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

BBA - Proteins and Proteomics 1869 (2021) 140693

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

Research Paper

In silico analysis of the aggregation propensity of the SARS-CoV-2 proteome: Insight into possible cellular pathologies

Manuel Flores-León [a,b,1], Diana F. Lázaro [a,1], Liana Shvachiy [a,c,1], Anita Krisko [a,*], Tiago F. Outeiro [a,d,e,f,*]

[a] Department of Experimental Neurodegeneration, Center for Biostructural Imaging of Neurodegeneration, University Medical Center Göttingen, Göttingen 37073, Germany
[b] Departamento de Medicina Genómica y Toxicología Ambiental, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, AP 70-228, 04510, México, DF, Mexico
[c] Cardiovascular Autonomic Function Lab, Centro Cardiovascular da Universidade de Lisboa, Lisbon, Portugal
[d] Max Planck Institute for Experimental Medicine, Göttingen, Germany
[e] Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom
[f] Scientific Employee With a Honorary Contract at Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Göttingen, Germany.

ARTICLE INFO

Keywords:
SARS-CoV-2
Protein aggregation
Proteostasis
Bioinformatics

ABSTRACT

The SARS-CoV-2 virus causes the coronavirus disease 19 emerged in 2020. The pandemic triggered a turmoil in public health and is having a tremendous social and economic impact around the globe. Upon entry into host cells, the SARS-CoV-2 virus hijacks cellular machineries to produce and maintain its own proteins, spreading the infection. Although the disease is known for prominent respiratory symptoms, accumulating evidence is also demonstrating the involvement of the central nervous system, with possible mid- and long-term neurological consequences. In this study, we conducted a detailed bioinformatic analysis of the SARS-CoV-2 proteome aggregation propensity by using several complementary computational tools. Our study identified 10 aggregation prone proteins in the reference SARS-CoV-2 strain: the non-structural proteins Nsp4, Nsp6 and Nsp7 as well as ORF3a, ORF6, ORF7a, ORF7b, ORF10, CovE and CovM. By searching for the available mutants of each protein, we have found that most proteins are conserved, while ORF3a and ORF7b are variable and characterized by the occurrence of a large number of mutants with increased aggregation propensity. The geographical distribution of the mutants revealed interesting differences in the localization of aggregation-prone mutants of each protein. Aggregation-prone mutants of ORF7b were found in 7 European countries, whereas those of ORF3a in only 2. Aggregation-prone sequences of ORF7b, but not of ORF3a, were identified in Australia, India, Nepal, China, and Thailand. Our results are important for future analysis of a possible correlation between higher transmissibility and infection, as well as the presence of neurological symptoms with aggregation propensity of SARS-CoV-2 proteins.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a positive-sense single-stranded RNA virus from the *Coronaviridae* family [1] that causes the coronavirus disease 2019 (COVID-19). In January 2020, human-to-human transmission was confirmed, with respiratory symptoms being the most prominent effect. [2] Other COVID-19 symptoms were also reported and included shortness of breath, sore throat, sneezing, fever, exhaustion and, in some cases, gastrointestinal and neurological effects. [3] In fact, around 80% of the patients develop

Abbreviations: APR, aggregation-prone regions; A3D, Aggrescan3D; CoVex, Coronavirus Explorer; COVID-19, coronavirus disease 19; NSP, non-structural proteins; ORF, open reading frame; PASTA, prediction of amyloid structural aggregation; PDB, Protein Data Bank; SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2.

* Corresponding authors at: Department of Experimental Neurodegeneration, Center for Biostructural Imaging of Neurodegeneration, University Medical Center Göttingen, Göttingen 37073, Germany.
E-mail addresses: anita.krisko@med.uni-goettingen.de (A. Krisko), tiago.outeiro@med.uni-goettingen.de (T.F. Outeiro).
[1] Equal contribution.

https://doi.org/10.1016/j.bbapap.2021.140693
Received 25 May 2021; Received in revised form 29 June 2021; Accepted 30 June 2021
Available online 5 July 2021

Table 1
PDB codes from the available SARS-CoV-2 proteins. The chain used for the structural analyses of the crystallized structure is shown along with the complex that the main structure was crystallized with.

Code	Chain	Protein	Complex
6YB7	A	Main Protease (Active site)	None
6 W75	A	NSP16	NSP10
6 W75	B	NSP10	NSP16
7BV1	B	NSP8	NSP7 - NSP12
7BV1	C	NSP7	NSP8 - NSP12
6WLC	A	NSP15 (Endoribonuclease)	UMP
6WOJ	A	NSP3 (macro domain)	ADP ribose
6WXD	A	NSP9 (RNA replicase)	None
6WJI	A	Nucleocapsid phosphoprotein (C-terminal)	None
6YI3	A	Nucleocapsid phosphoprotein (N-Terminal)	None
6 W37	A	ORF7A	None
6W9C	A	Papain-like protease	None
6YYT	A	Polymerase	None
6VSB	A	Spike glycoprotein	Prefusion with a single receptor-binding domain
6XDC	A	ORF3a	None
7JTL	A	ORF8	None
6ZLW	i	Nsp1	Bound to human 40S ribosomal unit
6ZSL	A	Helicase	None

some type of neurological manifestation throughout the infection, ranging from anosmia (11% of the cases) to encephalopathy (up to 32% of the cases). [4–6] Therefore, alongside with other coronaviruses, SARS-CoV-2 is also considered a neurotropic virus. [7,8] Concerningly, several reports link the infection with cognitive impairment and even with neurodegeneration. [9,10] Therefore, the long-term consequences of the infection for the central nervous system need to be further investigated, and targets for therapeutic intervention need to be defined. [11,12]

The SARS-CoV-2 genome comprises an open reading frame (ORF) 1a/b encoding for non-structural proteins (NSP), as well as for structural proteins (spike, envelope, membrane, and nucleocapsid proteins), and for accessory proteins such as ORF 3, 6, 7a, 7b, 8 and 9b. [13,14] The NSPs form the replicase-transcriptase complex, which include two viral proteases, Nsp3 (papain-like protease) and Nsp5 (chymotrypsin-like, 3C-like), the primase complex (NSP7-NSP8), the primary RNA-dependent RNA polymerase (Nsp12), an helicase (Nsp13), an exoribonuclease (Nsp14), an endonuclease (Nsp15) and N7- and 2′*O*-methyltransferases (Nsp10 and Nsp16). [15,16] The spike protein is one of the essential proteins for the infection success of the virus since it mediates the virus-host cell surface attachment. Once inside the host cell, the replication cycle is a regulated process where viral proteins are expressed, folded, and assembled, allowing the production of new viral particles. [17,18] As the virus hijacks the host-cell protein homeostasis (proteostasis) network to ensure the folding, assembly, and release of new viral particles, the cell becomes overloaded with both viral proteins and also endogenous proteins. At some point, the accumulation of proteins may exceed the capacity of cellular quality control systems, resulting in protein misfolding and aggregation, which may, in turn, exacerbate cellular pathologies.

For a long time, protein aggregation was considered an unspecific process. However, it is now known that it relies on a combination of physicochemical parameters within protein sequences and structure. [19] Therefore, several experimental-based models enabled the creation of databases and bioinformatic tools and algorithms to predict and identify aggregation-prone regions (APR) in proteins. These models explore the concept that aggregation is enhanced by the presence of amino acids with higher hydrophobicity and β-sheet propensity, and with lower net charge. [19,20] Therefore, an important aspect to consider in the SARS-CoV-2 proteome is the aggregation-potential of the viral proteins, as this may dramatically change due to mutations.

Here, we employed computational APR-prediction tools to assess the aggregation propensity of the SARS-CoV-2 proteins. We found that 10 proteins can be classified as aggregation-prone. While some of these proteins display low or no variability across geographical regions, others accommodate a large number of mutant sequences. For all the analyzed proteins, we identified mutations that increase the aggregation-propensity of the protein. Interestingly, aggregation-prone variants of ORF3a and ORF7b seem to display specific geographic distribution.

In total, our study provides the foundation for future experimental studies focusing on the analysis of cellular pathologies associated with the aggregation of selected candidates of the SARS-CoV-2 proteome and may enable the identification of novel targets for therapeutic intervention.

2. Methods

2.1. Generation of the SARS-CoV-2 protein sequence set

A full proteome of SARS-CoV-2 virus isolated in Wuhan, China and released on January 17th, 2020 was obtained from the NCBI Virus

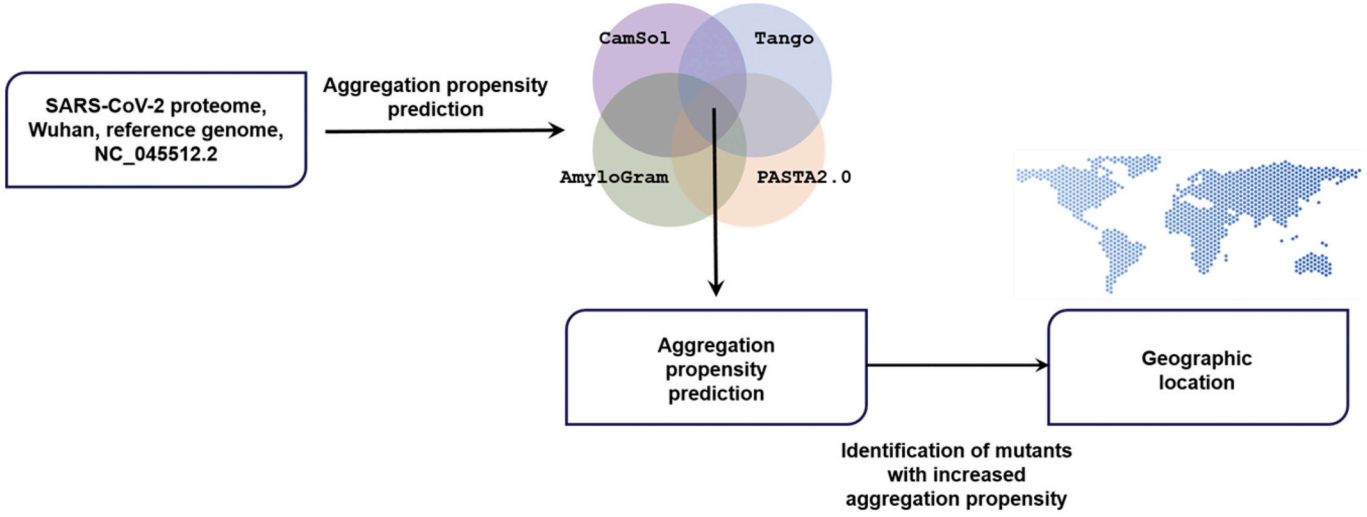

Fig. 1. Workflow of the study. Aggregation prone proteins were predicted in the SARS-CoV-2 reference strain. Among all mutant sequences available in August 2020, mutations increasing the aggregation propensity were identified and mapped according to their geographical location.

Table 2
Full set of values/scores of aggregation propensity predictions from all SARS-CoV-2 proteins using 4 different algorithms.

Protein	CamSol Intrinsic solubility	Tango (beta aggregation propensity)	Amylogram	PASTA (Pairing energy, kJ)	Prone to aggregate
BSA	0.479889	6.513342	0.67476910427	−9.230629	Negative control
Beta-Amyloid	1.103345	22.4587	0.964286175387	−20.171292	Positive control
just Aβ fragment	0.655945	3.635395	0.5936772918	−6.888103	Positive control
3′-to-5′ exonuclease	−1.365414	3.540996	0.780836059694905	−10.987355	No
2-O-Ribose Methyltransferase	−0.503960	7.303587	0.917586033222489	−7.926963	No
3C Like Proteinase	−1.094241	5.254686	0.866995916960981	−12.714527	No
endoRNAse	−0.225170	5.126858	0.889817000611829	−11.885505	No
Envelope protein	−2.915438	25.39225	0.914774818	−29.181144	Yes
Helicase	−1.206905	2.752962	0.885824647794826	−8.055405	No
Leader protein	1.394877	0.2026	0.674893206737557	−5.359897	No
Membrane glycoprotein	−2.887348	26.75686	0.862681744	−34.513196	Yes
Nsp2	−0.623592	4.99456	0.810156395859043	−12.65233	No
Nsp3	−3.745792	5.9662	0.888048791900172	−8.066283	No
Nsp4	−4.727092	19.00763	0.855524114	−32.389486	Yes
Nsp6	−6.013864	36.30086	0.887343189	−32.815411	Yes
Nsp7	0.231983	17.76802	0.889817001	−7.767171	Yes
Nsp8	0.816455	2.594571	0.672595099068358	−6.916269	No
Nsp9	1.063820	7.472212	0.691380284008589	−5.15149	No
Nsp10	−0.040202		0.665628490992486	−5.345617	No
Nsp11	0.964712	0.071538	0.52229465066422	−2.681862	No
Nucleocapsid phosphoprotein	2.216086	4.942186	0.850755764881348	−6.850497	No
ORF3a	−3.400784	19.49843	0.888048792	−26.32989	Yes
ORF6	−0.804210	23.35634	0.866119264	−21.15957	Yes
ORF7a	−1.501395	22.73566	0.816087315	−16.067259	Yes
ORF7b	−3.341397		0.858967046	−21.061641	Yes
ORF8	−1.072771	10.9687	0.83488134871516	−10.071886	No
ORF10	−1.501395	18.52229	0.804993185	−10.10498	Yes
RNA-dependent RNA polymerase	−1.702360		0.6386355287	−9.090557	No
Surface Glycoprotein	−3.379700	26.76	0.862681743572924	−16.434159	No

database. [15,21–23] These protein sequences were set as reference and were used for the prediction of aggregation propensity. As positive controls for aggregation-prone proteins, human aSyn and Aβ peptide 1–42 were used. Bovine serum albumin was used as a negative control as it is a soluble globular protein. Only complete sequences were included for each protein.

2.2. Prediction of aggregation propensity of individual SARS-CoV-2 proteins

The aggregation propensity of SARS-CoV-2 proteins was predicted using four different methods. The cutoffs for each method were determined based on the results obtained for the positive (strongly aggregation prone proteins) and negative (soluble globular protein) controls used in the study.

2.2.1. PASTA 2.0 (prediction of amyloid structural aggregation)

The full proteomes of SARS-CoV-2, with the controls, were analyzed using the PASTA 2.0 server. [24] As an output, the method provides the prediction of the aggregation prone proteins and protein fragments, as well as the analysis of protein disorder and secondary structure. For further analysis, we focused solely on the prediction of the aggregation-prone proteins, displayed as aggregation pairing energy, whereby the lowest aggregation pairing energy signifies the highest aggregation propensity. All proteins smaller than 50 residues with the aggregation pairing energy below -7 kJ, and proteins larger than 1000 residues with the aggregation pairing energy below −20 kJ were considered aggregation prone and were included into further analysis. For proteins of sizes between 100 and 1000 residues, the limit of aggregation pairing energy was −15 kJ.

2.2.2. AmyloGram

Briefly, this approach predicts amyloid proteins using random forests algorithm trained on the n-gram encoded peptides. [25] Using R package, [26] the set of SARS-CoV-2 proteins was analyzed by AmyloGram and all proteins with the score above 0.85 were considered aggregation-prone and included in further analysis.

2.2.3. CamSol

CamSol is a free online software that predicts the solubility of proteins. [27] The use of this method requires either the knowledge of the native structure of the target protein to calculate the structure-corrected solubility profile or the intrinsic solubility profile can be predicted based on the amino acid sequence.

The profiles consist of a score for each residue and it characterizes their impact on the overall solubility of the protein. This sequence-based method predicts protein solubility and generic aggregation propensity. The lower the score is, the lower the solubility of the protein, i.e. the higher the aggregation propensity. Proteins with the intrinsic solubility lower than −1.5 (a.u.) were deemed aggregation prone.

2.2.4. Tango

TANGO is a statistical mechanics algorithm that predicts protein aggregation based on the physico-chemical principles of β-sheet formation. [19,28,29] It predicts secondary structure content of analyzed proteins. A high propensity to form β-sheet structure and the burial of the residues of the β-region in the hydrophobic interior means that a particular amino acid sequence is aggregation-prone. The prediction of aggregation propensity is done in a sequence specific manner.

SARS-CoV-2 full protein sequences were imported to TANGO in FASTA format. For each aminoacid of each analyzed protein, the percentage of beta turn, beta strand, beta sheet and alpha helix aggregation were obtained and exported in a .csv file. Subsequently, for each protein the mean beta sheet aggregation propensity was calculated by normalization to the length of the protein. Any analyzed protein with an aggregation tendency above 10% over 5–6 residues was assigned as a potential aggregation prone protein. The graphs for each protein were generated and the data compared with controls. All the acquired data was statistically analyzed by GraphPad Prism 8 software.

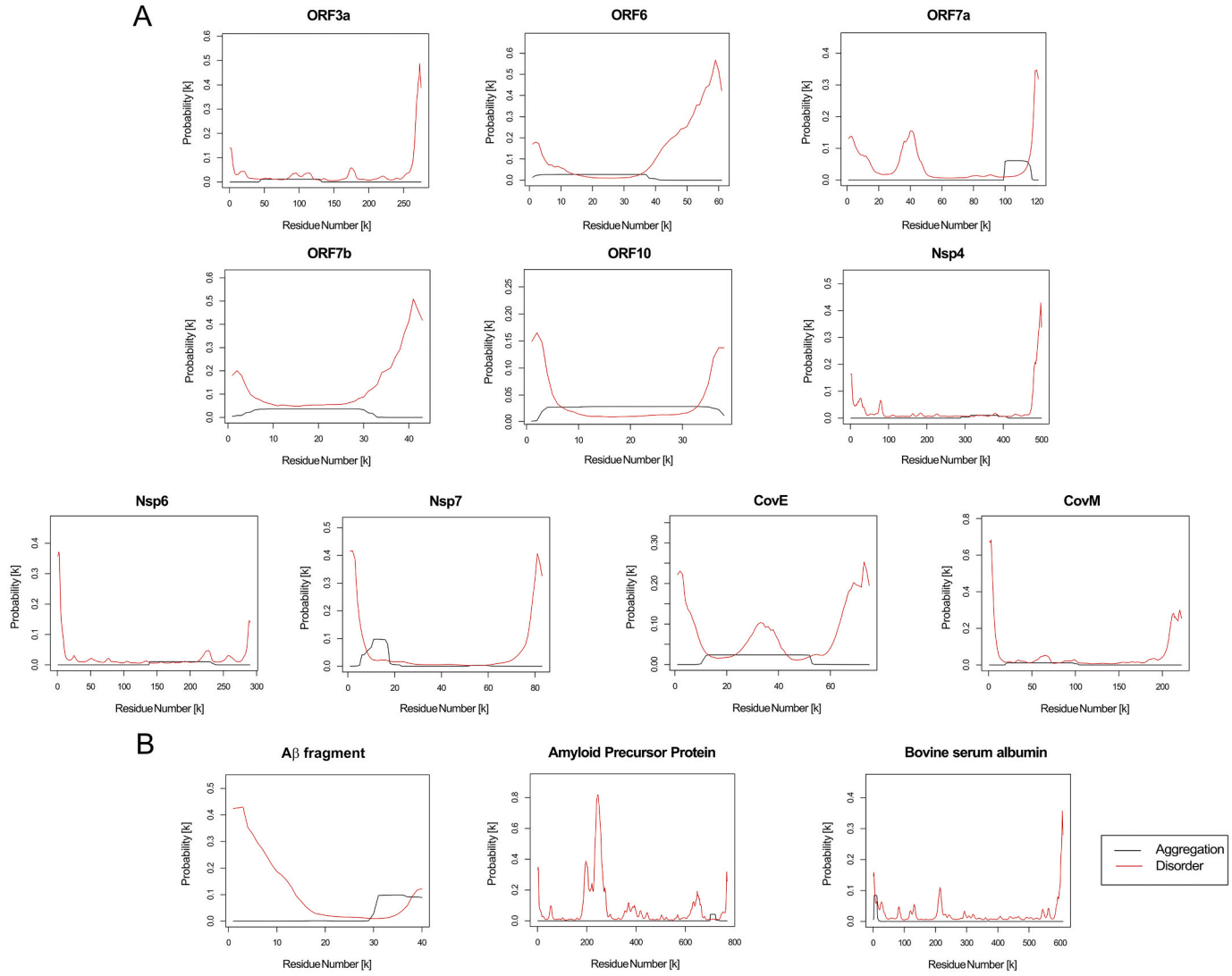

Fig. 2. Aggregation probability calculated using PASTA2.0. The graphs displays the probability of each amino acid to be part of a prone to aggregate region (black curve) and to be a part of disordered regions for: a) each SARS-CoV-2 aggregation-prone protein; b) positive and negative controls proteins used in the study.

2.3. Selection criteria

Proteins predicted as likely to aggregate with at least three of the methods applied were employed in further analysis. A second set of sequences was generated with a non-redundant list of all mutants of each SARS-CoV-2 protein from all available geographic locations until August 23rd, 2020 (NCBI Virus database). Starting with this set of sequences, from each occurring mutant, only a unique representative was included in the further analysis. The aggregation propensity was then computed with the methods mentioned above for the representative sequences of each mutant. All mutations that increased the aggregation propensity were included in further analysis.

To evaluate if a particular mutation increased the aggregation propensity of a protein variant, we established the following procedure: the output score of each method was normalized to the value of the same protein from the reference strain by dividing the values of each studied protein with the value of the same protein from the reference strain. Next, the log2 of the obtained number was calculated and all protein variants with log2 > 0 had increased aggregation propensity. A sum, mean and median of this score were calculated and each was used to perform the hierarchical clustering of sequences. All the protein variants with sum, mean and median larger than 0 were selected as those where the mutation triggered an increase in aggregation propensity. Finally, all geographical locations with mutations that increased the protein aggregation propensity were marked on maps using R software, package 'maps'.

2.4. SARS-CoV-2 protein structural analysis

SARS-CoV-2 protein structures were obtained from the Protein Data Bank (PDB). [30] The curation of the redundant structures was performed taking into account the following parameters: when possible, monomers and/or non-conjugated or bound proteins were used, crystalized proteins with the highest resolution were included regardless of the method used as well as the structure with the most complete amino-acid sequence. More detailed characteristics are summarized in Table 1.

2.5. Aggrescan3D (A3D)

The Aggrescan3D 2.0 server was used to evaluate the aggregation propensity of the proteins according to their tertiary structure. [31] PDB codes (from the curation process) were submitted from the available crystalized proteins. The chain identifier used for each PDB entry is listed in Table 1.

Additional parameters that were selected as part of the analysis: Stability calculations - on, Dynamic mode - on, and a distance of

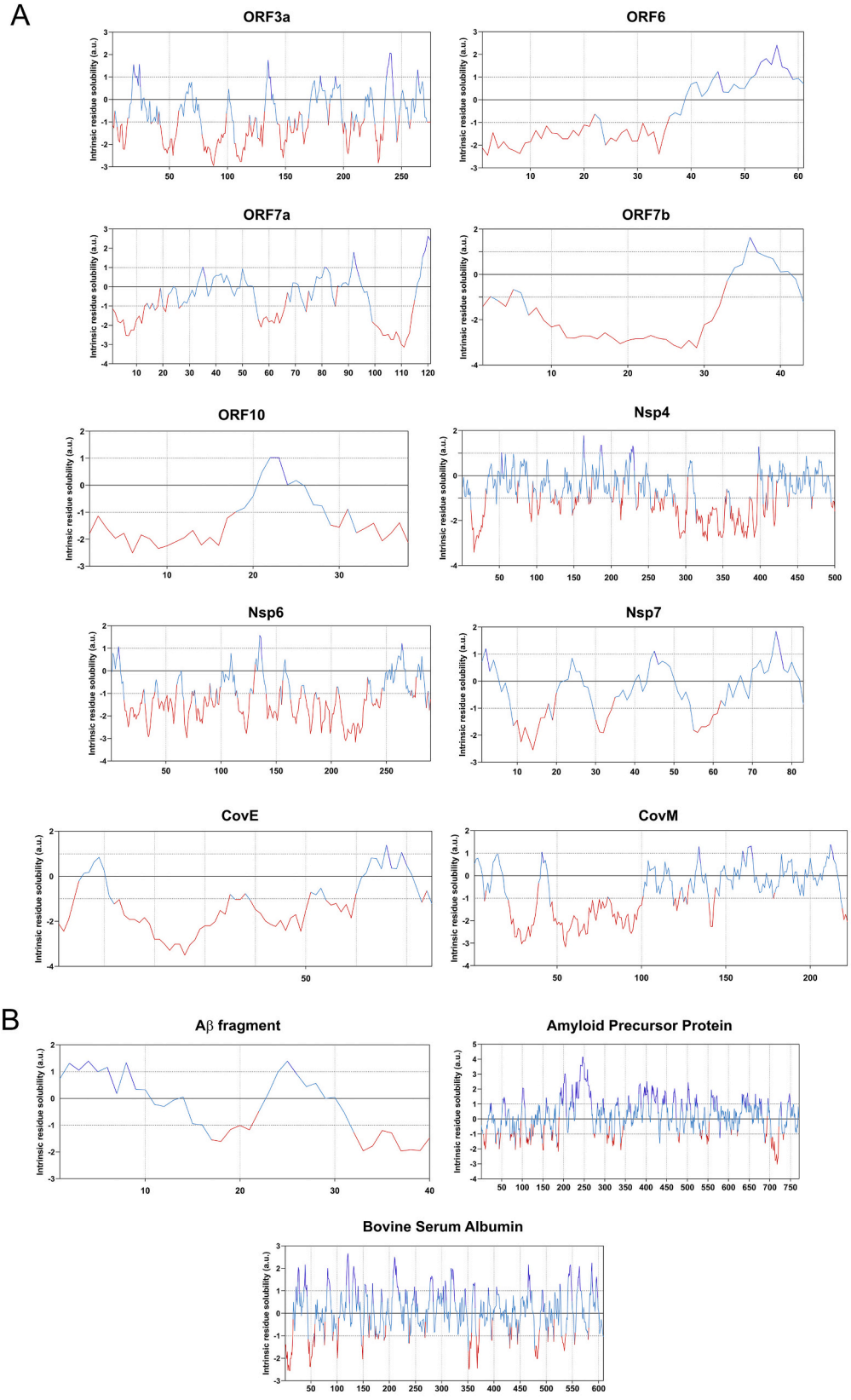

Fig. 3. CamSol based analyses of protein solubility. Each graph displays the sequence based solubility profile for each query protein. The part of the curve corresponding to the most soluble protein fragments are shown in blue, and the regions of low solubility are shown in red for: a) SARS-CoV-2 aggregation prone protein solubility graphs; and b) positive and negative control proteins used in the study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

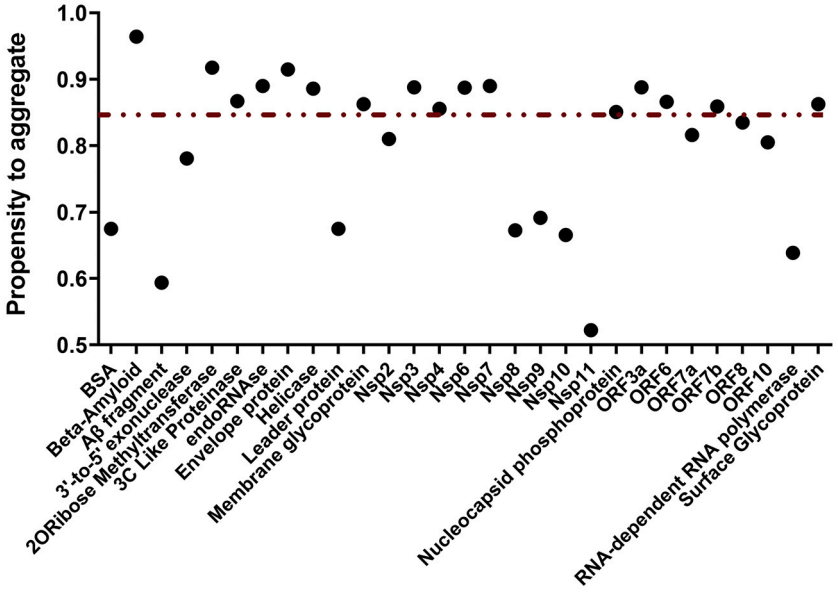

Fig. 4. Amylogram based calculations of aggregation propensity. The complete set of the SARS-CoV-2 proteome and the positive and negative control proteins are displayed. The non-continuous line marks the threshold (>0.85) considered as "high propensity to aggregate". The beta-amyloid peptide has the highest value, which relates to its propensity to aggregate.

Aggregation of 10 Å. Once the analysis was finalized, the .csv file that had the score of each amino acid was downloaded and processed in Microsoft Excel®. Using GraphPad Prism 8® the aggregation propensity graph was obtained, and it was analyzed to obtain the Area under the curve. Amino acids below an Aggregation score of 1 were excluded as well as peaks that contained less than 2 continuous amino-acids prone to aggregate.

To evaluate the mutant forms reported in the literature, the same tool was used to mutate the residue in each variant that the PDB sequence allowed us to work with. The same parameters and exclusion criteria were used as in the analysis of the wild type variant.

2.6. Coronavirus explorer (CoVex)

The Coronavirus Explorer Online tool was used to generate a string analysis of the human proteins that could interact with the SARS-CoV-2 proteins. [32,33] To generate the specific network, the desired SARS-CoV-2 protein was selected in the "Filter Viral Protein" section and "Brain - Substantia nigra" was selected in the "Tissue" section.

3. Results

3.1. Aggregation propensity of the reference SARS-CoV-2 proteome

In order to analyze the aggregation propensity of the full proteome of the SARS-CoV-2 virus, the protein sequences of the Wuhan reference strain were analyzed using four different bioinformatic tools: Pasta2.0, CamSol, Amylogram and Tango (Fig. 1). The proteins were then sorted according to their likelihood to aggregate (as likely or unlikely). At a false positive rate of <5%, Pasta2.0 has a sensitivity of 40%, and at a false positive rate of <10%, it has a sensitivity of 30% for predicting APRs, making it the most sensitive tool compared to the others. Therefore, to minimize false APR predictions, the results obtained with Pasta2.0 were selected as our reference predictions.

To further strengthen our aggregation propensity prediction, we then analyzed the selected proteins using the other 3 algorithms (Table 2). If the protein was predicted as aggregation-prone using at least 2 other algorithms, we classified it as aggregation-prone. The analyses showed that 10 proteins from the SARS-CoV-2 proteome were prone to aggregate. The non-structural proteins Nsp4, Nsp6 and Nsp7 as well as ORF3a, ORF6, ORF7a, ORF7b, ORF10, CovE and CovM were predicted as aggregation-prone. Thus, we continued our study with these 10 selected proteins.

Using PASTA, we obtained the probability of the individual residues of each studied protein sequence to aggregate (black line) and to be disordered (red line), and found that proteins predicted to be prone to aggregate typically had APRs between 30 and 100 residues long (Fig. 2). Interestingly, these regions were also recognized as regions of low disorder. On the other hand, CamSol informed on the intrinsic residue solubility for each residue in the analyzed sequences. According to this parameter, the aggregation prone proteins were characterized according to large domains with low solubility (Fig. 3).

Using Amylogram we obtained the propensity for aggregation into amyloid structure of each protein from the reference SARS-CoV-2 strain. The results showed a wide range of aggregation propensities. However, only the proteins characterized by a score higher than 0.85 were considered prone to aggregate into amyloid structure (Fig. 4). Furthermore, TANGO provided an average aggregation score per residue for every analyzed sequence, and the output displayed the likelihood of each amino acid to adopt certain conformation. In our case, we focused on aggregation into amyloid-like structure (Fig. 5).

3.2. Effect of mutations on the aggregation of selected SARS-CoV-2 proteins

Since the sequence of the reference strain was reported, numerous mutants of SARS-CoV-2 virus were reported during 2020, from different geographical locations. Therefore, we collected and analyzed the mutant sequences for each protein, reported until August 2020, for their likelihood to aggregate. First, we curated the set of unique mutant sequences available per protein and kept only unique mutant sequences to avoid redundancy (Table 3). This showed that some proteins, like ORF7b or ORF10, appeared to be rather conserved, while others, like ORF3a or Nsp4, displayed a larger variability in their sequence, with numerous mutants occurring in different geographical regions. Next, using a set of unique sequences for each protein, we predicted the aggregation propensity for each mutant sequence using all 4 algorithms employed in our pipeline (full set of results are available in supplementary data, appendix

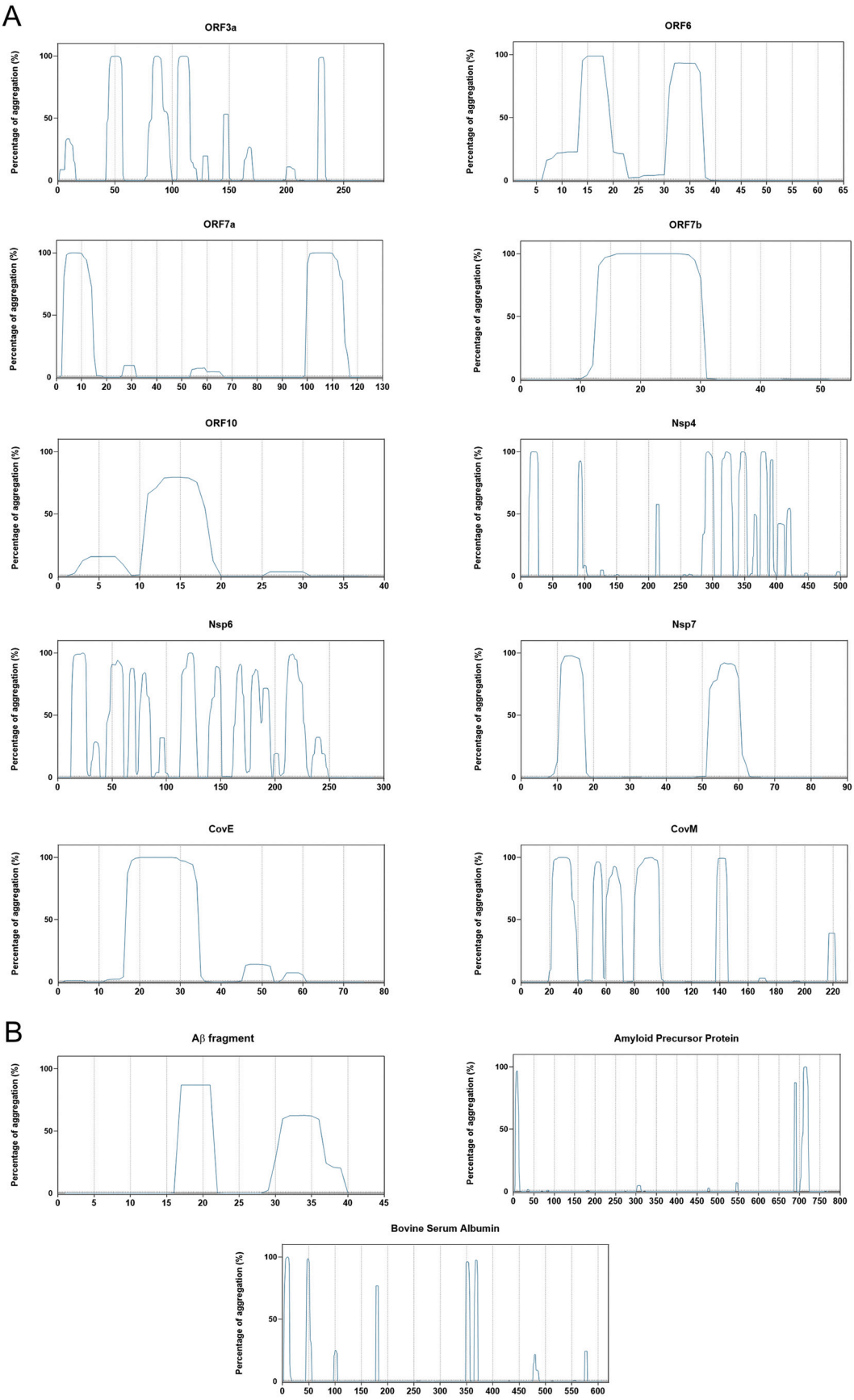

Fig. 5. Aggregation propensity calculated using TANGO. Graphs display the percentage of aggregation, signifying the probability of aggregation throughout the queried amino acid sequence for: a) SARS-CoV-2 proteins characterized by high aggregation propensity; b) positive and negative control proteins used in the study.

Table 3
Number of unique mutant sequences reported per protein. Proteins shown are the ones analyzed according to the established aggregation-prone criteria.

	ORF3a	ORF6	ORF7a	ORF7b	ORF10	Nsp4	Nsp6	Nsp7	E protein	M protein
Total number of unique sequences	102	21	32	6	12	52	47	24	12	25

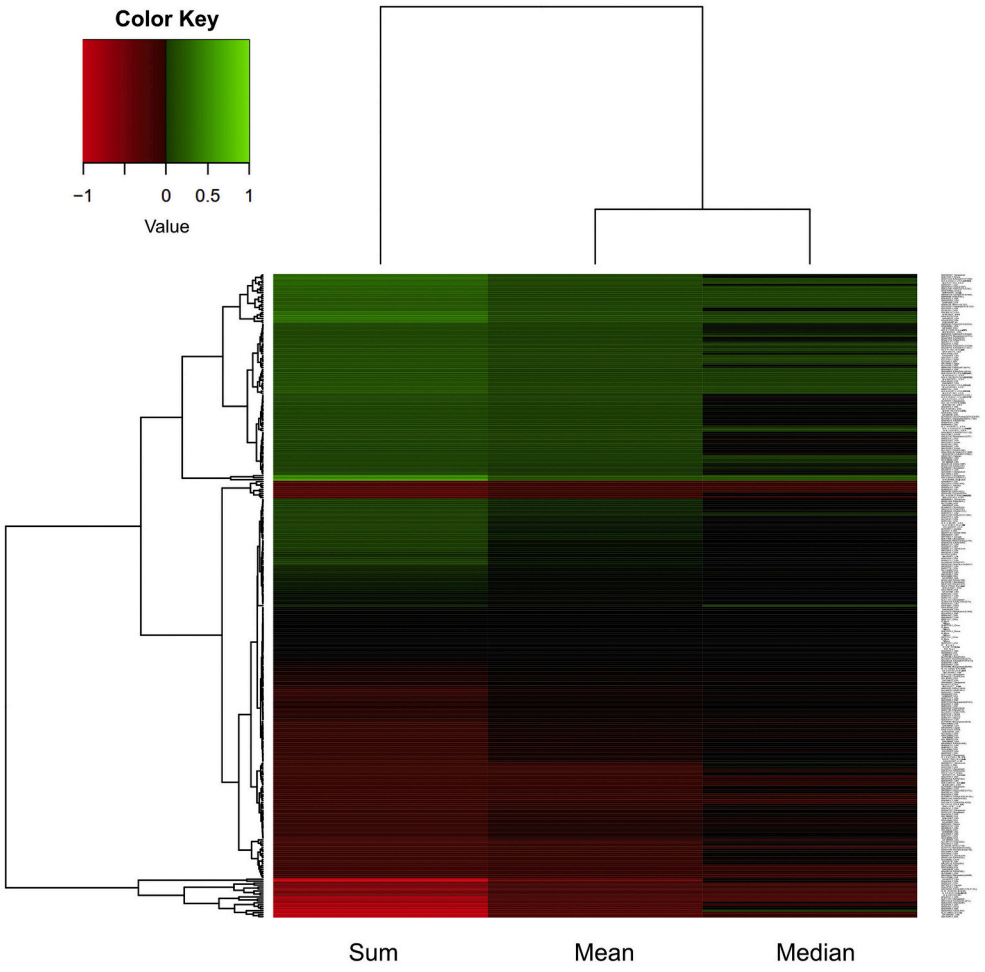

Fig. 6. Hierarchical clustering heatmap of the sequence variants from different geographical locations based on aggregation propensity. The sum, mean and median value of the log2 of their normalized aggregation score was used to evaluate their aggregation propensity and to cluster them. Green labels an increase and red a decrease in aggregation propensity.

A). In most cases, single point mutations did not alter the protein's susceptibility to aggregate. However, ~40% of the mutant sequences displayed an increased aggregation proneness score in at least one of the algorithms employed.

In order to reduce the possibility of false positive predictions, our selection criteria was to choose mutants with increased aggregation propensity when compared to the reference sequence. First, the output score of each method used for the predictions of the likelihood of the protein to aggregate was normalized to the score of the reference strain. Then, the log2 of each value was calculated and this number was termed AggreScore. AggreScore equaled to 0 for the variants with unchanged aggregation propensity compared to the reference strain. All sequences with the AggreScore greater than 0 were considered to be prone to aggregate. Next, we calculated the sum, mean and median of AggreScores for each sequence and performed hierarchical clustering (Fig. 6). The sum, mean and median were used to integrate the information in the AggreScore obtained by the four employed methods and build selection criteria stringent enough to avoid false positives. Therefore, we selected all sequences with sum, mean and median greater than 0 for subsequent analyses.

3.3. Geographical distribution of aggregation-prone variants of selected SARS-CoV-2 proteins

To assess a possible correlation between geographical distribution and SARS-CoV-2 mutations, we retrieved the available sequences of each mutant and plotted on the world map according to the countries where they emerged. 3948 sequences were available for ORF7b, and 1981 sequences for ORF3a. A significantly lower number of sequences was available for the other studied proteins (supplementary data, appendix A) and, therefore, we could not proceed with the analysis of their geographical distribution. The geographical distribution of aggregation-prone mutants of ORF7b and ORF3a is presented in Fig. 7. For ORF7b (Fig. 7a) and ORF3a (Fig. 7b), over 80% of mutant sequences were found in the USA followed by Bangladesh, Egypt and Saudi Arabia. Interestingly, we detected several differences among less abundant mutant sequences: aggregation-prone mutants of ORF7b appeared in 7 different European countries, whereas those of ORF3a appeared in only 2.

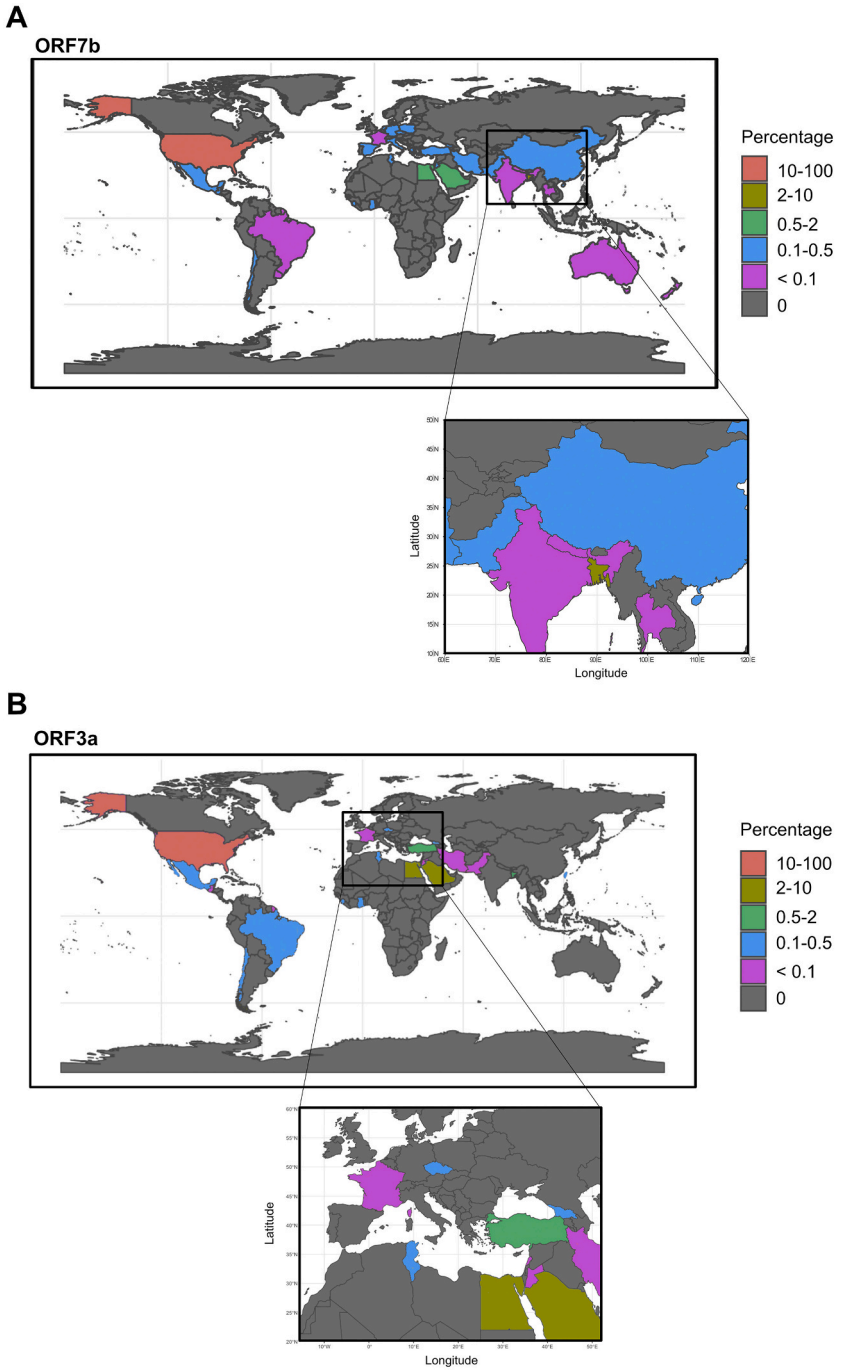

Fig. 7. Geographical distribution of variants of (A) ORF7b and (B) ORF3a characterized by increased aggregation propensity. The location of the SARS-CoV-2 variants are shown as a percentage marked by different colors (key, right side) in each reported country.

Moreover, aggregation-prone sequences of ORF7b, but not of ORF3a, were also identified in Australia. Aggregation-prone variants of ORF3a were absent in India, Nepal, China, and Thailand, regions where aggregation-prone ORF7b were detected. ORF3a variants were, however, present in Taiwan.

3.4. Structural alterations in ORF3a mutants and possible biological effects

To assess the effects of mutations at the level of the tertiary structure of the protein, we combined data available in the PDB and the use of the online tool A3D to visualize the protein topology, APR locations, and how the mutations affected them. For this, we used two criteria: (i) that the PDB data was available and (ii) a high ΔG difference between the variants and the reference sequence (wild type) (predicted by A3D) (Table 4). PDB data was only available for ORF3a. From the 17 variants predicted by A3D to affect the structure, we selected the top 5 for subsequent analysis: USA (Q57H, Y160H), Bangladesh (S171L), USA (G172C), Egypt (V55F, S171L), and USA (L94P, V97A, F120L). ORF3a is a 275 amino acid transmembrane protein (Fig. 8a) with 5 helixes, 8 β-strands and 3 transmembrane domains (Fig. 8a & b). Most of the mutations found in these variants were located in structural domains of the ORF3a protein (Fig. 8a) and, therefore, the structural destabilization might be due to an increase in the aggregation score of the APR. To evaluate this, quantification of the area under the curve for the wild type and the variants with higher ΔG was performed (Fig. 8c) and Table 4).

Table 4
ORF3a mutants compared to the wild type. The percentage of the peak area of the APRs is shown. The ΔG difference is associated with the destabilization of the protein's structure. Underlined ΔG show the analyzed variants.

ORF3a Variant	% Peak Area	Energy Difference (kcal/mol)
Wild type	65.91	0.0
QNG41554.1 USA (L53F)	72.24	2.2044
QNG41770.1 USA (L140F)	66.80	2.1808
QND77223.1 USA (Q57H, P104S)	67.97	2.4939
QNA38024.1 USA (Q57H, Y160H)	69.82	5.0052
QNA38792.1 USA (Q57H, I123V)	70.09	1.5448
QMU94777.1 Bangladesh (S171L)	69.53	11.2476
QMI98320.1 USA (Q57H, Q185H)	69.55	1.3686
QLM05656.1 USA (G172C)	68.89	4.3654
QLM05764.1 USA (L53H, Q57H)	67.44	2.9557
QLH56255.1 Saudi Arabia (Q57H, S216P)	66.35	3.3689
QKS66941.1 Egypt (V55F, S171L)	69.13	9.9923
QKN20824.1 USA (Q57H, A59V)	65.81	1.0337
QKG90399.1 USA (Q57H, R134C)	71.74	1.5122
QKG64052.1 USA (F56C)	71.57	1.2334
QKE44990.1 USA (L94P, V97A, F120L)	67.98	5.4940
QJD47849.1 Taiwan (Q57H, M125I)	74.34	1.2187
QJD23418.1 USA (Q57H, A99D)	67.94	1.0058

This curve showed an increased aggregation score for some amino acids (black arrows) when compared to the wild type protein. Nevertheless, the % of the peak area did not change drastically (Table 4). The 3D structural analysis of the APR showed that the regions that harbor the mutations did not change their propensity to aggregate (Fig. 8d), yellow circles), but that the greater change was in nearby regions. These regions (Fig. 8c) corresponded to the marked APR (Fig. 8d, black circles) in the protein structure. Interestingly, two of the variants with the highest ΔG (Egypt and Bangladesh) had a higher APR score on two different regions of the protein. This might be related to the predicted destabilization of the protein structure.

Finally, in order to address the possible impact in the host cell biological functions, we used Coronavirus Explorer (CoVex) in order to identify human proteins that could possibly interact with ORF3a. This interaction network identified 8 proteins that can potentially interact with ORF3a (Fig. 8e). Among the predicted proteins, four (VPS39, VPS11, HMOX1 and CLCC1) have been associated with inflammatory processes, endoplasmic reticulum stress, protein misfolding/clearance and autophagy. [34–37]

4. Discussion

Cellular proteostasis is maintained through a well-orchestrated network that ensures proteins are produced, folded, maintained, and degraded according to quality control standards that enable proper cellular function. External stimuli, such as temperature or chemical stress, mutations, or age-associated alterations can affect the activity of the various surveillance components of the proteostasis network, leading to protein misfolding and accumulation, and eventually to cellular pathologies.

Viruses are infectious agents that hijack different cellular components to produce and process their own proteins, often in high amounts, thereby affecting the overall cellular proteostasis. This can lead to the aggregation of endogenous cellular proteins, or to the aggregation of viral proteins, creating a vicious cycle that can further disrupt the proteostasis network and cause disease. [38]

In the context of the current SARS-CoV-2 pandemic, we reasoned that it would be important to assess whether certain viral proteins might be more aggregation-prone, thereby compromising the proteostasis of the cells they infect. In our study, we took advantage of several available algorithms (TANGO, CamSol and AmyloGram) to predict/assess the aggregation potential of the whole reference viral-proteome from the original Wuhan strain. The validity of the predicted APRs is based on the reliability of the employed tools, as well as of the controls used in the study. The use of different algorithms for aggregation propensity prediction is essential for increasing the confidence in the predictions, as the algorithms use slightly different parameters. We found that the aggregation-prone regions are usually between 30 and 100 residues long and are located in the low disorder regions of the proteins, which provides insight into the aggregation propensity propagation and stability of the proteins throughout the life cycle of the virus.

From the various algorithms employed, we used PASTA 2.0 for the first selection, as this seems the most accurate in the prediction of aggregation propensity. Our results showed that different types of proteins had the highest aggregation propensity scores: Nsp 4, Nsp 6, Nsp7, ORF 3a, ORF 6, ORF 7a, ORF 7b, ORF10, CoVE and CovM (Fig. 9a). Interestingly, NSPs synthesis requires 2/3 of the genome of the virus and these are of upmost importance for the viral replication cycle within the host cell. [39] Likewise, ORF proteins are essential for promoting the membrane rearrangement and cell death in the host cells, as well as other structural changes that lead to the propagation of the virus and the demise of the host cell. [40,41] The CovE and CovM proteins are known to play essential roles in virus morphogenesis and assembly through interactions with other viral proteins. [42,43] Therefore, understanding the propensity of these proteins for aggregation is of great importance for understanding the long-term effects in the host cells. The hijacking of the cellular machinery to produce viral proteins in high amounts may, on one hand, result in the aggregation of aggregation-prone proteins and, in parallel, overwhelm the proteostasis network, leading to additional protein aggregation (Fig. 9).

Accessory proteins are considered non-essential for the viral structure but, in some viral families, they are important for the virulence or evasion of the host immune responses. Among these, ORF3a, ORF7a, and ORF7b show a high propensity for aggregation. The molecular function of the accessory proteins, such as ORF3a and ORF7b, has been characterized in SARS-CoV and they seem to be important for the virulence and host interaction. Given that ORF3a and ORF7b share approximately 80% homology with those of SARS-CoV-2 [44], their function might be similar in this virus. ORF3a is an ion-channel viroporin through where the virus might be released. [45,46] In mice, deletion of ORF3a reduces viral replication [47] suggesting that this accessory protein plays an important role in the virus' life cycle. Less is known regarding the function of ORF7b. It has been characterized as an important multimeric structural component of the virions and seems to be necessary for the infection. [48,49] Thus, it is possible that mutations or conditions that alter the function of ORF3a or ORF7b may affect the spreading of infection by SARS-CoV-2.

Subsequently, after identifying the 10 proteins most prone to aggregate, we asked whether mutations (reported around the world until August of 2020) might preferentially affect their aggregation. We identified several mutations predicted to increase protein aggregation and, strikingly, we found a correlation with the localization in different geographical regions. Future studies of the symptomatology of neurological disorders may reveal whether a connection exists with the presence of aggregation prone SARS-CoV-2 variants. It should be noted that due to possible sequencing bias, these are the only conclusions that can reliably be drawn from the geographical spreading of aggregation prone SARS-CoV-2 variants.

Thus far, we have not analyzed the variants of the virus that emerged in Brazil, South Africa and UK in the late 2020, beginning of 2021, which showed to be more transmissible and infectious, leading to a higher death toll. Therefore, in a future study, it will be of great interest to assess these new mutations for their effect in the aggregation propensity of the proteins in which they occur, and to investigate a possible correlation between higher transmissibility (and/or infection) and aggregation propensity. [50]

Protein aggregation is a major pathological hallmark in many

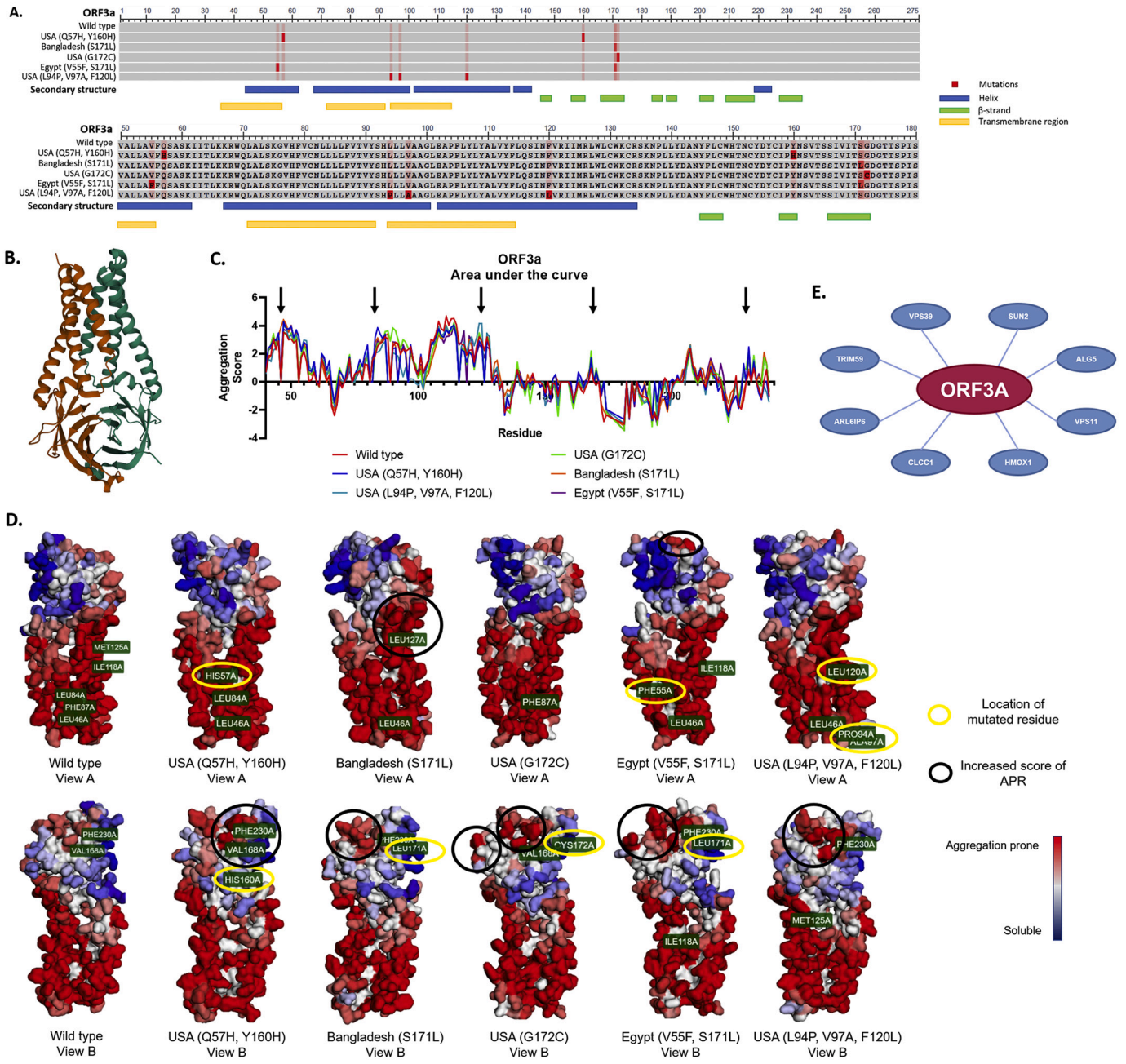

Fig. 8. ORF3a variant structural analysis. a) COBALT sequence alignment analysis of the mutant variants vs. wild type. The bright red boxes represent the mutation in each variant and the light red boxes represent the residue that was changed in another variant. Below the sequence, the important characteristics of the secondary structure are shown; the blue boxes are alpha-helixes, the green boxes represent the beta-strands and the yellow boxes the transmembrane regions. b) Secondary structure of the ORF3a protein wild type reported in the PDB (6XDC). c) Area under the curve analysis of the aggregation score of each amino acid from each ORF3a variant. Black arrows point towards the peaks that showed a higher increase in the aggregation score compared to the wild type. d) Aggrescan 3D 2.0 reconstruction of the variant structures. Yellow circles show where the mutated amino acid is located, black circles show the important changes in the aggregation prone regions. e) CoVex Protein-Protein interaction network analysis. Red circle is the viral protein, blue circles are the host proteins. The interaction prediction tool is based on the human cell line HEK-293 T.

neurodegenerative disorders. [38,51] Interestingly, several studies have hinted at possible connections between viral infections and the onset of diseases such as Parkinson's disease [11,52] or Alzheimer's disease. [12,53–55] Therefore, it is of great interest to investigate whether this connection results from direct interactions between host and viral aggregation-prone proteins, or from a more general effect at the level of the proteostasis network. Consistently, SARS-CoV-2 mRNA is present in brain samples of COVID-19 patients, [56] suggesting that the virus may also directly infect the brain. Moreover, it was demonstrated that the virus can infect neurons through the ACE2 receptor and that the infection modifies the brain vascular topology, ultimately affecting the microenvironment in ways that may be hazardous for the neighboring cells. [57] Astrocytes can also be infected by SARS-CoV2, reducing their energy metabolism, and triggering an inflammatory response that reduces neuronal viability. [58] These reports support the idea that understanding the infection mechanisms and consequences for neuronal and glial cells will be important as they may inform on possible strategies for therapeutic intervention.

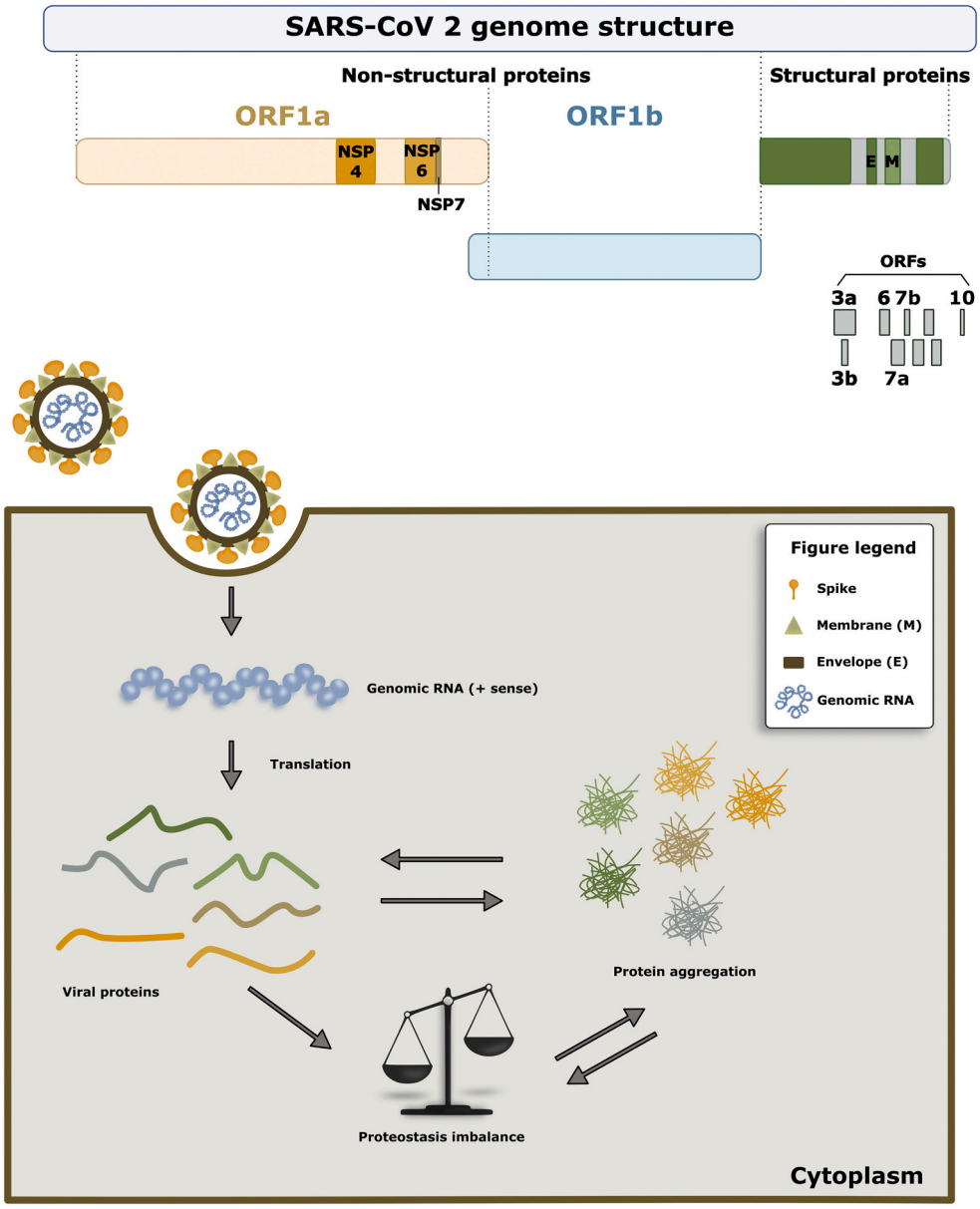

Fig. 9. Proteostasis imbalance due to SARS-CoV-2 aggregation-prone proteins. a) In our study, we identified 10 proteins that showed an increase in the predicted aggregation propensity: non-structural proteins Nsp4, Nsp6 and Nsp7, as well as ORF3a, ORF6, ORF7a, ORF7b, ORF10, CovE and CovM. b) One of the common consequences of protein aggregation is, for example, impairment of the protein degradation systems. This extra burden on the degradation system can promote additional protein accumulation, ultimately causing a proteostasis imbalance. This may result in the accumulation of viral proteins that can engage in promiscuous interactions with other proteins, and may increase the virulence of the virus.

5. Conclusions

In the present study, we conducted a detailed bioinformatic analysis of the SARS-CoV-2 proteome aggregation propensity by using several complementary computational tools such as PASTA2.0, CamSol, AmyloGram and TANGO. To our knowledge, this is the first study to analyze and compare the SARS-CoV-2 protein variants according to their geographical location. First, we found that the aggregation-prone regions are usually between 30 and 100 residues long and are located in the low disorder regions of these proteins. Second, the analyses showed that 10 proteins were predicted as aggregation prone. Furthermore, the number of reported variants showed that some proteins, like ORF7b or ORF10, appeared to be rather conserved with a few reported variants, while others, like ORF3a or Nsp4, displayed a larger variability. From these mutants, ~40% displayed an increased aggregation propensity score in at least one of the algorithms employed. Additionally, the aggregation-prone mutants of ORF7b were identified in 7 different European countries, Australia, India, Nepal, China, and Thailand. Compared to this, the ORF3a variants were detected in 2 European countries and in Taiwan. Given that the immunological response and its components play an important role in the SARS-CoV-2 infection, this may vary depending on specific genetic backgrounds of different populations. Consistently, several studies have reported a link between the immunotype and the severity of the disease, mortality and/or prognosis. [59–62] However, how the immunotype affects the cellular machinery, and which are the mechanisms involved, are still important and open questions that need to be addressed in the context of aggregation-prone proteins.

Finally, to shed light on the possible cellular mechanisms affected by these prone-aggregation proteins, like ORF3a, 8 proteins that can potentially interact with ORF3a were identified. These have been associated with inflammatory processes, endoplasmic reticulum stress, protein misfolding/clearance and autophagy. Taking all of these into account, this work provides new insights into the importance of understanding and studying the long-term effects of the SARS-CoV-2 infection. Additionally, it reinforces the emerging idea that it is important to address the effects of the infection on neuronal and glial cells in order to develop future therapeutic strategies.

Data availability

All data analyzed (accession numbers) during this study are included in this published article in the supplementary data, appendix B.

Funding

Not applicable.

Declaration of Competing Interest

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

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Acknowledgements

The authors acknowledge funding by the Sino-German Center for Research Promotion in Beijing (GZ C-0054). AK is funded by the Heisenberg program of the German Research Foundation (DFG). The authors are grateful to Dr. Nicolás Lemus Diaz for his help with data curation, analysis, and presentation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbapap.2021.140693.

References

[1] A.E. Gorbalenya, S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A. Gulyaeva, B. L. Haagmans, C. Lauber, A.M. Leontovich, B.W. Neuman, D. Penzar, S. Perlman, L. L.M. Poon, D.V. Samborskiy, I.A. Sidorov, I. Sola, J. Ziebuhr, The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, Nat. Microbiol. 5 (2020) 536–544, https://doi.org/10.1038/s41564-020-0695-z.

[2] J.F.W. Chan, S. Yuan, K.H. Kok, K.K.W. To, H. Chu, J. Yang, F. Xing, J. Liu, C.C. Y. Yip, R.W.S. Poon, H.W. Tsoi, S.K.F. Lo, K.H. Chan, V.K.M. Poon, W.M. Chan, J. D. Ip, J.P. Cai, V.C.C. Cheng, H. Chen, C.K.M. Hui, K.Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, Lancet. 395 (2020) 514–523, https://doi.org/10.1016/S0140-6736(20)30154-9.

[3] J. Machhi, J. Herskovitz, A.M. Senan, D. Dutta, B. Nath, M.D. Oleynikov, W. R. Blomberg, D.D. Meigs, M. Hasan, M. Patel, P. Kline, R.C.C. Chang, L. Chang, H. E. Gendelman, B.D. Kevadiya, The natural history, pathobiology, and clinical manifestations of SARS-CoV-2 infections, J. NeuroImmune Pharmacol. 15 (2020) 359–386, https://doi.org/10.1007/s11481-020-09944-5.

[4] E.M. Liotta, A. Batra, J.R. Clark, N.A. Shlobin, S.C. Hoffman, Z.S. Orban, I. J. Koralnik, Frequent neurologic manifestations and encephalopathy-associated morbidity in Covid-19 patients, Ann. Clin. Transl. Neurol. 7 (2020) 2221–2230, https://doi.org/10.1002/acn3.51210.

[5] G.C. Román, P.S. Spencer, J. Reis, A. Buguet, M.E.A. Faris, S.M. Katrak, M. Láinez, M.T. Medina, C. Meshram, H. Mizusawa, S. Öztürk, M. Wasay, The neurology of COVID-19 revisited: a proposal from the environmental neurology specialty Group of the World Federation of neurology to implement international neurological registries, J. Neurol. Sci. 414 (2020) 116884, https://doi.org/10.1016/j.jns.2020.116884.

[6] G. Nepal, J.H. Rehrig, G.S. Shrestha, Y.K. Shing, J.K. Yadav, R. Ojha, G. Pokhrel, Z. L. Tu, D.Y. Huang, Neurological manifestations of COVID-19: A systematic review, Crit. Care 24 (2020), https://doi.org/10.1186/s13054-020-03121-z.

[7] M. Desforges, A. Le Coupanec, É. Brison, M. Meessen-Pinard, P.J. Talbot, Neuroinvasive and neurotropic human respiratory coronaviruses: potential neurovirulent agents in humans, Adv. Exp. Med. Biol. 807 (2014) 75–96, https://doi.org/10.1007/978-81-322-1777-0_6.

[8] A. Ramani, L. Müller, P.N. Ostermann, E. Gabriel, P. Abida-Islam, A. Müller-Schiffmann, A. Mariappan, O. Goureau, H. Gruell, A. Walker, M. Andrée, S. Hauka, T. Houwaart, A. Dilthey, K. Wohlgemuth, H. Omran, F. Klein, D. Wieczorek, O. Adams, J. Timm, C. Korth, H. Schaal, J. Gopalakrishnan, <scp>SARS</scp> -CoV-2 targets neurons of 3D human brain organoids, EMBO J. 39 (2020), e106230, https://doi.org/10.15252/embj.2020106230.

[9] M.E. Cohen, R. Eichel, B. Steiner-Birmanns, A. Janah, M. Ioshpa, R. Bar-Shalom, J. J. Paul, H. Gaber, V. Skrahina, N.M. Bornstein, G. Yahalom, A case of probable Parkinson's disease after SARS-CoV-2 infection, Lancet Neurol. 19 (2020) 804–805, https://doi.org/10.1016/S1474-4422(20)30305-7.

[10] H. Adam, W. Trender, S.R. Chamberlain, A. Jolly, J.E. Grant, F. Patrick, N. Mazibuko, S. Williams, J.M. Barnby, P. Hellyer, M.A. Mehta, Cognitive deficits in people who have recovered from COVID-19 relative to controls: An N=84,285 online study, MedRxiv. (2020), https://doi.org/10.1101/2020.10.20.20215863.

[11] A. Lippi, R. Domingues, C. Setz, T.F. Outeiro, A. Krisko, SARS-CoV-2: at the crossroad between aging and Neurodegeneration, Mov. Disord. 35 (2020) 716–720, https://doi.org/10.1002/mds.28084.

[12] R. Domingues, A. Lippi, C. Setz, T.F. Outeiro, A. Krisko, SARS-CoV-2, immunosenescence and inflammaging: Partners in the COVID-19 crime, Aging (Albany NY) 12 (2020) 18778–18789, https://doi.org/10.18632/aging.103989.

[13] A.R. Fehr, S. Perlman, Coronaviruses: An overview of their replication and pathogenesis, in: Coronaviruses Methods Protoc., Springer, New York, 2015, pp. 1–23, https://doi.org/10.1007/978-1-4939-2438-7_1.

[14] J.F.W. Chan, K.H. Kok, Z. Zhu, H. Chu, K.K.W. To, S. Yuan, K.Y. Yuen, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, Emerg. Microbes Infect. 9 (2020) 221–236, https://doi.org/10.1080/22221751.2020.1719902.

[15] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E. C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature. 579 (2020) 265–269, https://doi.org/10.1038/s41586-020-2008-3.

[16] Y.A. Malik, Properties of Coronavirus and SARS-CoV-2, n.d.

[17] S. Lukassen, R.L. Chua, T. Trefzer, N.C. Kahn, M.A. Schneider, T. Muley, H. Winter, M. Meister, C. Veith, A.W. Boots, B.P. Hennig, M. Kreuter, C. Conrad, R. Eils, SARS -CoV-2 receptor ACE 2 and TMPRSS 2 are primarily expressed in bronchial transient secretory cells, EMBO J. 39 (2020), https://doi.org/10.15252/embj.20105114.

[18] T.K. Mukherjee, P. Malik, R. Maitra, J.R. Hoidal, Ravaging SARS-CoV-2: rudimentary diagnosis and puzzling immunological responses, Curr. Med. Res. Opin. 37 (2021) 207–217, https://doi.org/10.1080/03007995.2020.1862532.

[19] R. Linding, J. Schymkowitz, F. Rousseau, F. Diella, L. Serrano, A comparative study of the relationship between protein structure and β-aggregation in globular and intrinsically disordered proteins, J. Mol. Biol. 342 (2004) 345–353, https://doi.org/10.1016/j.jmb.2004.06.088.

[20] S. Deforte, V.N. Uversky, Intrinsically disordered proteins in PubMed: what can the tip of the iceberg tell us about what lies below? RSC Adv. 6 (2016) 11513–11521, https://doi.org/10.1039/c5ra24866c.

[21] G.D. Williams, R.-Y. Chang, D.A. Brian, A Phylogenetically conserved hairpin-type 3′ Untranslated region Pseudoknot functions in coronavirus RNA replication, J. Virol. 73 (1999) 8349–8355, https://doi.org/10.1128/jvi.73.10.8349-8355.1999.

[22] P.V. Baranov, C.M. Henderson, C.B. Anderson, R.F. Gesteland, J.F. Atkins, M. T. Howard, Programmed ribosomal frameshifting in decoding the SARS-CoV genome, Virology. 332 (2005) 498–510, https://doi.org/10.1016/j.virol.2004.11.038.

[23] M.P. Robertson, H. Igel, R. Baertsch, D. Haussler, M. Ares, W.G. Scott, The structure of a rigorously conserved RNA element within the SARS virus genome, PLoS Biol. 3 (2005), https://doi.org/10.1371/journal.pbio.0030005.

[24] I. Walsh, F. Seno, S.C.E. Tosatto, A. Trovato, PASTA 2.0: An improved server for protein aggregation prediction, Nucleic Acids Res. 42 (2014), https://doi.org/10.1093/nar/gku399.

[25] M. Burdukiewicz, P. Sobczyk, S. Rödiger, A. Duda-Madej, P. MacKiewicz, M. Kotulska, Amyloidogenic motifs revealed by n-gram analysis, Sci. Rep. 7 (2017) 1–10, https://doi.org/10.1038/s41598-017-13210-9.

[26] R: The R Project for Statistical Computing, (n.d.). https://www.r-project.org/ (accessed April 7, 2021).

[27] P. Sormanni, F.A. Aprile, M. Vendruscolo, The CamSol method of rational design of protein mutants with enhanced solubility, J. Mol. Biol. 427 (2015) 478–490, https://doi.org/10.1016/j.jmb.2014.09.026.

[28] A.M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins, Nat. Biotechnol. 22 (2004) 1302–1306, https://doi.org/10.1038/nbt1012.

[29] F. Rousseau, J. Schymkowitz, L. Serrano, Protein aggregation and amyloidosis: confusion of the kinds? Curr. Opin. Struct. Biol. 16 (2006) 118–126, https://doi.org/10.1016/j.sbi.2006.01.011.

[30] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242, https://doi.org/10.1093/nar/28.1.235.

[31] A. Kuriata, V. Iglesias, J. Pujols, M. Kurcinski, S. Kmiecik, S. Ventura, Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility, Nucleic Acids Res. 47 (2019) W300–W307, https://doi.org/10.1093/nar/gkz321.

[32] S. Sadegh, J. Matschinske, D.B. Blumenthal, G. Galindez, T. Kacprowski, M. List, R. Nasirigerdeh, M. Oubounyt, A. Pichlmair, T.D. Rose, M. Salgado-Albarrán, J. Späth, A. Stukalov, N.K. Wenke, K. Yuan, J.K. Pauling, J. Baumbach, Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing, Nat. Commun. 11 (2020) 1–9, https://doi.org/10.1038/s41467-020-17189-2.

[33] CoVex, (n.d.). https://exbio.wzw.tum.de/covex/explorer (accessed April 7, 2021).

[34] X. Zeng, C.R. Carlin, Host cell autophagy modulates early stages of adenovirus infections in airway epithelial cells, J. Virol. 87 (2013) 2307–2319, https://doi.org/10.1128/jvi.02014-12.

[35] G. Miao, H. Zhao, Y. Li, M. Ji, Y. Chen, Y. Shi, Y. Bi, P. Wang, H. Zhang, ORF3a of the COVID-19 virus SARS-CoV-2 blocks HOPS complex-mediated assembly of the SNARE complex required for autolysosome formation, Dev. Cell 56 (2021), https://doi.org/10.1016/j.devcel.2020.12.010, 427–442.e5.

[36] M. Rossi, M. Piagnerelli, A. Van Meerhaeghe, K. Zouaoui Boudjeltia, Heme oxygenase-1 (HO-1) cytoprotective pathway: a potential treatment strategy against coronavirus disease 2019 (COVID-19)-induced cytokine storm syndrome, Med. Hypotheses 144 (2020) 110242, https://doi.org/10.1016/j.mehy.2020.110242.
[37] Y. Jia, T.J. Jucius, S.A. Cook, S.L. Ackerman, Loss of Clcc1 results in ER stress, misfolded protein accumulation, and neurodegeneration, J. Neurosci. 35 (2015) 3001–3009, https://doi.org/10.1523/JNEUROSCI.3678-14.2015.
[38] R. Marreiros, A. Müller-Schiffmann, S.V. Trossbach, I. Prikulis, S. Hänsch, S. Weidtkamp-Peters, A.R. Moreira, S. Sahu, I. Soloviev, S. Selvarajah, V. R. Lingappa, C. Korth, Disruption of cellular proteostasis by H1N1 influenza a virus causes α-synuclein aggregation, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 6741–6751, https://doi.org/10.1073/pnas.1906466117.
[39] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, J. Med. Virol. 92 (2020) 418–423, https://doi.org/10.1002/jmv.25681.
[40] H.S. Hillen, G. Kokic, L. Farnung, C. Dienemann, D. Tegunov, P. Cramer, Structure of replicating SARS-CoV-2 polymerase, Nature. 584 (2020) 154–156, https://doi.org/10.1038/s41586-020-2368-8.
[41] D. Jose, SARS-Cov-2 systems biology, Ann. Syst. Biol. (2020) 029–032, https://doi.org/10.17352/asb.000009.
[42] D.X. Liu, Q. Yuan, Y. Liao, Coronavirus envelope protein: a small membrane protein with multiple functions, Cell. Mol. Life Sci. 64 (2007) 2043–2048, https://doi.org/10.1007/s00018-007-7103-1.
[43] B.W. Neuman, G. Kiss, A.H. Kunding, D. Bhella, M.F. Baksh, S. Connelly, B. Droese, J.P. Klaus, S. Makino, S.G. Sawicki, S.G. Siddell, D.G. Stamou, I.A. Wilson, P. Kuhn, M.J. Buchmeier, A structural analysis of M protein in coronavirus assembly and morphology, J. Struct. Biol. 174 (2011) 11–22, https://doi.org/10.1016/j.jsb.2010.11.021.
[44] G. Mariano, R.J. Farthing, S.L.M. Lale-Farjat, J.R.C. Bergeron, Structural characterization of SARS-CoV-2: where we are, and where we need to be, Front. Mol. Biosci. 7 (2020), https://doi.org/10.3389/fmolb.2020.605236.
[45] W. Lu, B.J. Zheng, K. Xu, W. Schwarz, L. Du, C.K.L. Wong, J. Chen, S. Duan, V. Deubel, B. Sun, Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 12540–12545, https://doi.org/10.1073/pnas.0605402103.
[46] E. Issa, G. Merhi, B. Panossian, T. Salloum, S. Tokajian, SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis, MSystems 5, 2020, https://doi.org/10.1128/msystems.00266-20.
[47] C. Castaño-Rodriguez, J.M. Honrubia, J. Gutiérrez-Álvarez, M.L. DeDiego, J. L. Nieto-Torres, J.M. Jimenez-Guardeño, J.A. Regla-Nava, R. Fernandez-Delgado, C. Verdia-Báguena, M. Queralt-Martín, G. Kochan, S. Perlman, V.M. Aguilella, I. Sola, L. Enjuanes, Role of severe acute respiratory syndrome coronavirus viroporins E, 3a, and 8a in replication and pathogenesis, MBio 9 (2018), https://doi.org/10.1128/mBio.02325-17.
[48] S.R. Schaecher, J.M. Mackenzie, A. Pekosz, The ORF7b protein of severe acute respiratory syndrome coronavirus (SARS-CoV) is expressed in virus-infected cells and incorporated into SARS-CoV particles, J. Virol. 81 (2007) 718–731, https://doi.org/10.1128/JVI.01691-06.
[49] M.-L. Fogeron, R. Montserret, J. Zehnder, M.-H. Nguyen, M. Dujardin, L. Brigandat, L. Cole, M. Ninot-Pedrosa, L. Lecoq, A. Böckmann, SARS-CoV-2 ORF7b: is a bat virus protein homologue a major cause of COVID-19 symptoms? BioRxiv (2021) https://doi.org/10.1101/2021.02.05.428650.
[50] O.T.R. Toovey, K.N. Harvey, P.W. Bird, J.W.T. Tang, Introduction of Brazilian SARS-CoV-2 484K.V2 related variants into the UK, J. Inf. Secur. (2021), https://doi.org/10.1016/j.jinf.2021.01.025.
[51] A. Müller-Schiffmann, S.V. Trossbach, V.R. Lingappa, C. Korth, Viruses as 'Truffle Hounds': molecular tools for untangling brain cellular pathology, Trends Neurosci. (2020), https://doi.org/10.1016/j.tins.2020.11.004.
[52] T. Yamada, Viral etiology of Parkinson's disease: focus on influenza a virus, Parkinsonism Relat. Disord. 2 (1996) 113–121, https://doi.org/10.1016/1353-8020(96)00006-5.
[53] D.H. Cribbs, B.Y. Azizeh, C.W. Cotman, F.M. LaFerla, Fibril formation and neurotoxicity by a herpes simplex virus glycoprotein B fragment with homology to the Alzheimer's Aβ peptide, Biochemistry. 39 (2000) 5988–5994, https://doi.org/10.1021/bi000029f.
[54] W.A. Eimer, D.K. Vijaya Kumar, N.K. Navalpur Shanmugam, A.S. Rodriguez, T. Mitchell, K.J. Washicosky, B. György, X.O. Breakefield, R.E. Tanzi, R.D. Moir, Alzheimer's disease-associated β-amyloid is rapidly seeded by herpesviridae to protect against brain infection, Neuron 99 (2018), https://doi.org/10.1016/j.neuron.2018.06.030, 56–63.e3.
[55] M. Sochocka, K. Zwolińska, J. Leszek, The infectious etiology of Alzheimer's disease, Curr. Neuropharmacol. 15 (2017), https://doi.org/10.2174/1570159x15666170313122937.
[56] V.G. Puelles, M. Lütgehetmann, M.T. Lindenmeyer, J.P. Sperhake, M.N. Wong, L. Allweiss, S. Chilla, A. Heinemann, N. Wanner, S. Liu, F. Braun, S. Lu, S. Pfefferle, A.S. Schröder, C. Edler, O. Gross, M. Glatzel, D. Wichmann, T. Wiech, S. Kluge, K. Pueschel, M. Aepfelbacher, T.B. Huber, Multiorgan and renal tropism of SARS-CoV-2, N. Engl. J. Med. 383 (2020) 590–592, https://doi.org/10.1056/nejmc2011400.
[57] E. Song, C. Zhang, B. Israelow, A. Lu-Culligan, A.V. Prado, S. Skriabine, P. Lu, O. El Weizman, F. Liu, Y. Dai, K. Szigeti-Buck, Y. Yasumoto, G. Wang, C. Castaldi, J. Heltke, E. Ng, J. Wheeler, M.M. Alfajaro, E. Levavasseur, B. Fontes, N. G. Ravindra, D. van Dijk, S. Mane, M. Gunel, A. Ring, S.A. Jaffar Kazmi, K. Zhang, C.B. Wilen, T.L. Horvath, I. Plu, S. Haik, J.L. Thomas, A. Louvi, S.F. Farhadian, A. Huttner, D. Seilhean, N. Renier, K. Bilguvar, A. Iwasaki, Neuroinvasion of SARS-CoV-2 in human and mouse brain, J. Exp. Med. 218 (2021), https://doi.org/10.1084/JEM.20202135.
[58] F. Crunfli, V.C. Carregari, F.P. Veras, P.H. Vendramini, A.G. Fragnani Valença, A.S. L. Marcelo Antunes, C. Brandão-Teles, G. de S. Zuccoli, G. Reis De Oliveira, L. C. Silva-Costa, V.M. Saia-Cereda, A.C. Codo, P.L. Parise, D.A. Toledo Teixeira, G. F. de Souza, S.P. Muraro, B.M. Silva Melo, G.M. Almeida, E.M. Silva Firmino, R. G. Ludwig, G.P. Ruiz, T.L. Knittel, G.G. Davanzo, J.A. Gerhardt, P.B. Rodrigues, J. Forato, M.R. Amorim, N.B. Silva, M.C. Martini, M.N. Benatti, S. Batah, L. Siyuan, R.E.M. Pereira Silva, R.B. João, L.S. Silva, M.H. Nogueira, Í.K. Aventurato, M.R. de Brito, M.K. Machado Alvim, J.R. da Silva, L.L. Damião, M.E.D.P. Castilho Stefano, I. M.P. De Sousa, E.D. da Rocha, S.M. Gonçalves, L.H.L. da Silva, V. Bettini, B.M. de Campos, G. Ludwig, R.M. Mendes Viana, R. Martins, A.S. Vieira, J.C. Alves-Filho, E. Arruda, A.S. Sebollela, F. Cendes, F.Q. Cunha, A. Damásio, M.A. Ramirez Vinolo, C.D. Munhoz, S.K. Rehen, T. Mauad, A.N. Duarte-Neto, L.F.F. da Silva, M. Dolhnikoff, P. Saldiva, A.T. Fabro, A.S. Farias, P.M.M. Moraes-Vieira, J.L. Proença Módena, C.L. Yasuda, M.A. Mori, T.M. Cunha, D. Martins De Souza, SARS-CoV-2 infects brain astrocytes of COVID-19 patients and impairs neuronal viability, MedRxiv. 16 (2020), https://doi.org/10.1101/2020.10.09.20207464.
[59] C. Maucourant, I. Filipovic, A. Ponzetta, S. Aleman, M. Cornillet, L. Hertwig, B. Strunz, A. Lentini, B. Reinius, D. Brownlie, A. Cuapio, E.H. Ask, R.M. Hull, A. Haroun-Izquierdo, M. Schaffer, J. Klingström, E. Folkesson, M. Buggert, J. K. Sandberg, L.I. Eriksson, O. Rooyackers, H.G. Ljunggren, K.J. Malmberg, J. Michaëlsson, N. Marquardt, Q. Hammer, K. Strålin, N.K. Björkström, Natural killer cell immunotypes related to COVID-19 disease severity, Sci. Immunol. 5 (2020) 6832, https://doi.org/10.1126/SCIIMMUNOL.ABD6832.
[60] D. Mathew, J.R. Giles, A.E. Baxter, D.A. Oldridge, A.R. Greenplate, J.E. Wu, C. Alanio, L. Kuri-Cervantes, M.B. Pampena, K. D'Andrea, S. Manne, Z. Chen, Y. J. Huang, J.P. Reilly, A.R. Weisman, C.A.G. Ittner, O. Kuthuru, J. Dougherty, K. Nzingha, N. Han, J. Kim, A. Pattekar, E.C. Goodwin, E.M. Anderson, M. E. Weirick, S. Gouma, C.P. Arevalo, M.J. Bolton, F. Chen, S.F. Lacey, H. Ramage, S. Cherry, S.E. Hensley, S.A. Apostolidis, A.C. Huang, L.A. Vella, M.R. Betts, N. J. Meyer, E.J. Wherry, Z. Alam, M.M. Addison, K.T. Byrne, A. Chandra, H. C. Descamps, Y. Kaminskiy, J.T. Hamilton, J.H. Noll, D.K. Omran, E. Perkey, E. M. Prager, D. Pueschl, J.B. Shah, J.S. Shilan, A.N. Vanderbeck, Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications, Science (80) 369 (2020), https://doi.org/10.1126/SCIENCE.ABC8511.
[61] E. Bernal, L. Gimeno, M.J. Alcaraz, A.A. Quadeer, M. Moreno, M.V. Martínez-Sánchez, J.A. Campillo, J.M. Gomez, A. Pelaez, E. García, M. Herranz, M. Hernández-Olivo, E. Martínez-Alfaro, A. Alcaraz, Á. Muñoz, A. Cano, M. R. McKay, M. Muro, A. Minguela, Activating killer-cell immunoglobulin-like receptors are associated with the severity of coronavirus disease 2019, J. Infect. Dis. (2021), https://doi.org/10.1093/infdis/jiab228.
[62] E. Martín-Sánchez, J.J. Garcés, C. Maia, S. Inogés, A. López-Díaz de Cerio, F. Carmona-Torre, M. Marin-Oto, F. Alegre, E. Molano, M. Fernandez-Alonso, C. Perez, C. Botta, A. Zabaleta, A.B. Alcaide, M.F. Landecho, M. Rua, T. Pérez-Warnisher, L. Blanco, S. Sarvide, A. Vilas-Zornoza, D. Alignani, C. Moreno, I. Pineda, M. Sogbe, J. Argemi, B. Paiva, J.R. Yuste, Immunological biomarkers of fatal COVID-19: a study of 868 patients, Front. Immunol. 12 (2021), https://doi.org/10.3389/fimmu.2021.659018.