


# Genetic and Structural Analysis of SARS-CoV-2 Spike Protein for Universal Epitope Selection

Christopher Markosian,<sup>1,2</sup> Daniela I. Staquicini,<sup>1,2</sup> Prashant Dogra,<sup>3,4</sup> Esteban Dodero-Rojas,<sup>5</sup> Joseph H. Lubin,<sup>6</sup> Fenny H.F. Tang,<sup>1,2</sup> Tracey L. Smith,<sup>1,2</sup> Vinícius G. Contessoto,<sup>5</sup> Steven K. Libutti,<sup>7,8</sup> Zihui Wang,<sup>3,4</sup> Vittorio Cristini,<sup>3,9,10,11</sup> Sagar D. Khare,<sup>6,7,12</sup> Paul C. Whitford,<sup>13</sup> Stephen K. Burley ,<sup>\*,6,7,12,14,15</sup> José N. Onuchic,<sup>\*,5,16,17,18</sup> Renata Pasqualini,<sup>\*,†1,2</sup> and Wadih Arap<sup>\*,†1,19</sup>

<sup>1</sup>Rutgers Cancer Institute of New Jersey, Newark, NJ 07101, USA

<sup>2</sup>Division of Cancer Biology, Department of Radiation Oncology, Rutgers New Jersey Medical School, Newark, NJ 07103, USA

<sup>3</sup>Mathematics in Medicine Program, Houston Methodist Research Institute, Houston, TX 77030, USA

<sup>4</sup>Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY 10022, USA

<sup>5</sup>Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA

<sup>6</sup>Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

<sup>7</sup>Rutgers Cancer Institute of New Jersey, New Brunswick, NJ 08901, USA

<sup>8</sup>Department of Surgery, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ 08901, USA

<sup>9</sup>Neal Cancer Center, Houston Methodist Research Institute, Houston, TX 77030, USA

<sup>10</sup>Department of Imaging Physics, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

<sup>11</sup>Physiology, Biophysics, and Systems Biology Program, Graduate School of Medical Sciences, Weill Cornell Medicine, New York, NY 10022, USA

<sup>12</sup>Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

<sup>13</sup>Department of Physics and Center for Theoretical Biological Physics, Northeastern University, Boston, MA 02115, USA

<sup>14</sup>RCSB Protein Data Bank, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

<sup>15</sup>RCSB Protein Data Bank, San Diego Supercomputer Center, and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92067, USA

<sup>16</sup>Department of Biosciences, Rice University, Houston, TX 77005, USA

<sup>17</sup>Department of Chemistry, Rice University, Houston, TX 77005, USA

<sup>18</sup>Department of Physics and Astronomy, Rice University, Houston, TX 77005, USA

<sup>19</sup>Division of Hematology/Oncology, Department of Medicine, Rutgers New Jersey Medical School, Newark, NJ 07103, USA

\***Corresponding authors:** E-mails: wadih.arap@rutgers.edu (W.A.); renata.pasqualini@rutgers.edu (R.P.); stephen.burley@rcsb.org (S.K.B.); jonuchic@rice.edu (J.N.O.).

†These authors jointly supervised this work.

**Associate editor:** Banu Ozkan

## Abstract

**Evaluation of immunogenic epitopes for universal vaccine development in the face of ongoing SARS-CoV-2 evolution remains a challenge. Herein, we investigate the genetic and structural conservation of an immunogenically relevant epitope (C662–C671) of spike (S) protein across SARS-CoV-2 variants to determine its potential utility as a broad-spectrum vaccine candidate against coronavirus diseases. Comparative sequence analysis, structural assessment, and molecular dynamics simulations of C662–C671 epitope were performed. Mathematical tools were employed to determine its mutational cost. We found that the amino acid sequence of C662–C671 epitope is entirely conserved across the observed major variants of SARS-CoV-2 in addition to SARS-CoV. Its conformation and accessibility are predicted to be conserved, even in the highly mutated Omicron variant. Costly mutational rate in the context of energy expenditure in genome replication and translation can explain this strict conservation. These observations may herald an approach to developing vaccine candidates for universal protection against emergent variants of coronavirus.**

**Key words:** COVID-19, epitope, variants.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

## Introduction

The ongoing Coronavirus disease 2019 (COVID-19) global pandemic continues to pose an unprecedented health threat to humankind and the potential for evolution of more infectious and/or virulent SARS-CoV-2 variants with resistance to vaccines remains a considerable concern. Within 1 year after SARS-CoV-2 first emerged clinically, the United States Food & Drug Administration (FDA) began issuing Emergency Use Authorization (EUA) for COVID-19 vaccines for the general population. Less than 1 year out from their initial deployment, we began experiencing two setbacks in the fight against the virus. First, the Delta and Omicron “Variants of Concern” (VOCs) have proven to be more transmissible and more virulent than other previously described variants. Second, immunological protection unfortunately appears to wane in the months following either initial infection or full vaccination. Although the FDA granted full approval for the Moderna and Pfizer-BioNTech vaccines, COVID-19 vaccines in general have been reported to have decreased effectiveness against Delta and Omicron variant infection and symptomatic illness (Fowlkes et al. 2021; Andrews et al. 2022).

The spike (S) protein, which is one of the main structural proteins of SARS-CoV-2 (Cao et al. 2021), is the primary target of neutralizing antibodies generated following infection (Liu et al. 2020) and has formed the basis of nearly all first-generation COVID-19 vaccines (Dai and Gao 2021; Jalkanen et al. 2021). A recent analysis of more than 300,000 viral genomes suggested that mutations within the S protein represent one of the main pathways of adaptive evolution in SARS-CoV-2 (Rochman et al. 2021), thereby raising concerns about the potential for resistance to neutralizing antibodies acquired through either vaccination or previous COVID-19 infection (Chen et al. 2021; Wang et al. 2021; Zhou and Wang, 2021). Despite recent developments supporting the concept of “inescapable” monoclonal antibodies with broad activity against multiple variants of SARS-CoV-2 (Starr et al. 2021) and the FDA-approved antiviral drug, remdesivir, for treatment of infected patients, the ongoing genetic evolution of the virus continues to raise uncertainty regarding the possibility of new viral mutations that allow immune escape. Both the World Health Organization (WHO, 2022) and the United States Centers for Disease Control & Prevention (CDC, 2021) actively monitor new variants; by definition, the ones with greater numbers of mutations and high transmissibility have been labeled VOCs (Chen et al. 2021; Harvey et al. 2021). In the face of the highly transmissible Delta and Omicron VOCs, the FDA now recommends an additional dose (booster) of vaccine for all eligible individuals at least 6 months post full vaccination.

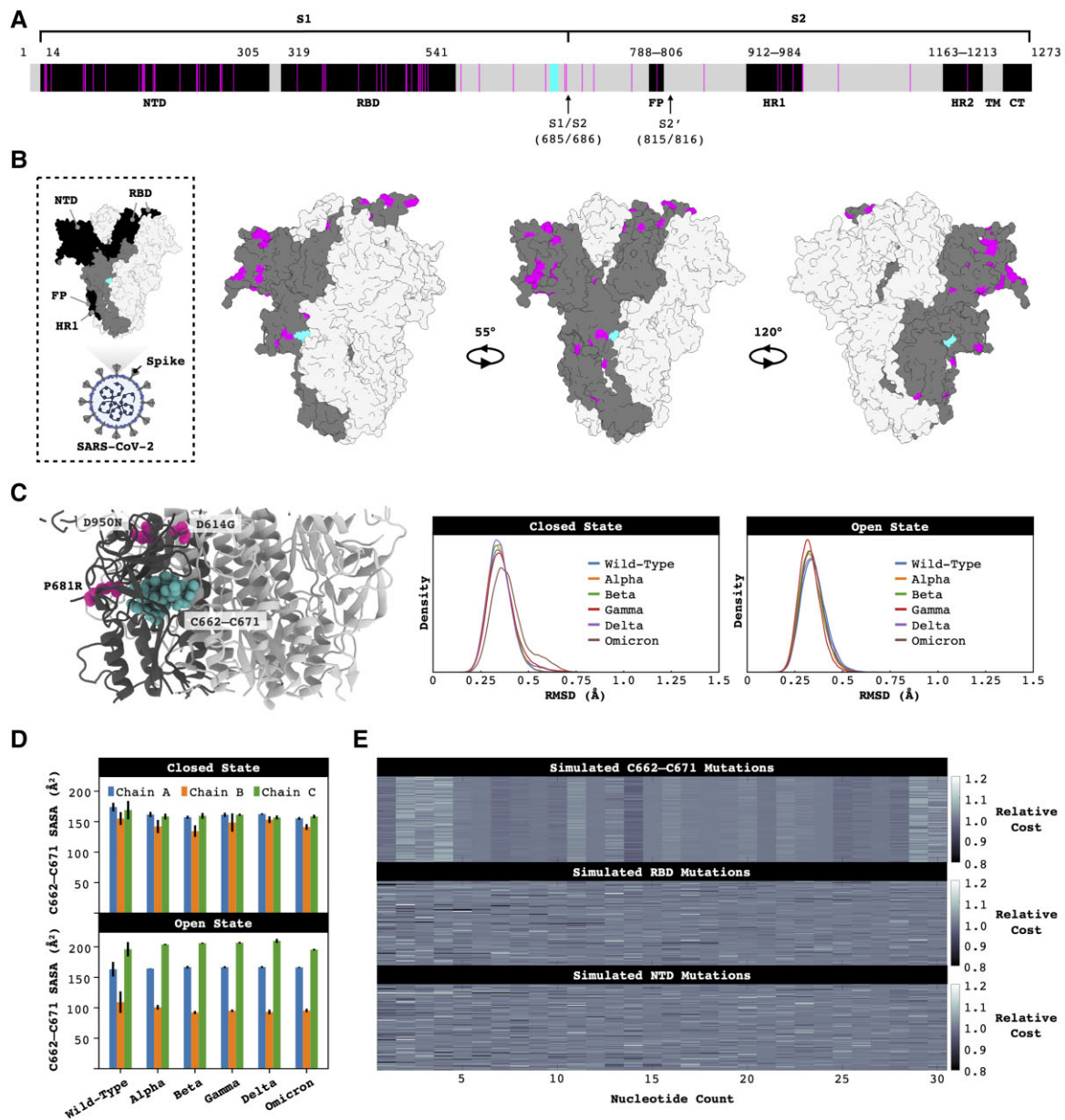
The persistence of the COVID-19 pandemic strongly re-emphasizes the need for adaptable strategies for immunization that enhance efficacy and broaden the spectrum of protection beyond those elicited by the vaccines currently in clinical use and under investigational research and development. In previous work, we have established

alternative vaccine strategies based on targeted pulmonary immunization with ligand-directed phage constructs (Staquicini, Barbu, et al. 2021). Phage particles are generally harmless to humans, but induce a potent nonspecific humoral response; they are inexpensively produced and distributed at an industrial scale with no temperature-controlled supply cold-chain requirement (Staquicini, Tang, et al. 2021), which represents one of the major impediments to moving first-generation COVID-19 vaccines into resource-limited settings. To drive the phage-based vaccine concept towards clinical translation, phage particles were engineered to display a highly immunogenic cyclic decapeptide epitope of the SARS-CoV-2 S protein (residues C662–C671, sequence CDIPIGAGIC) (hereafter referred to as the C662–C671 epitope) fused with a lung-directed motif that facilitates uptake and distribution after targeted aerosol delivery (Staquicini, Barbu, et al. 2021; Staquicini, Tang, et al. 2021). The cyclization of this epitope within the native full-length S protein is enabled by a disulfide bridge between the flanking cysteine residues. The C662–C671 epitope closely recapitulates its native cyclic conformation with the same disulfide bridge when displayed on the surface of the phage and elicits a strong and specific antibody response in mice. This experimental evidence strongly suggests that such a strategy fulfills the functional attributes for induction of a broad immune response to SARS-CoV-2 (Staquicini, Tang, et al. 2021). In order to assess the utility of our approach against existing and/or new variants of SARS-CoV-2, we herein investigated the C662–C671 epitope further by serially performing comparative sequence, structural, simulation, and computational analyses in the context of the emerging mutational impact of the five current and former VOCs.

## Results

### C662–C671 Sequence is Conserved across Major Variants

To examine the nature of mutational selective pressure on SARS-CoV-2, we retrieved S protein sequences from 14 major variants that have arisen throughout the pandemic (supplementary table S1, Supplementary Material online). We observed that although these variants together harbor substitutions or deletions at 86 different amino acid positions distributed throughout the 1,273-residue protein, the C662–C671 epitope is strictly conserved (supplementary fig. S1, Supplementary Material online), an attribute that could be exploited towards the development of a broad-spectrum, possibly universal, COVID-19 vaccine. Of note, mutations in the current and former VOCs (Alpha, Beta, Gamma, Delta, and Omicron) are primarily clustered in the N-terminal domain (NTD) and the receptor-binding domain (RBD) of S protein (fig. 1A). Based on the antibody repertoires identified in the plasma of SARS-CoV-2-infected individuals, both RBD and NTD are the targets of various neutralizing antibodies, with RBD being immunodominant (Harvey et al. 2021).



**Fig. 1.** Sequence and structural conservation of the C662–C671 epitope of SARS-CoV-2 S protein across variants. (A) Domain representation of the residue sites affected by missense mutations or deletions present throughout the five current and former VOCs (Alpha, Beta, Gamma, Delta, and Omicron) (magenta) and the unaffected C662–C671 epitope (cyan). Amino acid insertions (i.e., ins214EPE for Omicron) are not represented. The two arrows signify cleavage sites. NTD, N-terminal domain; RBD, receptor-binding domain; FP, fusion peptide; HR, heptapeptide repeat sequence; TM, transmembrane; CT, cytoplasmic tail. (B) Atomic structural representation of one monomer (dim gray) showing all residue sites affected by VOC mutations or deletions (magenta) and the C662–C671 epitope (cyan). SARS-CoV-2 particle created with BioRender.com. (C) The closed- and open-state (one RBD up) S protein structures of each VOC were simulated to assess the impact of mutations on the conformation of the C662–C671 epitope. The structural model of the Delta VOC that was used for simulations is shown on the left as a representative example. Shown (right) are the probability distributions of the spatial RMSD values (with respect to the wild-type strain conformation) of the C662–C671 epitope backbone atoms in each variant, calculated from 500-ns explicit-solvent simulations without glycans. (D) SASA of C662–C671 of each chain in the context of the trimeric S protein structure. Top: closed conformation; bottom: one-open (chain A) conformation. Bars represent the average SASA computed for the five lowest-energy models out of 100 generated for each variant for each conformation. Error bars indicate standard deviation. (E) Mutation effects analysis of the C662–C671 epitope, RBD, and NTD presented as a heatmap of the combined average costs of single nucleotide and amino acid synthesis for the three sequences of interest. For RBD and NTD, 100 sub-sequences of 30-nucleotide length were sampled from the wild-type sequences to generate mutants. The color legend denotes the total cost of nucleotide and amino acid production for mutants normalized to the cost of wild-type production. Each bar of the heatmap represents a single mutant.

Moreover, selection pressure on the virus from neutralizing antibodies generated *via* natural infection, vaccines, or antibody-based therapies is also consistent with the

rapid rise in mutations within the RBD, which is highly tolerant to immune-evading protein changes based on evolutionary modeling (Van Egeren et al. 2021). The C662–C671



epitope, which is located outside the NTD and the RBD, has also been shown to be targeted by neutralizing antibodies (Farrera-Soler et al. 2020; Li et al. 2020; Heffron et al. 2021). While studies with patient convalescent serum samples suggest that antigen escape from neutralizing antibodies likely drives the rise and dominance of new variants (Harvey et al. 2021), conserved epitopes across different strains, such as C662–C671, may help guide the development of a vaccine candidate against different SARS-CoV-2 variants. Moreover, such mutations within C662–C671 were only identified in 1,852 instances of 9,205,762 sequenced (a total of ~0.02% for the 10-residue epitope) SARS-CoV-2 isolates worldwide as of March 22, 2022 according to COVID-19 virusMutation Tracker (Alam et al. 2021). It is also remarkable that the C662–C671 epitope shares 100% identity with the S protein of the previous zoonotic coronavirus (SARS-CoV) that caused the original SARS outbreak in 2003 and partial homology with the corresponding S protein of the Middle East Respiratory Syndrome coronavirus from 2012 (supplementary fig. S1, Supplementary Material online), indicating the potential role of this epitope for immunization against not only SARS-CoV-2, but also other pathogenic coronaviruses. Collectively, these empirical results suggest that C662–C671 epitope might represent a privileged motif for vaccine development and perhaps even a potential target for inescapable antibodies (Starr et al. 2021).

### Epitope Conformation and Accessibility are Predicted to be Stable Despite Adjacent Mutations

To gain further insight into the structural properties of C662–C671, we performed comparative structural analysis (fig. 1B) and molecular dynamics simulations (fig. 1C and supplementary figs. S2–S6, Supplementary Material online). Prior simulations suggest that the disulfide bridge-enabled cyclization acts as a structural constraint that enables the short peptide to retain a conformation with minimal variation (Staquicini, Tang, et al. 2021). In addition, all-atom explicit-solvent simulations of the S protein indicate that mutations associated with VOCs do not impact the structural properties of this epitope. For each individual variant, 500 ns simulations were performed, and the spatial root-mean-square deviation (RMSD) of the C662–C671 epitope was calculated, relative to its configuration in the closed- and open-state structures of the wild-type sequence (fig. 1C). For each variant, the probability distribution as a function of RMSD (backbone atoms) was very similar and peaked around 0.35 Å. RMSD values that included the side-chain atoms were also very low and centered around 0.4–0.6 Å (supplementary fig. S2, Supplementary Material online). These observations strongly suggest that this region remains conformationally stable despite the presence of numerous mutations acquired elsewhere in the protein during the global COVID-19 pandemic. Further, this observation appears independent of the glycosylation state of the S protein,

though the sampled RMSD values are slightly smaller when glycans were included (supplementary figs. S3–S5, Supplementary Material online). The C662–C671 epitope is partially located on the solvent-exposed surface in both the closed- and open-state S protein (Staquicini, Tang, et al. 2021), a central structural property for vaccine development, because antibody binding requires access to epitopes. To further evaluate this empirical observation, we performed solvent accessibility surface area (SASA) calculations of the C662–C671 epitope for the wild-type strain and major variants. We demonstrate that opening of the RBD of one protomer (chain A) results in a marked increase in epitope accessibility of a separate protomer (chain C); furthermore, S protein mutants of the current and former VOCs do not display major changes in SASA patterns for each protomer compared to the wild-type strain (fig. 1D and supplementary table S2, Supplementary Material online), revealing that epitope accessibility is likely not affected.

### High Mutational Cost Supports Epitope Conservation

To understand the mechanistic basis of the relatively low mutational rate of the C662–C671 epitope in comparison to the RBD and NTD, we have reasoned that mutations within the two latter protein domains—in addition to providing a functional advantage to the virus—may also be energetically more favorable with respect to the cost of genome replication and translation of mutant sequences. Based on the observation that genome replication and translation combined are the most expensive processes of virus production for the host cells, energy limitation could induce selection pressure and genetic drift on newly incorporated elements in viral genomes (Mahmoudabadi et al. 2017). Thus, the incorporation of “lower-cost” nucleotides or amino acid residues [based on adenosine triphosphate (ATP) utilization] resulting in energy conservation could be advantageous for SARS-CoV-2 (Chen et al. 2016). To test this possibility, we used the Monte Carlo technique to generate mutants of the C662–C671 epitope, RBD, and NTD and to calculate the cost of synthesis of nucleotides and amino acids relative to the wild-type sequences of these three regions (Supplementary Material online) (Dogra et al. 2020). We have found that the evolutionary costs of production of C662–C671 mutants (per nucleotide and per residue combined) are significantly higher than those of RBD and NTD mutants ( $P < 0.0001$ ; one-way ANOVA and Tukey’s test) (fig. 1E and supplementary fig. S7, Supplementary Material online). At specific nucleotide positions, point mutations can cause up to a 4% increase in the relative cost of synthesis in the C662–C671 epitope, unlike in the RBD or NTD (supplementary fig. S7, Supplementary Material online). This value is based on the relative cost of synthesis of individual nucleotides in the mutants compared to the wild-type sequence. While these increments in the relative cost may seem marginal, when scaled to the infection level, the total cost of synthesis of millions

of copies of such mutated sequences can be quite large. Overall, this result indicates that point mutations in the C662–C671 epitope may be evolutionarily costly for the virus compared to mutations in the RBD and NTD, thereby making the former less favorable for natural selection. Such empirical analysis, while admittedly lacking an intimate description of molecular mechanisms involved in virus production, might have captured an essential phenomenon in the natural selection of mutations.

## Discussion

The rapid evolution of the SARS-CoV-2 genome (Harvey et al. 2021) highlights the need for developing a single vaccine (or a small number of vaccines) with a broad, ideally universal, spectrum of activity against viral variants. To achieve these desirable attributes, a candidate vaccine antigen should be: (1) immunogenic, (2) highly conserved (i.e., free of non-conserved missense mutations and/or single-residue deletions), (3) capable of recapitulating its endogenous conformation when displayed in heterologous contexts, and (4) surface solvent-exposed (either partially or fully) and thus suitable for recognition by antibodies and other ligands from either the humoral or cellular immune response. Several recent studies with different strategies such as artificial intelligence/machine learning (AI/ML) (Malone et al. 2020), phage display-based immunoprecipitation and sequencing technology (Shrock et al. 2020), or systematic site-directed mutagenesis (Starr et al. 2020) scanned and predicted mutational hotspot regions for the design of broad-spectrum vaccines, diagnostics, and antibody-based therapies. Notwithstanding insights into epitope mapping, it remains unclear whether these polypeptide chain segments are evolutionary stable and, therefore, suitable to control the spread of emerging COVID-19 variants worldwide.

We recently reported that the C662–C671 epitope of SARS-CoV-2 elicits the generation of anti-S protein antibodies when incorporated into an aerosol-delivered targeted phage-based vaccine prototype (Staquicini, Barbu, et al. 2021; Staquicini, Tang, et al. 2021). Here, we refined and integrated structural, biophysical, and mathematical tools to present additional experimental evidence that C662–C671 meets all the desirable criteria of an epitope for a potential broad-spectrum, hopefully universal, vaccine against SARS-CoV-2 variants. Indeed, computational methods, including but not limited to AI/ML-based tools, can facilitate timely insight into the impact of novel viral mutations during the COVID-19 pandemic (Burley et al. 2021). As of March 2022, the Omicron sub-variant BA.2, which also does not contain mutations within the C662–C671 epitope, is rapidly spreading across the globe, emphasizing the importance of our integrated strategies. In the context of phage-based vaccines, such a versatile approach is highly amenable to simple cloning strategies that would allow the display of other exposed and preserved immunogenic epitopes from SARS-CoV-2 or other viral surface glycoproteins. Targeted phage-based vaccines could, therefore, be designed to include admixtures

(cocktails) of distinct epitopes of S protein, which may help suppress the development of viral resistance or immune evasion (Baum et al. 2020). Finally, targeted phage-based vaccines are affordable to manufacture and robust in cold-free supply chain conditions, which would be particularly impactful in low-income countries with vaccine shortages and distribution challenges. From a broad perspective, this work provides an experimental framework and rational tools for the discovery and evaluation of SARS-CoV-2 epitopes with potentially universal features along with the possible generation of inescapable monoclonal antibodies as early therapeutics against COVID-19.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

We thank Dr Milka Kostic (Life Science Editors) for professional editorial services. We thank Dr Intikhab S. Alam (King Abdullah University of Science and Technology) for his assistance with querying the COVID-19 virusMutation Tracker. This work was supported in part by core services from Rutgers Cancer Institute of New Jersey (NCI Cancer Center Support Grant number P30CA072720); by the RCSB Protein Data Bank, which is jointly funded by the National Science Foundation (NSF) (grant number DBI-1832184), the US Department of Energy (grant number DOE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health (NIH) (grant number R01GM133198); by NSF grants (grant numbers CHE-1614101 and PHY-1522550 to J.N.O. and grant number MCB-1915843 to P.C.W.); and by awards from the Levy-Longenbaugh Donor-Advised Fund (to R.P. and W.A.), and the Welch Foundation (grant number C-1792 to J.N.O.). The work at the Center for Theoretical Biological Physics was also supported by the NSF (grant number PHY-2019745). J.N.O. is a Cancer Prevention Research in Texas Scholar in Cancer Research. The work at Houston Methodist Research Institute was partly supported by the Cockrell Foundation (to P.D.) and the NIH (grant number 1R01CA253865 to Z.W. and V.C.; grant number 1R01CA222007 to Z.W. and V.C.; grant number 1R01CA226537 to Z.W., V.C., R.P., and W.A.). We are grateful for the computational resources and support provided by the AMD COVID-19 High Performance Computing (HPC) Fund program, the Northeastern University Discovery cluster, and the Northeastern University Research Computing staff.

## Author Contributions

C.M., D.I.S., P.D., E.D.-R., J.H.L., F.H.F.T., V.G.C., V.C., S.D.K., P.C.W., S.K.B., J.N.O., R.P., and W.A. designed the research;

C.M., D.I.S., P.D., E.D.-R., J.H.L., F.H.F.T., T.L.S., V.G.C., and P.C.W. performed the research; S.D.K., S.K.L., and Z.W. contributed conceptual insights, reagents, or analytical tools; C.M., D.I.S., P.D., E.D.-R., J.H.L., F.H.F.T., T.L.S., V.G.C., Z.W., V.C., S.D.K., P.C.W., S.K.B., J.N.O., R.P., and W.A. analyzed the data; C.M., D.I.S., P.D., F.H.F.T., P.C.W., S.K.B., J.N.O., R.P., and W.A. wrote the initial draft of the manuscript, to which all of the authors contributed edits. J.N.O., R.P., and W.A. supervised the overall project.

## Data Availability

The data underlying this article are either available in the article and in its online supplementary material or will be shared on reasonable request to the corresponding author.

**Conflict of Interest.** D.I.S., R.P., and W.A. are listed as inventors on a patent application related to immunization strategies (International Patent Application PCT/US2020/053758, entitled *Targeted Pulmonary Delivery Compositions and Methods Using Same*). C.M., D.I.S., F.H.F.T., T.L.S., S.K.L., R.P., and W.A. are inventors on International Patent Application PCT/US2021/040392, entitled *Enhancing Immune Responses Through Targeted Antigen Expression*, which describes immunization technology adapted for COVID-19. PhageNova Bio has licensed these intellectual properties and C.M., D.I.S., F.H.F.T., T.L.S., S.K.L., R.P., and W.A. may be entitled to standard royalties. S.K.L., R.P., and W.A. are founders and equity stockholders of PhageNova Bio. S.K.L. is a Board Member and R.P. is Chief Scientific Officer and a paid consultant of PhageNova Bio. R.P. and W.A. are founders and equity shareholders of MBrace Therapeutics; R.P. is a Board Member and paid consultant and W.A. is a Scientific Advisor at MBrace Therapeutics. These arrangements are managed in accordance with the established institutional conflict-of-interest policies of Rutgers, The State University of New Jersey.

## References

- Alam I, Radovanovic A, Incitti R, Kamau AA, Alarawi M, Azhar EI, Gojobori T. 2021. CovMT: an interactive SARS-CoV-2 mutation tracker, with a focus on critical variants. *Lancet Infect Dis*. **21**:602.
- Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, Gower C, Kall M, Groves N, O'Connell AM, et al. 2022. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *N Engl J Med*. **386**:1532–1546.
- Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, Giordano S, Lanza K, Negron N, Ni M, et al. 2020. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science*. **369**:1014–1018.
- Burley SK, Arap W, Pasqualini R. 2021. Predicting proteome-scale protein structure with artificial intelligence. *N Engl J Med*. **385**: 2191–2194.
- Cao C, Cai Z, Xiao X, Rao J, Chen J, Hu N, Yang M, Xing X, Wang Y, Li M, et al. 2021. The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat Commun*. **12**:3917.
- Chen WH, Lu G, Bork P, Hu S, Lercher MJ. 2016. Energy efficiency trade-offs drive nucleotide usage in transcribed regions. *Nat Commun*. **7**:11334.
- Chen RE, Zhang X, Case JB, Winkler ES, Liu Y, VanBlargan LA, Liu J, Errico JM, Xie X, Suryadevara N, et al. 2021. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat Med*. **27**:717–726.
- Dai L, Gao GF. 2021. Viral targets for vaccines against COVID-19. *Nat Rev Immunol*. **21**:73–82.
- Dogra P, Ruiz-Ramírez J, Sinha K, Butner JD, Peláez MJ, Rawat M, Yellepeddi VK, Pasqualini R, Arap W, Sostman HD, et al. 2020. Innate immunity plays a key role in controlling viral load in COVID-19: mechanistic insights from a whole-body infection dynamics model. *ACS Pharmacol Transl Sci*. **4**:248–265.
- Farrera-Soler L, Daguier JP, Barluenga S, Vadas O, Cohen P, Pagano S, Yerly S, Kaiser L, Vuilleumier N, Winssinger N. 2020. Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma. *PLoS One*. **15**:e0238089.
- Fowlkes A, Gaglani M, Groover K, Thiese MS, Tyner H, Ellingson K, HEROES-RECOVER Cohorts. 2021. Effectiveness of COVID-19 vaccines in preventing SARS-CoV-2 infection among frontline workers before and during B.1.617.2 (Delta) variant predominance—Eight U.S. locations, December 2020–August. *MMWR Morb Mortal Wkly Rep*. 1167–1169.
- Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Peacock SJ, et al. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*. **19**:409–424.
- Heffron AS, McIlwain SJ, Amjadi MF, Baker DA, Khullar S, Armbrust T, Halfmann PJ, Kawaoka Y, Sethi AK, Palmenberg AC, et al. 2021. The landscape of antibody binding in SARS-CoV-2 infection. *PLoS Biol*. **19**:e3001265.
- Jalkanen P, Kolehmainen P, Häkkinen HK, Huttunen M, Tähtinen PA, Lundberg R, Maljanen S, Reinholm A, Tauriainen S, Pakkanen SH, et al. 2021. COVID-19 mRNA vaccine induced antibody responses against three SARS-CoV-2 variants. *Nat Commun*. **12**: 3991.
- Li Y, Lai D-Y, Zhang H-N, Jiang H-W, Tian X, Ma M-L, Qi H, Meng Q-F, Guo S-J, Wu Y, et al. 2020. Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol Immunol*. **17**:1095–1097.
- Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, Luo Y, Chan JF-W, Sahi V, Figueroa A, et al. 2020. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature*. **584**: 450–456.
- Mahmoudabadi G, Milo R, Phillips R. 2017. Energetic cost of building a virus. *Proc Natl Acad Sci USA*. **114**:E4324–E4333.
- Malone B, Simovski B, Moliné C, Cheng J, Gheorghie M, Fontenelle H, Vardaxis I, Tennøe S, Malmberg JA, Stratford R, et al. 2020. Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2 leading to universal blueprints for vaccine designs. *Sci Rep*. **10**:22375.
- Rochman ND, Wolf YI, Faure G, Mutz P, Zhang F, Koonin EV. 2021. Ongoing global and regional adaptive evolution of SARS-CoV-2. *Proc Natl Acad Sci USA*. **118**:e2104241118.
- Centers for Disease Control and Prevention (CDC). 2021. SARS-CoV-2 variant classifications and definitions. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>. Accessed 22 March 2022.
- Shrock E, Fujimura E, Kula T, Timms RT, Lee IH, Leng Y, Robinson ML, Sie BM, Li MZ, Chen Y, et al. 2020. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science*. **370**:eabd4250.
- Staquicini DI, Barbu EM, Zemans RL, Dray BK, Staquicini FI, Dogra P, Cardó-Vila M, Miranti CK, Baze WB, Villa LL, et al. 2021. Targeted phage display-based pulmonary vaccination in mice and non-human primates. *Med (NY)*. **2**:321–342.
- Staquicini DI, Tang FH, Markosian C, Yao VJ, Staquicini FI, Rojas ED, Contessoto VG, Davis D, O'Brien P, Habib N, et al. 2021. Design and proof of concept for targeted phage-based COVID-19 vaccination strategies with a streamlined cold-free supply chain. *Proc Natl Acad Sci USA*. **118**:e2105739118.

- Starr TN, Czudnochowski N, Liu Z, Zatta F, Park Y-J, Addetia A, Pinto D, Beltramello M, Hernandez P, Greaney AJ, *et al.* 2021. SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**:97–102.
- Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, *et al.* 2020. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **3**: 1295–1310.e20.
- World Health Organization (WHO). 2022. Tracking SARS-CoV-2 variants. Available from: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>. Accessed 22 March 2022.
- Van Egeren D, Novokhodko A, Stoddard M, Tran U, Zetter B, Rogers M, Pentelute BL, Carlson JM, Hixon M, Joseph-McCarthy D, *et al.* 2021. Risk of rapid evolutionary escape from biomedical interventions targeting SARS-CoV-2 spike protein. *PLoS One* **16**: e0250780.
- Wang N, Sun Y, Feng R, Wang Y, Guo Y, Zhang L, Deng Y-Q, Wang L, Cui Z, Cao L, *et al.* 2021. Structure-based development of human antibody cocktails against SARS-CoV-2. *Cell Res.* **31**: 101–103.
- Zhou W, Wang W. 2021. Fast-spreading SARS-CoV-2 variants: challenges to and new design strategies of COVID-19 vaccines. *Signal Transduct Target Ther.* **6**:226.