Heliyon 8 (2022) e09222

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Unique peptide signatures of SARS-CoV-2 virus against human proteome reveal variants' immune escape and infectiveness

Vasileios Pierros [a,1], Evangelos Kontopodis [a,b,1], Dimitrios J. Stravopodis [b], George Th. Tsangaris [a,*]

[a] Proteomics Research Unit, Biomedical Research Foundation of the Academy of Athens, 11527, Athens, Greece
[b] Section of Cell Biology and Biophysics, Department of Biology, School of Science, National and Kapodistrian University of Athens, 15701, Athens, Greece

ARTICLE INFO

Keywords:
Core unique peptide
Covid-19
Delta variant
Immune escape
NF9 peptide
SARS-CoV-2
Spike protein
Uniquome

ABSTRACT

SARS-CoV-2 pandemic has necessitated the identification of sequence areas in the viral proteome that are capable to serve as antigenic sites and treatment targets. In the present study, we have applied a novel approach for mechanistically illuminating the virus-host organism interactions, by analyzing the Unique Peptides (UPs) of the virus featured by a minimum amino acid sequence length being defined as Core Unique Peptides (CrUPs), not of the virus *per se*, but against the entire proteome of the host organism. This approach resulted in the identification of CrUPs of the virus itself, which could not be recognized in the host organism proteome. Thereby, we analyzed the SARS-CoV-2 proteome for identification of CrUPs against the human proteome, which have been defined as C/H-CrUPs. We herein reveal that SARS-CoV-2 include 7.503 C/H-CrUPs, with the SPIKE_SARS2 being detected as the protein with the highest density of C/H-CrUPs. Extensive analysis has indicated that the critical P681R mutation produces new C/H-CrUPs around the R685 cleavage site, while the L452R mutation causes loss of antigenicity of the NF9 peptide and strong(er) binding of the virus to its ACE2 receptor protein. Simultaneous formation of these mutations in detrimental variants like Delta leads to the immune escape of the virus, its massive entrance into the host cell, a notable increase in virus formation, and its massive release and thus elevated infectivity of human target cells.

1. Introduction

Covid-19 pandemic has emerged the urgent necessity of the identification of sequence sites of the SARS-CoV-2 viral proteome that can serve as appropriate treatment targets and antigenic positions suitable for production of therapeutic vaccines.

As we have recently described, a Unique Peptide (UP) is defined as the peptide carrying an amino acid sequence that appears only in one of all proteins in a particular proteome. To this direction, our team has also introduced, for the first time, the concept of Core Unique Peptide (CrUP), which represents the peptide bearing a minimum length of amino acid sequence that resides solely in one of all proteins in a profiled proteome, thereby rendering it a unique signature for identification and differential recognition of a given protein (Alexandridou et al., 2009; Kontopodis et al., 2019). Hence, to thoroughly map the UP-specific landscape of a proteome of interest, we have developed a novel bioinformatics tool that is based on advanced algorithms being dedicated to big-data analysis. Its engagement to deep and accurate processing of the 20.430 reviewed *Homo sapiens* (human) proteins led to the recognition and identification of more than 7×10^6 CrUPs, which represent the backbone of human Uniquome that is mainly described as the voluminous collection of UPs shaping the human proteome (Kontopodis et al., 2022 and Kontopodis et al. manuscript in preparation).

Most importantly, to further illuminate the mechanisms controlling virus-host interactions, we have recently developed a novel, dynamic and advanced bioinformatics platform to thoroughly analyze and compare virus-derived CrUPs against host-organism proteome(s). This unique collection contains peptides that notably differ from the virus-specific CrUPs themselves, with each one of them being described as the peptide carrying an amino acid sequence of minimum length that is accommodated exclusively in one out of all proteins throughout the viral proteome. This virus against host CrUPs bear two cardinal properties: first, they are unique in virus proteome and, second, they do not exist in host-organism proteome. Therefore, the virus against host proteome-

* Corresponding author.
E-mail address: gthtsangaris@bioacademy.gr (G.Th. Tsangaris).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2022.e09222
Received 14 October 2021; Received in revised form 19 November 2021; Accepted 21 March 2022

Table 1. Viral CrUPs against Human proteome (C/H-CrUPs).

VIRUS	Proteins (number)	Total number of AA	Total C/H-CrUPs (number)	C/H-CrUPs appeared 1 time (number)	C/H-CrUPs appeared 2 times (number)	C/H-CrUPs appeared 3 times (number)	C/H-CrUPs Density
SARS-CoV-2	16	14.401	7.503	4.213	3.289	1	75%
SARS-CoV	15	14.396	7.534	4.236	3.298	0	75%
MERS	10	14.216	7.413	4.077	3.336	0	76%

Viral proteomes of the β coronavirus group SARS-CoV-2, SARS-CoV and MERS-CoV were analyzed for core unique peptides (CrUPs) against the human proteome. The identified CrUPs of each virus against the human proteome are presented (C/H-CrUPs). C/H-CrUPs were further analyzed for the times by which they appear in each viral proteome. C/H-CrUP density is defined as the percentage of total amino acids contained in C/H-CrUPs of each virus to the total number of the virus amino acids.

derived CrUPs can advance our knowledge and understanding of virus-host interactions, and virus infectiveness and pathogenicity dynamics. Furthermore, they can be used as diagnostic and antigenic peptides, and likely therapeutic targets, as well. Altogether, these CrUPs seem to represent a completely new entity of peptides capable to significantly improve our view and comprehension regarding the structuring, functioning and mapping of virus and human Uniquomes, and their proteomic "cross-talks", towards immune escape and infectiveness (Kontopodis et al., 2022).

Since human cells can host the SARS-CoV-2 virus, we have herein engaged our novel bioinformatics platform not only for the profiling of CrUPs in the SARS-CoV-2 proteome *per se*, but, most importantly, for their identification against the human proteome (C/H-CrUPs). Remarkably, C/H-CrUPs can likely serve as targets for the immune response upon infection, and antigenic sites with major pharmaceutical and diagnostic potential, for the successful clinical management of Covid-19 pandemic.

2. Results and discussion

2.1. SARS-CoV-2 core unique peptides against human proteome

The SARS-CoV-2 proteome is structurally quite simple. In the UNIPROT database (version 7/2021), 16 reviewed and 75.714 unreviewed proteins have been included (Jungreis et al., 2021). For the present study, only the 16 reviewed proteins are examined, since the unreviewed proteome components contain (among others) duplicate registrations, and unverified sequences and protein fragments, which could lead to unreliable data regarding the uniqueness of a protein sequence.

To recognize all the CrUPs being embraced in the SARS-CoV-2 proteome against the human proteome, we *in silico* constructed a new, artificial, "hybrid-proteome" that contained all the reviewed human proteins (20.430 proteins), plus the one protein derived from the SARS-CoV-2 viral proteome (20.431 proteins). Thus, 16 "hybrid proteomes" including the 16 SARS-CoV-2 proteins were constructed. Hence, these "hybrid proteomes" were bioinformatically searched one by one for the identification of SARS-CoV-2-specific CrUPs in human protein sequence environments (C/H-CrUPs).

Strikingly, 7.503 C/H-CrUPs were detected, with 4.213 of them being presented one time in the SARS-CoV-2 proteome, 3.289 being observed two times in the viral proteome and only one peptide ("VNNATN") with a 6 amino acid length being recognized three times (Table 1). Data processing and analysis unveiled that C/H-CrUPs retain a length range from 4 to 9 amino acids, while longer peptides could not be identified in the SARS-CoV-2 virus proteome. Length distribution showed that the majority of C/H-CrUPs have a 6 amino acid length, whereas only one with 4 amino acids and only two with 9 amino acids C/H-CrUPs were observed (Figure 1).

The distribution of C/H-CrUPs across SARS-CoV-2 proteins demonstrated that the Replicase Polyprotein 1ab (R1AB_SARS2), which is the longest viral protein consisted of 7.096 amino acids, produces almost half of the identified C/H-CrUPs (5.334; 49,3%) (Table 2). On the other hand, the Putative ORF3b protein (ORF3B_SARS2), with a length of 22 amino acids, produces only 15 C/H-CrUPs that show a protein density of 68%. Notably, Spike glycoprotein (SPIKE_SARS2) is presented with the highest C/H-CrUPs density (78%), thus indicating its intriguing feature to carry the highest number of C/H-CrUPs (987), in terms of their physical length,

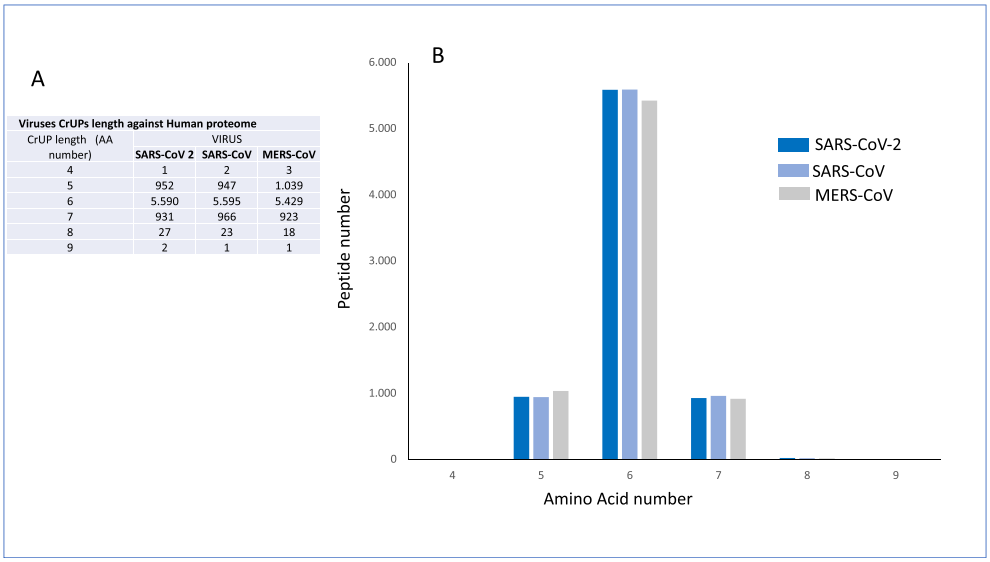


Figure 1. Amino acid length distribution of virus Core Unique Peptides (CrUPs) against human proteome. A) Set of CrUPs derived from SARS-CoV-2, SARS-CoV and MERS-CoV viruses against the human proteome. The CrUPs were identified, listed and grouped according to their amino acid length. B) Graphical presentation of CrUPs amino acid length across β coronavirus group.

Table 2. Virus detailed analysis.

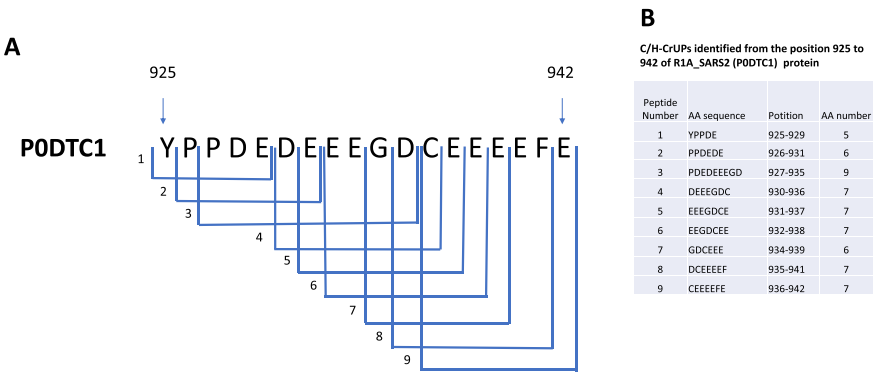
SARS-CoV-2					
Entry ID	Entry Name	Protein Name	Length (AA number)	C/H-CrUPs (number)	C/H-CrUPs Density
P0DTD1	R1AB_SARS2	Replicase polyprotein 1ab	7096	5334	75%
P0DTC1	R1A_SARS2	Replicase polyprotein 1a	4405	3294	75%
P0DTC2	SPIKE_SARS2	Spike glycoprotein	1273	987	**78%**
P0DTC9	NCAP_SARS2	Nucleoprotein	419	308	74%
P0DTC3	AP3A_SARS2	ORF3a protein	275	210	76%
P0DTC5	VME1_SARS2	Membrane protein	222	171	77%
P0DTC7	NS7A_SARS2	ORF7a protein	121	90	74%
P0DTC8	NS8_SARS2	ORF8 protein	121	82	68%
P0DTD2	ORF9B_SARS2	ORF9b protein	97	69	71%
P0DTD3	ORF9C_SARS2	Putative ORF9c protein	73	50	68%
P0DTC4	VEMP_SARS2	Envelope small membrane protein	75	48	64%
P0DTC6	NS6_SARS2	ORF6 protein	61	44	72%
P0DTG0	ORF3D_SARS2	Putative ORF3d protein	57	40	70%
P0DTD8	NS7B_SARS2	ORF7b protein	43	29	67%
P0DTG1	ORF3C_SARS2	ORF3c protein	41	23	56%
P0DTF1	ORF3B_SARS2	Putative ORF3b protein	22	15	68%
SARS-CoV					
Entry ID	**Entry Name**	**Protein Name**	**Length (AA number)**	**S/H-CrUPs (number)**	**S/H-CrUPs Density**
P0C6X7	R1AB_SARS	Replicase polyprotein 1ab	7.073	5.346	76%
P0C6U8	R1A_SARS	Replicase polyprotein 1a	4.382	3.301	75%
P59594	SPIKE_SARS	Spike glycoprotein	1.275	970	**76%**
P59595	NCAP_SARS	Nucleoprotein	422	319	76%
P59632	AP3A_SARS	ORF3a protein	274	208	76%
P59596	VME1_SARS	Membrane protein	221	162	73%
P59633	NS3B_SARS	ORF3b protein	154	113	73%
P59635	NS7A_SARS	ORF7a protein	122	93	76%
P59636	ORF9B_SARS	ORF9b protein	98	71	72%
Q80H93	NS8B_SARS	ORF8b protein	84	59	70%
P59637	VEMP_SARS	Envelope small membrane protein	75	47	63%
Q7TLC7	Y14_SARS	Uncharacterized protein 14	70	45	64%
P59634	NS6_SARS	ORF6 protein	63	44	70%
Q7TFA1	NS7B_SARS	Protein non-structural 7b	44	27	61%
Q7TFA0	NS8A_SARS	ORF8a protein	39	27	69%
MERS					
Entry ID	**Entry Name**	**Protein Name**	**Length (AA number)**	**M/H-CrUPs (number)**	**M/H-CrUPs Density**
K9N7C7	R1AB_MERS1	Replicase polyprotein 1ab	7.078	5.364	76%
K9N638	R1A_MERS1	Replicase polyprotein 1a	4.391	3.338	76%
K9N5Q8	SPIKE_MERS1	Spike glycoprotein	1.353	1.024	**76%**
K9N4V7	NCAP_MERS1	Nucleoprotein	411	301	73%
K9N643	ORF4B_MERS	Non-structural protein ORF4b	246	185	75%
K9N7D2	ORF5_MERS1	Non-structural protein ORF5	224	169	75%
K9N7A1	VME1_MERS1	Membrane protein	219	158	72%
K9N4V0	ORF4A_MERS1	Non-structural protein ORF4a	109	77	71%
K9N796	ORF3_MERS1	Non-structural protein ORF3	103	74	72%
K9N5R3	VEMP_MERS1	Envelope small membrane protein	82	59	72%

Analysis of the SARS-CoV-2, SARS-CoV and MERS-CoV virus is presented. All viruses' proteins have been *in silico*analyzed and each protein is shown by its Entry-ID, Entry Name and Protein Name according to the UNIPTOT database. The amino acid length of each protein and the number along with density of CrUPs per protein against the human proteome are shown. Density is defined as the percentage of total amino acids contained in CrUPs of each protein to the total number of the protein's amino acids.

as opposed to the ORF3c protein (ORF3C_SARS2), which is characterized by a respective density of only 56% (Table 2). A typical example for the construction of C/H-CrUPs is the peptide "PDEDEEEGD". This peptide is a 9 amino acid in length C/H–CrUP that belongs to Replicase polyprotein 1a (R1A_SARS2), starting at position 927 and ending at position 935 (Figure 2). Around this peptide, 8 C/H-CrUPs were recognized with a 5–7 amino acid length range.

2.2. Comparative analysis of SARS-CoV-2, SARS-CoV and MERS-CoV core unique peptides against human proteome

In order to illuminate the mechanisms orchestrating the differential pathologies of SARS-CoV-2 compared to other coronavirus family members, we, next, applied the same strategy to other two similar viruses, the Severe Acute Respiratory Syndrome CoronaVirus (SARS-



Peptide Number	AA sequence	Potition	AA number
1	YPPDE	925-929	5
2	PPDEDE	926-931	6
3	PDEDEEEGD	927-935	9
4	DEEEGDC	930-936	7
5	EEEGDCE	931-937	7
6	EEGDCEE	932-938	7
7	GDCEEE	934-939	6
8	DCEEEEF	935-941	7
9	CEEEEFE	936-942	7

Figure 2. Identification of C/H-CrUPs around amino acid position 925-942 of the SARS-CoV-2 protein R1A_SARS2 (P0DTC1). In between these amino acid positions one of the two C/H-CrUPs with a 9 amino acid length is included (927–935). A) Schematic representation of the C/H-CrUPs included in that peptide (925-942), B) Table of C/H-CrUPs.

CoV) and the Middle East Respiratory Syndrome-related CoronaVirus (MERS-CoV). Among human viruses, SARS-CoV-2 (C) together with SARS-CoV (S) and MERS-CoV (M) constitute the β coronavirus group, and they use the same cellular receptor, the Angiotensin-Converting Enzyme 2 (ACE2), with SARS-CoV-2 sharing approximately 80 and 70% amino acid sequence identity with SARS-CoV and MERS-CoV, respectively (Saputri et al., 2020; Walls et al., 2020). SARS-CoV viral proteome includes 15 reviewed proteins, while MERS-CoV contains 10 reviewed proteins in the UNIPROT database. Our findings confirm the strong similarities among these three coronaviruses at the level of CrUP structure and architecture against human proteome. Interestingly, a more comprehensive analysis of CrUPs per protein has revealed significant differences between them. The density of M/H-CrUPs per protein ranges between 71-76% (5% range), the density of S/H-CrUPs per protein varies between 61-76% (15% range) and the density of C/H-CrUPs per protein fluctuates between 56-78% (22% range) (Table 2), thus indicating the comparatively more heterogenous CrUPs density in the SARS-CoV-2 coronaviral proteome.

2.3. Comparative analysis of viruses spike protein

Among all SARS-CoV-2 proteins, the SPIKE_SARS2 (P0DTC2) one (Spike) has received the greatest attention as a key element for virus attachment to the host cell, and as such it has become a principal target for therapeutic vaccine development (Papa et al., 2021; Xia 2021). To mechanistically couple protein's molecular features with virus pathology at the level of C/H-CrUPs, we comparatively analyzed the Spike proteins of the three coronaviruses, and, next, we projected the findings onto SPIKE_SARS2 mutation map. Spike glycoprotein presents a length of 1.273 amino acids in SARS-CoV-2, 1.275 amino acids in SARS-CoV and 1.373 amino acids in MERS-CoV (Agrawal et al., 2021). Their densities in CrUPs against the human proteome are measured as 78%, 76% and 76%, respectively, exhibiting the highest CrUP density values among all proteins for each virus herein studied (Table 2). Amino acid sequence alignment of SPIKE_SARS2 (P0DTC2), SPIKE_SARS (P59594) and R9UQ53_MERS (R9UQ53) proved that these three viral Spike proteins share a group of 12 regions, herein defined as Universal Peptides (UnPs) (Figure 3 and Table 3). The majority of coronaviral UnPs are clustered in

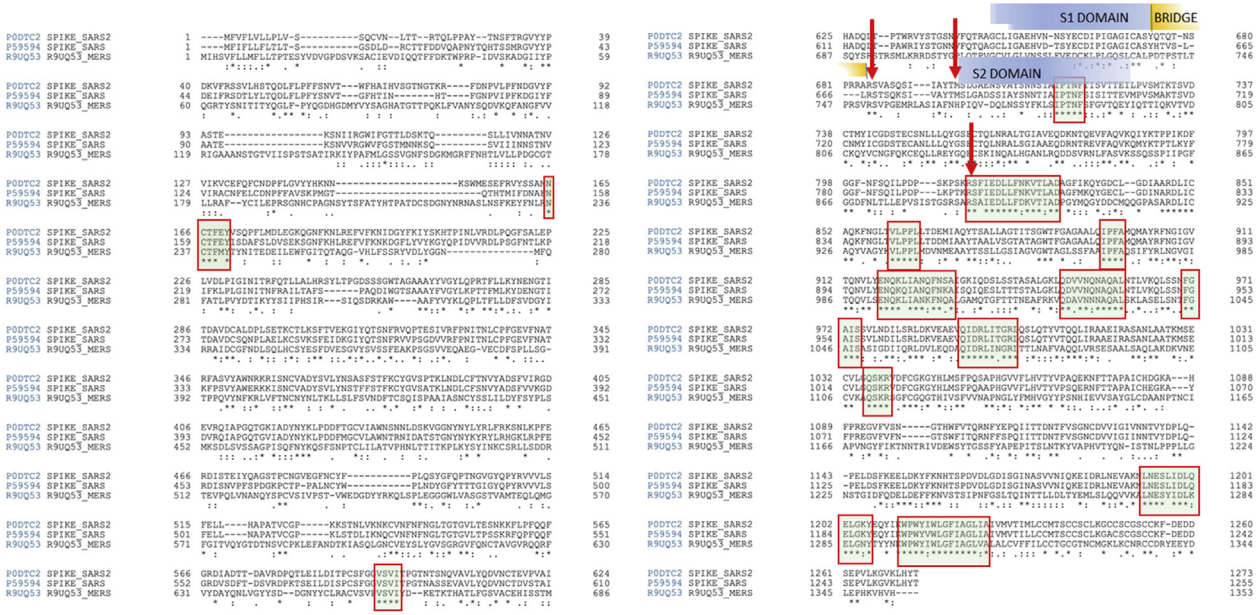
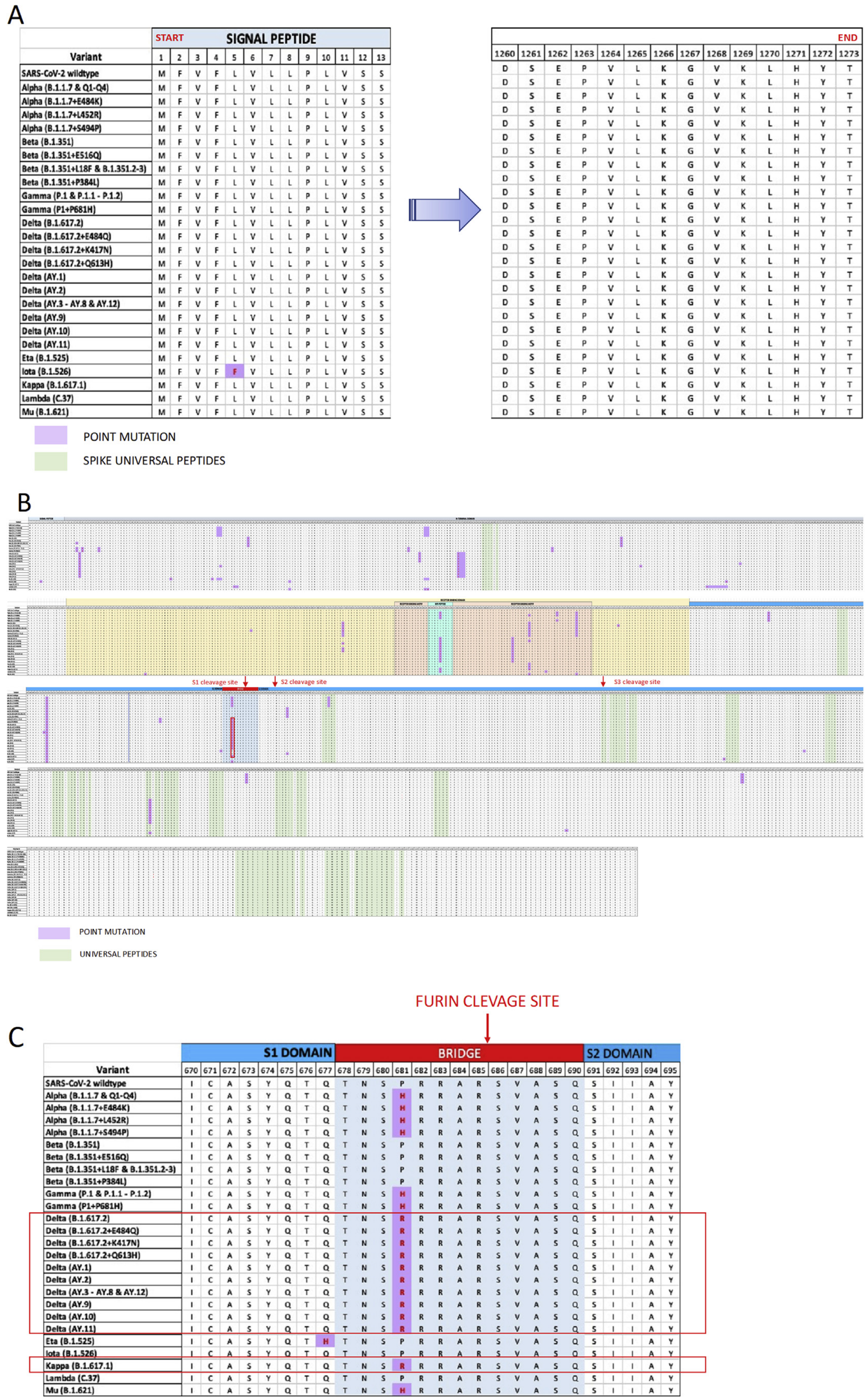


Figure 3. Alignment of the SARS-CoV-2, SARS-CoV and MERS-CoV Spike proteins. The amino acid sequence of sSpike proteins P0DTC2, P59594 and R9UQ53 derived from the SARS-CoV-2, SARS-CoV and MERS-CoV viruses, respectively, were obtained for Uniprot database and subsequently aligned, according to an online available bioinformatic tool in that database. Green blocks with red outline mark the identical peptidic sequences between the alignment sequences. The identical peptide sequences are considered as Universal Peptides (UnPs). Red arrows indicate the cleavage sites of the SARS-CoV-2 Spike protein.

Table 3. Spike-derived Universal Peptides (UnPs) and their residing CrUPs against the human proteome.

SITE				SEQUENCE	C/H-CrUP	DOMAIN	
165-168	170			NCTF*Y	CTFEY	S1 Domain	N-Terminal domain
595-598				VSVI	VSVITP	S1 Domain	
714-718				IPTNF	IPTNFT	S1 Domain	
815-816	818-823	825-827	829-830	RS*IEDLLF*KVT*AD	RSFIED	S2 Domain	S3 Cleavage site (Furin)
					SFIEDL		
					FIEDLL		
					IEDLLF		
					EDLLFN		
					DLLFNK		
					LLFNKV		
					LFNKVT		
860-864				VLPPL	VLPPLLT	S2 Domain	
896-899				IPFA	IPFAMQ	S2 Domain	Internal fusion peptide
918-921	923-925	927-928	930	ENQK*IAN*FN*A	ENQKLI	S2 Domain	
					QKLIAN		
					KLIANQ		
					LIANQF		
					IANQFN		
					ANQFNS		
					NQFNSA		
949-950	952-953	955-959		QD*VN*NAQAL	DVVNQN	S2 Domain	Heptad Repeats 1
					VVNQNA		
					NQNAQA		
					NAQALN		
970-974				FGAIS	FGAISSV	S2 Domain	Heptad Repeats 1
992-997	999-1001			QIDRLI*GRL	**QIDRLI**	S2 Domain	
					IDRLIT		
					DRLITG		
					RLITGR		
1036-1039				QSKR	QSKRVD	S2 Domain	
1193-1204	1206			LNESLIDLQELG*Y	**NESLID**	S2 Domain	Heptad Repeats 2
					SLIDLQ		
					LIDLQE		
					IDLQELG		
					DLQELGK		
					QELGKY		
1211-1216	1217-1224		1226	KWPWY*WLGFIAGL*A	**WPWY**	S2 Domain	Trans membrane domain
					PWYIWL		

Collection of the Universal peptides of SARS-CoV-2, SARS-CoV and MERS-CoV spike proteins according to Figure 4B alignement. The position in each protein sequence and the peptide sequence are shown. "*" symbol indicates positions with different amino acids residues among the examined proteins. CrUPs being contained in Universal Peptides (UnPs) are recorded. Notably, they are followed by the domain of Spike protein which they belong in. Yellow blocks indicate complete sequence CrUPs that appear in the Universal Peptides (UnPs) in all Spike proteins alignment.



(caption on next page)

the S2 domain of each Spike protein, with a critical one of them (UnPs) containing the Furin cleavage site 3 (R^{815}↓S).

2.4. Analysis of SARS-CoV-2 variants spike protein

Most importantly, SARS-CoV-2 Spike protein has presented a significant mutational diversity (Sanches et al., 2021; Tzou et al., 2020). Hitherto, 9 main variants with adaptive mutations and high spread to human populations, named from Alpha to Lambda, respectively, have been thoroughly mapped and characterized. These 9 variants are divided in 39 sub-variants, while other 32 sporadic variants have also been described (Tzou et al., 2020). To investigate the association of mutational profiling with C/H–CrUP landscaping of SARS-CoV-2 Spike protein, the 39 sub-variants together with the wild-type Spike protein (SPIKE_SARS2, P0DTC2) were suitably aligned (Figure 4). This multiple alignment illustrates all the herein identified Universal Peptides (UnPs) (Table 3) and all the mutations previously announced per isolated variant (Figure 4B). Notably, it seems that almost all the hitherto characterized mutations are identified in regions being located outside the UnPs group. Their majority are clustered in the S1 domain of Spike protein, with two critical mutations being detected in the S1–S2 bridge region, at the amino acid residue 681 that resides in proximity to the first cleavage position by Furin protease, in between the 685th and 686th amino acid residue (Figure 4C) (Davidson et al., 2020; Coutard et al., 2020).

Remarkably, all the examined mutations herein prove to create new CrUPs against the human proteome compared to the wild-type Spike protein, thus indicating that the mutant virus strains need novel clinical treatments. This is an important finding, since these new C/H-CrUPs do not exist in the human proteome, but are observed exclusively in the mutant virus proteomes, thereby justifying the great attention Alpha, Delta, Kappa, Lambda and Mu variants have recently received at the worldwide level (Tzou et al., 2020). Table 4 lists all the novel C/H-CrUPs being created by the hitherto reported mutations in coronavirus variants. These variants include 25 mutations, which produce 44 new CrUPs against the human proteome. It may be these novel C/H-CrUPs that give rise to formation of new Intrinsically Disordered Regions (IDRs) and Small Linear Motifs (SLiMs) in the SARS-CoV-2 Spike protein mutant versions (van der Lee et al., 2014; Hraber et al., 2020).

The molecular mechanism of Spike protein's proteolytic activation has been shown to play a crucial role in the selection of host species, virus binding to the ACE2 receptor, virus-cell fusion, and viral infection of human lung cells (Peacock et al., 2021; Whittaker 2021; Shang et al., 2020a). Spike protein contains three cleavage sites: the R^{685}↓S and the R^{815}↓S positions that serve as direct targets of Furin protease, and the T^{696}↓M position that can be recognized by TMPRSS2 protease (Hoffmann et al., 2020a, 2020b; Takeda, 2021). Analysis of the wild-type C/H-CrUPs and the new formed, mutation-induced, C/H-CrUPs in Spike protein unveiled that the mutation-driven, novel, peptides are created exclusively around the critical R^{685}↓S cleavage site by the two pathogenic mutations P681H and P681R (Table 5).

2.5. Analysis of C/H-CrUPs around the R^{685}↓S cleavage site

Notably, among these four new peptides (Table 5), the only one that embraces Furin's cleavage site is the "SRRRAR↓S" C/H–CrUP, which is solely generated by the P681R mutation carried by the Delta and Kappa coronavirus variants, while at the same time the "PRRARSV" peptide maintains its uniqueness even after the replacement of Proline (P) with Arginine (R) and its transformation to "RRRARSV" (Figure 5A,B).

The Furin cleavage site R^{685}↓S has been characterized as a 20 amino acid sequence motif that corresponds to the amino acid sequence A672-S691 of the Spike protein (Figure 4A,B) (Wu and Zhao, 2020). The 8 amino acid sequence peptide "SPRRAR↓SV" (S680–V687) serves as the core region of the motif, while two flanking solvent-accessible regions of 8 amino acids (A672-N679) and 4 amino acids (A688-S691) long, respectively, are recognized (Takeda, 2021; Wu and Zhao, 2020).

Pro-protein Convertase (PC) Furin and/or Furin-like PCs act as sequence-specific proteases and can cleave the Spike protein in a position recognizing the unique, and positively charged by the Arginine, motif "R-x-x-R↓S" (Wu and Zhao, 2020). Since Furin and/or Furin-like PCs are secreted from host cells and bacteria in the airway epithelium, while other PCs, such as PC5/6A and PACE4, exhibit widespread tissue distribution, it is likely that their activities may be critically implicated in the SARS-CoV-2-induced damage and pathology of multiple infected organs (Örd et al., 2020). It seems that Furin's cleavage site essentially contributes to the infection process and disease progression, and offers a powerful target for immunogenetic, antigenic and therapeutic interventions, as strongly supported by the recently developed new antibody against Furin's cleavage site (Braun and Sauter, 2019; Zahradník et al., 2021; Wu et al., 2020).

Most importantly, the SARS-CoV-2 Delta variant that carries the critical mutation P681R seems to be more infectious and pathogenic than the wild-type virus form, while the importance of this mutation has very recently begun to be recognized (Wu et al., 2020). Replacement of Proline (P) with Arginine (R) at position 681 causes the loss of amino acid sequence uniqueness that characterizes the wild-type "PRRARSV" C/H–CrUP and likely increases the possibility of Furin's cleavage site (core region) to be significantly stabilizing its conformation, thus facilitating a more efficient Spike protein cleavage process by the Furin protease (Whittaker, 2021; Callaway, 2021).

To the same direction, novel SLiMs, such as "SRRR", "RRR", "RRRAR" and "RRRARS", can be produced by the mutant C/H-CrUPs, which may act as specific targets of other than Furin PCs, thereby enabling the stronger (and quicker) binding of the mutant virus to its host ACE2 receptor, which likely leads to a comparatively more generalized infection and massive mutant virus production (Table 6) (Shorthouse and Hall, 2021; Davey et al., 2015). This finding seems to be evidenced by the remarkable increase of the total number of motifs created by the P681R mutation identified within the human proteome (Table 6). Of note, the mutant C/H–CrUP-derived new SLiMs, in the SARS-CoV-2 Delta variant, could render Spike protein antigenically weak or defective, fostering it to lose its capacity to serve as antibody target and thus promoting the virus immune escape (Davey et al., 2015; Almehdi et al., 2021).

2.6. Analysis of C/H-CrUPs around the ACE2 receptor site

An important issue for viral infectivity and pathogenesis is the receptor recognition and binding of the virus to the host cell surface. SARS-CoV-2 belongs to the β coronavirus group and, like SARS-CoV, uses the same cellular receptor, the Angiotensin-Converting Enzyme 2 (ACE2) (Walls et al., 2020; Wang et al., 2020). The SARS-CoV-2 Spike protein attaches to ACE2 receptor by a Receptor-Binding Domain (RBD) defined

Figure 4. Alignment of the SARS-CoV-2 Spike protein (SPIKE_SARS2, P0DTC2) of the 25 sup-variants belonging to the major 9 virus variants, together with the native (wild-type) Spike Protein. A) N-terminal and C-terminal areas of the native (wild-type) Spike protein, and the 25 sup-variants are presented. B) Complete Spike protein sequence alignment. Purple blocks mark the point mutation sites in variants; green color indicates the Unique Peptides (UnPs) of the Spike proteins from Figure 3; yellow color denotes the Receptor-Binding Domain (RBD) of Spike protein to ACE2; pink color indicates the Receptor-Binding Motif (RBM); cyan color marks the NF9 peptide; light blue color indicates the bridge between S1 and S2 domains; red arrows denote the cleavage sites. Different domains of the Spike protein are marked with different colors in the upper side of the alignment. C) The Spike protein alignment around the bridge domain (light blue color) between the S1 and S2 domains is presented. Red arrow denotes the Furin cleavage site R^{685}↓S. Purple blocks mark the point mutations around this position, while red outline indicates the Delta and Kappa variants carrying the critical mutation P681R.

Table 4. New C/H-CrUPs of SARS-CoV-2 Spike protein in Alpha, Delta, Kappa and Lambda variants.

Mutations position	Mutation	Variant	New C/H-CrUPs first AA position	New C/H-CrUPs
19	T19R	Delta_P0DTC2	-	-
70	V70F	Delta_P0DTC2	69	HFSGTN
			70	FSGTNG
75 - 76	G75V&T76I	Lambda_P0DTC2	71	SGTNVI
			75	VIKRFD
222	A222V	Delta_P0DTC2	218	QGFSVL
258	W258L	Delta_P0DTC2	-	-
417	K417N	Delta_P0DTC2	413	GQTGNI
			414	QTGNIA
452	L452R	Delta_P0DTC2	449	YNYRY
		Kappa_P0DTC2		
		Alpha_P0DTC2		
	L452Q	Lambda_P0DTC2	448	NYNYQ
			449	YNYQY
478	T478K	Delta_P0DTC2	474	QAGSKP
			478	KPCNG
484	E484Q	Kappa_P0DTC2	481	NGVQG
			483	VQGFN
			484	QGFNC
	E484K	Alpha_P0DTC2	484	KGFNC
490	F490S	Lambda_P0DTC2	487	NCYSP
494	S494P	Alpha_P0DTC2	-	-
501	N501Y	Alpha_P0DTC2	498	QPTY
			499	PTYG
			500	TYGV
			501	YGVG
570	A570D	Alpha_P0DTC2	568	DIDDTT
614	D614G	Delta_P0DTC2	609	AVLYQG
		Kappa_P0DTC2		
		Alpha_P0DTC2		
		Lambda_P0DTC2		
		Delta_P0DTC2	610	VLYQGV
		Kappa_P0DTC2		
		Alpha_P0DTC2		
		Lambda_P0DTC2		
681	P681R	Delta_P0DTC2	680	SRRRARS
		Kappa_P0DTC2		

	P681H	Alpha_P0DTC2	677	QTNSH
			678	TNSHR
			680	SHRRAR
716	T716I	Alpha_P0DTC2	714	IPINF
859	T859N	Lambda_P0DTC2	855	FNGLNV
			857	GLNVLP
950	D950N	Delta_P0DTC2	946	GKLQN
			947	KLQNVV
			948	LQNVVN
			949	QNVVNQ
982	S982A	Alpha_P0DTC2	978	NDILAR
1071	Q1071H	Kappa_P0DTC2	1067	YVPAH
			1069	PAHEKN
			1071	HEKNF
1118	D1118H	Alpha_P0DTC2	1113	QIITTH
			1115	ITTHN
			1116	TTHNT
			1117	THNTF
			1118	HNTFV

The new C/H-CrUPs of SARS-CoV-2 spike protein (SPIKE_SARS2, P0DTC2) across the variants Alpha, Delta, Kappa and Lambda are presented. In the first column, the position of each mutation in the Spike protein sequence is shown. In the second column the mutation is recorded. In the third column, the SARS-CoV-2 main variant which each mutation is appeared in, is recorded. In the fourth column, the position of the first amino acid residues of the new C/H–CrUP created by each mutation is shown. In the last column, the new created C/H-CrUPs by each mutation is recorded. Each mutant amino acid residue in the new C/H-CrUPs is denoted by red color. Mutations that not create new C/H-CrUPs are indicated by the symbol '-'. Some mutations produce multiple new C/H-CrUPs, while 4 new C/H-CrUPs are created in more than one variant.

in the Spike protein from positions F318 up to F541 (Shang et al., 2020b). Nowadays, this region has received great attention, as it seems to be the target of antibodies against the virus and other therapeutic interventions (Chen et al., 2021; Zahradník et al., 2021; Hastie et al., 2021). Additional studies have shown that from the amino acid residue W436 up to the Q506 one the RBD contains the Receptor-Binding Motif (RBM), which carries 12 contact positions with ACE2 (Hatmal et al., 2020). Mutation analysis revealed that in 10 positions of the RBD region 13 mutations were described (Figure 4 and Table 7). In RBM, 10 mutations in 6 sequence positions were reported for different virus variants (Table 7), while from the 10 contact positions only the P501Y in Alpha, Beta, Gamma and Mu variants was found to be mutated.

2.7. C/H-CrUPs around the NF9 peptide

The most important region in RBM is the peptide "NYNYLYRLF" (from 448 to 456 position). This Tyrosine (Y)-enriched peptide contains two contact site (Y449 and Y453) and it is known as the NF9 peptide (Motozono et al., 2021). It seems to affect antigen recognition, by being an immunodominant HLA*24:02-restricted epitope identified by the $CD8^+$ T-cells. Furthermore, NF9 stimulation also increases cytokine production by the $CD8^+$ T-cells, such as IFN-γ, TNF-α and IL-2 (Kared et al., 2021). Analysis of C/H-CrUPs that are being associated with the NF9 peptide showed that it contains 3 UPs (Figure 5D,E, and Table 7). Mutation analysis indicated that in the NF9 peptide the mutation L452R is carried by the variants Alpha, Delta, Lamda and Kappa, while the mutation L452Q appears in the variant Lambda. Further analysis unveiled that these mutations are observed in the amino acid that resides at position 5, exactly in the middle of the peptide, creating 3 and 4 new C/H CrUPs, respectively (Table 8). These mutations have a dramatic effect in the uniqueness of the NF9 peptide(s). Namely, the 6 amino acid length C/H-CrUPs "NYNYLY" lose their uniqueness against the human proteome, while only by the mutation L452Q a new CrUP with 5 amino acid length is surprisingly created (Figure 5D,E). The loss of uniqueness of this peptide, which notably is located at the beginning of NF9 peptide, seems to be crucial, as it leads to the loss of antigenic capacity of the NF9 peptide, thus evading the HLA-A24-restricted immunity and inducing the immune escape of the virus. Interestingly, related studies have shown that the L452R mutation (and subsequently the new created C/H-CrUPs herein characterized) increases the infectiveness of SARS-CoV-2, by strengthening the electrostatic interactions of this region on Spike protein with the ACE2 virus receptor (Motozono et al., 2021).

Hitherto, epidemiological data indicated that the dominant variant of SARS-CoV-2 is the Delta variant (Micochova et al., 2021). Under the light of the aforementioned findings, variant's enhanced pathogenicity seems to be the outcome of the simultaneous presence (accumulation) of two critical mutations, the L452R and P681R ones, in Delta variant. The mutation L452R, through the loss of NF9 peptide uniqueness, causes virus immune escape and strong(er) binding of the virus to its cognate receptor, while at the same time the mutation P681R facilitates the Spike protein cleavage process by different proteases, inducing a generalized infection and a massive virus release. Therefore, the Delta variant gains a significant advantage of escape from the immune system *per se*, as well as from the vaccination-induced immunity, together with an increased infectiveness as a result of virus entrance into the host cell, and an increase of virus formation and its massive release.

Table 5. New C/H-CrUPs around the SARS-CoV-2 Spike protein cleavage sites.

Cleavage site	Mutation	Variant	New C/H-CrUPs first AA position	New C/H-CrUP
$R^{685}\downarrow S$	P681R	Delta & Kappa	680	SRRRAR↓S
	P681H	Alpha & Gamma	677	QTNSH
			678	TNSHR
			680	SHRRAR
$T^{696}\downarrow M$	A701V	Beta	None	
$R^{815}\downarrow S$	None		None	

The new C/H-CrUPs created by the mutations around the SARS-CoV-2 spike protein (SPIKE_SARS2, P0DTC2) are identified. Fist column: The cleavage site of SARS-CoV-2 Spike protein. Second column: The mutation identified around the cleavage site. Third column: The virus variants in which the mutation appears in. Fourth column: The position in the SARS-CoV-2 Spike protein sequence which the first amino acid of the C/H–CrUP appears in. Fifth column: The sequence of the new C/H–CrUP. "↓" symbol indicates the cleavage site within this peptide.

3. Conclusion

Since mutations outside the Spike protein locus in SARS-CoV-2 coronavirus genome have not been yet completely mapped, in a systematic manner, our study importantly reveals novel and useful information of all the remaining, Spike protein-independent, C/H-CrUPs that seem to hold strong promise and open new therapeutic windows for the Covid-19 pandemic. Finally, the approach of virus-host UP-specific signature identification could prove a useful tool for the elucidation of virus infectiveness, prevention of virus immune escape, domination of pathogenic variants, and identification of new antigenic and pharmacological targets.

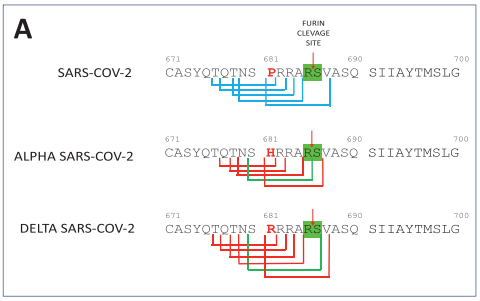
4. Materials and methods

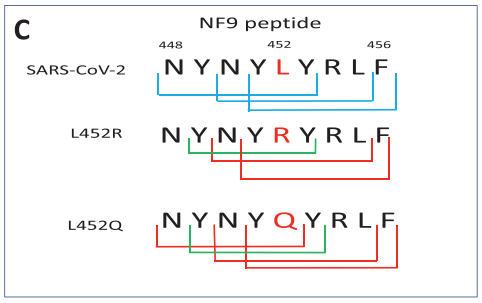
4.1. Methods

A new bioinformatics tool that has been recently built on an advanced big-data algorithm was herein developed to extract CrUP collections from proteomes of interest and, thereby, create organism-specific Uniquomes. The user can specify the min and max peptide lengths that the tool will analyze. The tool will split each protein to all possible peptides of length min to length max, thus generating a very large set of peptides (for a protein of length **L** with a window of size **W**, a set of "$\mathbf{C = L - W + 1}$" will be generated). In the next step, all these peptides, starting from smallest and ending to largest, will be searched against the rest of the proteome to decide whether the peptide exists on another protein or not. Since the search is dedicated for the smallest possible peptide (Core Unique Peptide: CrUP), the tool will first make sure that the peptide under examination does not already contain a smaller CrUP. This is ensured by examining if any of the already identified CrUPs of the protein is contained within the peptide under examination. All peptides that conform to these two rules are considered as CrUPs. Figure 6 describes the algorithm we have herein developed and used to recognize these novel CrUPs.

In Figure 7, a sliding window of 9 amino acids is applied on O00400 ACATN_HUMAN protein, generating the candidate peptides "VYVKNFGRR" and "YVKNFGRRK". These peptides will be searched against the rest of the proteome, to determine their uniqueness once we have ensured that they do not already contain a smaller CrUP. The latter is determined by examining whether an already defined CrUP is included within the peptide.

To address the question of the present study, the aforementioned tool was expanded by developing a new feature, where the user can give a reference and a target proteome. This new feature allows the tool to search all the peptides of the target proteome against the reference





B C/H-CrUPs of wild-type and mutant SARS-CoV-2 around the $R^{685}\downarrow S$ cleavage site.

		PEPTIDES	
		MUTATION	
POSITION	SARS-CoV-2	P681R	P681H
676	TQTNSP	TQTNSR	
677	QTNSPR	QTNSRR	QTNSH
678	TNSPRR	TNSRRR	TNSHR
679	NSPRRA	NSRRRA	NSHRRA
680		SRRRAR↓S	SHRRAR
681	PRRAR↓SV	RRRAR↓SV	HRRAR↓SV

D NF9 wild-type and mutant C/H-CrUPS.

		PEPTIDES	
		MUTATION	
POSITION	SARS-CoV-2	L452R	L452Q
448	NYNYLY		NYNYQ
449		YNYRY	YNYQY
450	NYLYRL	NYRYRL	NYQYRL
451	YLYRLF	YRYRLF	YQYRLF

Figure 5. C/H-CrUPs residing around the $R^{685}\downarrow S$ cleavage site and belonging to the NF9 peptide of Spike protein (SPIKE_SARS2, P0DTC2). A) Amino acid sequences of Spike protein between position 671 and 700 in wild-type, Alpha and Delta variants of SARS-CoV-2 virus are shown. In each variant, the identified C/H-CrUPs are marked. Blue lines indicate C/H-CrUPs derived from wild-type protein around the $R^{685}\downarrow S$ cleavage site. Red lines denote C/H-CrUPs produced by the P681H and P681R mutations. Green lines indicate the new created mutant C/H-CrUPs that derive from the P681H and P681R mutations in Alpha and Delta variants, respectively. B) Set of C/H-CrUPs generated around the $R^{685}\downarrow S$ cleavage site of wild-type and mutant Spike protein forms. C) Amino acid sequences of the NF9 peptide between positions 448 and 456 in wild-type Spike protein, before and after creation of the L452R and L452Q mutations. Blue lines indicate C/H-CrUPs that belong to the NF9 peptide. Red lines denote C/H-CrUPs that are produced by the L452R and L452Q mutations. Green lines indicate the new generated mutant collection of C/H-CrUPs derived from the L452R and L452Q mutations. D) Set of C/H-CrUPs residing in the NF9 peptide in wild-type, and L452R and L452Q mutated protein forms.

Table 6. Small Linear Motifs (SLiMs) of wild-type C/H-CrUPs and C/H-CrUPs created by the critical mutation P681R being detected in human proteome.

Motif	Number of proteins in UNIPROT contain the motif	Motif found	Protein Entry ID	Protein Entry Name	Protein full Name
PRRARSV	0	-	-		
XRRARSV	1	ARRARSV	P37088	SCNNA_HUMAN	Amiloride-sensitive sodium channel subunit alpha
PXRARSV	0	-	-		
PRXARSV	1	PRPARSV	Q96PD5	PGRP2_HUMA	N-acetylmuramoyl-L-alanine amidase
PRRXRSV	1	PRRSRSV	Q9UQ35	SRRM2_HUMAN	Serine/arginine repetitive matrix protein 2
PRRAXSV	2	PRRASSV	Q04844	ACHE_HUMAN	Acetylcholine receptor subunit epsilon
		PRRALSV	Q5VZ46	K1614_HUMAN	Uncharacterized protein KIAA1614
PRRARXV	0	-	-		
PRRARSX	1	PRRARSS	Q92902	HPS1_HUMAN	Hermansky-Pudlak syndrome 1 protein
TOTAL TIMES	**6**				

Motif	Number of proteins in UNIPROT contain the motif	Motif found	Protein Entry ID	Protein Entry Name	Protein full Name
SRRRARS	0	-	-		
XRRRARS	6	RRRRARS	Q8WUQ7	CATIN_HUMAN	Cactin
		RRRRARS	P18825	ADA2C_HUMAN	Alpha-2C adrenergic receptor
		DRRRARS	Q96QZ7	MAGI1_HUMAN	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 1
		PRRRARS	C9J069	ALM1_HUMAN	Apical junction component 1 homolog
		WRRRARS	O00198	HRK_HUMAN	Activator of apoptosis harakiri
		PRRRARS	Q9NZV5	SELN_HUMAN	Selenoprotein N
SXRRARS	3	SDRRARS	Q8N2C7	UNC80_HUMAN	Protein unc-80 homolog
		SPRRARS	Q92902	HPS1_HUMAN	Hermansky-Pudlak syndrome 1 protein

		SCRRARS	Q5T4W7	ARTN_HUMAN	Artemin
SRXRARS	1	SRDRARS	Q92917	GPKOW_HUMAN	G-patch domain and KOW motifs-containing protein
SRRXARS	1	SRRQARS	Q9NSI2	F2007A_HUMAN	Protein FAM207A
SRRRXRS	10	SRRRPRS	Q70EL4	UBP43_HUMAN	Ubiquitin carboxyl-terminal hydrolase 43
		SRRRIRS	P05198	IF2A_HUMAN	Eukaryotic translation initiation factor 2 subunit 1
		SRRRRRS	P18583	SOV_HUMAN	Protein SON
		SRRRSRS	Q8N2M8	CLASR_HUMAN	CLK4-associating serine/arginine rich protein
		SRRRSRS	Q15058	KIF_HUMAN	Kinesin-like protein KIF14
		SRRRRRS	Q5M9Q1	NKAPL_HUMAN	NKAP-like protein
		SRRRSRS	Q14498	RBM39_HUMAN	NA-binding protein 39
		SRRRSRS	Q96T37	RBM15_HUMAN	RNA-binding protein 15
		SRRRSRS	Q13247	SRSF6_HUMAN	Serine/arginine-rich splicing factor 6
		SRRRQRS	Q9UQ35	SRRM2_HUMAN	Serine/arginine repetitive matrix protein 2
SRRRAXS	6	SRRRAIS	Q00987	MDM2_HUMAN	3 ubiquitin-protein ligase Mdm2
		SRRRAQS	Q9NQU5	PAK_HUMAN	Serine/threonine-protein kinase PAK 6
		SRRRADS	Q53GL0	PKHO1_HUMAN	Pleckstrin homology domain-containing family O member 1
		SRRRAWS	Q9UKN7	MYO15_HUMAN	Unconventional myosin-XV
		SRRRAFS	Q9GZK7	011A1_HUMAN	Olfactory receptor 11A1
		SRRRAVS	Q9BYX2	TBD2A_HUMAN	TBC1 domain family member 2A
SRRRARX	3	SRRRARR	O95450	ATS2_HUMAN	A disintegrin and metalloproteinase with thrombospondin motifs 2
		SRRRARD	Q8N5L8	RP25L_HUMAN	Ribonuclease P protein subunit p25-like protein
		SRRRARV	Q9GZQ6	NPFF1_HUMAN	Neuropeptide FF receptor 1
TOTAL TIMES	**30**				

Motif	Number of proteins in UNIPROT contain the motif	Motif found	Protein Entry ID	Protein Entry Name	Protein full Name
RRRARSV	0	-	-		
XRRARSV	1	ARRARSV	P37088	SCNNA_HUMAN	Amiloride-sensitive sodium channel subunit alpha
RXRARSV	1	RPRARSV	Q86X29	LSR_HUMAN	Lipolysis-stimulated lipoprotein receptor
RRXARSV	2	RRDARSV	Q8WWN8	ARAP3_HUMAN	Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 3
		RRPARSV	Q9H427	KCNKF_HUMAN	Potassium channel subfamily K member 15
RRRXRSV	3	RRRSRSV	P18583	SON_HUMAN	Protein SON
		RRRKRSV	P49685	GPR15_HUMAN	G-protein coupled receptor 15
		RRRASSV	O14681	EI24_HUMAN	Etoposide-induced protein 2.4 homolog
RRRAXSV	3	RRRAQSV	Q7LDG7	GRP2_HUMAN	RAS guanyl-releasing protein 2
		RRRAPSV	P21333	FLNA_HUMAN	Filamin-A
		RRRARPV	Q7RTU4	BHA09_HUMAN	Class A basic helix-loop-helix protein 9
RRRARXV	4	RRRARQV	Q8N9Z2	CC71L_HUMAN	Coiled-coil domain-containing protein 71L
		RRRARAV	Q6NUJ1	SAPL1_HUMAN	Proactivator polypeptide-like 1
		RRRARVV	Q9GZQ6	NPFF1_HUMAN	Neuropeptide FF receptor 1
		RRRARSW	Q8WUQ7	CATIN_HUMAN	Cactin
RRRARSX	5	RRRARSS	P18825	ADA2C_HUMAN	Alpha-2C adrenergic receptor
		RRRARSP	Q96QZ7	MAGI1_HUMAN	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 1
		RRRARSK	C9J069	AJM1_HUMAN	Apical junction component 1 homolog
		RRRARSR	O00198	HRK_HUMAN	Activator of apoptosis harakiri
		RRRARSL	Q9NZV5	SELN_HUMAN	Selenoprotein N
TOTAL TIMES	**19**				

Motif	Number of proteins in UNIPROT contain the motif	Motif	Number of proteins in UNIPROT contain the motif	Motif	Number of proteins in UNIPROT contain the motif
XRRRARX	47	SRRXXRS	46	RXXRS	3774
XRRRAXS	44	SRRRXRX	53	**TOTAL TIMES**	**3774**
XRRRXRS	139	SRRRXXS	50		
XRRXARS	30	SRRRAXX	25		
XRXRARS	22	**TOTAL TIMES**	**174**		
XXRRARS	27				
SXRRARX	29				
SXRRAXS	29				
SXRRXRS	72				
SXRXARS	17				
SXXRARS	20				
SRXRARX	19				
SRXRAXS	35				
SRXRXRS	175				
SRXXARS	16				
SRRXARX	19				
SRRXAXS	24				
TOTAL TIMES	**735**				

The list of SLiMs of wild-type and mutant C/H-CrUPs produced by the critical mutation P681R in SPIKE_SARS2, and being detected in the human proteome, are presented. Green block indicates the C/H–CrUP in wild-type protein; blue block denotes the mutant C/H–CrUP peptide derived from the P681R mutation; yellow block descibed the newly created C/H–CrUP by the same mutation. X (in red color) is used for the position within the peptide to create the motif. In the third column, the detected motif is recorder, and is followed by the Protein Entry-ID and the protein name it is detected in. "Total" summarizes the time for which the motifs related to C/H–CrUP are recorded in the human proteome.

proteome, thus creating a set of CrUPs of target versus reference proteome. To this direction, the tool (similar to the initial implementation) will split all proteins in the target proteome to all possible peptides of length min to length max. Now, instead of searching for the uniqueness of each peptide within the same proteome, it performs that search against the reference proteome. Like before, the peptide under examination must not contain any smaller peptides already identified as CrUPs. The algorithm we have employed to identify these CrUPs is described diagrammatically in Figure 6.

4.2. Motifs and SLiMs search

For Motif and SLiM identification, and search, the tool offers the user the ability to perform a motif search to identify putative SLiMs. User gives an N-length peptide, as well as the number of amino acids that can vary in the given peptide. Then, the tool creates all possible combinations of peptides that can be produced by considering in each combination exactly N-amino acid(s) as unknown. Once these combinations are produced, an exhaustive search using regular expressions is performed against the reference proteome, to locate all possible proteins containing such peptides. To better highlight the process, if the user provides the peptide "TQYILG" and N = 2, the following combinations will be generated:

- ??YILG
- ?Q?ILG
- ?QY?LG
- ?QYI?G

Table 7. C/H-CrUPs of wild-type and mutant Receptor-Binding Domain (RBD) of SARS-Cov-2 Spike protein.

SARS-CoV-2 SPIKE PROTEIN RECEPTOR BINDING DOMAIN															
WILDTYPE								MUTANT							
Position	C/H-CrUP						Peptide number / peptide length	Mutation	VARIANT	NEW C/H-CrUP					Peptide number/ peptide length
346	VFNATR	FNATRF	NATRFA	ATRFAS	TRFASV	RFASVY	6/6AA	R346K	Mu	VFNATK	FNATKF	ATKFAS	TKFASV	KFASVY	5/6AA
384	YGVSPT	GVSPTK	VSPTKL	SPTKLN	PTKLND		5/6AA	P384L	Beta	YGVSLT	GVSLTK	SLTKLN	LTKLND		4/6AA
417	PGQTGKI	GQTGKIA	TGKIAD	GKIADY			2/7AA, 2/6AA	K417N	Beta, Delta	GQTGNI	QTGNIA	TGNIAD	GNIADY		4/6AA
								K417T	Gamma	PGQTGT	GQTGTI	QTGTIA	TGTIAD		4/6AA
452	GNYNYL	NYNYLY	NYLYRL	YLYRLF	LYRLFR		5/6AA	L452R	Alpha, Delta, Iota, Kappa	GNYNYR	YNYRY	NYRYRL	YRYRLF		1/5AA, 3/6AA
								L452Q	Lambda	NYNYQ	YNYQY	NYQYRL	YQYRLF	QYRLFR	2/5AA, 3/6AA
478	YQAGST	AGSTPC	STPCN				1/5AA, 2/6AA	T478K	Delta	YQAGSK	QAGSKP	AGSKPC	GSKPCN	KPCNG	1/5AA, 4/6AA
484	CNGVEG	NGVEGF	GVEGFN				3/6AA	E484K	Alpha, Beta, Gamma, Eta, Mu	CNGVKG	NGVKGF	GVKGFN	KGFNC		1/5AA, 3/6AA
								E484Q	Kappa	NGVQG	VQGFN	QGFNC			3/5AA
490	FNCYF	CYFPLQ	YFPLQS	FPLQSY			1/5AA, 3/6AA	F490S	Lambda	FNCYS	NCYSP	CYSPLQ	SPLQSY		2/5AA, 2/6AA
494	YFPLQS	FPLQSY	PLQSYG	QSYGF	SYGFQP		1/5AA, 4/6AA	S494P	Alpha	PLQP	LQPY	QPYG	PYGF		4/4AA
501	GFQPTN	FQPTNG	QPTNGV	PTNGVG	TNGVGY	NGVGYQ	6/6AA	N501Y	Alpha, Beta, Gamma, Mu	GFQPTY	QPTYG	YGVGY			2/5AA, 1/6AA
516	VVLSFE	VLSFELL	LSFELLH	FELLHA	ELLHAP		2/7AA, 3/6AA	E516Q	Beta	VVVLSFQ	VLSFQL	SFQLLH	FQLLHA	QLLHAP	4/6AA, 1/7AA

Novel C/H-CrUPs created by critical mutations in the Receptor-Binding (RBD) domain of SARS-CoV-2 wild-type and mutant Spike protein (SPIKE_SARS2, P0DTC2) amino acid sequence are identified. Peptide number/peptide length is the number of a given length C/H–CrUP around the position. By red color the amino acids in wild-type C/H-CrUPs, which will be modified, and the mutated amino acids in the new C/H-CrUPs are marked. Light blue color indicates the peptides which disappear from the wild-type viral proteome by the mutation, yellow color shows the completely new created C/H-CrUPs peptides by the mutation.

Table 8. NF9-specific C/H–CrUPS.

	PEPTIDES		
		MUTATION	
POSITION	SARS-CoV-2	L452R	L452Q
448	NYNYLY		NYNYQ
449		YNYRY	YNYQY
450	NYLYRL	NYRYRL	NYQYRL
451	YLYRLF	YRYRLF	YQYRLF

The C/H-CrUPs in wild-type and mutant NF9 peptide are listed. By red color the mutant amino acids are marked.

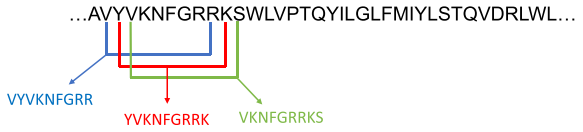


Figure 7. Presentation of the bioinformatic process developed for the identification of the CrUPs peptides, performed amino acid by amino acid residue.

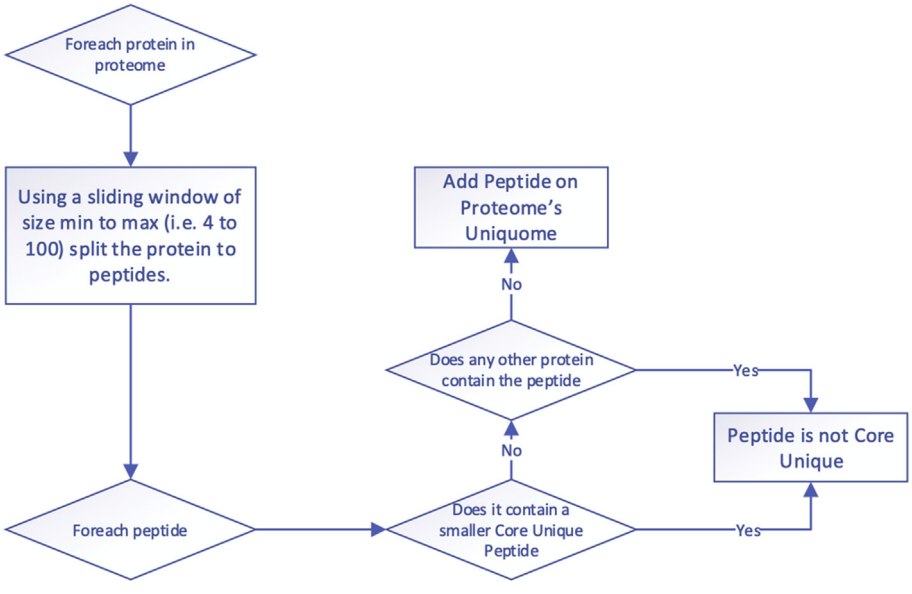


Figure 6. Schematic presentation of the algorithm herein developed for the identification of Core Unique Peptides (CrUPs).

- ?QYIL?
- T??ILG
- T?Y?LG
- T?YI?G
- T?YIL?
- TQ??LG
- TQ?I?G
- TQ?IL?
- TQY??G
- TQY?L?
- TQYI??

User will receive a list of all proteins containing peptides that match the criteria, including the motif against which the peptide was matched, and all the positions within the protein sequence where that peptide can be found. All proteomes were taken from the UNIPROT database.

4.3. Algorithm's application to the identification of virus CrUPs against human proteome

To recognize all the CrUPs being embraced in a virus proteome against the human proteome, we *in silico* constructed a new, artificial, "hybrid-proteome" that contained all the reviewed human proteins (20.430 proteins), plus the one protein derived from the viral proteome (20.431 proteins). Thereby, n "hybrid proteomes", including the n viral proteins, were constructed, with n representing the number of viral proteins. Hence, these "hybrid proteomes" were bioinformatically searched one by one for the identification of virus-specific CrUPs in human protein sequence environments.

4.4. Databases

All proteomes and proteins were obtained from UNIPROT [https://www.uniprot.org]. SARS-CoV-2 wild-type and variant/mutated sequences derived from Stanford COVID database [https://covdb.stanford.edu/page/mutation-viewer/]. Motifs were taken from the Eukaryotic Linear Motif resource for Functional Sites in Proteins [http://elm.eu.org/index.html] and KEGG/GenomeNet/MOTIF2 [https://www.genome.jp/tools/motif/MOTIF2.html]. SLiM-containing proteins were taken from Davey lab SLiM servers (The Institute of Cancer Research {ICR}, UK) [http://slim.icr.ac.uk/slimsearch/] and [http://slim.icr.ac.uk/index.php?page=tools].

Declarations

Author contribution statement

Vasileios Pierros: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data.

Evangelos Kontopodis: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Dimitrios J. Stravopodis, George Th. Tsangaris: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

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

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Agrawal, A., Varshney, R., Pathak, M., Patel, S.K., Rai, V., Sulabh, S., Gupta, R., Solanki, K.S., Varshney, R., Nimmanapalli, R., 2021. Exploration of antigenic determinants in spike glycoprotein of SARS-CoV2 and identification of five salient potential epitopes. Virusdisease 1–10.

Alexandridou, A., Tsangaris, G. Th., Vougas, K., Nikita, K., Spyrou, G., 2009. UniMaP: finding unique mass and peptide signatures in the human proteome. Bioinformatics 25, 3035–3037.

Almehdi, A.M., Khoder, G., Alchakee, A.S., Alsayyid, A.T., Sarg, N.H., Soliman, S.S.M., 2021. SARS-CoV-2 spike protein: pathogenesis, vaccines, and potential therapies. Infection 49, 855–876.

Braun, E., Sauter, D., 2019. Furin-mediated protein processing in infectious diseases and cancer. Clin. Transl. Immunol. 8, e1073.

Callaway, E., 2021. The mutation that helps Delta spread like wildfire. Nature 596, 472–473.

Chen, Y., Zhang, Y.N., Yan, R., Wang, G., Zhang, Y., Zhang, Z.R., Li, Y., Ou, J., Chu, W., Liang, Z., Wang, Y., Chen, Y.L., Chen, G., Wang, Q., Zhou, Q., Zhang, B., Wang, C., 2021. ACE2-targeting monoclonal antibody as potent and broad-spectrum coronavirus blocker. Signal Transduct. Targeted Ther. 6 (1), 315.

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G., Decroly, E., 2020. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antivir. Res. 176, 104742.

Davey, N.E., Cyert, M.S., Moses, A.M., 2015. Short linear motifs - ex nihilo evolution of protein regulation. Cell Commun. Signal. 13, 43.

Davidson, A.D., Williamson, M.K., Lewis, S., Shoemark, D., Carroll, M.W., Heesom, K.J., Zambon, M., Ellis, J., Lewis, P.A., Hiscox, J.A., Matthews, D.A., 2020. Characterization of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. Genome Med. 12, 68.

Hastie, K.M., Li, H., Bedinger, D., Schendel, S.L., Dennison, S.M., Li, K., Rayaprolu, V., Yu, X., Mann, C., Zandonatti, M., et al., 2021. Defining variant-resistant epitopes targeted by SARS-CoV-2 antibodies: a global consortium study. Science eabh2315.

Hatmal, M.M., Alshaer, W., Al-Hatamleh, M.A.I., Hatmal, M., Smadi, O., Taha, M.O., Oweida, A.J., Boer, J.C., Mohamud, R., Plebanski, M., 2020. Comprehensive structural and molecular comparison of spike proteins of SARS-CoV-2, SARS-CoV and MERS-CoV, and their interactions with ACE2. Cells 9, 2638.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al., 2020a. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181, 271–280.

Hoffmann, M., Kleine-Weber, H., Pöhlmann, S., 2020b. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol. Cell 78, 779–784.

Hraber, P., O'Maille, P.E., Silberfarb, A., Davis-Anderson, K., Generous, N., McMahon, B.H., Fair, J.M., 2020. Resources to discover and use short linear motifs in viral proteins. Trends Biotechnol. 38, 113–127.

Jungreis, I., Sealfon, R., Kellis, M., 2021. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. Nat. Commun. 12, 2642.

Kared, H., Redd, A.D., Bloch, E.M., Bonny, T.S., Sumatoh, H., Kairi, F., Carbajo, D., Abel, B., Newell, E.W., Bettinotti, M.P., et al., 2021. SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-19 individuals. J. Clin. Invest. 131, e145476.

Kontopodis, E., Pierros, V., Anagnostopoulos, A., Stravopodis, D., Papassideri, I., Vorgias, C., Tsangaris, G.T., 2019. Data processing approach for the construction and evaluation of an organism's UNIQUOME with comparative analysis for the Human, Rat and Mouse Uniquomes. Paper presented at XIII. In: Annual Congress of the European Proteomics Association: from Genes via Proteins and Their Interactions to Functions, Potsdam, Germany. March 24-28, P194.

Kontopodis, E., Pierros, V., Stravopodis, D., Tsangaris, G.T., 2022. Prediction of SARS-CoV-2 Omicron variant immunogenicity, immune escape and pathogenity, through the analysis of spike protein-specific core unique peptides. Vaccines 10, 357.

Mlcochova, P., Kemp, S., Dhar, M.S., Papa, G., Meng, B., Ferreira, I., Datir, R., Collier, D.A., Albecka, A., Singh, S., et al., 2021. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. Nature 599, 114–119.

Motozono, C., Toyoda, M., Zahradnik, J., Saito, A., Nasser, H., Tan, T.S., Ngare, I., Kimura, I., Uriu, K., Kosugi, Y., Yue, Y., Shimizu, R., Ito, J., Torii, S., Yonekawa, A., Shimono, N., Nagasaki, Y., Minami, R., Toya, T., Sekiya, N., Sato, K., 2021. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. Cell Host Microbe 29 (7), 1124–1136.e11.

Örd, M., Faustova, I., Loog, M., 2020. The sequence at Spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. Sci. Rep. 10, 16944.

Papa, G., Mallery, D.L., Albecka, A., Welch, L.G., Cattin-Ortolá, J., Luptak, J., Paul, D., McMahon, H.T., Goodfellow, I.G., Carter, A., Munro, S., James, L.C., 2021. Furin

cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. PLoS Pathog. 17, e1009246.
Peacock, T.P., Goldhill, D.H., Zhou, J., Baillon, L., Frise, R., Swann, O.C., Kugathasan, R., Penn, R., Brown, J.C., Sanchez-David, R.Y., et al., 2021. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. Nat. Microbiol. 6, 899–909.
Sanches, P., Charlie-Silva, I., Braz, H., Bittar, C., Freitas Calmon, M., Rahal, P., Cilli, E.M., 2021. Recent advances in SARS-CoV-2 Spike protein and RBD mutations comparison between new variants Alpha (B.1.1.7, United Kingdom), Beta (B.1.351, South Africa), Gamma (P.1, Brazil) and Delta (B.1.617.2, India). J. Virus Erad. 7, 100054.
Saputri, D.S., Li, S., van Eerden, F.J., Rozewicki, J., Xu, Z., Ismanto, H.S., Davila, A., Teraguchi, S., Katoh, K., Standley, D.M., 2020. Flexible, functional, and familiar: characteristics of SARS-CoV-2 spike protein evolution. Front. Microbiol. 11, 2112.
Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., Li, F., 2020a. Cell entry mechanisms of SARS-CoV-2. Proc. Natl. Acad. Sci. U. S. A. 117, 11727–11734.
Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., Li, F., 2020b. Structural basis of receptor recognition by SARS-CoV-2. Nature 581, 221–224.
Shorthouse, D., Hall, B.A., 2021. SARS-CoV-2 variants are selecting for spike protein mutations that increase protein stability. J. Chem. Inf. Model. 61, 4152–4155.
Takeda, M., 2021. Proteolytic activation of SARS-CoV-2 spike protein. Microbiol. Immunol. 66.
Tzou, P.L., Tao, K., Nouhin, J., Rhee, S.-Y., Hu, B.D., Pai, S., Parkin, N., Shafer, R.W., 2020. Coronavirus antiviral research database (CoV-RDB): an online database designed to facilitate comparisons between candidate anti-coronavirus compounds. Viruses 12, 1006.
van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al., 2014. Classification of intrinsically disordered regions and proteins. Chem. Rev. 114, 6589–6631.
Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 181, 281–292.
Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K.Y., et al., 2020. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell 181, 894–904.e9.
Whittaker, G.R., 2021. SARS-CoV-2 spike and its adaptable furin cleavage site. Lancet Microbe 2, e488–e489.
Wu, Y., Zhao, S., 2020. Furin cleavage sites naturally occur in coronaviruses. Stem Cell Res. 50, 102115.
Wu, C., Zheng, M., Yang, Y., Gu, X., Yang, K., Li, M., Liu, Y., Zhang, Q., Zhang, P., Wang, Y., et al., 2020. Furin: a potential therapeutic target for COVID-19. iScience 23, 101642.
Xia, X., 2021. Domains and functions of spike protein in SARS-CoV-2 in the context of vaccine design. Viruses 13, 109.
Zahradník, J., Marciano, S., Shemesh, M., Zoler, E., Harari, D., Chiaravalli, J., Meyer, B., Rudich, Y., Li, C., Marton, I., Dym, O., et al., 2021. SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. Nat. Microbiol. 6, 1188–1198.