





Serodominant SARS-CoV-2 Nucleocapsid Peptides Map to Unstructured Protein Regions

Jacob P. Vandervaart,^{a,b*}  Nicole L. Inniss,^{a,c} Ted Ling-Hu,^d George Minasov,^{a,c} Grant Wiersum,^{a,c} Monica Rosas-Lemus,^{a,c§} Ludmilla Shuvalova,^{c,e}  Chad J. Achenbach,^d  Judd F. Hultquist,^d  Karla J. F. Satchell,^{a,b,c}  Kelly E. R. Bachta^d

^aDepartment of Microbiology-Immunology, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA

^bRobert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA

^cCenter for Structural Biology of Infectious Diseases, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA

^dDepartment of Medicine, Division of Infectious Diseases, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA

^eDepartment of Pharmacology, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA

Jacob P. Vandervaart, Nicole L. Inniss, and Ted Ling-Hu contributed equally to this work. The order of authors is based on contribution over the longest span of time.

Judd F. Hultquist, Karla J. F. Satchell, and Kelly E. R. Bachta are co-senior authors. The order of authors is based on direct supervision for data collection and analysis by the first authors in reverse order.

ABSTRACT The SARS-CoV-2 nucleocapsid (N) protein is highly immunogenic, and anti-N antibodies are commonly used as markers for prior infection. While several studies have examined or predicted the antigenic regions of N, these have lacked consensus and structural context. Using COVID-19 patient sera to probe an overlapping peptide array, we identified six public and four private epitope regions across N, some of which are unique to this study. We further report the first deposited X-ray structure of the stable dimerization domain at 2.05 Å as similar to all other reported structures. Structural mapping revealed that most epitopes are derived from surface-exposed loops on the stable domains or from the unstructured linker regions. An antibody response to an epitope in the stable RNA binding domain was found more frequently in sera from patients requiring intensive care. Since emerging amino acid variations in N map to immunogenic peptides, N protein variation could impact detection of seroconversion for variants of concern.

IMPORTANCE As SARS-CoV-2 continues to evolve, a structural and genetic understanding of key viral epitopes will be essential to the development of next-generation diagnostics and vaccines. This study uses structural biology and epitope mapping to define the antigenic regions of the viral nucleocapsid protein in sera from a cohort of COVID-19 patients with diverse clinical outcomes. These results are interpreted in the context of prior structural and epitope mapping studies as well as in the context of emergent viral variants. This report serves as a resource for synthesizing the current state of the field toward improving strategies for future diagnostic and therapeutic design.

KEYWORDS SARS-CoV-2, nucleocapsid, three-dimensional structure, immune response, epitope, variants of concern

The ongoing coronavirus disease 2019 (COVID-19) pandemic has resulted in over 6.7 million deaths globally, with more than 1 million in the United States alone (<https://coronavirus.jhu.edu>). This severe respiratory disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel coronavirus that emerged in late 2019 (1). As of early 2022, large-scale seroprevalence studies have estimated that more than 60% of Americans have been infected (2). These studies are largely based on serological testing for antibodies against the SARS-CoV-2 nucleocapsid (N) protein. This viral protein is highly abundant during acute infection and highly immunogenic, resulting in a strong and relatively ubiquitous antibody response, even in asymptomatic or mild cases (2–5). Critically,

Editor Samuel K. Campos, University of Arizona

Copyright © 2023 Vandervaart et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Karla J. F. Satchell, k-satchell@northwestern.edu.

*Present address: Jacob P. Vandervaart, Evozyne, Chicago, Illinois.

§Present address: Monica Rosas-Lemus, University of New Mexico, Albuquerque, New Mexico.

The authors declare a conflict of interest. K.J.F.S. has significant financial interest in Situ Biosciences LLC, a contract research organization that conducts coronavirus research unrelated to this work. K.J.F.S. also received payments from Buoy Health during the time of this investigation for consultation on COVID-19 topics unrelated to this work. All other authors declare no conflicts of interest.

Received 20 January 2023

Accepted 28 April 2023

Published 16 May 2023

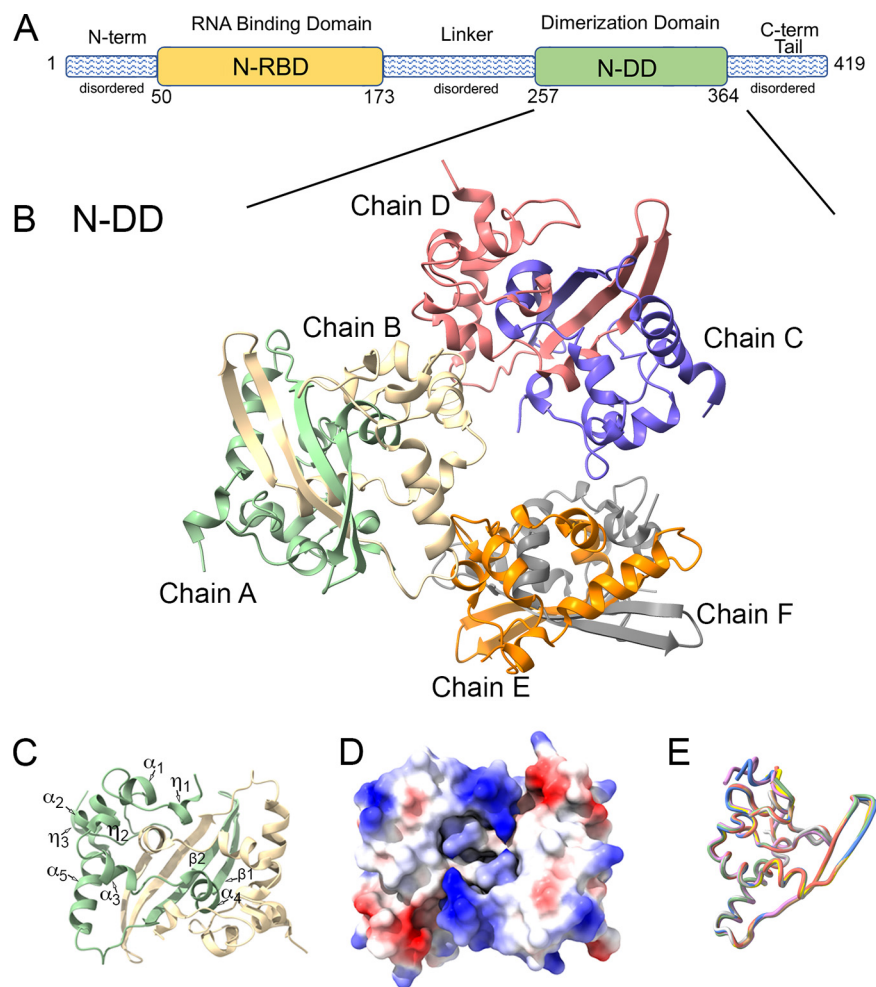


FIG 1 X-ray structure of N protein C-terminal dimerization domain. (A) Schematic diagram of the N protein with distinct domains. (B) X-ray structure of N protein N-DD solved at 2.05 Å shown as a cartoon with three dimers in the asymmetric unit (PDB accession no. 6WJI). (C) Chains A (green) and B (beige) shown as a dimer. Chain A is labeled with secondary structure elements. (D) Electrostatic projection of dimer in the same orientation as panel C. (E) Overlay licorice diagram of structure PDB accession no. 6WJI chain A (green) with other N-DD protein structures, PDB accession no. 6WZQ (blue), 6WZO (tan), 6YUN (violet), 7CE0 (yellow), 7C22 (orange), and 7DE1 (gray)

the N protein is not present in the COVID-19 vaccines authorized for use in the United States, and thus, the detection of anti-N antibodies is specifically indicative of prior infection (2).

The SARS-CoV-2 N protein binds to viral genomic RNA and oligomerizes around it to form a closed capsule that both protects the genome from antiviral responses and directs its packaging into new virions (6, 7). Beyond its role in nucleocapsid assembly and packaging, the N protein is also important during viral RNA synthesis, where it binds double-stranded RNA during viral genome replication and participates in the discontinuous transcription process necessary to generate subgenomic mRNAs (8, 9). The protein is comprised of five domains (Fig. 1A). There are two stable, proteolysis-resistant domains, the N-terminal RNA binding domain (N-RBD), which binds to virus genomic RNA, and the dimerization domain (N-DD), which facilitates protein oligomerization and has some non-specific RNA binding activity. These two domains are interspersed with intrinsically disordered domains designated the N-terminal arm (NTD), linker region (linker), and C-terminal tail (Fig. 1A) (10–13).

Antibodies targeting the N protein or the spike (S) protein can be detected in as few as 1 to 2 weeks following symptom onset and generally remain detectable up to a year after infection, with estimated seroconversion rates of roughly 50% after 2 years (14), though

TABLE 1 All publicly available X-ray structures of N protein-stable domains^a

Domain	PDB accession no.	Resolution (Å)	pH	Public release date (yr-mo-day)	Reference or source
N-RBD	6VYO	1.7	6.0	2020-03-11	Unpublished
	6M3M	2.7	6.5	2020-03-18	49
	6WKP	2.67	6.5	2020-04-29	Unpublished
	7CDZ	1.8		2020-09-02	50
	7VNU	1.95	7.5	2021-10-27	Unpublished
	7VBD	1.94	8.0	2022-08-31	Unpublished
N-DD	6WJI	2.05	7.5	2020-04-22	This study
	7C22	2.0		2020-05-20	51
	6YUN	1.44	7.8	2020-05-20	33
	6WZO	1.42		2020-05-27	52
	6WZQ	1.45	8.5	2020-05-27	52
	6ZCO	1.361		2020-07-01	33
	7CE0	1.5	9.0	2020-09-02	50
	7DE1	2.0	9.3	2021-01-27	53
	7F2B	2.0	6.2	2021-09-01	54
	7F2E	3.1	4.5	2021-10-20	54
	7O05	1.94		2022-04-13	55
	7UXX	1.85	8.3	2022-06-22	56
	7VBE	1.59	5.0	2022-07-06	Unpublished
	7VBF	1.3	8.5	2022-07-06	Unpublished

^aExcluding proteins in complex with antibodies or ligands.

this varies widely with detection method (15). Several studies have reported that higher anti-SARS-CoV-2 antibody titers are associated with more severe outcomes (16–18), though this may be dependent on time since symptom onset, immunoglobulin, and antigen (17, 19). A few studies have reported that higher anti-N antibody responses are associated with more severe outcomes, particularly with hospitalizations of long duration (20–22). While several reports have mapped or predicted antigenic regions within the N protein, these often lack consensus, and only a handful have compared antigenic regions by patient outcome (23–32).

In April 2020, we reported the first crystal structure of the N protein N-DD from SARS-CoV-2, which was made immediately accessible through the RCSB Protein Data Bank, along with structures solved by others of the same protein domain (Table 1). In the manuscript, we report the methods and statistics of this crystal structure that has been widely used in efforts to understand N protein structure and function. Furthermore, with sera collected from 34 COVID-19 patients between March and April of 2020 during the first wave of the pandemic; we mapped highly prevalent, serodominant antigenic epitopes in the N protein using a peptide-based enzyme-linked immunosorbent assay (ELISA) with 10-amino-acid resolution. These data were used to identify patient groups who produced distinct antibody responses to various peptide signatures within the N protein. Combined analysis of these peptide signatures with the structure revealed that all immunodominant peptides were derived from unstructured regions of the protein or surface-localized loops of the stable domains, including an epitope in N-DD not recognized by prior studies. These data are contextualized by a literature review of N protein epitope mapping and amino acid (aa) variation efforts to date summarizing the current state of the field.

RESULTS

Description of the first publicly released X-ray crystal structure for SARS-CoV-2 N-DD. Since the start of the COVID-19 pandemic, there has been a worldwide effort to structurally characterize each of the SARS-CoV-2 viral proteins. While the structure of the full-length N protein is yet to be fully resolved, several structures of each of the stable, proteolysis-resistant domains have been solved (Table 1). We determined the first publicly available structure of SARS-CoV-2 N-DD (residues 247 to 364), which was released on 22 April 2020 to the Protein Data Bank with accession no. 6WJI (Table 1). The crystal belonged to the $P2_12_12_1$ space group (cell parameters, $a = 43.61$, $b = 122.32$, $c = 130.63$) and was refined to 2.05 Å resolution. The asymmetric unit consisted

of six polypeptide chains, which formed three homodimers (Fig. 1B). Each protomer contained 108 residues (257 to 364) that formed 5 α helices, α_1 (269 to 275), α_2 (288 to 295), α_3 (301 to 306), α_4 (310 to 318), and α_5 (345 to 357); 3 η helices, η_1 (258 to 262), η_2 (296 to 298), and η_3 (358 to 363); and 2 β -strands, β_1 (319 to 325) and β_2 (328 to 338), all of which were connected by five loops (Fig. 1C).

The SARS-CoV-2 N-DD homodimer resembles a rectangular prism (Fig. 1C) with a buried surface area of $\sim 5,274 \text{ \AA}^2$. One large interface is formed by a four-stranded, antiparallel β -sheet, comprised of the paired protomer β -hairpins (β_1 - β_2 - β'_1 - β'_2) (best depicted between chains A and B [Fig. 1B]). A second large interface is formed by interlocked helices and loops (α_4 - η'_1) (best depicted between chains C and D [Fig. 1B]). The structure has nine chloride ions bound (see Table S1 in the supplemental material), which were positioned on the positively charged helical face of each dimer (Fig. 1D).

The overall topology observed in our structure was essentially identical to the 13 subsequently deposited structures of the SARS-CoV-2 N-DD (Table 1). The structural alignment of our structure with other representative structures showed that the backbones were nearly identical (Fig. 1E). Our structure was most similar to the high-resolution structure (PDB accession no. 6YUN) solved by Zinzula et al. (33), with a root mean square deviation (RMSD) of 0.230 \AA across all 108 pairs.

Identification of public epitopes in patient sera by tiled peptide ELISAs. The relatively high abundance and immunogenicity of the SARS-CoV-2 N protein make it a major antigen presented during acute infection (3, 24). To better understand the antigenic regions of the protein, we used an ELISA-based assay to test for the presence of serum antibodies against a tiled array of 59 overlapping linear peptides spanning the N protein (Table S2). Each array was independently probed in duplicate with serum collected from a total of 34 individuals hospitalized with COVID-19 between March and April of 2020 (cohort previously reported by Simons et al. [34]). Specimens from early in the COVID-19 pandemic were chosen for three reasons, (i) these early specimens precluded potential confounding effects due to vaccination or reinfection, (ii) circulating clades at this time had limited diversity, which minimized confounding effects due to mismatches between the peptide array and the patient SARS-CoV-2 isolate, and (iii) these specimens were previously shown to harbor antibodies against the N protein. Study participants included 20 patients admitted to the intensive care unit (ICU) and 14 patients who were admitted to the hospital but never admitted to the ICU. Serum was collected in a range from 3 to 44 days following symptom onset (median, 12.5 days; interquartile range, 9 to 16.75 days). Sera collected from three healthy individuals prior to the pandemic served as negative controls alongside a blank “no serum” control.

Previous reports have shown that several immunodominant regions of the N protein result in an antibody response in a majority of COVID-19 patients (deemed “public” epitopes), while other regions elicit a response in only a few individuals (deemed “private” epitopes) (24). To identify public epitopes detectable in participant sera, the normalized absorbance across the 34 serum samples for each peptide was plotted and interpreted relative to the mean and standard deviation of all absorbance data in the data set (see Materials and Methods). Peptides with the upper quartile of the normalized absorbance greater than two standard deviations above the mean of all data were considered highly prevalent and highly reactive (Fig. 2A, magenta bars; $n = 2$ peptides). Peptides with an upper quartile greater than one standard deviation above the mean of all data were considered highly prevalent and moderately reactive (Fig. 2A, blue bars; $n = 8$ peptides). All 10 peptides were deemed public epitopes.

Two peptides in the N protein were highly prevalent and highly reactive, peptide 34 (aa 232 to 248, SKMSGKGQQQQQTVTK) and peptide 41 (aa 281 to 297, QTQGNFGDQELIRQGT) (Fig. 2B, magenta sequences). Peptide 34 bridges the end of the linker domain and the beginning of the N-DD. The neighboring peptides to either side of peptide 34 (peptides 33 and 35) were less reactive, indicating that the most antigenic sequence within this region was contained solely within peptide 34 in the nonoverlapping sequence (Fig. 2B, underlined, aa 235 to 245, SGKGQQQQGT). However, two highly prevalent and moderately reactive peptides,

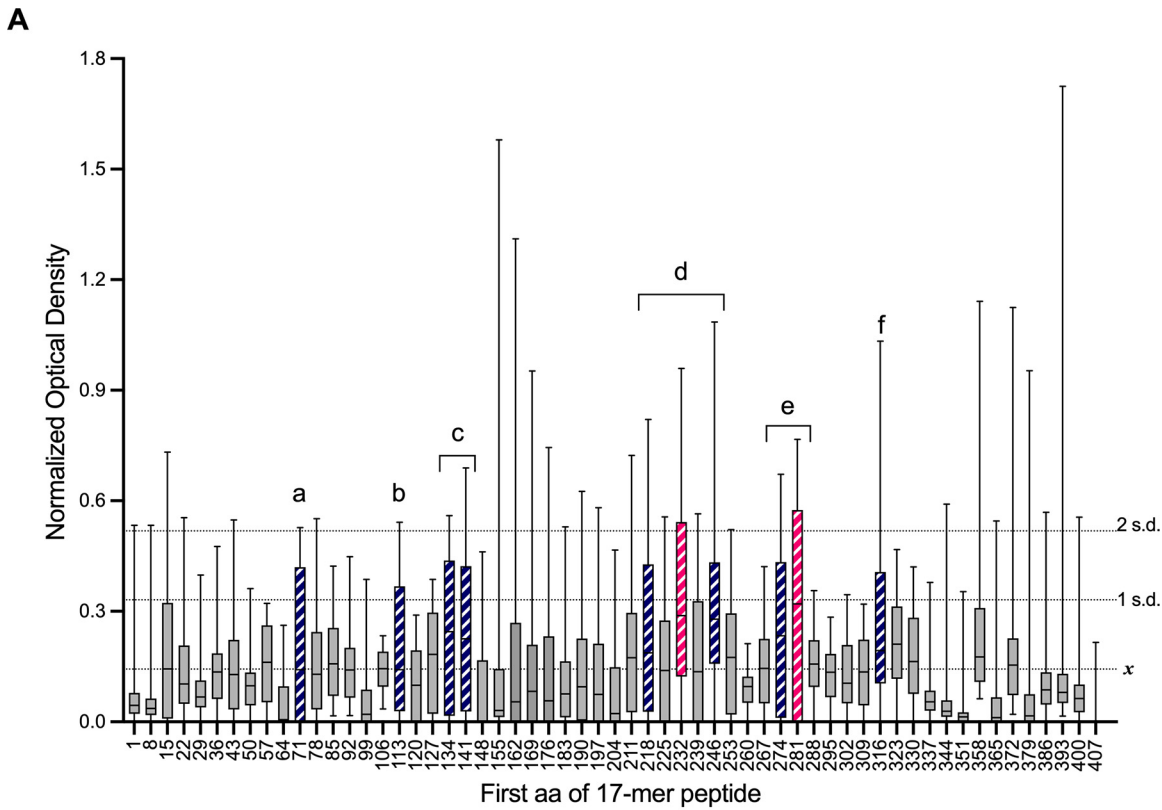


FIG 2 Public epitopes in the SARS-CoV-2 N protein. (A) Normalized serum ELISA data (optical density) from 34 patients shown as a whisker plot from minimum to maximum, with boxes representing the 25% and 75% quartiles with the bar set at the median. Magenta hatched bars indicate that the median optical density was at least 1 standard deviation (1 SD) above the mean (x) of all data, and the upper quartile exceeded 2 SDs above the mean. Blue bars indicate that the median was above x , and the upper quartile exceeded 1 SD above the mean. (B) Sequences of antigenic peptides are listed with peptides resulting in a modest response in blue and a strong response in magenta. Overlapping 10- or 11-mer sequences are underlined. Regions a to f are labeled in order of sequence of N.

peptides 32 (aa 218 to 234) and 36 (aa 246 to 262), are closely adjacent, suggesting the entire region may be antigenic (Fig. 2A, region d). The other highly prevalent, highly antigenic peptide, peptide 41 (aa 281 to 297, QTQGNFGDQELIRQGTD), was also located in the N-DD. The neighboring peptides (peptides 42 and 43, aa 288 to 311) to the C-terminal side did not elicit a strong response, though moderately reactive peptide 40 (aa 274 to 290) overlaps the N-terminal side (Fig. 2A and B, region e). This suggests that the most antigenic sequence may be in the 10-aa overlap between the peptides (Fig. 2B, underlined, aa 281 to 290, QTQGNFGDQE).

Five additional highly prevalent and moderately reactive public epitopes were also identified, including peptides 11 (aa 71 to 87, region a), 17 (aa 113 to 129, region b), 20 (aa 134 to 150, region c), and 21 (aa 141 to 157, also region c) in the N-RBD, as well as peptide 46 (aa 316 to 332, region f) in the N-DD (Fig. 2A and B).

Clustering of patient antibody profiles identified three additional private immunogenic regions. In addition to the public epitopes, there were also several instances of peptides that induced strong antibody reactions, but in fewer patient samples. To better characterize these highly reactive but less prevalent “private” epitopes, we used principal-component analysis (PCA) and *k*-means clustering to identify groups of participants with similar antibody responses and visualized these responses in a heatmap.

First, we performed a PCA on the peptide profile from each serum sample to reduce noise before clustering (Fig. 3A). A *k*-means clustering algorithm was used on the top two principal components to separate the 34 participants (and controls) into 4 groups that represent 4 distinct antibody response profiles (Fig. 3A). The optimal number of clusters was determined using the elbow method. Participant outcomes (hospitalized without ICU admission [non-ICU], hospitalized with ICU admission [ICU], and death) were overlaid (Fig. 3B), and antibody response by cluster was visualized as a heatmap (Fig. 3C). The mean responses across the peptides for each cluster are also shown as a line plot (Fig. S1).

Group 1 (Fig. 3A, orange, $n = 12$) displayed low antibody titers against nearly all peptides within the N protein (Fig. 3C; Fig. S1) and clustered tightly with the negative-control samples (Fig. 3A, group 0, gray). This group also contained the highest proportion of non-ICU participants (Fig. 3B, $n = 8/12$).

Group 2 (Fig. 3A, blue, $n = 14$) was the largest group and had antibody responses that were dominated by public epitopes identified in our prior analysis (Fig. 2, Fig. 3C, and Fig. S1). This group also contained the highest proportion of ICU participants (Fig. 3B, $n = 10/14$).

Group 3 (Fig. 3A, green, $n = 4$) followed a similar serodominant antibody response pattern to group 2, but with higher average absorbance values and unique responses to a few additional peptides (Fig. 3C; Fig. S1). This included a higher overall response to peptide 3 (aa 15 to 31, designated region g) in the N-terminal domain and to peptides 23 and 24 (aa 155 to 178, designated region h) corresponding to the N-RBD. Group 3 also showed a strong response to the linker region covered by peptides 25 to 26 and 28 to 29 (aa 169 to 213, designated region i), with the highest overall response for peptides 26 (aa 176 to 192) and 28 (aa 190 to 296). In addition, group 3 patients had uniquely robust antibody responses to several C-terminal tail peptides 52 to 58 (aa 358 to 409, designated region j). Overall, group 3 patients, in addition to the immunodominant profile captured in group 2, reacted more broadly to peptides of the N protein specifically in the unstructured regions.

Group 4 (Fig. 3A, red, $n = 4$) had the most unique response profile, with low reactivity to nearly all peptides, with the exception of a very strong response to peptides 23 and 24 (region h, aa 162 to 171), which share the overlapping sequence PQTTLPKGF (Fig. 3C and D). Group 4, and to a lesser extent, group 3, was characterized by the dominance of the response to region h in the N-RBD, as most patients who were represented in groups 1 and 2 showed no response to this epitope (Fig. 3C; Fig. S1). These responses seemed to occur in participants who experienced more severe outcomes (2 non-ICU, 5 ICU, 1 death between groups 3 and 4), though more participants would be needed to assess significance (Fig. 3B).

Antigenic peptides are distinct from RNA binding regions of the N-RBD and N-DD. To structurally contextualize the epitopes of the N protein, we mapped the immunogenic peptides to the crystal structures (Fig. 4). Four of the antigenic regions were derived from the disordered, flexible regions of the protein (Fig. 4A), including the N-terminal arm (region g, aa 15 to 31), the linker region (regions i and d, aa 169 to 213 and 218 to 269, respectively), and the C-terminal tail (region j, aa 358 to 409). The peptide from aa 235 to 245 (region d) was particularly immunogenic, eliciting the strongest response in more than 50% of patients. Unfortunately, no structures are yet available for these flexible regions.

Six of the immunogenic peptides (regions a, b, c, h, e, and f), as well as part of region d, mapped to either the N-RBD or N-DD (Fig. 4A), consistent with our prior finding that both

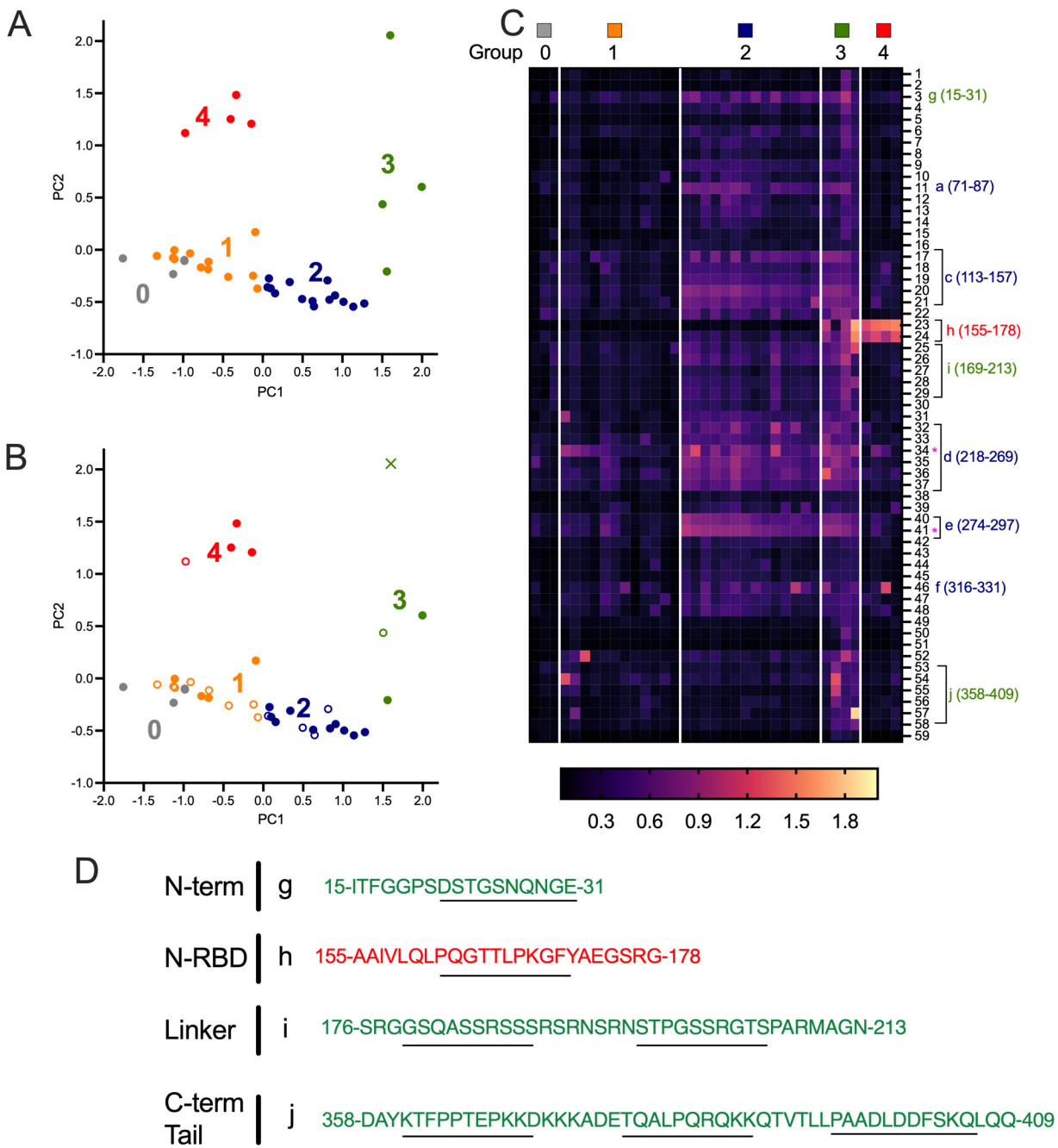


FIG 3 Private epitope identification by *k*-means clustering. (A) PCA representation of antibody profiles based on raw ELISA abundance data with *k*-means cluster labels overlaid. Group 0 represents control sera. (B) PCA representation of ELISA data with *k*-means clustering depicting patients treated in the ICU (solid circles) versus those not treated in the ICU (hollow circles). One patient death is designated with an X in group 3. (C) Magenta heatmap of raw optical read data with peptides in rows and each column representing a single patient. The most prevalent public peptides are marked with magenta asterisk. Peptide regions of interest are indicated at the right as detailed in Fig. 1B and panel D. (D) Sequences of antigenic peptides are listed with newly identified peptides resulting from group 3 in green and from group 4 in red. Overlapping 10- or 11-mer sequences are underlined. Regions g to j are labeled in order of sequence of N.

stable domains are antigenic (34). Visualization of these peptide positions on the crystal structures revealed that most of these mapped to flexible loops in the N-RBD and N-DD (Fig. 4B and C), and they were all solvent exposed (Fig. 4D and E). Furthermore, the peptides mapped to regions of the N-RBD distinct from the RNA binding regions, which were identified previously by nuclear magnetic resonance (35) (Fig. 4F). In the N-DD, the immunogenic peptides mapped to regions that projected away from the dimer interface and thus are exposed even in the dimerized form (Fig. 4G).

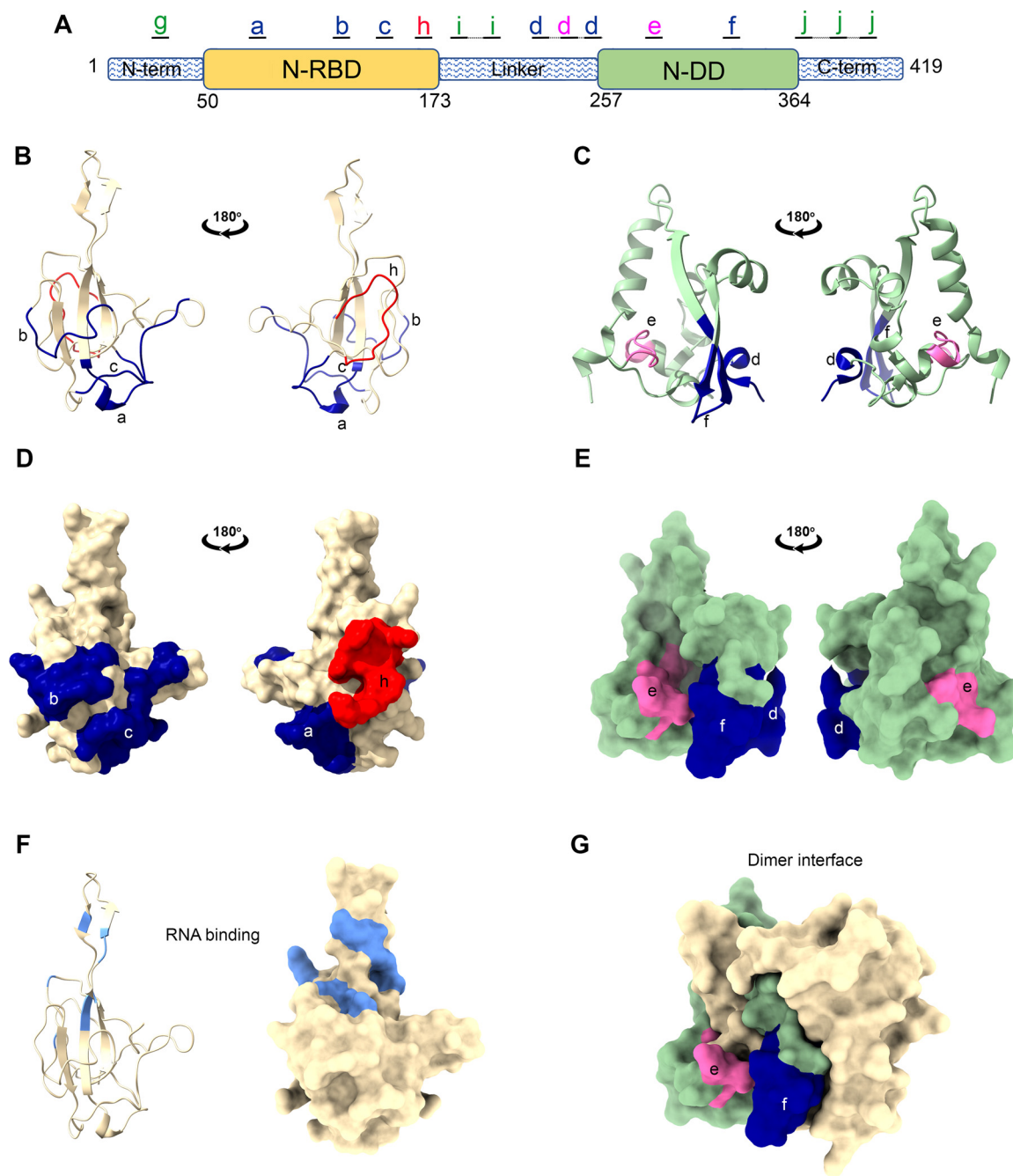


FIG 4 Mapping of antigenic peptides to structures of N-RBD and N-DD. Locations of peptides are shown on the schematic diagram (A) and ribbon (B and C) and surface (D and E) representations of the structure of the N-RBD (B and D, PDB accession no. 6VYO) and N-DD (E and F, PDB accession no. 6WJI). Colored regions are labeled as in Fig. 2B and Fig. 3D, with peptide regions that generated modest responses shown in blue (peptide regions a, b, c, d, and f), high responses shown in pink (region e), and the highly reactive private epitope seen in group 4 shown in red (region h). (F) The residues shown by nuclear magnetic resonance to bind RNA are highlighted in light blue in the ribbon. Surface projection diagrams of the N-RBD in the same orientation as shown at the left of panels A and C. (F) The surface projection of the dimer of N-DD is shown with both chains A (green) and B (tan). Chain A is in the same orientation as the left of panel E.

Altogether, these data and analyses indicate that the most immunogenic peptides map to either the flexible linker regions of the N protein or map to regions within the N-RBD and N-DD domains that are freely accessible and not bound by RNA or buried by dimerization.

Amino acid changes in common SARS-CoV-2 variants occur in N antigenic regions. The ongoing evolution of SARS-CoV-2 has resulted in the periodic emergence of variants with enhanced fitness that have spread broadly throughout the population. Early

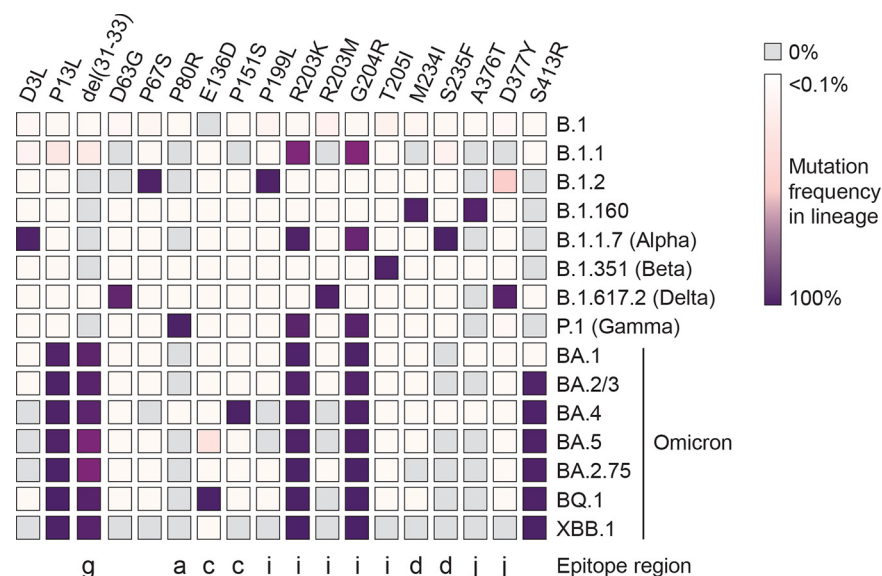


FIG 5 Mutations in N protein across major SARS-CoV-2 lineages. Mutational frequency in the N protein across 15 SARS-CoV-2 lineages is depicted as a heatmap. Only mutations that achieve at least 75% prevalence in at least one of the above-described Pango lineages are included.

SARS-CoV-2 lineages were defined by only a few mutations relative to the reference viral genome reported from Wuhan, China (36). While most early lineages had no mutations in the N protein, lineage B.1.1 arose with two adjacent mutations in the linker region, R203K and G204R (Fig. 5), which have been associated with enhanced infectivity and viral fitness (37). Late in 2020, regionally specific lineages carrying a couple of additional N protein mutations became predominant, including B.1.2 (P67S and P199L) in North America and B.1.160 (M234I and A376T) in Europe. These lineages would later be displaced by variants of concern (VOCs) that arose independently in different parts of the globe, including the Alpha variant in the United Kingdom (with N mutations D3L, R203K, G204R, and S235F), the Beta variant in South Africa (with N mutation T205I), and the Gamma variant in Brazil (with N mutations P80R, R203K, and G204R). The Delta variant, which was first reported in India, would become predominant worldwide by the end of 2021, most sublineages of which carried three N mutations (D63G, R203M, and D377Y). The R203M mutation was found to drastically enhance viral infectivity, similar to the R203K and G204R mutations found in other VOCs (38).

In late 2021, the emergence of the Omicron variant led to the rapid displacement of all former lineages. The Omicron VOCs all have N mutation P13L, deletion of amino acids 31 to 33, R203K, and G204R. The Omicron variant has diversified substantially since its emergence into additional sublineages, most of which carried an additional N protein mutation S413R (BA.2, BA.4, BA.5, etc.). Sublineages of the Omicron variant have been predominant worldwide since then and currently represent all circulating VOCs, with some sublineages carrying additional N mutations (i.e., P151S in BA.4 and E136D in BQ.1) (Fig. 5).

The emergence of N mutations can impact antibody responses and could impact antigen-based monitoring. Notably, many of the lineage-defining mutations in the N protein map within one of the previously described, highly prevalent public epitopes (P80R in region a, E136D and P151S in region c, and M234I and S235F in region d), with others mapping to one of the patient-specific private epitopes (31 to 33 deletion in region g; P199L, R203K/M, G204R, and T205I in region i; and A376T and D377Y in region j) (Fig. 5). While R203K/M and G204R have been previously linked to enhanced infectivity and viral fitness (37, 38), it remains unknown if these or other mutations mapping to N epitopes result in altered immunogenicity or immune escape.

TABLE 2 Comparison of antigenic N peptides identified in this study with prior studies

Peptide no.	Amino acids	Sequence (N→C)	Region	Domain	Group(s)	Data for peptides identified in other studies	
						No.	Reference(s)
3	22–31	DSTGSNQNGE	g	NTD	2, 3	4	23–25
11	74–84	INTNSSPDDQI	a	N-RBD	2, 3	2	23, 25
17	117–126	PEAGLPYGAN	b	N-RBD	2, 3	1	25
20, 21	141–150	TPKDHIGTRN	c	N-RBD	2, 3	2	25, 27
23, 24	162–171	PQGTTLPKGF	h	N-RBD	3, 4	7	24–26, 28–30, 32
25, 26	169–178	KGFYAEGRG	i	Linker	3	6	23–26, 32
28, 29	197–206	STPGSSRGTS	i	Linker	3	4	23–25, 31
32	221–231	LLLLDRLNQL	d	Linker	2, 3	5	24, 27–29, 31
34	235–245	SGKGQQQGGT	d	Linker	1, 2, 3, 4	3	23–25
36	249–259	KSAAEASKKPR	d	Linker	2, 3	5	23–25, 27, 29
40, 41	281–290	QTQGNFGDQE	e	N-DD	2, 3	1	25
46	323–332	EVTPSGTWLT	f	N-DD	2, 3, 4	1	25
52	361–371	KTFPPTEPKKD	j	C-term tail	3	9	23–30, 32
54, 55	379–388	TQALPQRQKK	j	C-term tail	3	8	24–30, 32
57	396–406	PAADLDDFSKQ	j	C-term tail	3	6	24, 26–29, 32

DISCUSSION

Numerous studies have established that patients with COVID-19 infection develop serum responses to the SARS-CoV-2 N protein (3, 4, 24, 39), and we previously showed that these antibodies can develop to both the full-length N protein and the structured regions, N-RBD and N-DD (34). Diagnostic tests that depend on identification of antibodies to the N protein have been widely deployed to detect evidence of prior SARS-CoV-2 infection. Unlike antibody responses to the S protein, which is included in most SARS-CoV-2 vaccines, N protein responses can more readily discriminate individuals with prior infection from those previously vaccinated. For the purposes of vaccine development and for understanding potential changes in antigenicity in SARS-CoV-2 variants or cross-reactivity with other coronaviruses, there is ongoing interest in identifying strong immunogenic epitopes within the N protein using smaller peptide arrays.

In our study, we used a library of overlapping 17-mer peptides (see Table S2 in the supplemental material) to map linear N protein epitopes with 10-aa resolution (Table 2) and analyzed the data to identify both public epitopes that induced an antibody response in most participants (Fig. 2) and private epitopes that resulted in high reactivity, but in only a subpopulation of patients (Fig. 3). This study revealed that the most prevalent highly reactive peptides were located toward the end of the linker domain (aa 235 to 245, region d) and within the N-DD (aa 281 to 290, region e). Region d and three other immunogenic regions (g, i, and j) identified in this study map to the unstructured regions of the N protein, including the N-terminal, linker, and C-terminal tail regions, respectively.

In addition to epitopes in the flexible interdomain regions, we also identified epitopes mapping within the unstructured surface-exposed loops of the stable folded N-RBD (Fig. 4B, D, and F). The most reactive peptide in N-RBD was located at aa 162 to 171 (region h), although antibodies against this peptide were identified in only a small fraction of patients (groups 3 and 4). Unique to our study, we identified two broadly reactive epitopes within the N-DD, a highly reactive peptide at aa 281 to 290 (region e) and a moderately reactive peptide at aa 323 to 332 (region f). Immunogenic epitopes have not been previously experimentally mapped to this region. We found that these epitopes are also surface exposed in regions that are not masked by the dimer interface. The identification of immunogenic epitopes in this domain is consistent with our prior published studies that antibody responses occur against both the N-RBD and the N-DD (34).

To place our results into context, we conducted a literature review examining studies that either predicted or experimentally determined antibody responses to the SARS-CoV-2 N protein (Table 2). Several previous publications have used immunoinformatic approaches

to predict epitopes (23–26). Early in the pandemic, Grifoni et al. (26) used a computational approach and predicted aa 43 to 65, 154 to 175, and 356 to 404 as potential B cell epitopes. Two of these regions were identified by our study as private epitopes present in a subset of our cohort. Amino acids 154 to 175 corresponded to region h, the highly reactive peptide from group 4, and amino acids 356 to 404 corresponded to the C-terminal tail region we designated j, a region associated with strong responses in group 3. Additionally, we noted antibody responses to five of the six N protein epitopes predicted by Oliveira et al. (23) as probable B-cell epitopes, including aa 21 to 32 (region g), 76 to 82 (region a), 176 to 206 (region i), 235 to 243, 249 to 263 (region d), and 363 to 379 (region j). Overall, our data most closely mirrored the epitope predictions of Crooke et al. (25). Of the 14 potential epitopes identified by computational approaches, we experimentally identified nine of them. Notably, Crooke et al. (25) was the only study of those reviewed that predicted responses to our two unique reactive epitopes within the N-DD, aa 281 to 290 (region e) and 323 to 332 (region f).

Shrock et al. (24) deployed a phage display platform (VirScan) that encoded 20-mer peptides tiled every five amino acids across the entire SARS-CoV-2 proteome to profile antibody responses. Their study revealed that most patients seroconvert and produce antibodies against the N protein and that responses to the N protein were stronger in hospitalized patients. Using the 20-mer library, four highly reactive epitopes were identified that were recurrently targeted by antibodies in at least 15% of patients, aa 222 to 242, 240 to 260, 365 to 385, and 386 to 406, which corresponded to public epitopes from our study in region d and private epitopes in region j. All of the highly reactive peptides map to the unstructured linker or C-terminal tail regions. Finally, Shrock et al. (24) also identified stronger antibody responses in hospitalized patients against amino acids 161 to 181. These responses corresponded to regions h and i in our study, which represent surface-exposed regions of the N-RBD and the beginning of the unstructured linker domain.

Studies that have evaluated proteome or linear peptide overlapping arrays are most similar to the study we conducted and have yielded the most granular insights into N protein antibody responses (27 to 32). Amrun et al. (30) scanned an 18-mer peptide library, identified aa 153 to 170 as reactive against pooled plasma from 18 patients, and found that a strong response to this peptide correlated with more severe disease. This peptide overlaps with our region h, identified as a private epitope in groups 3 and 4. These data are consistent with our finding that 5 of 8 patients in these groups were treated in the ICU, and group 3 contained the only patient death in our cohort. In a smaller cohort of patients ($n = 10$), Wang et al. used a linear epitope array of 15-mer peptides with a five-residue overlap to define antibody responses to the N protein (29). This study found eight distinct peptides which elicited strong IgG antibody responses, including three (aa 166 to 170, 226 to 230, and 366 to 400) which overlapped with regions h, d, and j in our study. Using an identical peptide-overlapping strategy in a larger cohort ($n = 89$ patients) (29), Voss et al. found a total of 26 N protein peptides that elicited an antibody response (27). The strongest response was to eight peptides that mapped to the C-terminal tail (aa 355 to 400), with the most robust responses to aa 361 to 375 and aa 381 to 395, consistent with our highly reactive private group 3 epitopes in the C-terminal tail region j. Epitopes with medium response intensity were identified for nine peptides mapping to the linker domain, with aa 221 to 235 and 246 to 260 showing the strongest responses. These peptides match our study's most highly reactive public epitope in region d. The final epitope with a strong response was aa 156 to 170 within the N-RBD, corresponding to region h. In contrast to the study by Amrun et al. (30) and our data, a strong reaction to this peptide correlated with less severe disease in this cohort. In a similar study, Holenya et al. (28) mapped high-resolution epitopes from 24 patients using a 15-mer peptide array with an 11-amino-acid overlap. They, too, identified similar immunoreactive epitopes at aa 161 to 171 (region h), 221 to 231 (region d), and three peptides within the C-terminal tail domain (region j). Finally, Hotop et al. (31) used a 15-mer overlapping peptide microarray to investigate responses in 36 SARS-CoV-2 patients. They noted that antibody responses to aa 220 to 234 (region d) within the linker domain were associated with more severe disease. Region d

was one of the two most prevalent immunodominant regions in our study, with brisk antibody responses within groups 2, 3, and 4, though our cohort was too small to draw clear conclusions on the association between antibody response and patient outcomes. Camerini et al. (32) took a slightly different approach and investigated serum responses to the N protein using the full-length N protein as well as larger 100-, 50-, and 30-amino-acid peptide fragments. Significant responses occurred in aa 301 to 400 that were resolved to the last 69 amino acids (350 to 419). This result parallels our findings that the C-terminal tail (region j) has strong epitopes and that responses to this region are variable across patient samples. Unlike our peptide array, Camerini et al. detected no responses to peptide fragments smaller than 30 amino acids.

Ultimately, the antigenicity of linear peptides mapping to the central linker region of the N protein, including peptide d, was consistent with nearly all other epitope-mapping studies (27–32). In total, these data support our finding that region d is a recurring highly prevalent epitope from the linker region that was identified in five other studies. The broad computational and *in vitro* support of serological responses to region d suggest that SARS-CoV-2 diagnostic test accuracy could be improved if diagnostics included peptides within this region. The recurring identification of region h in the N-RBD in nine other studies and j in the C-terminal tail in eight other studies confirms these are strong epitopes (Table 2), although all studies indicate these are strong epitopes in only a fraction of patients (private epitopes). Similar to region d, including peptides within regions h and j may improve diagnostic test accuracy.

The most significant difference compared to other studies is our identification of highly reactive peptides in the N-DD. We identified a predominant public epitope in region e that led to strong antibody responses in both groups 2 and 3 and region f that led to positive responses in more than 50% of all patients and high responses in a portion of patients in groups 2, 3, and 4. No other experimental studies identified epitopes in the N-DD, and only one computational study predicted epitopes in this stable folded domain (25). The nature of this discrepancy remains unclear, especially given the paucity of N mutations that map to this region (Fig. 5). Determination of the role of antibody responses to N-DD regions e and f and their impact on disease outcome would benefit from sampling larger, longitudinal patient cohorts.

The limitations of our study are the small cohort size, delay between sample acquisition and analysis, and the exclusive use of linear peptide fragments. Despite these limitations, we defined the dominant antibody response profile of our cohort and clustered patients into four unique groups based on distinct antibody response profiles. Given the small size of our cohort, we were not able to correlate specific antibody responses with patient outcome. However, most of the published studies likewise examined small patient cohorts, such that our study revealed the benefit of an aggregate analysis confirming the identification of dominant epitopes and suggested that region h private peptides are correlated with poor patient outcomes. Another limitation of our study is the long lag between data collection and analysis, as in the interval, our structural biology group focused on other COVID priorities, including drug development. This has not impacted our results, as a literature review demonstrated that our data are highly consistent with other published studies, and our analysis benefited from comparison to the larger aggregate sample size from 10 studies. Given the constant, albeit slow, evolution of the SARS-CoV-2 virus, an ideal data set would compare serological responses longitudinally over the course of the pandemic to assess for changing responses to the N protein. Inclusion of patients with a wider range of clinical symptoms and with and without history of prior vaccination would also provide important comparisons. Last, our study focused exclusively on linear peptide fragments to assess antibody response. This type of analysis excludes the possibility that conformational epitopes are important for immunogenicity. However, in a prior study, we previously showed that the level of the antibody response to full-length N-protein, N-RBD, and N-DD also did not correlate with patient outcome (34).

Ultimately, our study adds to the growing body of literature on antibody responses to the SARS-CoV-2 N protein and is unique in that it places those responses into the context

of the structure of the protein. Such analyses are critical for efforts to produce subunit vaccines as well as improved sensitivity and specificity of newly developed diagnostics, which may use peptide-specific antibodies rather than full-length N protein as the basis for seroconversion. Clearly, the SARS-CoV-2 N protein is highly and broadly immunogenic, and antibody responses are robust and durable despite the duration of the current pandemic. Though much attention has been devoted to the SARS-CoV-2 spike protein, studies such as ours suggest that broadening the evaluation of patient antibody responses to include the N protein will undoubtedly inform diagnostic development and inform treatment approaches. Although short of the penultimate goal of developing accurate tools to predict patient outcomes, studies that combine structural biology with peptide-tiled arrays such as ours continue to be powerful tools to study and compare antibody responses to evolving variants of concern of SARS-CoV-2.

MATERIALS AND METHODS

Protein expression, purification, and structure determination. Methods for expression and purification of the N-DD (amino acids 247 to 364, previously referred to as the N-CTD) as recombinant protein in *Escherichia coli* were previously described (34). The pure N-DD protein was set up as 2- μ L crystallization sitting drops (1 μ L protein/1 μ L reservoir solution) in 96-well plates and equilibrated with commercially available Classics II, JCSG+, MPD, and CompPAS suites. Diffraction-quality crystals appeared in 6 days from condition D8 of the Classics II screen (0.1 M HEPES, pH 7.5, and 25% [wt/vol] polyethylene glycol 3350 [PEG 3350]). Diffraction data were collected at the 21-ID-F Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source, Argonne National Laboratory. The data set was processed and scaled with the HKL-3000 suite (40). The structure was solved by molecular replacement with Phaser (41) from the CCP4 suite (42) using the crystal structure of the oligomerization domain from SARS coronavirus TW1 (PDB accession no. 2CJR) (43) as a search model. The initial solution went through several rounds of refinement in REFMAC v5.8.0258 (44) and manual model corrections using Coot (45). The water molecules were generated using ARP/wARP (46), and chloride ions were added to the model manually during visual inspection in Coot. Translation-Libration-Screw (TLS) groups were created by the TLSMD server (47) (<http://skuld.bmsc.washington.edu/~tlsmd/>), and TLS corrections were applied during the final stages of refinement. Molprobrity (48) (<http://molprobrity.biochem.duke.edu/>) was used for monitoring the quality of the model during refinement and for the final validation of the structure.

The final models and diffraction data were deposited to the Protein Data Bank (<https://www.rcsb.org/>) with the assigned PDB accession no. 6WJI. The plasmid used for protein production of N-DD is available at BEI Resources (catalog no. NR-52434).

Serum collection and processing from COVID-19 patients. Individuals over the age of 18 admitted to Northwestern Memorial Hospital with a positive PCR-based COVID-19 diagnostic test, who provided informed consent themselves or through an appropriate surrogate, were enrolled in the study per institutional review board (IRB) approval no. STU00206850. Methods and procedures for specimen collection, processing, and deposition to the Northwestern COVID-19 BioBank have been previously described (34). For this study, only the first available serum specimen from 34 participants was used. All available clinical data from the Northwestern Medicine Enterprise Data Warehouse (NMEDW) and from electronic chart review were pulled for this subset of hospitalized participants per institutional review board approval no. STU00212267. These NMEDW data were utilized to determine demographics, clinical assessments, symptom onset, laboratory measurements, COVID-19 disease severity, and hospital outcomes (intensive care unit [ICU] care and death) for the study analyses.

ELISA. A 59-peptide array covering the amino acid sequence of the N protein as 17 or 13 mer with 10-amino-acid overlaps was obtained from BEI Resources (catalog no. NR-52404). One milligram of lyophilized peptide was resuspended in 1 mL phosphate-buffered saline (PBS) and diluted to 5 μ g/mL in PBS, and 100 μ L was added to each well of a 96-well microtiter plate (Nunc-Immuno MicroWell 96-well solid plates; Fisher Scientific). Plates were stored at 4°C for 24 to 48 h and washed three times with 250 μ L PBS plus 0.05% Tween 20 (PBS-T). Plates were blocked with 200 μ L PBS-T plus 2% bovine serum albumin for 4 h. Blocking buffer was removed and replaced with 100 μ L patient sera diluted 1:1,000 in blocking buffer. Plates were incubated for 1 h at room temperature and then washed three times with PBS-T. One hundred microliters of 1 μ g/mL horseradish peroxidase (HRP)-conjugated anti-human IgG in blocking buffer was added to wells and incubated for 1 h at room temperature with shaking. The plate was washed three times with 250 μ L PBS-T. We added 100 μ L 3,3',5,5'-tetramethylbenzidine (TMB) substrate to wells for 1 min, and then reactions were stopped with 100 μ L stop solution. Absorbance was read at 450 nm on a Molecular Devices SpectraMax M3 spectrophotometer using SoftMax Pro v6.5.1 software.

Clustering analysis. Absorbance values were first normalized to the control and then subjected to a principal-component analysis (PCA). *k*-Means is an algorithm that seeks to partition samples into clusters using Euclidean distance. We performed *k*-means in the PCA feature space to identify distinct clusters of patients with similar absorbance profiles. The number of optimal clusters was determined via the elbow method, i.e., a method that visually identifies a point at which additional clusters no longer reduce the within-cluster distances. All analyses were performed in Python (v3.8.8). PCA and *k*-means were run using PCA and KMeans modules from the sklearn package (v0.24.1).

N protein mutational analysis. The frequency of mutations in the N protein across 15 prominent SARS-CoV-2 lineages was calculated using available whole-genome sequences from the Global Initiative on Sharing Avian Influenza Data (GISAID) database at <https://nextstrain.org/>, accessed 2 December 2022. Only mutations that reached at least 75% frequency in at least one of the examined lineages were included in the analysis.

Data availability. The structure of N protein N-DD was deposited to the Protein Data Bank (<https://www.rcsb.org/>) with accession no. 6WJ1. The plasmid for expression of the N-DD (catalog no. NR-52434) and the peptide library used in this study (catalog no. NR-52404) are available from BEI Resources (www.beiresearch.org).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Lacy M. Simons for her assistance with specimen biobanking. Nicole Palacio and Pablo Penaloza-MacMaster are thanked for technical advice in the development of the ELISA assay. Clinical data collection was supported in part by the Northwestern Medicine Enterprise Data Warehouse.

This project has been funded in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Department of Health and Human Services, under contract no. HHSN272201700060C and 75N93022C00035 (to K.J.F.S.), NIH COVID-19 supplements to the National Cancer Institute-funded Northwestern University Robert H. Lurie Comprehensive Cancer Center (P30CA060553 to K.J.F.S. and J.F.H.), and supplemental funding through the National Center for Advancing Translational Sciences awards (UL1TR001422 and UL1TR002389 to J.F.H.). Additional funding for this work was provided by an American Cancer Society (ACS) Clinician Scientist Development Grant (134251-CSDG-30-053-01-MPC to K.E.R.B.); a Dixon Translational Research Grant made possible by the generous support of the Dixon Family Foundation (J.F.H.); NIAID grants R21AI163912, U19AI135964, and U19AI171110 (to J.F.H.); and through a generous contribution from the Walder Foundation Foundation's Chicago Coronavirus Assessment Network (Chicago CAN) Initiative (to J.F.H.). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. Use of the LS-CAT sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). The funding sources had no role in the study design, data collection, analysis, interpretation, or writing of the report.

Conceptualization, N.L.I., M.R.-L., J.F.H., K.J.F.S., and K.E.R.B.; Data Curation, T.L.-H., C.J.A., J.F.H., and K.E.R.B.; Formal Analysis, N.L.I., T.L.-H., G.M., J.F.H., K.J.F.S., and K.E.R.B.; Funding Acquisition, J.F.H., K.J.F.S., and K.E.R.B.; Investigation, J.P.V., T.L.-H., G.M., G.W., M.R.-L., L.S., and K.E.R.B.; Methodology, J.P.V., T.L.-H., M.R.-L., J.F.H., and K.E.R.B.; Project Administration, N.L.I., J.F.H., K.J.F.S., and K.E.R.B.; Resources, C.J.A. and J.F.H.; Software, T.L.-H.; Supervision, J.F.H., K.J.F.S., and K.E.R.B.; Validation, T.L.-H. and G.M.; Visualization, N.L.I., T.L.-H., J.F.H., K.J.F.S., and K.E.R.B.; Writing – Original Draft, N.L.I., T.L.-H., M.R.-L., J.F.H., K.J.F.S., and K.E.R.B.; Writing – Review & Editing, N.L.I., T.L.-H., G.M., J.F.H., K.J.F.S., and K.E.R.B.

REFERENCES

- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus Investigating and Research Team. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382:727–733. <https://doi.org/10.1056/NEJMoa2001017>.
- Clarke KEN, Jones JM, Deng Y, Nycz E, Lee A, Iachan R, Gundlapalli AV, Hall AJ, MacNeil A. 2022. Seroprevalence of infection-induced SARS-CoV-2 antibodies—United States, September 2021–February 2022. *MMWR Morb Mortal Wkly Rep* 71:606–608. <https://doi.org/10.15585/mmwr.mm7117e3>.
- de Assis RR, Jain A, Nakajima R, Jasinskas A, Felgner J, Obiero JM, Norris PJ, Stone M, Simmons G, Bagri A, Irsch J, Schreiber M, Buser A, Holbro A, Battagay M, Hosimer P, Noesen C, Adenaiye O, Tai S, Hong F, Milton DK, Davies DH, Contestable P, Corash LM, Busch MP, Felgner PL, Khan S. 2021. Analysis of SARS-CoV-2 antibodies in COVID-19 convalescent blood using a coronavirus antigen microarray. *Nat Commun* 12:6. <https://doi.org/10.1038/s41467-020-20095-2>.
- Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, Liao P, Qiu JF, Lin Y, Cai XF, Wang DQ, Hu Y, Ren JH, Tang N, Xu YY, Yu LH, Mo Z, Gong F, Zhang XL, Tian WG, Hu L, Zhang XX, Xiang JL, Du HX, Liu HW, Lang CH, Luo XH, Wu SB, Cui XP, Zhou Z, Zhu MM, Wang J, Xue CJ, Li XF, Wang L, Li ZJ, Wang K, Niu CC, Yang QJ, Tang XJ, Zhang Y, Liu XM, Li JJ, Zhang DC, Zhang F, Liu P, Yuan J, Li Q, Hu JL, Chen J, et al. 2020. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med* 26:845–848. <https://doi.org/10.1038/s41591-020-0897-1>.

5. Okba NMA, Muller MA, Li W, Wang C, GeurtsvanKessel CH, Corman VM, Lamers MM, Sikkema RS, de Bruin E, Chandler FD, Yazdanpanah Y, Le Hingrat Q, Descamps D, Houhou-Fidouh N, Reusken C, Bosch BJ, Drosten C, Koopmans MPG, Haagmans BL. 2020. Severe acute respiratory syndrome coronavirus 2-specific antibody responses in coronavirus disease patients. *Emerg Infect Dis* 26:1478–1488. <https://doi.org/10.3201/eid2607.200841>.
6. Zhang B, Tian J, Zhang Q, Xie Y, Wang K, Qiu S, Lu K, Liu Y. 2022. Comparing the nucleocapsid proteins of human coronaviruses: structure, immunoregulation, vaccine, and targeted drug. *Front Mol Biosci* 9:761173. <https://doi.org/10.3389/fmolb.2022.761173>.
7. Vabret N, Britton GJ, Gruber C, Hegde S, Kim J, Kuksin M, Levantovsky R, Malle L, Moreira A, Park MD, Pia L, Risson E, Saffern M, Salomé B, Esai Selvan M, Spindler MP, Tan J, van der Heide V, Gregory JK, Alexandropoulos K, Bhardwaj N, Brown BD, Greenbaum B, Gümüş ZH, Homann D, Horowitz A, Kamphorst AO, Curotto de Lafaille MA, Mehandru S, Merad M, Samstein RM, Sinai Immunology Review Project. 2020. Immunology of COVID-19: current state of the science. *Immunity* 52:910–941. <https://doi.org/10.1016/j.immuni.2020.05.002>.
8. Grellet E, L'Hôte I, Goulet A, Imbert I. 2022. Replication of the coronavirus genome: a paradox among positive-strand RNA viruses. *J Biol Chem* 298:101923. <https://doi.org/10.1016/j.jbc.2022.101923>.
9. Sola I, Almazan F, Zuniga S, Enjuanes L. 2015. Continuous and discontinuous RNA synthesis in coronaviruses. *Annu Rev Virol* 2:265–288. <https://doi.org/10.1146/annurev-virology-100114-055218>.
10. Cubuk J, Alston JJ, Incicco JJ, Singh S, Stuchell-Breteron MD, Ward MD, Zimmerman MI, Vithani N, Griffith D, Wagoner JA, Bowman GR, Hall KB, Soranno A, Holehouse AS. 2021. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* 12:1936. <https://doi.org/10.1038/s41467-021-21953-3>.
11. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. 2014. The SARS coronavirus nucleocapsid protein-forms and functions. *Antiviral Res* 103:39–50. <https://doi.org/10.1016/j.antiviral.2013.12.009>.
12. Jayaram H, Fan H, Bowman BR, Ooi A, Jayaram J, Collisson EW, Lescar J, Prasad BV. 2006. X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J Virol* 80:6612–6620. <https://doi.org/10.1128/JVI.00157-06>.
13. Chang CK, Sue SC, Yu TH, Hsieh CM, Tsai CK, Chiang YC, Lee SJ, Hsiao HH, Wu WJ, Chang WL, Lin CH, Huang TH. 2006. Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci* 13:59–72. <https://doi.org/10.1007/s11373-005-9035-9>.
14. Loesche M, Karlson EW, Talabi O, Zhou G, Boutin N, Atchley R, Loevinsohn G, Chang JBP, Hasdianda MA, Okenla A, Sampson E, Schram H, Magsipoc K, Goodman K, Donahue L, MacGowan M, Novack LA, Jarolim P, Baden LR, Nilles EJ. 2022. Longitudinal SARS-CoV-2 nucleocapsid antibody kinetics, seroreversion, and implications for seroepidemiologic studies. *Emerg Infect Dis* 28:1859–1862. <https://doi.org/10.3201/eid2809.220729>.
15. Harris RJ, Whitaker HJ, Andrews NJ, Aiano F, Amin-Chowdhury Z, Flood J, Borrow R, Linley E, Ahmad S, Stapley L, Hallis B, Amirthalingam G, Hoschler K, Parker B, Horsley A, Brooks TJG, Brown KE, Ramsay ME, Ladhani SN. 2021. Serological surveillance of SARS-CoV-2: six-month trends and antibody response in a cohort of public health workers. *J Infect* 82:162–169. <https://doi.org/10.1016/j.jinf.2021.03.015>.
16. Lynch KL, Whitman JD, Lacanienta NP, Beckerdite EW, Kastner SA, Shy BR, Goldgof GM, Levine AG, Bapat SP, Stramer SL, Esensten JH, Hightower AW, Bern C, Wu AHB. 2021. Magnitude and kinetics of anti-severe acute respiratory syndrome coronavirus 2 antibody responses and their relationship to disease severity. *Clin Infect Dis* 72:301–308. <https://doi.org/10.1093/cid/ciaa979>.
17. Laing AG, Lorenc A, Del Molino Del Barrio I, Das A, Fish M, Monin L, Munoz-Ruiz M, McKenzie DR, Hayday TS, Francos-Quijorna I, Kamdar S, Joseph M, Davies D, Davis R, Jennings A, Zlatareva I, Vantourout P, Wu Y, Sofra V, Cano F, Greco M, Theodoridis E, Freedman JD, Gee S, Chan JNE, Ryan S, Bugallo-Blanco E, Peterson P, Kisand K, Haljasmagi L, Chadli L, Moingeon P, Martinez L, Merrick B, Bisnauthsing K, Brooks K, Ibrahim MAA, Mason J, Lopez Gomez F, Babalola K, Abdul-Jawad S, Cason J, Mant C, Seow J, Graham C, Doores KJ, Di Rosa F, Edgeworth J, Shankar-Hari M, Hayday AC. 2020. A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat Med* 26:1623–1635. <https://doi.org/10.1038/s41591-020-1038-6>.
18. Rijkers G, Murk JL, Wintermans B, van Looy B, van den Berge M, Veenemans J, Stohr J, Reusken C, van der Pol P, Reimerink J. 2020. Differences in antibody kinetics and functionality between severe and mild severe acute respiratory syndrome coronavirus 2 infections. *J Infect Dis* 222:1265–1269. <https://doi.org/10.1093/infdis/jiaa463>.
19. Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, Xie C, Ma K, Shang K, Wang W, Tian DS. 2020. Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China. *Clin Infect Dis* 71:762–768. <https://doi.org/10.1093/cid/ciaa248>.
20. Guthmiller JJ, Stovicek O, Wang J, Changrob S, Li L, Halfmann P, Zheng NY, Utset H, Stamper CT, Dugan HL, Miller WD, Huang M, Dai YN, Nelson CA, Hall PD, Jansen M, Shanmugarajah K, Donington JS, Krammer F, Fremont DH, Joachimiak A, Kawaoka Y, Tesic V, Madariaga ML, Wilson PC. 2021. SARS-CoV-2 infection severity is linked to superior humoral immunity against the spike. *mBio* 12:e02940-20. <https://doi.org/10.1128/mBio.02940-20>.
21. Atyeo C, Fischinger S, Zohar T, Slein MD, Burke J, Loos C, McCulloch DJ, Newman KL, Wolf C, Yu J, Shuey K, Feldman J, Hauser BM, Caradonna T, Schmidt AG, Suscovich TJ, Linde C, Cai Y, Barouch D, Ryan ET, Charles RC, Lauffenburger D, Chu H, Alter G. 2020. Distinct early serological signatures track with SARS-CoV-2 survival. *Immunity* 53:524–532.e4. <https://doi.org/10.1016/j.immuni.2020.07.020>.
22. Ogata AF, Lazarovits R, Uwamanzu-Nna A, Gilboa T, Cheng CA, Walt DR. 2022. Severe acute respiratory syndrome coronavirus 2 antigens as targets of antibody responses. *Clin Lab Med* 42:97–109. <https://doi.org/10.1016/j.cll.2021.10.002>.
23. Oliveira SC, de Magalhaes MTQ, Homan EJ. 2020. Immunoinformatic analysis of SARS-CoV-2 nucleocapsid protein and identification of COVID-19 vaccine targets. *Front Immunol* 11:587615. <https://doi.org/10.3389/fimmu.2020.587615>.
24. Shrock E, Fujimura E, Kula T, Timms RT, Lee I-H, Leng Y, Robinson ML, Sie BM, Li MZ, Chen Y, Logue J, Zuiani A, McCulloch D, Lelis FJN, Henson S, Monaco DR, Travers M, Habibi S, Clarke WA, Caturegli P, Laeyendecker O, Piechocka-Trocha A, Li JZ, Khatri A, Chu HY, Villani A-C, Kays K, Goldberg MB, Hachohen N, Filbin MR, Yu XG, Walker BD, Wesemann DR, Larman HB, Lederer JA, Elledge SJ, Lavin-Parsons K, Parry B, Lilley B, Lodenstein C, McKaig B, Charland N, Khanna H, Margolin J, Gonye A, Gushterova I, Lasalle T, Sharma N, Russo BC, Rojas-Lopez M, et al. 2020. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* 370:eabd4250. <https://doi.org/10.1126/science.abd4250>.
25. Crooke SN, Ovsyannikova IG, Kennedy RB, Poland GA. 2020. Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome. *Sci Rep* 10:14179. <https://doi.org/10.1038/s41598-020-70864-8>.
26. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. 2020. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* 27:671–680.e2. <https://doi.org/10.1016/j.chom.2020.03.002>.
27. Voss C, Esmail S, Liu X, Knauer MJ, Ackloo S, Kaneko T, Lowes L, Stogios P, Seitova A, Hutchinson A, Yusifov F, Skarina T, Evdokimova E, Loppnau P, Ghiabi P, Haijan T, Zhong S, Abdoh H, Hedley BD, Bhayana V, Martin CM, Slessarev M, Chin-Yee B, Fraser DD, Chin-Yee I, Li SS. 2021. Epitope-specific antibody responses differentiate COVID-19 outcomes and variants of concern. *JCI Insight* 6:e148855. <https://doi.org/10.1172/jci.insight.148855>.
28. Holenya P, Lange PJ, Reimer U, Woltersdorf W, Panterodt T, Glas M, Wasner M, Eckey M, Drosch M, Hollidt JM, Naumann M, Kern F, Wenschuh H, Lange R, Schnatbaum K, Bier FF. 2021. Peptide microarray-based analysis of antibody responses to SARS-CoV-2 identifies unique epitopes with potential for diagnostic test development. *Eur J Immunol* 51:1839–1849. <https://doi.org/10.1002/eji.202049101>.
29. Wang H, Wu X, Zhang X, Hou X, Liang T, Wang D, Teng F, Dai J, Duan H, Guo S, Li Y, Yu X. 2020. SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. *ACS Cent Sci* 6:2238–2249. <https://doi.org/10.1021/acscentsci.0c00742>.
30. Amrun SN, Lee CY, Lee B, Fong SW, Young BE, Chee RS, Yeo NK, Torres-Ruesta A, Carissimo G, Poh CM, Chang ZW, Tay MZ, Chan YH, Chen MI, Low JG, Tarbyah PA, Kalimuddin S, Pada S, Tan SY, Sun LJ, Leo YS, Lye DC, Renia L, Ng LFP. 2020. Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. *EBioMedicine* 58:102911. <https://doi.org/10.1016/j.ebiom.2020.102911>.
31. Hotop SK, Reimering S, Shekhar A, Asgari E, Beutling U, Dahlke C, Fathi A, Khan F, Lutgehetmann M, Ballmann R, Gerstner A, Tegge W, Cicin-Sain L, Billitewski U, McHardy AC, Bronstrup M. 2022. Peptide microarrays coupled to machine learning reveal individual epitopes from human antibody responses with neutralizing capabilities against SARS-CoV-2. *Emerg Microbes Infect* 11:1037–1048. <https://doi.org/10.1080/22221751.2022.2057874>.
32. Camerini D, Randall AZ, Trappi-Kimmons K, Oberai A, Hung C, Edgar J, Shandling A, Huynh V, Teng AA, Hermanson G, Pablo JV, Stumpf MM, Lester SN, Harcourt J, Tamin A, Rasheed M, Thornburg NJ, Satheshkumar

- PS, Liang X, Kennedy RB, Yee A, Townsend M, Campo JJ. 2021. Mapping SARS-CoV-2 antibody epitopes in COVID-19 patients with a multi-coronavirus protein microarray. *Microbiol Spectr* 9:e01416-21. <https://doi.org/10.1128/Spectrum.01416-21>.
33. Zinzula L, Basquin J, Bohn S, Beck F, Klumpe S, Pfeifer G, Nagy I, Bracher A, Hartl FU, Baumeister W. 2021. High-resolution structure and biophysical characterization of the nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute respiratory syndrome coronavirus 2. *Biochem Biophys Res Commun* 538:54–62. <https://doi.org/10.1016/j.bbrc.2020.09.131>.
 34. Simons LM, Lorenzo-Redondo R, Gibson M, Kinch SL, Vandervaart JP, Reiser NL, Eren M, Lux E, McNally EM, Tambur AR, Vaughan DE, Bachtá KER, Demonbreun AR, Satchell KJF, Achenbach CJ, Ozer EA, Ison MG, Hultquist JF. 2022. Assessment of virological contributions to COVID-19 outcomes in a longitudinal cohort of hospitalized adults. *Open Forum Infect Dis* 9:ofac027. <https://doi.org/10.1093/ofid/ofac027>.
 35. Dinesh DC, Chalupska D, Silhan J, Koutna E, Nencka R, Veverka V, Boura E. 2020. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog* 16:e1009100. <https://doi.org/10.1371/journal.ppat.1009100>.
 36. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>.
 37. Wu H, Xing N, Meng K, Fu B, Xue W, Dong P, Tang W, Xiao Y, Liu G, Luo H, Zhu W, Lin X, Meng G, Zhu Z. 2021. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. *Cell Host Microbe* 29:1788–1801.e6. <https://doi.org/10.1016/j.chom.2021.11.005>.
 38. Syed AM, Taha TY, Tabata T, Chen IP, Ciling A, Khalid MM, Sreekumar B, Chen PY, Hayashi JM, Soczek KM, Ott M, Doudna JA. 2021. Rapid assessment of SARS-CoV-2-evolved variants using virus-like particles. *Science* 374:1626–1632. <https://doi.org/10.1126/science.abc16184>.
 39. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, Dela Cruz CS, Wang Y, Wu C, Xiao Y, Zhang L, Han L, Dang S, Xu Y, Yang Q-W, Xu S-Y, Zhu H-D, Xu Y-C, Jin Q, Sharma L, Wang L, Wang J. 2020. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clin Infect Dis* 71:778–785. <https://doi.org/10.1093/cid/ciaa310>.
 40. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. 2006. HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* 62:859–866. <https://doi.org/10.1107/S0907444906019949>.
 41. McCoy AJ. 2007. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63:32–41. <https://doi.org/10.1107/S0907444906045975>.
 42. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242. <https://doi.org/10.1107/S0907444910045749>.
 43. Chen CY, Chang CK, Chang YW, Sue SC, Bai HI, Riang L, Hsiao CD, Huang TH. 2007. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J Mol Biol* 368:1075–1086. <https://doi.org/10.1016/j.jmb.2007.02.069>.
 44. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355–367. <https://doi.org/10.1107/S0907444911001314>.
 45. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501. <https://doi.org/10.1107/S0907444910007493>.
 46. Cohen SX, Ben Jelloul M, Long F, Vagin A, Knipscheer P, Lebbink J, Sixma TK, Lamzin VS, Murshudov GN, Perrakis A. 2008. ARP/wARP and molecular replacement: the next generation. *Acta Crystallogr D Biol Crystallogr* 64:49–60. <https://doi.org/10.1107/S0907444907047580>.
 47. Painter J, Merritt EA. 2006. TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr* 39:109–111. <https://doi.org/10.1107/S0021889805038987>.
 48. Chen VB, Arendall WB, III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12–21. <https://doi.org/10.1107/S0907444909042073>.
 49. Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X, He S, Zhou Z, Zhou Z, Chen Q, Yan Y, Zhang C, Shan H, Chen S. 2020. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B* 10:1228–1238. <https://doi.org/10.1016/j.apsb.2020.04.009>.
 50. Peng Y, Du N, Lei Y, Dorje S, Qi J, Luo T, Gao GF, Song H. 2020. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *EMBO J* 39:e105938. <https://doi.org/10.15252/embj.2020105938>.
 51. Zhou R, Zeng R, von Brunn A, Lei J. 2020. Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Mol Biomed* 1:2. <https://doi.org/10.1186/s43556-020-00001-4>.
 52. Ye Q, West AMV, Silletti S, Corbett KD. 2020. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci* 29:1890–1901. <https://doi.org/10.1002/pro.3909>.
 53. Yang M, He S, Chen X, Huang Z, Zhou Z, Zhou Z, Chen Q, Chen S, Kang S. 2020. Structural insight into the SARS-CoV-2 nucleocapsid protein C-terminal domain reveals a novel recognition mechanism for viral transcriptional regulatory sequences. *Front Chem* 8:624765. <https://doi.org/10.3389/fchem.2020.624765>.
 54. Jia Z, Liu C, Chen Y, Jiang H, Wang Z, Yao J, Yang J, Zhu J, Zhang B, Yuchi Z. 2022. Crystal structures of the SARS-CoV-2 nucleocapsid protein C-terminal domain and development of nucleocapsid-targeting nanobodies. *FEBS J* 289:3813–3825. <https://doi.org/10.1111/febs.16239>.
 55. Rafael Ciges-Tomas J, Franco ML, Vilar M. 2022. Identification of a guanine-specific pocket in the protein N of SARS-CoV-2. *Commun Biol* 5:711. <https://doi.org/10.1038/s42003-022-03647-8>.
 56. Mercaldi GF, Bezerra EHS, Batista FAH, Tonoli CCC, Soprano AS, Shimizu JF, Nagai A, da Silva JC, Filho HVR, do Nascimento Faria J, da Cunha MG, Zeri ACM, Nascimento AFZ, Proenca-Modena JL, Bajgelman MC, Rocco SA, Lopes-de-Oliveira PS, Cordeiro AT, Bruder M, Marques RE, Sforca ML, Franchini KG, Benedetti CE, Figueira ACM, Trivella DBB. 2022. Discovery and structural characterization of chicoric acid as a SARS-CoV-2 nucleocapsid protein ligand and RNA binding disruptor. *Sci Rep* 12:18500. <https://doi.org/10.1038/s41598-022-22576-4>.