EQE
 EDWLDDDSQQTGQDGSSEDNQTTITQIEVQPLE**ELTPV**QTIENVSFSG
 YLKLTDNVYIKNADIVEAKVKPTVVNAANVYLKHGGVAGALNKATNNAM
 QVESDDYIATN**GPLKVG**SCV**LSGHNL**AKHCLHVGPVNK**EDIQL**LKSAYE
NFNQHEVLLAP**LLSAGIF**GADPIHSLRVCVDTVRTNVYLAVFDKNLYD**KLVS**SFL
 EMKSEKQVEQKIAE**IPKEEV**KPFITESKPSVEQRKQDDKKIKACVEEVTTLTEE
 TKFLTENLLYID**INGNLHPD**SATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIP
 TTKAGGTTEMLAKALRKVPTDNYITTPGQGLNGYVEEAKTVLKKCSAFYI
 LPSIISNEKQELGTVSWNLREMLAHAEETRLMPVCVETKAIVSTIQRKYKGIK
 IQEGVVDYGARFYFYT**SKTTVAS**INTL**NDLNET**LVMTPLGYVTHGLNEEAAR
 YMRSL**KVPATVS**VSSPDVATAYNGYLTSSSKT**PEEHFIE**TISLAGSYKDWSSYG
 QSTQ**LGI**EFLKRGDKSVYYSNPTTFHLDGEVIT**FDNLKTLL**SLREVRTIKVFTT
 VDNINLHTQVVDMSMTYGGQFGPTYLDGADVTKIKPHNSH**EKGT**FYVLPND
 TLRVEAFEYHTDPSFLGRYMSALNHTKKWKYPQVNGLTSLIKWADNNCYL
ATALLTQQIELKFNPALQDAYRRAR**AGEANF**CALILAYCNKTVGELGDVRE
 TMSYLFQHANLDSCKRVLNVVCKTCGQQQT**LKGV**EAVMYMGTLSEYQFKKG
 VQIPCTCGKQATKYLQVESPFVMSAPPAQYELKHGTFTCASEYTGNYQCG
 HYKHITSKETLYCIDGALLTKSSEYKGPITDVYKENSYTTTIKPVYKLDGCVCT
 EIDPKLDNYKKDNYSYFTEQPIDLVNPQYPNASFDNFKVCDNIKFADDLNLQ
 TGYKKPASRELKVTFFDLNGDVDAIDYKH**YTPSF**KK**GAKLL**HKPIVWHVNNAT
 NKATYKPNWTCIRCLWSTKPVET**SNSF**DVLKSEDAQGMND**LACED**LKPVSEE
 VVENPTIQKDVLECNVKTTEVVGDIILKPAN**NSLKITE**EVGHDTLMAAY**VDNSS**LT
 IKKP**NELSRVL**GLK**TLATHGLA**AVNSVPWDTIANYAKPFLNKVSTTTNIVTRCL
 NRVCNTNMPYFFTLQLCTFTRSTNSRIKASMPPTIAKNVTKSVGKFCLEASF
 YLKSPNFSKLINIIWFLLLSV**CLGSLI**Y**STAALGV**LSNLMGMPSYCTGYREGYLN
 STNVTIATYCTGSIPCSVCLSGDLSLDTYPSLETIQITISSFKWDLTAFGLVAEWFL
 AYILFT**RRFFYVL**GLAAIMQLFFSYFAVHFISNSWLMWLIINLVQMAPISAMVRMYIF
 FASFYVWVKSYYVHVVDGCSNSTCMCYKRNRRATRVECTIVNGV**RRSFYVYA**
 NGGKGFCCLHNWNCVNDTFCAGSTFISDEVARD**LSLQF**KRPINPTDQSSYIV
 DSV**TVKNG**SIHL**YFD**KAGQKTYERHSLSHFVNLDNLRANNTKGSPLINIVFDG
 KSKCEESSAKSASVYYSQLMCQPILLDQALVSDVGDSAEVAVKMFDAYVNTF
 SSTFNVP**MEK**LK**TLVATA**EAE**LAKNV**SLDNVLSTFISAARQG**FVDS**D**VET**KDV
 VECLKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVA
 KSHNIALIWNVKDFMSLSEQLRKQIRSA**AKKN**LPFKLTCAATRQVVNVVTTK
IALKGGKIVNNWLKQLIK**VT**LVFLVAA**IF**Y**LIT**PVHVMSKHTDFSSEIIGYKA
 IDGGVTRDIASDTCTFANKHADFDWFSSQGGSYTNDKACPLIAAVITREV
 GFVVPGL**PGTIL**RRTNGDF**LHFLPR**VSAVGNICYTPSKLIEYTDFAVSACVL

AAECTIFKDASGKPVPCYDTNVLEGSVAYESLRPD**TRYVLM**DGSIIFPNT
 YLEGSVRVTT**FDSEYCR**HGTCERSEAGVCVSTSGRWVLLNNDYYRSLPGV
 FCGVDAVNLLTN**FTPLIQ**IGALDIS**SASIVAG**GIVAVVTCLAYFMRFRRAF
 GEYSHVAVNTLLFLMSFTVLCCTPVYSFLPGVYSVIYLYLTFTYLTNDVSFLA
 HIQWMVMFTPLVPFWITAIYIICISTKHFYWFSSNYLKRRRVFNGVSFSTF
 EEAALCTFLLNKEMYLKLRSDV**LLPLTQ**YNNRYALYNKYKYFSGAMDTT
 SYRE**AACCHLA**KALNDFS**NSGSDV**LYQPPQTSITSAVLQSGFRKMAFP
 SGKVEGCMVQVTCGTTTLNGLWL**DDVVYCP**RHVICTSEDMLNPYEDLLI
 RKSNNHFLVQAGNVQLRVIGHSMQNCVLKLVDTANPKTPKYKFVRIQPG
 QTFSVLACYNGSP**SGVYQC**AMRPNFTIKGSFLNGSCGSGVGNIDYDCVSF
 CYMHMELPTGVHAGTD**LEGNFYG**PFVDRQTAQAAG**TDTTITV**NVLAWLYAA
 VINGDRWFLNRFTTTLNDFNLVAMK**YNYEPLTQ**DHVDILGPLSAQTGIAVLDM
 CAS**LKELLQNGMNGRTIL**GSALLEDEFTPFDDVVRQCSGVTFSQSAVKRTIKGT
 HHWLLLT**ILLSLV**LQSTQWSLFFLYENAFLPFAMGIAMSAFAMMFVKHKKH
 AFL**CLFLLPS**LATVAFNMVMPASWVMRIMTWLDM**DTSLSG**FKLKDCVMY
ASAVVLLILMTARTVYDDGARRVWTLMNVLTVYKYVYGNALDQAISMWALISV
 TSNY**SGVVTVM**FLARGIVMCVEYCPFITGNTLQCIMLVYCFGLYFCTCYFGL
 FCLLNRYFRLTLGVYDYLSTQEFYRMNS**QGLLP**PKNSIDAFKLNKLLGVGGK
 PCIKVATVQSKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDT
 TEAF**EKMVS**LLSVLLSMQGAVDINKLCEEMLDNRATLQAI**ASEFSS**LPSSAAAFAT
 AQEAYEQAVANG**DSEVVLK**LKSLNVAKSEFDRDA**AMQRKLE**KMADQAMTQ
 MYK**QARSED**KRAKVTSAMQTMFTMLRKLNDALNNIINNARDGCVPLNIPLT
 TAAKLMVIPDYNTYKNTCD**GTTFTYA**SALWEIQVVDADSKIVQLSEISMDNSP
 NLAWPLIVTALRANSVKLQNNELSPVALRQMSCAAGTTQTACTDDNALAYYN
 TTKGGRFVL**ALLSDLQ**DLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPVKV
 YLYFIKGLNNLRGMVLSLAATVRLQAGNATEVPANSTVLSFCFA**VDAAKAY**
 KDYLASGGQIPITNCVKMLCTHTGTGQAI**ITVTP**EA**NMDQ**ESFGGASCLYCRCH
 IDHPNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKG**YGCS**CDQL
 REPMLQSADAQSFLNGFAV

In the RPP 1a molecule, there are eleven octamers (Table 13) and more than a hundred heptamers homologous to human proteins.

Some of the octamers are found in more than one human protein, some fold into long n-mers, for example 16-mer EDIQLLSAYENFNQH₁₁₂₆₋₁₁₄₁, 14-mer EVEKGVLPQLEQPY₅₅₋₆₈ and 13-mer SVEEVLSEARQHL₃₄₋₄₆. The question of the participation of this large molecule in provoking an autoimmune response requires further study.

Replicase polyprotein RPP 1ab

Replicase polyprotein 1ab (RPP 1ab) consists of 7096 amino acid residues. In the RPP 1ab molecule, there are 210 hepta- / octamer homologous to human proteins. Some of them fold into long (more than 15 amino acid residues) n-mers. The role of this huge molecule in provoking an autoimmune response also requires study.

EVASION AND PROVOCATION OF AUTOIMMUNE RESPONSE

Based on the fact that the external SARS CoV-2 proteins are the first to contact host's immune system, while the internal proteins are only the second, it would be reasonable to divide the proteins of the virus into *external* and *internal* proteins. The difference from the generally accepted classification (structural / nonstructural) is

Table 13

**Localization of homologous octamers
in RPP 1a and human proteins [3, adapted]**

In Replicase polyprotein 1a	In human proteins
SVEEVLS ₃₄₋₄₀	FLJ00176 protein (Fragment) ₂₆₀₋₂₆₆
SEARQHL ₄₀₋₄₆	Cytokine-inducible inhibitor of signaling type IV ₉₆₋₁₀₂
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
DIQLKSA 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

minimal. The first group includes proteins S, E and M, the second — all the others, including the N protein. We will consider proteins of the first group as participants in the processes of mimicry and the second as provocateurs of an autoimmune response.

External proteins and mimicry

In the IT terminology, the word *evasion* means bypassing an information security device to deliver malware without being detected by the recipient. Virologists have long been familiar with viral immune evasion, having a variety of expressions to describe it, such as: *to avoid the immune response*, *to outwit the immune system*, *to outmaneuver your hosts*, *to subvert the host cellular response*, *viral mimicry*, *camouflage*, *subversion* and *piracy*. The growing virulence of SARS CoV-2 indicates that the virus's ability to deceive the innate immune system improves with some new mutations.

To mislead the immune system, the virus could have hijacked some regions of the genetic code from previous hosts. That makes its proteins similar to human proteins. Knowledge of the homology between the virus and human proteins might help understand the mechanisms of mimicry in the moment of infection and during the subsequent autoimmune response.

For evasion to occur, the virus must appear in front of the immune system and tell it: *don't shoot! I am one of you!* The traditional naming of a password is not suitable for such a message,

since a password, by definition, must be known by a very limited number of people or devices. Virology needs a term for a universal password, known to an unlimited number of participants on both sides of the information exchange. There is a suitable term in IT — *shibboleth*. Reported to a computer security system without distortion, the shibboleth allows an intruder (person or device) to gain access to the desired resources.

The word shibboleth, borrowed from the Bible (*Judges 12: 5-6*), is used by linguists and literary men (see *Shitboleth* in episode 16 of *Ulysses* by James Joyce), from whom psychologists and psychiatrists adopted it. Dr Dmitry Kormilets in a private communication suggested using this term in virology as well. In our interpretation, shibboleth is an area of the surface of virion, according to which the immune system must mistakenly recognize the virus as a part of the host organism and turn off the mechanisms designed to inactivate and / or destroy the intruder.

Some viral proteins are homologues of human proteins. It appears they have been hijacked from the host and included in their own genomes [26].

Tables 3–6 show a lot of motifs common to the external SARS CoV-2 proteins and humans. Which of them are directly involved in mimicry, it is now impossible to say. One can only point to the motifs in the most functionally important regions of the external proteins of SARS CoV-2 and human proteins. In addition, those regions of the external proteins of SARS CoV-2, in which the frequency of occurrence of homologous regions is the highest, deserve special attention.

Perhaps this is how the SARS CoV-2 protects its most important site (RBD) from the immune system.

The narrow region of the E protein transmembrane domain contains a variety of motifs homologous to proteins from humans, food, and intestinal bacteria (Tables 4 and 5). In this regard, the participation of E protein in mimicry seems to be the most probable.

In the structure of M protein, there is also a high “concentration” of motifs homologous to human ones (Table 6). In the protein M, four heptamers homologues of human proteins are fused into a decamer V₁₀–Q₁₉. The hydrophilic composition indicates a possible contact with the extracellular environment and the host's immune system. This outer protein is the second most likely candidate for the role of mimicry organizer.

Provoking of Autoimmune Response

The most likely candidate is ORF9b protein. In its small molecule, regions which are homologous to human proteins account for 34.0% (33 out of 97), the highest value of this indicator among all internal proteins. The biggest structural difference is found between SARS CoV-2 RF9b protein and similar bat and pangolin proteins compared to other SARS CoV-2 proteins, which indicates active mutagenesis [27].

The polyprotein molecules RPP 1a and RPP 1a b are huge, as is the number of homologous motifs in them. The possibility that the motifs take part in provoking an autoimmune response has not yet been proven. Enzymes and especially enzymes of

the cell cycle are evolutionarily highly conserved. In the process of disintegration of microorganisms, permanent inhabitants of the human intestine, peptides homologous to human proteins must be released into the intestinal lumen. Perhaps it is they who interact with the host's immune system and tune it to non-resistance to virion proteins.

Dysfunction of olfactory and taste receptors

ACE2 protein has been found at high levels in the human olfactory epithelium. May this explain COVID-19-associated olfactory dysfunction [28]?

In the RPP 1a molecule, heptamers SCGNFKV₅₀₅₋₅₁₁ and AIFYLIT₂₇₈₅₋₂₇₉₁ are homologous to human Olfactory receptor proteins 52N2₁₉₀₋₁₉₆ and 2W1₃₂₋₃₈, respectively. ORF7b contains a heptamer homologous to the Olfactory receptor protein 7D4 and may be involved in provoking an autoimmune response, contributing to olfactory dysfunction. In the RPP 1a, a heptamer LKTLSSL₁₅₅₆₋₁₅₆₂ is homologous to the human Bitter taste receptor T2R55₁₈₁₋₁₈₇.

In S protein, the octamer RRARSVAS₆₈₂₋₆₈₉ is homologous to the Amiloride-sensitive sodium channel subunit alpha₂₀₁₋₂₀₈, which is involved in salt taste perception [29].

If homologous motifs in the SARS CoV-2 molecule can trigger an autoimmune response, then these facts may explain why COVID-19 disease so often affects the sense of smell and taste.

Mutations

Mutations are a mechanism for escaping immune responses [4]. Among the SARS CoV-2 proteins, protein S has been studied for mutations. Of the external proteins, it is the most susceptible to mutation [27].

The differences of human S-proteins from Asia, Africa, Europe, North America, South America and Oceania from the reference sequence of the SARS CoV-2 Wuhan-Hu-1 protein, China, are described. There were found 9654 mutations, which correspond to 400 different sites of mutations. RBD alone contained 44 mutations [30]. Of course, far from all of the effects of these mutations have now been studied. Fortunately, not all of them matter. In theory, the mutation can increase, decrease, or not affect the immune response to the S-protein [4].

The D₆₁₄G mutation changes the conformation of the S protein [4]. SARS CoV and SARS CoV-2 recognize the ACE2 receptor through their S proteins. In the N-terminal domain, the sequences MESEFR₁₅₃₋₁₅₈ and SYLTPG₂₄₇₋₂₅₂ are specific for human SARS CoV-2. In RBD, the structural determinants for recognizing human ACE2 are the VGGNY₄₄₅₋₄₄₉ and EIYQAGSTPCNGV₄₇₁₋₄₈₃ sequences, as well as the disulfide bridge connecting C₄₈₀ and C₄₈₈ [31]. Note that none of the motives mentioned above coincide with regions homologous to human proteins.

S protein of SARS-CoV-2 variant Delta contains eight mutations, namely T₁₉R, G₁₄₂D, Δ₁₅₆₋₁₅₇, R₁₅₈G, L₄₅₂R, T₄₇₈K, P₆₈₁R, and D₉₅₀N [32]. S protein of variant Omicron, the most aggressive, contains many mutations, namely V₇₀Δ, T₉₅I, G₁₄₂D, V₁₄₃Δ, Y₁₄₄Δ, Y₁₄₅Δ, G₃₃₉D, S₃₇₁L, S₃₇₃P, S₃₇₅F, K₄₁₇N, N₄₄₀K, G₄₄₆S, S₄₇₇N, T₄₇₈K, E₄₈₄A, Q₄₉₃R, G₄₉₆S, Q₄₉₈R, N₅₀₁Y, Y₅₀₅H, N₆₅₅Y, N₆₇₉K, and P₆₈₁H,

several of which overlap with those in the Alpha, Beta, Gamma, or Delta variants [33,34]. Due to the huge and continuous stream of data, the topic of mutations in SARS CoV-2 proteins can only be considered in periodicals for now.

CONCLUSION

Judging by the degree of homology between SARS CoV-2 proteins and humans, the main means of bypassing innate immunity (shibboleth) should be the E protein, while the main provocateur of the autoimmune response is the ORF9b protein. Accordingly, the attention of researchers and especially — developers of vaccines against SARS CoV-2 should be paid primarily to these two proteins. It also should be taken into consideration that vaccines affecting such homologous regions can damage proteins of the human body.

Supplementary Materials: None.

Authors Contributions: Equal.

Funding: None.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: None.

ЛИТЕРАТУРА / REFERENCES

1. Beaudoin C.A., Jamasb A.R., Alsulami A.F. et al. Predicted structural mimicry of spike receptor-binding motifs from highly pathogenic human coronaviruses. *Comput Struct Biotechnol J.* 2021; 19: 3938–53. DOI: 10.1016/j.csbj.2021.06.041. PMID: 34234921; PMCID: PMC8249111.
2. Maryanovich A.T., Kormilets D.Y., Polyanovsky A.D. Xenin, the oldest after insulin? *Mol Biol Rep.* 2018; 45: 143–50. DOI: 10.1007/s11033-018-4147-2. PMID: 29340900.
3. Khavinson V., Terekhov A., Kormilets D., Maryanovich A. Homology between SARS CoV-2 and human proteins. *Sci Rep.* 2021; 11: 17199. DOI: 10.1038/s41598-021-96233-7. PMID: 34433832; PMCID: PMC8387358.
4. Kwarteng A., Asiedu E., Sylverken A.A. et al. Molecular characterization of interactions between the D614G variant of SARS-CoV-2 S-protein and neutralizing antibodies, A computational approach. *Infect Genet Evol.* 2021; 91: 104815. DOI: 10.1016/j.meeid.2021.104815. PMID: 33774178; PMCID: PMC7987576.
5. Uniprot database/ <https://www.uniprot.org/uniprot/?query=proteome,UP000464024+AND+proteomecomponent,%22Genome%22&sort=score>. Uniprot database – Homo sapiens. <https://www.uniprot.org/proteomes/UP000005640> Homo sapiens, last change, 03 Sept 2020.
6. Yi C., Sun X., Ye J. et al. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol Immunol.* 2020; 17: 621–30. DOI: 10.1038/s41423-020-0458-z. PMID: 32415260; PMCID: PMC7227451.
7. Jaimes J.A., André N.M., Chappie J.S. et al. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop.

- J Mol Biol. 2020; 432: 3309–25. DOI: 10.1016/j.jmb.2020.04.009. PMID: 32320687; PMCID: PMC7166309.
8. Li S., Zhang Y., Guan Z. et al. SARS-CoV-2 triggers inflammatory responses and cell death through caspase-8 activation. *Signal Transduct Target Ther.* 2020; 5: 235. DOI: 10.1038/s41392-020-00334-0. PMID: 33037188; PMCID: PMC7545816.
 9. Adamo S., Chevrier S., Cervia C. et al. Profound dysregulation of T cell homeostasis and function in patients with severe COVID-19. *Allergy.* 2021; 76: 2866–81. DOI: 10.1111/all.14866. PMID: 33884644; PMCID: PMC8251365.
 10. Wang G.L., Gao H.X., Wang Y.L. et al. Serum IP-10 and IL-7 levels are associated with disease severity of coronavirus disease 2019. *Cytokine.* 2021; 142: 155500. DOI: 10.1016/j.cyto.2021.155500. PMID: 33810947; PMCID: PMC7973056.
 11. Laterre P.F., François B., Collienne C. et al. Association of Interleukin 7 Immunotherapy With Lymphocyte Counts Among Patients With Severe Coronavirus Disease 2019 (COVID-19). *JAMA Netw Open.* 2020; 3: e2016485. DOI: 10.1001/jamanetworkopen.2020.16485. PMID: 32697322; PMCID: PMC7376391.
 12. Bekele Y., Sui Y., Berzofsky J.A. IL-7 in SARS-CoV-2 Infection and as a Potential Vaccine Adjuvant. *Front Immunol.* 2021; 12: 737406. DOI: 10.3389/fimmu.2021.737406. PMID: 34603318; PMCID: PMC8484798.
 13. Jamilloux Y., Henry T., Belot A. et al. Should we stimulate or suppress immune responses in COVID-19? Cytokine and anti-cytokine interventions. *Autoimmun Rev.* 2020; 19: 102567. DOI: 10.1016/j.autrev.2020.102567. PMID: 32376392; PMCID: PMC7196557.
 14. Ma S., Sun S., Li J. et al. Single-cell transcriptomic atlas of primate cardiopulmonary aging. *Cell Res.* 2021; 31: 415–32. DOI: 10.1038/s41422-020-00412-6. PMID: 32913304; PMCID: PMC7483052.
 15. Vicenzi M., Di Cosola R., Ruscica M. et al. The liaison between respiratory failure and high blood pressure, evidence from COVID-19 patients. *Eur Respir J.* 2020; 56: 2001157. DOI: 10.1183/13993003.01157-2020. PMID: 32430432; PMCID: PMC7241109.
 16. Wei S., Li C., Yin Z. et al. Histone methylation in DNA repair and clinical practice, new findings during the past 5-years. *J Cancer.* 2018; 9: 2072–81. DOI: 10.7150/jca.23427. PMC: 6010677. PMID: 29937925.
 17. Tiwari R., Mishra A.R., Gupta A., Nayak D. Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis. *J Biomol Struct Dyn.* 2021; 1–12. <https://doi.org/10.1080/07391102.2021.1874532>.
 18. Rocheleau L., Laroche G., Fu K., et al. Identification of a High-Frequency Intrahost SARS-CoV-2 Spike Variant with Enhanced Cytopathic and Fusogenic Effects. *MBio.* 2021; 12: e0078821. DOI: 10.1128/mBio.00788-21. Epub 2021 Jun 29. PMID: 34182784; PMCID: PMC8262852.
 19. Hutchison J.M., Capone R., Luu D.D. et al. Recombinant SARS-CoV-2 envelope protein traffics to the trans-Golgi network following amphipol-mediated delivery into human cells. *J Biol Chem.* 2021; 297: 100940. <https://doi.org/10.1016/j.jbc.2021.100940>.
 20. Boson B., Legros V., Zhou B. et al. The SARS-CoV-2 envelope and membrane proteins modulate maturation and retention of the spike protein, allowing assembly of virus-like particles. *J Biol Chem.* 2021; 296: 100111. <https://doi.org/10.1074/jbc.RA120.016175>.
 21. Hackstadt T., Chiramel A.I., Hoyt F.H. et al. Disruption of the Golgi Apparatus and Contribution of the Endoplasmic Reticulum to the SARS-CoV-2 Replication Complex. *Viruses.* 2021; 13: 1798. DOI: 10.3390/v13091798. PMID: 34578379; PMCID: PMC8473243.
 22. Hatakeyama D., Masuda T., Miki R. et al. In-vitro acetylation of SARS-CoV and SARS-CoV-2 nucleocapsid proteins by human PCAF and GCN5. *Biochem Biophys Res Commun.* 2021; 557: 273–9. DOI: 10.1016/j.bbrc.2021.03.173. PMID: 33894414; PMCID: PMC8030717.
 23. MROH2B — Function. https://www.nextprot.org/entry/NX_Q7Z745.
 24. Sansone A., Mollaioli D., Ciocca G. et al. Addressing male sexual and reproductive health in the wake of COVID-19 outbreak. *J Endocrinol Invest.* 2021; 44: 223–31. DOI: 10.1007/s40618-020-01350-1.
 25. Mihaescu G., Chifiriuc M.C., Ilescu C. et al. SARS-CoV-2, From Structure to Pathology, Host Immune Response and Therapeutic Management. *Microorganisms.* 2020; 8: 1468. DOI: 10.3390/microorganisms8101468. PMID: 32987852; PMCID: PMC7600570.
 26. Haig D.M. Subversion and piracy, DNA viruses and immune evasion. *Res Vet Sci.* 2001; 70: 205–19. DOI: 10.1053/rvsc.2001.0462. PMID 11676616.
 27. Tiwari M., Mishra D. Investigating the genomic landscape of novel coronavirus (2019-nCoV) to identify non-synonymous mutations for use in diagnosis and drug design. *J Clin Virol.* 2020; 128: 104441. DOI: 10.1016/j.jcv.2020.104441. PMID: 32425659; PMCID: PMC7227581.
 28. Chen M., Shen W., Rowan N.R. et al. Elevated ACE-2 expression in the olfactory neuroepithelium, implications for anosmia and upper respiratory SARS-CoV-2 entry and replication. *Eur Respir J.* 2020; 56: 2001948. DOI: 10.1183/13993003.01948-2020. PMID: 32817004; PMCID: PMC7439429.
 29. 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.* 2009; 10: 19. DOI: 10.1186/1471-2202-10-19. PMID: 19284620; PMCID: PMC2661083.
 30. Guruprasad L. Human SARS CoV-2 spike protein mutations. *Proteins.* 2021; 89: 569–76. DOI: 10.1002/prot.26042. PMID: 33423311; PMCID: PMC8014176.
 31. Guruprasad L. Evolutionary relationships and sequence-structure determinants in human SARS coronavirus-2 spike proteins for host receptor recognition. *Proteins.* 2020; 88: 1387–93. DOI: 10.1002/prot.25967. PMID: 32543705; PMCID: PMC7323375.
 32. Planas D., Veyer D., Baidaliuk A. et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature.* 2021; 596: 276–80. DOI: 10.1038/s41586-021-03777-9. PMID: 34237773.
 33. Chen J., Wang R., Gilby N.B., Wei G.W. Omicron (B.1.1.529): Infectivity, vaccine breakthrough, and antibody resistance. *ArXiv [Preprint].* 2021: arXiv:2112.01318v1. PMID: 34873578; PMCID: PMC8647651.
 34. Karim S.S.A., Karim Q.A. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *Lancet.* 2021; 398: 2126–8. DOI: 10.1016/S0140-6736(21)02758-6. PMID: 34871545; PMCID: PMC8640673.