FULL-LENGTH PAPER

Exhaustive mutational analysis of severe acute respiratory syndrome coronavirus 2 ORF3a: An essential component in the pathogen's infectivity cycle

Amit Benazraf | Isaiah T. Arkin

Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

Correspondence
Isaiah T. Arkin, Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem 9190400, Israel.
Email: arkin@huji.ac.il

Review Editor: John Kuriyan

Abstract

Detailed knowledge of a protein's key residues may assist in understanding its function and designing inhibitors against it. Consequently, such knowledge of one of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)'s proteins is advantageous since the virus is the etiological agent behind one of the biggest health crises of recent times. To that end, we constructed an exhaustive library of bacteria differing from each other by the mutated version of the virus's ORF3a viroporin they harbor. Since the protein is harmful to bacterial growth due to its channel activity, genetic selection followed by deep sequencing could readily identify mutations that abolish the protein's function. Our results have yielded numerous mutations dispersed throughout the sequence that counteract ORF3a's ability to slow bacterial growth. Comparing these data with the conservation pattern of ORF3a within the coronavirinae provided interesting insights: Deleterious mutations obtained in our study corresponded to conserved residues in the protein. However, despite the comprehensive nature of our mutagenesis coverage (108 average mutations per site), we could not reveal all of the protein's conserved residues. Therefore, it is tempting to speculate that our study unearthed positions in the protein pertinent to channel activity, while other conserved residues may correspond to different functionalities of ORF3a. In conclusion, our study provides important information on a key component of SARS-CoV-2 and establishes a procedure to analyze other viroporins comprehensively.

KEYWORDS
evolutionary conservation, genetic selection, ion channel, vulnerability mapping

1 | INTRODUCTION

1.1 | Virus and mutations

We are surrounded by viruses that pose a continuous challenge to our well-being. A seemingly endless number of mutations constantly vitiate the benefit of vaccinations and the drugs we use. The recent coronavirus disease 2019 (COVID-19) outbreak epitomized the above traits, whereby reappearing disease waves were spurred by the emergence of new strains (Hossain et al., 2021; Zhou & Wang, 2021).

The underlying cause of rapid viral evolution is the low replication fidelity of the viral genome due to the poor ability of their polymerases to avoid or correct errors. The rapid evolution is particularly evident in RNA viruses characterized by their high mutation rate, between 0.4 and 1.1 nucleotide errors per genome per round of replication (Bradwell et al., 2013; Lauber et al., 2013). This high mutation rate leads to various mutations, and consequently, cognate antiviral agents may become ineffective (Assa et al., 2016; Drake et al., 1998).

Because of the antagonistic interplay between the quickly changing fitness requirements of the virus and the maintenance of its essential functions, viruses undergo strong and diverse selective forces, leading to new variants possessing mutations in the genome (Pal et al., 2021). The mutation rate is itself subject to natural selection, and the mutations are fundamental building blocks of the evolutionary process—they constitute the variation (alongside other factors) upon which natural selection can act and are the cause of much of the novelty we see occur in evolution (Owen, 2010; Upadhyay et al., 2021).

1.2 | Coronavirinae and severe acute respiratory syndrome coronavirus 2

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a member of the Coronavirinae subfamily of the Coronaviridae. The virus is the etiological agent of the unusual viral pneumonia that started in Wuhan, China, leading to the pandemic outbreak called COVID-19 (Cucinotta & Vanelli, 2020). The virus that causes COVID-19 has a close similarity in sequence to the SARS-CoV-1, which caused the emergence of SARS in southern China in 2002/2003 (Fowler et al., 2010; Issa et al., 2020). Furthermore, The genome of SARS-CoV-2 showed 96.2% sequence similarity to a bat SARS-related coronavirus [13]. Studies show that many viruses closely related to SARS-CoV-2 have been circulating in horseshoe bats, indicating that recombination probably played a role in the evolutionary history of SARS-CoV-2 (Boni et al., 2020; Lytras et al., 2022). Together, SARS-CoV-2 and similar viruses form the sarbecovirus lineage within betacoronavirus.

As of September 6, 2022, the World Health Organization reported 600,366,479 confirmed cases of COVID-19, including a death toll of 6,460,493 deaths, a number that is still growing (https://covid19.who.int). Due to the fast spread of the virus around the globe and the development of mutations in SARS-CoV-2—dozens of variants spread among us. The disease spread has led to the constant fear that variants may develop higher distribution capabilities and resistance to vaccines that currently exist in the market. In particular, it was shown that even single mutations in specific proteins could change the pathogenicity of these viruses (Luk et al., 2019).

1.3 | Viral ion channels and ORF3a protein

Viruses exploit and modify host-cell ion homeostasis in favor of viral infection by using various viral-encoded viroporins (Gonzalez & Carrasco, 2003; Madan et al., 2008). Viroporins are small virally encoded hydrophobic proteins that function as ion channels—a group of membrane-spanning proteins that allow an ion flux across the cell membrane (Gonzalez & Carrasco, 2003; Hille, 1992; Nieva et al., 2012). Viroporins constitute a large family of multifunctional proteins broadly distributed in different viral families and are mainly concentrated in RNA viruses (Nieva et al., 2012). Highly pathogenic human viruses, such as SARS-CoV-2, encode more than one viroporin (Ewart et al., 1996). Ion channel activity is relevant for the virus's spread and may significantly impact host-cell ionic milieus and physiology (Nieva et al., 2012; Steinmann & Pietschmann, 2010; Wozniak et al., 2010). The loss of ion homeostasis triggered by the ion channel activity may have deleterious consequences for the cell, from stress responses to apoptosis (Bhowmick et al., 2012; Madan et al., 2008; Nieva et al., 2012). Thus using ion channels as new targets for therapeutic intervention represents a promising strategy (Charlton et al., 2020; Hover et al., 2017). An example of this importance can be found in the fight against the influenza virus: Amantadine, the second antiviral compound ever approved (Davies et al., 1964), targets the M2 protein (Hay et al., 1985) by blocking its proton channel activity (Pinto et al., 1992).

ORF3a, or by its second name, 3a protein, is a viral protein located in the genome of SARS-CoV-2, between the S and E proteins, and encodes a protein of 275 amino acids (Lu et al., 2020; Wu et al., 2020). The protein forms

a homotetrameric potassium-selective viroporin and may modulate virus release (Bianchi et al., 2021; Lu et al., 2006). A recent study determined the structure of SARS-CoV-2 ORF3a protein using Cryo-EM (Kern et al., 2021; Mousavizadeh & Ghasemi, 2021).

Studies have shown six functional domains within the structure of the ORF3a protein: signal peptide, residues 0–15; TRAF3-binding motif, residues 36–40; ion channel activity, residues 91–133; caveolin-binding motif, residues 141–149; Yxxψ motif, residues 160–163; and Di-acidic motif, residues 171–173 (Issa et al., 2020). Moreover, position 81–160 was found to be a cysteine-rich region, where residue C133 was found to be important for homodimerization (Lu et al., 2006).

Great effort has been put into finding the effect of different mutations in ion channel proteins and SARS-CoV-2 ORF3a viroporin, thereby formulating an offensive strategy against the virus. Still, an effective method to exhaustively screen protein mutations and discover the mutations' effect on that protein has not been employed (Bianchi et al., 2021; Issa et al., 2020; Wu et al., 2021).

1.4 | Aim of this study

The importance of monitoring mutations in different variants has emphasized the need for safe, effective, and simple assays to identify said modifications. In this manner, we can predict areas that are important for drug development in addition to areas that are already resistant to the drugs and vaccines in current use. In this study, utilizing the power of bacterial genetic selection, we devise a process of comprehensive random mutagenesis to monitor the effect of mutations in viroporins in general and in ORF3a in particular (Cirino et al., 2003).

2 | RESULTS AND DISCUSSION

To that end, we employed an approach based on random mutagenesis coupled with genetic selection. Since the experiments are conducted on an isolated viral protein in *Escherichia coli*, they do not carry any medical risk whatsoever. Specifically, we have previously shown that ORF3a's channel functionality is deleterious to bacteria due to excess membrane permeabilization (Tomar et al., 2021a). Consequently, one can identify mutations harmful to the protein by inducing protein expression, selecting viable bacteria, and determining the specific mutations by deep sequencing. Finally, the mutations obtained in the study can be compared with those found in circulating viral strains, thereby gaining insight into the various functional roles the protein may have.

2.1 | Library construction

Random mutagenesis was employed to evaluate each position's importance to protein function exhaustively. To determine the extent of mutagenesis coverage, we sequenced the gene from colonies in which the protein was not induced and consequently did not alter bacterial viability. In other words, the resulting library is indifferent to the particular modifications and, as such, represents the genetic variation that can subsequently be analyzed.
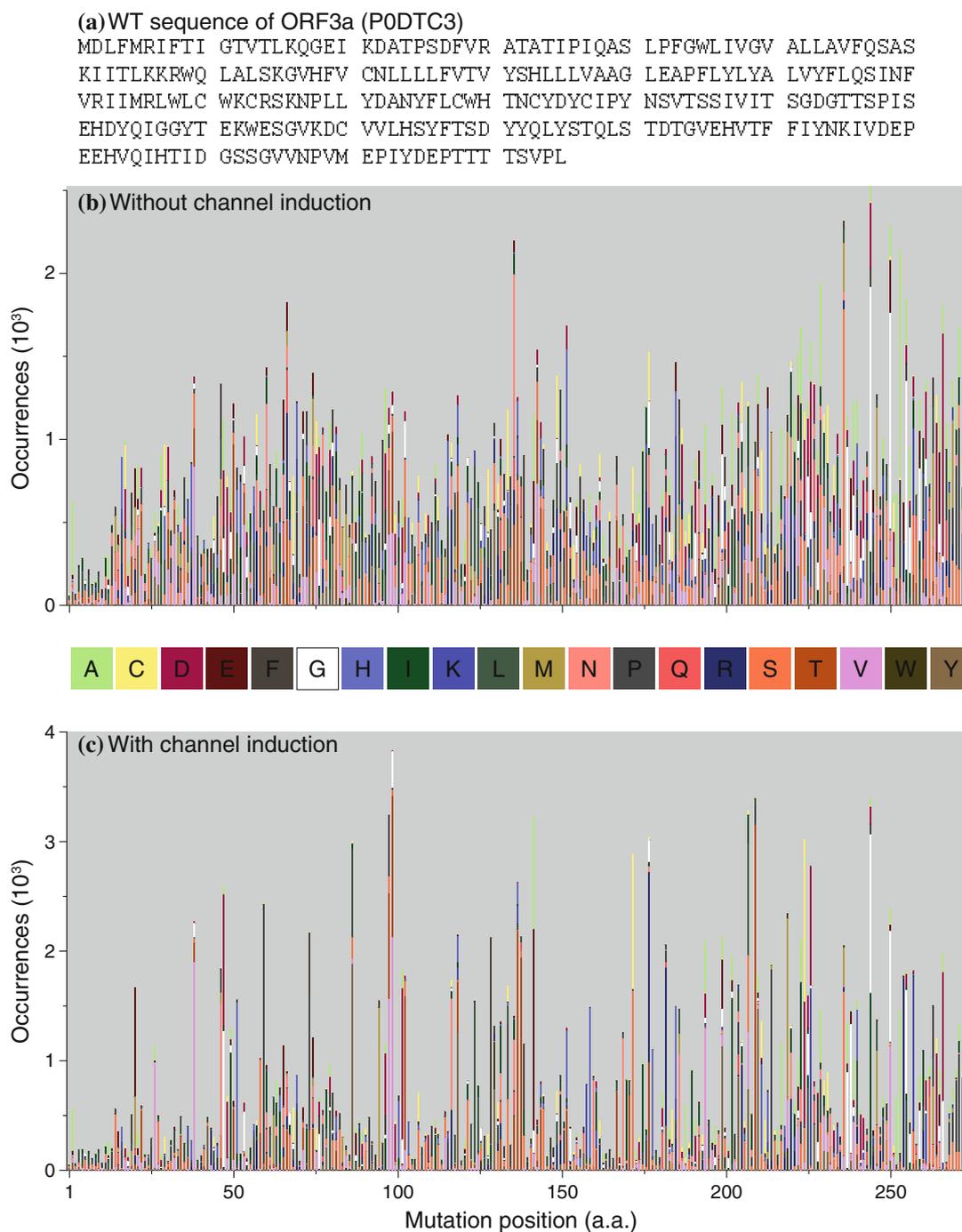
Results of 322,747 different sequences shown in Figure 1b yield an average of 105 mutations per residue with a standard deviation of 149. Furthermore, out of 5225 possible mutation types (19 mutations per each of the 275 amino acids), we obtained 2195 mutations, yielding a coverage of 42%. Together, both metrics indicate a comprehensive mutagenesis library. In further discussion, this library, where the protein was not induced, is eponymously termed an "uninduced library."

2.2 | Impact of selection

With an exhaustive library of protein variants in hand, we could proceed to genetic selection because ORF3a channel activity is deleterious to bacterial viability (Tomar et al., 2021a). Results of 305,333 mutations shown in Figure 1c exhibit a marked difference from the mutation library without protein expression. The library obtained upon protein induction (termed "induced library") is far more punctuated than the uninduced library. While the average number of mutations was similar, 108, the SD, 265, was significantly higher.

The uninduced and induced libraries shown in Figure 1 included an itemization and abundance of each mutation. For further analysis, we calculated a single value for each residue by summing each modification multiplied by its amino acid similarity according to an amino acid substitution matrix. The results shown in Figure S1 demonstrate once more the stark difference between the two libraries as a consequence of the protein's impact on bacteria phenotype.

To better evaluate our findings, we determined the over-representation of mutations in the induced library relative to the uninduced library: Mutations that abolish the protein's harmful activity are likely to be more prevalent when the protein is expressed. To that end, we calculated the number of occurrences of a specific mutation in the induced library divided by the number

(a) WT sequence of ORF3a (P0DTC3)

```
MDLFMRIFTI GTVTLKQGEI KDATPSDFVR ATATIPIQAS LPFGWLIVGV ALLAVFQSAS
KIITLKKRWQ LALSKGVHFV CNLLLLFVTV YSHLLLVAAG LEAPFLYLYA LVYFLQSINF
VRIIMRLWLC WKCRSKNPLL YDANYFLCWH TNCYDYCIPY NSVTSSIVIT SGDGTTSPIS
EHDYQIGGYT EKWESGVKDC VVLHSYFTSD YYQLYSTQLS TDTGVEHVTF FIYNKIVDEP
EEHVQIHTID GSSGVVNPVM EPIYDEPTTT TSVPL
```


(b) Without channel induction

A C D E F G H I K L M N P Q R S T V W Y

(c) With channel induction

Mutation position (a.a.)

FIGURE 1 Mutation libraries analyzed in the study representing amino acid substitution occurrences for each of the protein's 275 residues. Panels (b) and (c) depict the results without and with protein induction, respectively. The color coding for each amino reflects their physicochemical properties according to the "shapely models" as implemented in Rasmol (Sayle & Milner-White, 1995). The sequence of the wild-type protein is given in panel a

of representations of the same mutation in the uninduced library. The results, shown in Figure 2, point to several residues that are up to 20 times more common in the induced library than in the uninduced library. When taken into consideration with the data of the induced library, we can surmise that these mutations represent deleterious hotspots that most likely interfere with protein function.

We subsequently proceeded to project the deleterious mutation hotspots identified above on the protein structure. Interestingly, as shown in Figure 3, hotspots did not concentrate in one location in the protein or one secondary structure element. Instead, the harmful modification hotspots we found are dispersed along the ORF3a gene, even though studies show that the ion conductance pathway is concentrated only in a small part of the ORF3a gene (Issa et al., 2020).
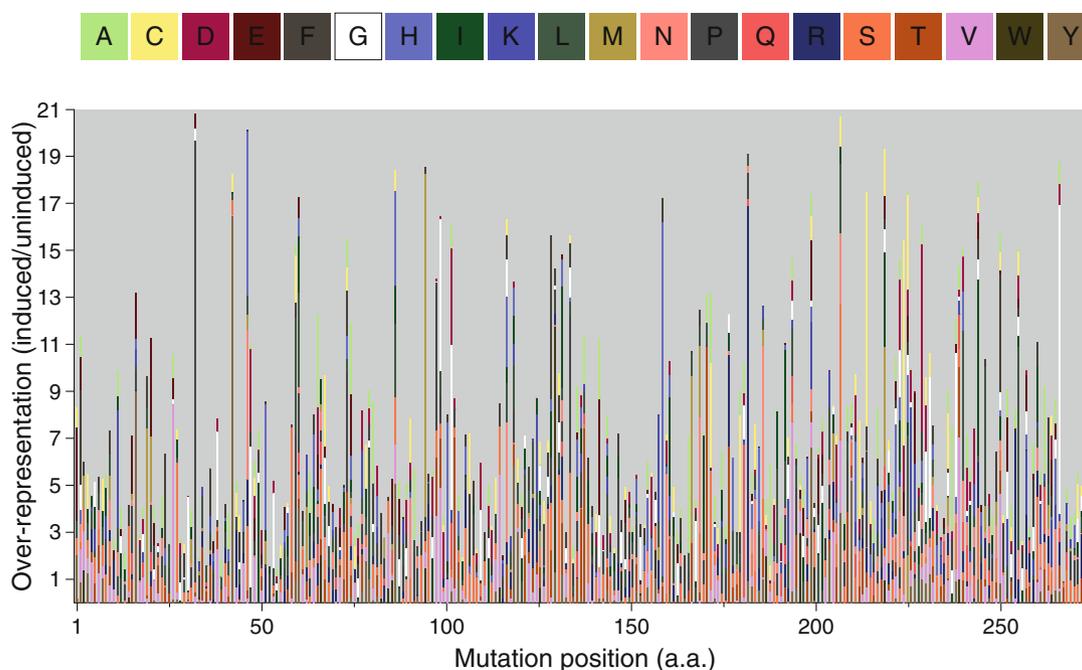
FIGURE 2 Mutation over-representation due to protein induction, presented as a function of amino acid mutation position, from 1 to 275. Values are calculated by dividing occurrences of specific mutations in the induced library (Figure 1c) by the occurrences of the same mutation in the uninduced library (Figure 1b). Amino acid color coding reflects their physicochemical properties according to the "shapely models" as implemented in Rasmol (Sayle & Milner-White, 1995)

2.3 | Existing ORF3a variations

The exhaustive search presented above has unearthed positions in the ORF3a protein whose mutations in our assay system are pernicious to the protein's activity. Such spots can be compared with variations in the protein sequence found in circulating strains of SARS-CoV-2 and to the same protein in different members of the coronavirinae. Those variations have evolved through natural evolution and thus are expected to be compatible with the protein's activity.

2.3.1 | Variation within the coronavirinae

Positions found as conserved in the coronavirinae subfamily represent sites in the ORF3a protein that are presumably essential for function. Such locations may overlap with the deleterious mutation hotspots we have identified experimentally. To that end, we used the UniProt database (UniProt Consortium, 2021) to download 11 sequences with similar functioning and a sequence identity of 80%–100% to the SARS-CoV-2 gene. Thereafter, multiple sequence alignment (MSA) of the aforementioned proteins yielded the results shown in Figure 4. Careful inspection of the outcome indicates that sequence conservation is observed throughout the protein, with 49% of residues found identical in all genes.

2.3.2 | Variation within SARS-CoV-2

Following the conservation analysis between ORF3a from different members of the coronavirinae, we turned to analyze conservation within a single species. Such an analysis benefits from the large-scale sequencing efforts of SARS-CoV-2 conducted due to the COVID-19 pandemic. For that reason, we compared 807,520 sequences of SARS-CoV-2 ORF3a, yielding the mutation frequencies seen in Figure 5. The detailed changes of every amino acid are depicted in Figure S2.

Close inspection of the mutations in the population highlights several residues that are particularly common, of which S26L is the most prevalent, seen in 205,955 samples (25% of our database). After it, Q57H, T223I, and G172V were the most common mutations, seen in 108,744, 48,561, and 38,043 of the sequences, respectively. Clearly, these common mutations are part of the natural viral evolution and may be directly beneficial to the virus's survival.

2.3.3 | Comparison of existing ORF3a variation with experimental results

Having determined the current 3a conservation patterns, we can compare these characteristics with our experimental results. We began by examining the similarity
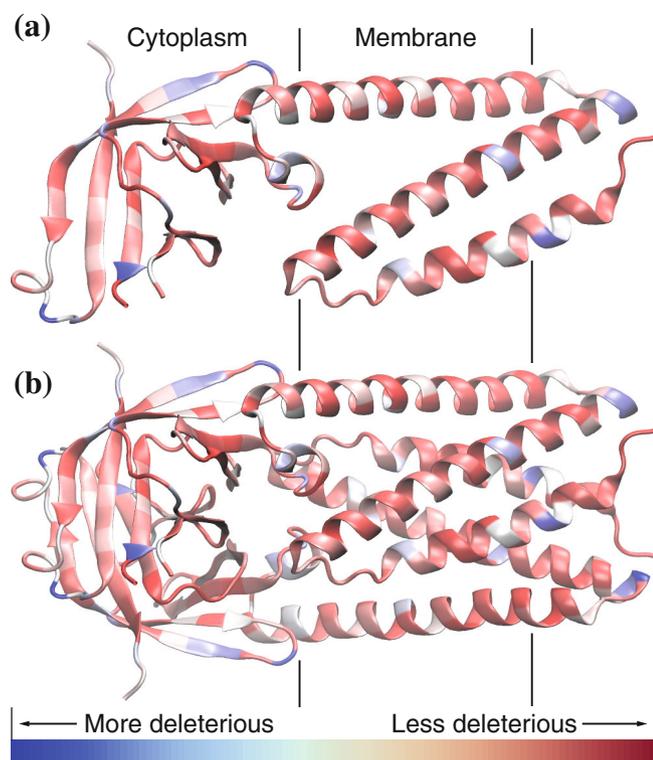
FIGURE 3 ORF3a monomeric (a) and dimeric (b) structures (Kern et al., 2021) colored according to the deleterious mutation hotspots identified in the study. The presumed locations of the lipid bilayer and cytoplasm are indicated. Each mutation in the induced library was multiplied by its PAM10 matrix score according to the corresponding position in the protein, and the sum of all the mutations in that position was calculated. The figure was generated by visual molecular dynamics (Humphrey et al., 1996)

between the conservation within the coronavirinae (Figure 4) and our study's outcome (Figure 1b). The results of this analysis, shown in Figure 6, are revealing. Poorly conserved residues are not frequently found in our experimental results (i.e., the right bottom triangle is empty). To reiterate, residues that change between different 3a sequences were not identified in our study as genetically selected mutations that impair protein function. Interestingly, however, the opposite correlation did not exhibit such an injective correlation with our results: While some conserved residues within the coronavirinae were identified in our study, others were not.

It is possible to rationalize the above results based on the different origins of the two data sets. Conservatively speaking, our genetic selection identified residues that negate 3a's ability to impair bacterial growth. The harm to bacteria derives from 3a's channel activity, and by deduction, our study selected residues that hinder the protein's conductivity. As a side note, we recognize the residues can impair channel activity directly, for example, by blocking the ion's pathway. However, residues may indirectly undermine conductivity, for example, by preventing the protein from folding correctly.

In contrast, the conservation observed within different members of the coronavirinae may reflect additional attributes of the protein. Therefore, it follows that our study would not uncover residues that underpin other functionalities beyond channel formation. Hence, it is tempting to speculate that conserved residues not found in our exhaustive study (>108 mutations per residue) are essential to additional functions of the proteins, as of yet undiscovered. In addition, one can speculate that a significant percentage of the similarity between these homologs rises due to their shared origin and not necessarily due to selective pressure. Hence, conservation per se may be routed in factors other than functional relevance.

Next, we compared the experimentally determined deleterious mutation hotspots with the conservation pattern in a single virus species. The scenario is more straightforward in this instance, as shown in Table 1. As expected, there is an inverse correlation between the two datasets, whereby highly mutable residues in SARS-CoV-2 variants (see Figure 5) were scarcely observed in our study. In contrast, the top deleterious mutations in our study were not observed in circulating strains. Hence, one may speculate that the divergence within a single species is insufficient to report on functions other than channel formation.

We also note that mutation T223I is an anomaly in the results since it is enriched in the selection but frequently occurs in the natural population. Due to the use of six primes when dividing the sequencing library (an inherent deficiency of our next-generation sequencing approach), we could not detect a combination of more than one mutation. Therefore, we can conclude that, in this case, an additional mutation may have compensated for the effect of the first mutation.

3 | CONCLUSIONS

The purpose of this study was to examine residues whose activity is essential to protein conductivity exhaustively. Using genetic selection, we could identify various mutations that do not congregate in one particular protein region but are dispersed throughout the sequence. Comparison to sequence conservation analysis within the coronavirinae was particularly insightful. While deleterious mutations found in the study were invariably conserved, not every conserved residue was found in our research. These insights may highlight residues that, while not crucial for channel formation, may be essential to other attributes. Furthermore, while our genetic selection technique is backed up by former studies on other
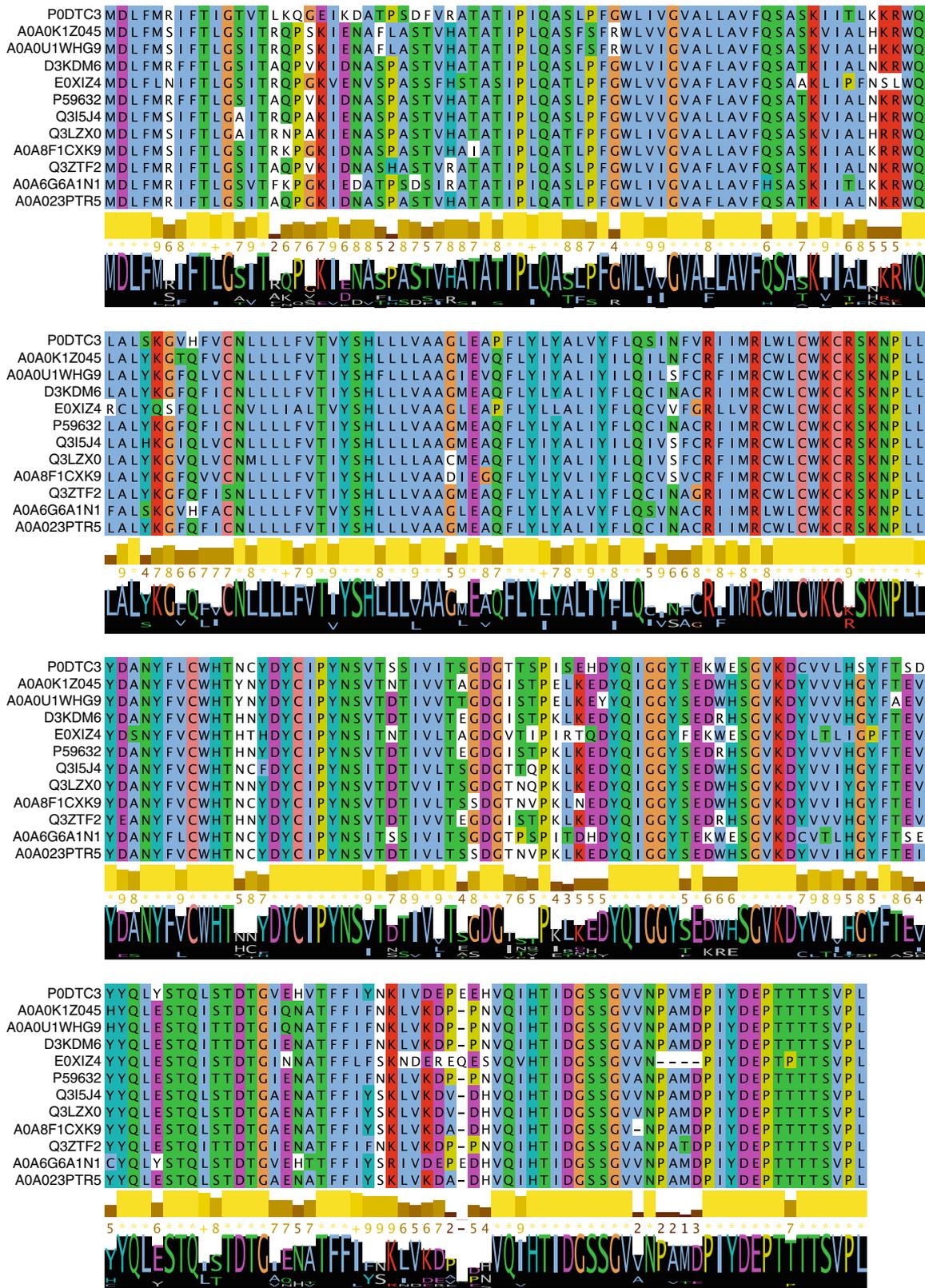
FIGURE 4 Multiple sequence alignment of the ORF3a family. The sequences from top to bottom are: P0DTC3, A0A0K1Z045, A0A0U1WHG9, D3KDM6, E0XIZ4, P59632, Q3I5J4, Q3LZX0, A0A8F1CXK9, Q3ZTF2, A0A6G6A1N1, and A0A023PTR5. The yellow diagram represents the calculated alignment score, and at the bottom the logo at that position. Amino acids are color-coded according to their physicochemical properties given by Jalview (Waterhouse et al., 2009), which was used to conduct the multiple sequence alignment using the default parameters
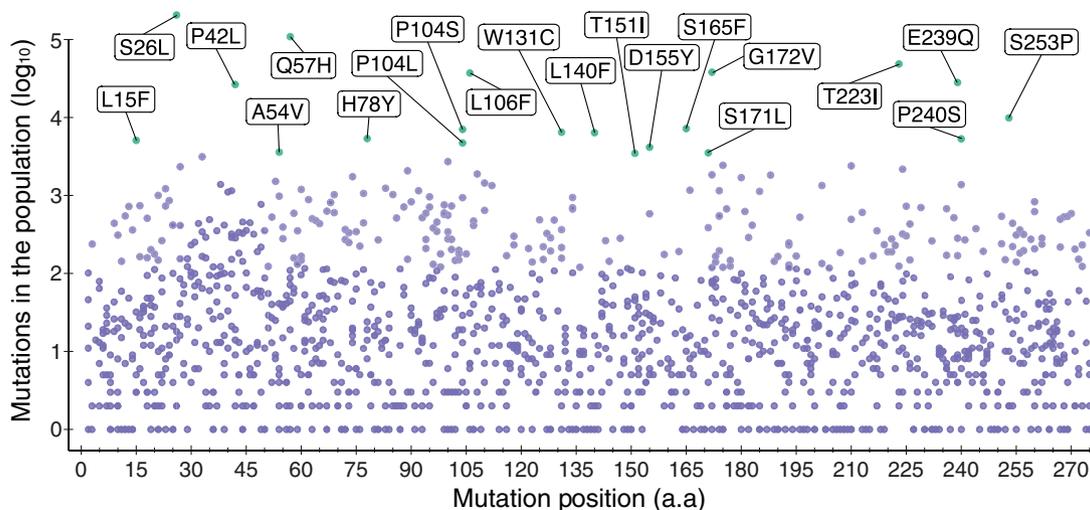
FIGURE 5 Mutations (relative to the P0DTC3 variant) found in 807,520 circulating variants of severe acute respiratory syndrome coronavirus 2 ORF3a. Particular abundant substitutions are indicated. The 20 most common mutations found in this study are presented as green dots with the mutation name. The database was downloaded from the NCBI virus database and analyzed by R studio (R Core Team, 2020)
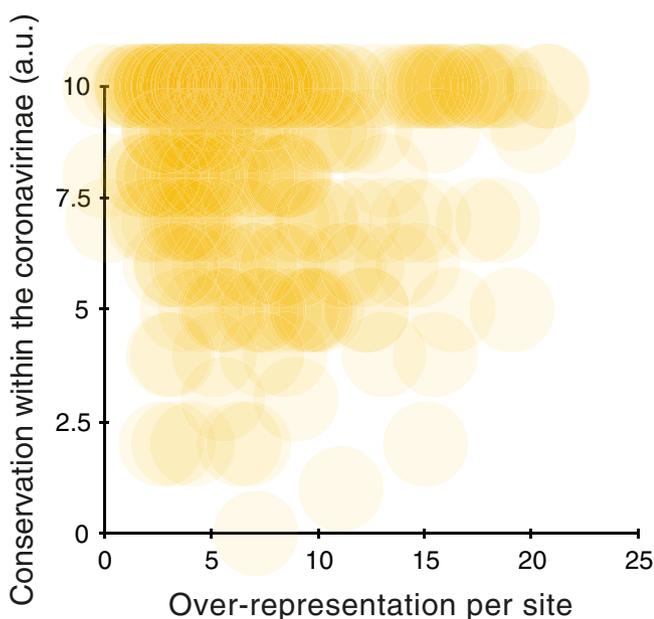
FIGURE 6 Comparison between 3a conservation within the coronavirinae and the results of this study. Each point in the graph represents a single residue in the protein, whose distance from the abscissa and ordinate corresponds to its prevalence in our induced library and conservation, respectively. Individual data are depicted in Figures 2 and 4.

proteins (Astrahan et al., 2011; Lahiri & Arkin, 2022; Taube et al., 2014; Tomar et al., 2021a, 2021b, 2022), we cannot fully discount the possibility that the *E. coli* system in this instance is not fully representative of the native environment in the viral host. Thus the functional threshold may not be the same as the physiological one.

4 | METHODS AND MATERIALS

4.1 | Library construction

4.1.1 | ORF3a cloning

ORF3a (gene number: P0DTC3) gene was ordered as a gBlock from IDT (Coralville, Iowa). The strain P0DTC3 of ORF3a was used as a reference genome because this strain was the first to be identified from the first patients in Wuhan, China (Wu et al., 2020). It was subsequently inserted using Gibson assembly (NEBL; Ipswich, Massachusetts) into the pMAL-p2X vector (NEBL), in which the protein is fused to the carboxy terminus of the maltose binding protein. The reaction was incubated for 30 min at 50°C. Subsequently, the plasmid was transformed into Escherichia coli DH10B cells (Durfee et al., 2008) and seeded on an LB agar plate with 100 μg/ml ampicillin at 37°C overnight. This construct has proven particularly suited for expressing numerous other viroporins (Astrahan et al., 2011; Taube et al., 2014; Tomar et al., 2019, 2021b, 2022).

4.1.2 | Random library construction

Random mutagenesis of the WT ORF3a gene was performed with the Genemorph II Random Mutagenesis Kit (Agilent; Santa Clara, California) according to the manufacturer's conditions for medium mutation frequency. The resulting PCR products were subsequently inserted in the pMAL-p2X vector (NEBL) using Gibson assembly.

TABLE 1 Correlation between ORF3a deleterious mutations found in this study and the protein sequence variation of circulating SARS-CoV-2 variants.

Mutation	Deleterious mutations found in study		Mutations found in circulating variants	
	Occurrences	Over-representation (induced/uninduced)	Occurrences	Abundance in population (%)
S209T	3152	3.8675	4	0.0005
S60P	2316	5.4112	95	0.0117
N137T	2179	3.8028	0	0
A99V	2129	4.883	757	0.0931
P138T	1940	7.4046	1	0.0001
A39V	1899	2.277	5	0.0006
F87Y	1889	5.0643	1	0.0001
L129F	1885	6.9557	483	0.0594
D142E	1843	4.9543	31	0.0038
L214P	1747	4.044	0	0
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
P104S	113	0.516	7029	0.86
S165F	9	1.125	7204	0.89
S253P	66	0.1451	9876	1.22
P42L	15	0.6	26,497	3.26
E239Q	3	1.5	28,099	3.46
L106F	38	0.4872	37,157	4.57
G172V	19	0.5135	38,043	4.68
T223I	970	2.1749	48,561	5.97
Q57H	59	0.4876	108,744	13.38
S26L	64	0.9552	205,955	25.34

Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

4.1.3 | Preliminary analysis using sanger sequencing

Confirmation of mutation rate was conducted by Sanger sequencing. Specifically, the transformation of the library was performed into Escherichia coli XL10-Gold bacteria (Agilent). Forty-four colonies were isolated and grown overnight in LB at 37°C. The growth culture was diluted, and the bacteria were grown in secondary culture in LB until reaching an O.D.600 of 0.07–0.1. The bacteria were then divided into 96-well flat-bottomed plates (Nunc; Roskilde, Denmark) to a final volume of 100 μl per well with 100 μM of isopropyl-β-ᴅ-1-thiogalactopyranoside (IPTG). The plate was incubated for 16 h at 30°C in a Synergy 2 multidetection microplate reader (Biotek; Winooski, Vermont) or Infinite 20 (Tecan Group; Männedorf, Switzerland) at a constant high shaking rate. O.D.600 readings were recorded every 15 min. For all measurements, duplicates were conducted. Ampicillin and D-glucose were added to all growth media to a final concentration of 100 μg/ml and 1%, respectively. Finally, the plasmid of each colony was isolated and sequenced by Sanger sequencing.

4.2 | Genetic selection

The genetic selection was conducted as described previously in detail (Tomar et al., 2021a). In brief, increased channel concentration results in membrane permeabilization and commensurate bacterial death. Control of protein expression is achieved by modifying the amount of the IPTG. Bacterial growth under this genetic assay allowed only colonies with specific mutations that impair protein function to survive and continue growing on the IPTG plate.

In total, 100 μl of DH10B bacteria transformed with the two libraries were seeded into two 30 ml agar plates

with 25 μl /100 μl with 100 μg /ml ampicillin and 1% D-glucose. In order to control our genetic selection, 100 μM of IPTG was added to the plate. After growing the plates overnight at 37°C, the colonies were collected, and the mutants′ plasmids were extracted from the colonies using a midi-prep kit (Biomiga; San Diego, California).

4.3 | Deep-sequencing

Next-generation sequencing employed Illumina dye sequencing technology (San Diego, California) following the manufacturer′s protocol. Due to the length of the gene, the sequence was read in three parts using six primers. Each primer was designed to have an overhang sequence according to the required protocol.

4.3.1 | Bioinformatic analyses

Initial data filtration was conducted using the following Galaxy Bioinformatic tools: Filter fastq-Quality filter—used to filter reads above the quality of 20 (Blankenberg et al., 2010); Fastq trimmer—used to trim the end or the start of reads with quality below average (Blankenberg et al., 2010); Reverse-complement—used to reverse complement the REV reads to match 5′–3′ reading (Gordon & Hannon, 2010). For visualization, we used FastQC. All further analyses were conducted with an in-house script using the R studio programming language (R Core Team, 2020), available upon request S1. Finally, the sequences were compared with the WT sequence using pairwise alignment, which excluded sequences with stop codons, reads containing missing nucleotides, or reads containing more than ten mutations per segment.

4.4 | ORF3a mutations in population

Two different strategies were employed to evaluate the sequence variation of ORF3a: diversity between various members of the coronaviridae and within variants of SARS-CoV-2.

Eleven different proteins with diverse sequence similarity (8%–100%) were selected for the ORF3a sequence conservation analysis. The sequences were compared by pairwise alignment with the EMBOSS needle tool employing the BLOSUM62 substitution matrix (Needleman & Wunsch, 1970). MSA was conducted using Jalview (Waterhouse et al., 2009).

The NCBI VIRUS database (https://www. ncbi.nlm. nih.gov/labs/virus/vssi/) was accessed on May 1, 2022, to download SARS-CoV-2 ORF3a sequences. Out of 2.3 million sequences, 807,520 complete sequences with a length of 275 amino acids were utilized for further analyses using in-house R programming scripts.

AUTHOR CONTRIBUTIONS
Amit Benazraf: Conceptualization (equal); data curation (equal); formal analysis (lead); investigation (lead); methodology (lead); software (lead); visualization (lead); writing – original draft (equal); writing – review and editing (equal). **Isaiah T. Arkin:** Conceptualization (equal); data curation (equal); funding acquisition (lead); investigation (equal); project administration (lead); supervision (lead); writing – original draft (equal); writing – review and editing (equal).

ORCID
Isaiah T. Arkin https://orcid.org/0000-0002-7659-1746

REFERENCES
Assa D, Alhadeff R, Krugliak M, Arkin IT. Mapping the resistance potential of Influenza's H$^+$ channel against an antiviral blocker. J Mol Biol. 2016;428(20):4209–17.

Astrahan P, Flitman-Tene R, Bennett ER, Krugliak M, Gilon C, Arkin IT. Quantitative analysis of influenza M2 channel blockers. Biochim Biophys Acta. 2011;1808(1):394–8.

Bhowmick R, Halder UC, Chattopadhyay S, Chanda S, Nandi S, Bagchi P, et al. Rotaviral enterotoxin nonstructural protein 4 targets mitochondria for activation of apoptosis during infection. J Biol Chem. 2012;287(42):35004–20.

Bianchi M, Borsetti A, Ciccozzi M, Pascarella S. SARS-Cov-2 ORF3a: mutability and function. Int J Biol Macromol. 2021;170:820–6.

Blankenberg D, Gordon A, Von Kuster G, et al. Manipulation of FASTQ data with galaxy. Bioinformatics. 2010;26(14):1783–5.

Boni MF, Lemey P, Jiang X, Lam TTY, Perry BW, Castoe TA, et al. Evolutionary origins of the SARS-CoV-2 Sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol. 2020;5(11):1408–17.

Bradwell K, Combe M, Domingo-Calap P, Sanjuán R. Correlation between mutation rate and genome size in Riboviruses: mutation rate of bacteriophage Qβ. Genetics. 2013;195(1):243–51.

Charlton FW, Pearson HM, Hover S, Lippiat JD, Fontana J, Barr JN, et al. Ion channels as therapeutic targets for viral

infections: further discoveries and future perspectives. Viruses. 2020;12(8):844.

Cirino PC, Mayer KM, Umeno D. Generating mutant libraries using error-prone PCR. New York: Springer; 2003. p. 3–9.

Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed. 2020;91(1):157–60.

Davies WL, Grunert RR, Haff RF, McGahen JW, Neumayer EM, Paulshock M, et al. Antiviral activity of 1-Adamantanamine (amantadine). Science. 1964;144(3620):862–3.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. Genetics. 1998;148(4):1667–86.

Durfee T, Nelson R, Baldwin S, Plunkett G III, Burland V, Mau B, et al. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. J Bacteriol. 2008;190(7):2597–606.

Ewart G, Sutherland T, Gage P, Cox G. The Vpu protein of human immunodeficiency virus type 1 forms cation-selective ion channels. J Virol. 1996;70(10):7108–15.

Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence-function relationships. Nat Methods. 2010;7(9):741–6.

Gonzalez ME, Carrasco L. Viroporins. FEBS Lett. 2003;552(1): 28–34.

Gordon A, Hannon G, Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished); 2010. p. 5. http://hannonlab. cshledu/fastx_toolkit

Hay AJ, Wolstenholme AJ, Skehel JJ, Smith MH. The molecular basis of the specific anti-influenza action of amantadine. EMBO J. 1985;4(11):3021–4.

Hille B. Ionic channels of excitable membranes. 2nd ed. Sunderland, Mass: Sinauer Associates; 1992.

Hossain MK, Hassanzadeganroudsari M, Apostolopoulos V. The emergence of new strains of SARS-CoV-2. What does it mean for COVID-19 vaccines? Expert Rev Vaccines. 2021;20(6): 635–8.

Hover S, Foster B, Barr JN, Mankouri J. Viral dependence on cellular ion channels–an emerging anti-viral target? J Gen Virol. 2017;98(3):345–51.

Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33–8.

Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. SARS-CoV-2 and ORF3a: nonsynonymous mutations, functional domains, and viral pathogenesis. Msystems. 2020;5(3):e00266-20.

Kern DM, Sorum B, Mali SS, Hoel CM, Sridharan S, Remis JP, et al. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. Nat Struct Mol Biol. 2021;28(7):573–82.

Lahiri H, Arkin IT. Searching for blockers of dengue and West Nile virus Viroporins. Viruses. 2022;14(8):1750.

Lauber C, Goeman JJ, Parquet MC, et al. The footprint of genome architecture in the largest genome expansion in RNA viruses. PLoS Pathog. 2013;9(7):e1003500.

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020; 395(10224):565–74.

Lu W, Zheng BJ, Xu K, Schwarz W, du L, Wong CKL, et al. Severe acute respiratory syndrome-associated coronavirus 3a protein forms an Ion Channel and modulates virus release. Proc Natl Acad Sci U S A. 2006;103(33):12540–5.

Luk HK, Li X, Fung J, Lau SK, Woo PC. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. Infect Genet Evol. 2019;71:21–30.

Lytras S, Hughes J, Martin D, Swanepoel P, de Klerk A, Lourens R, et al. Exploring the natural origins of SARS-CoV-2 in the light of recombination. Genome Biol Evol. 2022;14(2):evac018.

Madan V, Castelló A, Carrasco L. Viroporins from RNA viruses induce caspase-dependent apoptosis. Cell Microbiol. 2008; 10(2):437–51.

Mousavizadeh L, Ghasemi S. Genotype and phenotype of COVID-19: their roles in pathogenesis. J Microbiol Immunol Infect. 2021;54(2):159–63.

Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 1970;48(3):443–53.

Nieva JL, Madan V, Carrasco L. Viroporins: structure and biological functions. Nat Rev Microbiol. 2012;10(8):563–74.

Owen E. The importance of being Earnest about shank and thigh kinematics especially when using ankle-foot orthoses. Prosthet Orthot Int. 2010;34(3):254–69.

Pal A, Dobhal S, Dey KK, Sharma AK, Savani V, Negi VS. Polymorphic landscape of SARS-CoV-2 genomes isolated from Indian population in 2020 demonstrates rapid evolution in ORF3a, ORF8, Nucleocapsid phosphoprotein and spike glycoprotein. Comput Biol Chem. 2021;95:107594.

Pinto LH, Holsinger LJ, Lamb RA. Influenza virus M2 protein has Ion Channel activity. Cell. 1992;69(3):517–28.

R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.

Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. Trends Biochem Sci. 1995;20(9):374–6.

Steinmann E, Pietschmann T. Hepatitis C virus p7—a Viroporin crucial for virus assembly and an emerging target for antiviral therapy. Viruses. 2010;2(9):2078–95.

Taube R, Alhadeff R, Assa D, Krugliak M, Arkin IT. Bacteria-based analysis of HIV-1 Vpu channel activity. PLoS One. 2014;9(10): e105387.

Tomar PPS, Krugliak M, Arkin IT. Blockers of the Sars-CoV-2 3a channel identified by targeted drug repurposing. Viruses. 2021a;13(3):532.

Tomar PPS, Krugliak M, Arkin IT. Identification of SARS-CoV-2 E Channel blockers from a repurposed drug library. Pharmaceuticals (Basel). 2021b;14(7):604.

Tomar PPS, Krugliak M, Singh A, Arkin IT. Zika M-A potential Viroporin: mutational study and drug repurposing. Biomedicine. 2022;10(3):641.

Tomar PPS, Oren R, Krugliak M, Arkin IT. Potential Viroporin candidates from pathogenic viruses using bacteria-based bioassays. Viruses. 2019;11(7):632.

UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480–9.

Upadhyay V, Lucas A, Panja S, Miyauchi R, Mallela KM. Receptor binding, immune escape, and protein stability direct the natural selection of SARS-CoV-2 variants. J Biol Chem. 2021;297(4): 101208.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.

Wozniak AL, Griffin S, Rowlands D, Harris M, Yi MK, Lemon SM, et al. Intracellular proton conductance of the hepatitis C virus p7 protein and its contribution to infectious virus production. PLoS Pathog. 2010;6(9):e1001087.

Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–9.

Wu S, Tian C, Liu P, Guo D, Zheng W, Huang X, et al. Effects of SARS-CoV-2 mutations on protein structures and Intraviral protein–protein interactions. J Med Virol. 2021;93(4): 2132–40.

Zhou W, Wang W. Fast-spreading SARS-CoV-2 variants: challenges to and new design strategies of COVID-19 vaccines. Signal Transduct Target Ther. 2021;6(1):1–6.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Benazraf A, Arkin IT. Exhaustive mutational analysis of severe acute respiratory syndrome coronavirus 2 ORF3a: An essential component in the pathogen's infectivity cycle. Protein Science. 2023;32(1):e4528. https:// doi.org/10.1002/pro.4528