nature communications

Article

https://doi.org/10.1038/s41467-023-37786-1

High-throughput identification of prefusion-stabilizing mutations in SARS-CoV-2 spike

Received: 20 October 2022

Accepted: 31 March 2023

Published online: 10 April 2023

Check for updates

Timothy J. C. Tan[1], Zongjun Mou[2], Ruipeng Lei[3], Wenhao O. Ouyang[3], Meng Yuan[4], Ge Song[5,6,7], Raiees Andrabi[5,6,7], Ian A. Wilson[4,6,7,8], Collin Kieffer[9], Xinghong Dai[2], Kenneth A. Matreyek[10] & Nicholas C. Wu[1,3,11,12]

Designing prefusion-stabilized SARS-CoV-2 spike is critical for the effectiveness of COVID-19 vaccines. All COVID-19 vaccines in the US encode spike with K986P/V987P mutations to stabilize its prefusion conformation. However, contemporary methods on engineering prefusion-stabilized spike immunogens involve tedious experimental work and heavily rely on structural information. Here, we establish a systematic and unbiased method of identifying mutations that concomitantly improve expression and stabilize the prefusion conformation of the SARS-CoV-2 spike. Our method integrates a fluorescence-based fusion assay, mammalian cell display technology, and deep mutational scanning. As a proof-of-concept, we apply this method to a region in the S2 domain that includes the first heptad repeat and central helix. Our results reveal that besides K986P and V987P, several mutations simultaneously improve expression and significantly lower the fusogenicity of the spike. As prefusion stabilization is a common challenge for viral immunogen design, this work will help accelerate vaccine development against different viruses.

SARS-CoV-2 spike (S) glycoprotein, a homotrimeric class I fusion protein, naturally exists in a metastable, prefusion conformation on the virion surface[1]. Once the receptor-binding domain (RBD) of S transitions to an 'up' state and binds to the human angiotensin-converting enzyme II (hACE2) receptor[2–4], a cascade of conformational changes is triggered to promote virus-host membrane fusion, and hence virus entry[1,5–8]. This conformational change, which involves structural rearrangement of the first heptad repeat (HR1) and central helix (CH), as well as the shedding of the S1 subunit, converts S into the postfusion conformation[5–10]. To inhibit virus entry and fusion, neutralizing antibodies target a variety of mainly conformational epitopes on the prefusion conformation of S[11–15]. Many of these conformational epitopes disappear or rearrange in the postfusion conformation, which instead can expose non-neutralizing epitopes that are immunodominant[1]. Consistently, antibody titer to the prefusion conformation has a strong correlation with neutralization potency, whereas that to the postfusion conformation does not[16]. Therefore, effective COVID-19 vaccines require S to be locked in the prefusion conformation to preserve the neutralizing epitopes.

[1]Center for Biophysics and Quantitative Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [2]Department of Physiology and Biophysics, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA. [3]Department of Biochemistry, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [4]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [5]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA. [6]IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA 92037, USA. [7]Consortium for HIV/AIDS Vaccine Development (CHAVD), The Scripps Research Institute, La Jolla, CA 92037, USA. [8]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [9]Department of Microbiology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [10]Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA. [11]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [12]Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. e-mail: nicwu@illinois.edu

https://doi.org/10.1038/s41467-023-37786-1

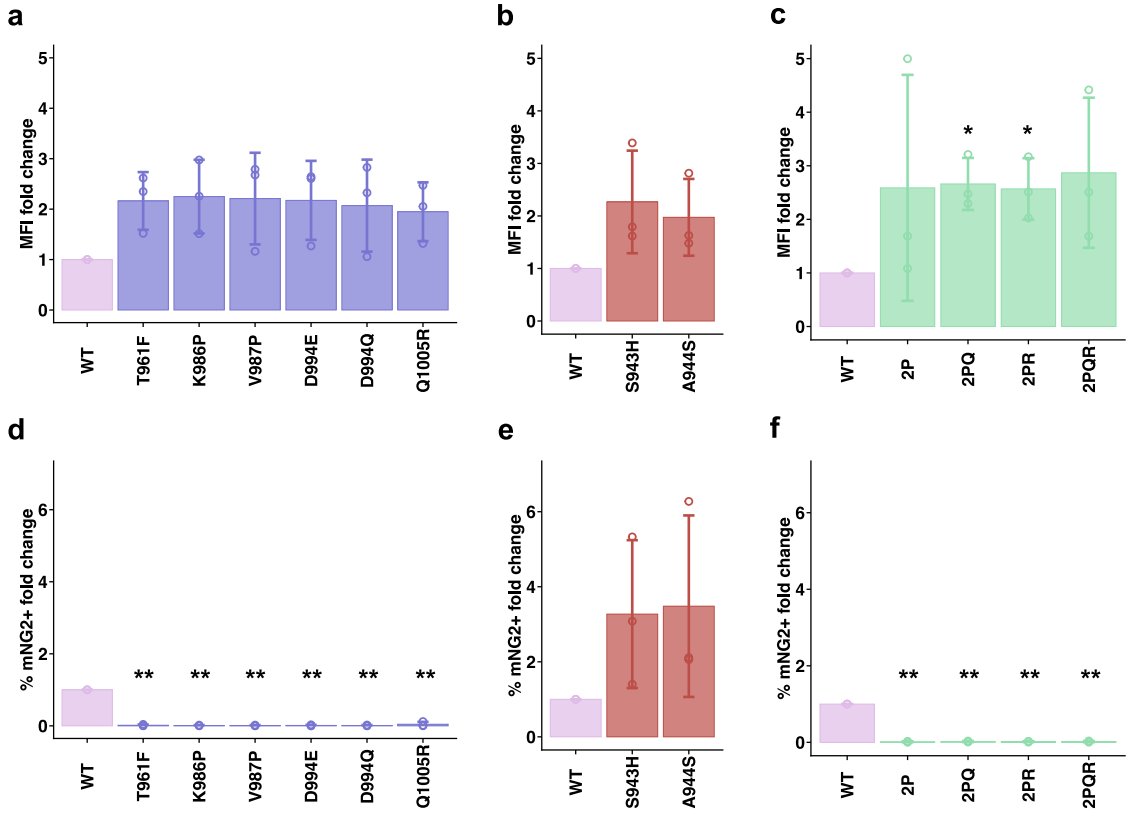

Fig. 3 | Validation of candidate prefusion-stabilizing mutations. a–c Expression of prefusion-stabilizing mutations (**a**), fusion-enhancing mutations (**b**), and combinations of candidate prefusion-stabilizing mutations of S (**c**) relative to WT. Of note, the numerical values of fold change in median fluorescence intensity (MFI) indicate relative and not absolute fold changes in surface expression levels of S. **d–f** Fold change in fusion activity of candidate prefusion-stabilizing mutations (**d**), fusion-enhancing mutations (**e**), and combinations of candidate prefusion-stabilizing mutations of S (**f**) relative to WT at 3 h post-mixing with hACE2-expressing cells. Abbreviations for combinatorial mutations are as follows: 2P, K986P/V987P; 2PQ, K986P/V987P/D994Q; 2PR, K986P/V987P/Q1005R; 2PQR, K986P/V987P/D994Q/Q1005R. Data are from $n = 3$ independent replicates, and shown as mean ± standard deviation. p-values were calculated from a two-sided Welch's t-test and are listed in the Source Data file; * $p < 0.05$; ** $p < 0.01$. Source data are provided as a Source Data file.

Although most SARS-CoV-2 neutralizing antibodies target RBD[61], recent studies have shown that antibodies to S2 can also neutralize, albeit often at a lower potency[45,62–65]. As a result, understanding the evolutionary constraints of S2 is relevant to SARS-CoV-2 antigenic drift and to design of more universal coronavirus vaccines. While many mutations in HR1 and CH, including those of major SARS-CoV-2 variants (Table S1), do not negatively impact the expression or fusogenicity of the S protein (Fig. 2b), HR1 and CH show high degrees of evolutionary conservation among betacoronaviruses (Fig. S10). This observation could be due to low levels of positive selection pressure on HR1 and CH, since most neutralizing antibodies are directed towards the RBD[61]. Alternatively, besides S protein expression and fusogenicity, other evolutionary constraints on HR1 and CH may be present in vivo. Future studies of the relationship among S protein expression, fusogenicity, and virus replication fitness will provide important biophysical insights into the evolution of SARS-CoV-2.

Since RBD is present in the prefusion conformation but not the postfusion conformation[5–7] and is the major target of neutralizing antibodies[61], this study used an RBD antibody to probe for surface expression. Nevertheless, we acknowledge that the folding of prefusion S can be more comprehensively probed by antibodies to conformational epitopes in the NTD and S2 subunit. Furthermore, due to the technical difficulties in sorting large syncytia, co-culturing of S-expressing cells and hACE2-expressing cells could only be performed for relatively short durations before sorting, leading to fewer syncytia and potentially lower reproducibility of results across replicates. Alternative strategies including microfluidics-based fusion experiments can be explored to obviate the kinetic limitations of the current fusion assay.

If the prefusion-stabilizing mutations of betacoronavirus S protein were not reported in late 2010s[17,66], it would not have been possible to develop an effective COVID-19 vaccine at the speed that occurred, even with the availability and utilization of the mRNA vaccine technology. It is unclear whether the next pandemic will be caused by a virus that we have prior knowledge about. Consequently, while the speed of vaccine manufacturing has been revolutionized by the mRNA vaccine technology[67], it is equally important to maximize the speed of immunogen design so that we are fully prepared for the next pandemic. We believe our work here provides an important step in that regard.

Methods

Cell culture

Human embryonic kidney 293T (HEK293T) landing pad cells obtained from Dr. Kenneth A. Matreyek (Case Western Reserve University) were grown and maintained in complete growth medium: Dulbecco's modified Eagle medium (DMEM) with high glucose (Gibco), supplemented with 10% v/v fetal bovine serum (FBS; VWR), 1× non-essential amino acids (Gibco), 100 U/mL penicillin and 100 µg/mL streptomycin (Gibco), 1× GlutaMAX (Gibco) and 2 µg/mL doxycycline (Thermo Scientific) at 37 °C, 5% CO_2 and 95% humidity. Expi293F cells (Gibco, catalog number A14527) were grown and maintained in Expi293 expression medium (Gibco) at 37 °C, 8% CO_2, 95% humidity and 125 rpm according to the manufacturer's instructions.

https://doi.org/10.1038/s41467-023-37786-1

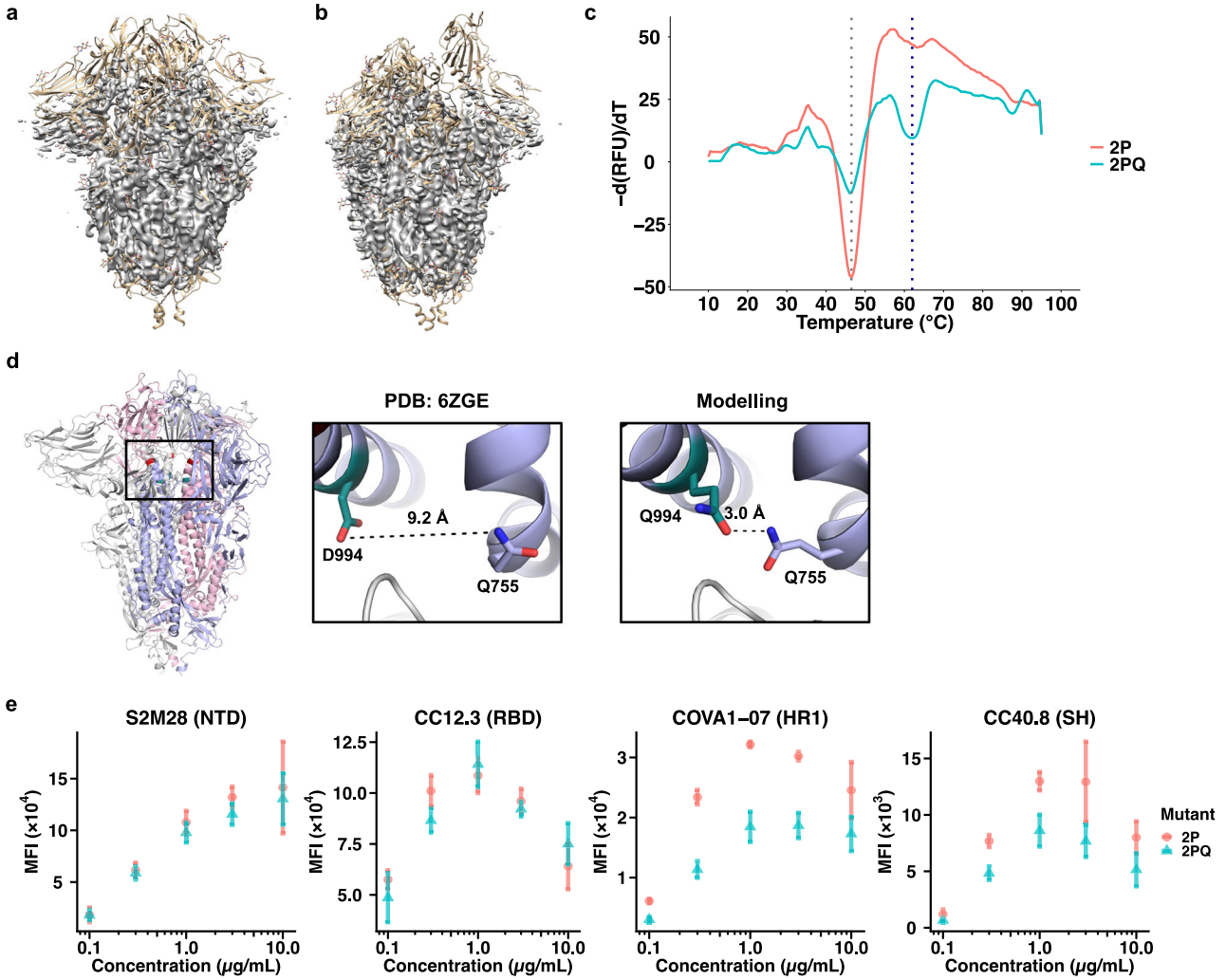

Fig. 4 | Biophysical characterization of 2PQ spike. a, **b** EM density map (colored grey) of 2PQ fitted on S with all-down RBD (PDB: 6VXX) (**a**), and one-up RBD (PDB: 6VYB) (**b**). **c** The first differential curves for the relative fluorescence unit (RFU) from differential scanning fluorimetry with respect to temperature are shown for soluble 2PQ and 2P. Grey dotted line indicates the first apparent melting temperature (T_m) of 2P and 2PQ at approximately 46.5 °C; blue dotted line indicates the second apparent T_m of 2PQ at approximately 62 °C. **d** D994Q allows formation of an additional intraprotomer hydrogen bond as shown by structural modelling. Distinct protomers are in grey, light blue and pink. The Q758 and Q994 side chains are shown as sticks representation. Hydrogen bond is indicated in black dashed line with the distance indicated. **e** Titration of S bearing 2P or 2PQ mutations with an N-terminal domain (NTD), receptor-binding domain (RBD), first heptad repeat (HR1), or stem helix (SH) antibody via flow cytometry. Median fluorescence intensities (MFI) are shown as mean ± standard deviation from $n = 3$ independent replicates. Source data are provided as a Source Data file.

Landing pad plasmids

attB plasmids each encoding (hACE2, an internal ribosomal entry site [IRES], and hygromycin resistance: attB-hACE2), (hACE2, an IRES, general control nondepressible 4 [GCN4] leucine zipper fused to $mNG2_{1\text{-}10}$, a (GSG) P2A self-cleaving peptide, and hygromycin resistance: attB-hACE2-mNG2-1-10), and (S with the PRRA motif in the furin cleavage site deleted, an IRES, GCN4 leucine zipper fused to $mNG2_{11}$, a (GSG) P2A self-cleaving peptide, and puromycin resistance: attB-S-mNG2-11) were constructed and assembled via polymerase chain reaction (PCR). The sequence of S used in this study was the ancestral (Wuhan-Hu-1) strain (GenBank accession ID: MN908947.3)[68]. The PRRA motif in the furin cleavage site was deleted to prevent spontaneous fusion of S-expressing cells with each other[69]. For experimental validation, mutants of S were individually constructed using PCR-based site-directed mutagenesis. Pairs of primers used for PCR-based site directed mutagenesis are listed in Table S2.

Deep mutational scanning library construction

Cassette primers for DMS library construction are listed in Table S3. Cassette primers were resuspended in MilliQ H_2O such that the final concentration of all primers is 10 µM. Forward cassette primers, named as CassetteX_N (X = 1, 2, ..., 19; N = 1, 2, ..., 8), that belong to the same cassette (i.e., the same value of X) were mixed in equimolar ratios. Each forward cassette primer also carried unique silent mutations (i.e. synonymous mutations) to help distinguish between sequencing errors and true mutations in downstream sequencing data analysis as described previously[70]. For the first round of PCR, two sets of reactions were set up. The first set had the mixed cassette primers and 5′-ACG ACG TCT CCT TCT CTA GGA AAG TGG GCT TTG C-3′ as forward and reverse primers, respectively. The second set had 5′-TGC TCG TCT CCA AAG TGA CAC TGG CCG ACG CCG G-3′ and CassetteX_Rprimers (X = 1, 2, ..., 19) as forward and reverse primers, respectively. Since we had 19 cassettes, there were 19 PCRs for each of the two sets of reactions. For both sets, the template used was attB-S-mNG2-11.

Thereafter, products corresponding to the correct size were excised and purified using Monarch DNA Gel Extraction kit (NEB). For the second round of PCR, 10 ng of PCR product from each of the first and second sets in the same cassette were mixed. 5′-ACG ACG TCT CCT TCT CTA GGA AAG TGG GCT TTG C-3′ and 5′-TGC TCG TCT CCA AAG TGA CAC TGG CCG ACG CCG G-3′ were used as the forward and reverse primers, respectively. PCR products corresponding to the correct size were excised and purified using DNA Gel Extraction kit (NEB). 100 ng of each gel-purified PCR products (total of 19) were mixed and digested with BsmBI restriction enzyme (NEB) for 2 h at 55 °C. Then, the product was purified using PureLink PCR Purification kit (Invitrogen) and served as the insert.

To amplify the vector, attB-S-mNG2-11, 5′-CAC TCG TCT CGA GAA GGC GTG TTC GTG TCC AAC G-3′, and 5′-GGC CCG TCT CAC TTT GTT GAA CAG CAG GTC CTC G-3′ were used as template, forward primer, and reverse primer, respectively. The PCR product was digested with DpnI (NEB) for 2 h at 37 °C, purified with PureLink PCR Purification kit (Invitrogen), digested with BsmBI restriction enzyme (NEB) for 2 h at 55 °C, and purified again using a PureLink PCR Purification kit (Invitrogen). All PCRs were performed using PrimeSTAR Max DNA Polymerase (Takara) according to the manufacturer's instructions.

BsmBI-digested vector and insert were ligated in a molar ratio of 1:100 to a total of 1 µg using T4 DNA ligase (NEB) for 2 h at room temperature. A control ligation reaction was set up by only having the BsmBI-digested vector (no insert). 1 µL ligation reaction products were transformed into chemically competent DH5α *Escherichia coli* cells and plated onto agar plates with 100 µg/mL ampicillin. The ligation mixture that contained vector and insert had at least 10 times more colonies than the control reaction. Subsequently, the ligation mixture was column-purified using a PureLink PCR Purification kit and eluted in 10 µL of MilliQ H_2O. 1 µL of the purified ligated product was mixed with 30 µL MegaX DH10β $T1^R$ electrocompetent *E. coli* cells (NEB) into an electroporation cuvette with a 1 mm gap (BTX). Electroporation was performed at 2.0 kV, 200 Ω and 25 µF using an ECM 830 square wave electroporation system (BTX). 1 mL of SOC recovery medium (NEB) was added immediately into cells after electroporation. Two electroporation reactions were performed. Cells were recovered for 1 h at 37 °C with shaking at 250 rpm. After recovery, cells were collected via centrifugation, resuspended in 400 µL lysogeny broth (LB), plated onto 150 mm agar plates supplemented with 100 µg/mL ampicillin, and incubated overnight at 37 °C. At least 1×10^6 colonies were scrape-harvested with LB broth and plasmids were extracted using a PureLink Plasmid Midiprep kit (Invitrogen).

Landing pad cell transfection

6.0×10^5 HEK293T landing pad cells in 1.35 mL of complete growth medium were seeded per well of a 6-well plate. 1.7 µg of attB-hACE2-mNG2-1-10 plasmid or attB-S-mNG2-11 plasmid were added into 5 µL FuGENE 6 transfection reagent (Promega) and OptiMEM (Gibco) to a total volume of 240 µL. The transfection mixture was subsequently added dropwise into cells. Transfection was carried out on the same day as seeding. One day post-transfection, 500 µL of complete growth medium was added to cells. Three days post-transfection, medium was discarded, cells were washed with 1× PBS, and incubated in negative selection medium (complete growth medium supplemented with 10 nM AP1903) for one day at 37 °C, 5% CO_2 and 95% humidity. Then, the medium was discarded, cells were washed with 1× PBS, and recovered in complete growth medium for two days at 37 °C, 5% CO_2 and 95% humidity. Cells were then trypsinized and grown in positive selection medium indefinitely: hACE2- and S-expressing cells were maintained in hygromycin medium (complete growth medium supplemented with 100 µg/mL hygromycin B [Invivogen]) and puromycin medium (complete growth medium supplemented with 1 µg/mL puromycin [Invivogen]), respectively.

To construct the S2 HR1/CH DMS cell line, the above protocol was used with modifications: 3.5×10^6 cells in 8 mL of complete growth medium in a T75 flask were transfected with 7.1 µg of the DMS plasmid library and 29 µL of FuGENE6 transfection reagent in 1.4 mL of Opti-MEM. For positive selection and regular maintenance, puromycin medium was used.

Flow cytometry

To validate hACE2 surface expression after transfection, landing pad cells were harvested via centrifugation at 300 × *g* for 5 min at 4 °C, resuspended in ice-cold FACS buffer (2% v/v FBS, 50 mM EDTA in DMEM supplemented with high glucose, L-glutamine and HEPES, without phenol red [Gibco]), and incubated with 2 µg/mL of SARS-CoV-2 S RBD-IgG Fc for 1 h at 4 °C. Then, cells were washed once, and resuspended with ice-cold FACS buffer. Cells were incubated with 1 µg/mL of phycoerythrin (PE)-conjugated anti-human IgG Fc (Bio-Legend, clone M1310G05, catalog number 410708). Cells were washed once and resuspended in ice-cold FACS buffer. Cells were analyzed using an Accuri C6 flow cytometer (BD Biosciences). Data was collected using BD Accuri C6 software v264 (BD Biosciences).

The above protocol for verification and quantification of S surface expression was used except cells were incubated with 5 µg/mL of CC12.3[44], an RBD antibody, instead of SARS-CoV-2 S RBD-IgG Fc, for 1 h at 4 °C. To quantify fold change in surface expression of S relative to WT based on median fluorescence intensity (MFI), Eq. (1) was used in the plot of FSC-A against PE:

$$\mathrm{MFI}_{\mathrm{FC}} = \frac{\mathrm{MFI}_{\mathrm{mutant}} - \mathrm{MFI}_{\mathrm{control}}}{\mathrm{MFI}_{\mathrm{WT}} - \mathrm{MFI}_{\mathrm{control}}} \quad (1)$$

MFI values were obtained after plotting data in FCS Express Flow Cytometry software v6 (De Novo Software). Gating strategy is shown in Fig. S11a.

To assess fusogenicity of S (WT or mutants), an equal number of hACE2, $mNG2_{1\text{-}10}$- and S, $mNG2_{11}$-expressing cells were mixed such that the total cell number is 5.0×10^5 cells per mL of complete growth medium. Cells were co-cultured for 3 h at 37 °C, 5% CO_2 and 95% humidity. Cells were then harvested and resuspended in ice-cold FACS buffer. Cells were analyzed using an Accuri C6 flow cytometer (BD Biosciences). Data was collected using BD Accuri C6 software v264 (BD Biosciences). Gating strategy is shown in Fig. S11b. The percentage of mNG2-positive events of mutants relative to that of WT S was calculated.

For titration of S-2P or S-2PQ, HEK293T landing pad cells stably expressing membrane-bound S-2P or S-2PQ were incubated with 0, 0.1, 0.3, 1.0, 3.0 or 10.0 µg/mL of S2M28, CC12.3, COVA1-07, or CC40.8 antibody for 1 h in ice-cold FACS buffer. Cells were washed and then incubated with 1 µg/mL PE-conjugated anti-human IgG Fc for 1 h at 4 °C. Cells were washed, resuspended in ice-cold FACS buffer, and analyzed for levels of PE using an Accuri C6 flow cytometer. Gating strategy is shown in Fig. S11a. MFI values were subtracted from those of negative control (0 µg/mL of antibody) and plotted against antibody concentration (Figs. 4e, S10).

Expression sorting

Cells expressing the S2 HR1/CH DMS library of S were harvested via centrifugation at 300 × *g* for 5 min at 4 °C. Supernatant was discarded, and cells were resuspended in ice-cold FACS buffer. Cells were incubated with 5 µg/mL of CC12.3 for 1 h at 4 °C. Then, cells were washed once, and resuspended with ice-cold FACS buffer. Cells were incubated with 2 µg/mL of PE anti-human IgG Fc. Cells were washed once, resuspended in ice-cold FACS buffer, and filtered through a 40 µm strainer. Cells were sorted via a four-way sort using a BD FACS Aria II cell sorter and BD FACS Diva software v8.0 (BD Biosciences), or a BigFoot spectral cell sorter and Sasquatch software firmware v888

(Invitrogen) according to PE fluorescence at 4 °C. Cells expressing the highest PE fluorescence were sorted into "bin 3", then the next highest into "bin 2", followed by "bin 1" and then "bin 0". Each bin had ~25% of the singlet population. Gating strategy is shown in Fig. S11c. Number of cells collected per bin per replicate is shown in Table S4. Of note, since CC12.3 binds to the RBD[44], an independently folded region of S that is present only in the prefusion but not postfusion conformation[1,71], our sort was based on the expression of prefusion S.

Fusion sorting

Cells expressing the HR1/CH DMS library of S, and cells expressing hACE2 were resuspended in complete growth medium and filtered through a 40 µm cell strainer to obtain single cell suspensions. 2.5×10^6 cells of each were mixed in a T-75 flask and complete growth medium added to a total volume of 10 mL. Six co-cultures were set up, with one of the co-cultures acting as a negative, non-fluorescent control by mixing hACE2- and S-expressing cells that do not have split mNG2. Co-cultures were incubated for 3 h at 37 °C, 5% CO_2 and 95% humidity. Subsequently, cells were harvested and pelleted via centrifugation at $300 \times g$ for 5 min at 4 °C. Supernatant was discarded, and cells were resuspended in ice-cold FACS buffer. Cells were sorted via a two-way sort using a BigFoot spectral cell sorter (Invitrogen) according to presence or absence of mNG2 fluorescence at 4 °C. Gating strategy is shown in Fig. S11d. Number of cells collected per bin per replicate is shown in Table S5.

Post-sorting genomic DNA extraction

After FACS, cell pellets were obtained via centrifugation at $300 \times g$ for 15 min at 4 °C, and the supernatant was discarded. Genomic DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's instructions with a modification: resuspended cells were incubated and lysed at 56 °C for 30 min instead of 10 min.

Deep sequencing

After genomic DNA extraction, the region of interest was amplified via PCR using 5′-CAC TCT TTC CCT ACA CGA CGC TCT TCC GAT CTA CAT CTG CCC TGC TGG CCG GCA CA-3′ and 5′-GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG CAA AAG TCC ACT CTC TTG CTC TG-3′ as forward and reverse primers, respectively. A maximum of 500 ng of genomic DNA per 50 µL PCR reaction was used as template; 4 µg of genomic DNA per expression or fusion bin, per replicate, was used as template. PCR was performed using KOD DNA polymerase (Takara) with the following settings: 95 °C for 2 min, 25 cycles of (95 °C for 20 s, 56 °C for 15 s, 68 °C for 20 s), 68 °C for 2 min, 12 °C indefinitely. All eight 50 µL reactions per bin per replicate were mixed after PCR. 100 µL of product per bin per replicate was used for purification using a PureLink PCR Purification kit. Subsequently, 10 ng of the purified PCR product per bin per replicate was appended with Illumina deep sequencing barcodes via PCR using KOD DNA polymerase with the following settings: 95 °C for 2 min, 9 cycles of (95 °C for 25 s, 56 °C for 15 s, 68 °C for 20 s), 68 °C for 2 min, 12 °C indefinitely. Barcoded products were mixed and sequenced with a MiSeq PE300 v3 flow cell (Illumina).

Analysis of deep sequencing data

Forward and reverse reads were merged via PEAR[72]. Using custom Python code, the merged reads were translated and matched to the corresponding mutant. Counts for expression and fusion bins for each replicate were tabulated. For each replicate, the frequency of each mutant was calculated as the count of that mutant divided by the total number of counts in that bin, as shown in Eq. (2):

$$F_{mut,\,binX} = \frac{C_{mut,binX}}{\Sigma C_{binX}} \text{ for } X = 0, 1, 2, 3, mNG2^-, mNG2^+ \quad (2)$$

For each replicate, the weighted expression score for each mutant (W_{mut}) was calculated using Eq. (3):

$$W_{mut} = \frac{(F_{mut,bin0} \times 0.25) + (F_{mut,bin1} \times 0.5) + (F_{mut,bin2} \times 0.75) + (F_{mut,bin3} \times 1)}{F_{mut,bin0} + F_{mut.bin1} + F_{mut,bin2} + F_{mut,bin3}} \quad (3)$$

The weighted expression scores were normalized (W_{mut}^{norm}) such that the average W_{mut} of nonsense mutations equals 0, and the average W_{mut} of silent mutations equals 1 using Eq. (4):

$$W_{mut}^{norm} = \frac{W_{mut} - W_{nonsense}^{avg}}{W_{silent}^{avg} - W_{nonsense}^{avg}} \quad (4)$$

The final expression score (W_{mut}^{avg}) for each mutant was calculated using Eq. (5):

$$W_{mut}^{avg} = \frac{1}{3} \times \left(W_{mut}^{norm,rep1} + W_{mut}^{norm,rep2} + W_{mut}^{norm,rep3} \right) \quad (5)$$

Fusion scores (U_{mut}) were calculated for each replicate by the formula shown in Eq. (6):

$$U_{mut} = \log_{10}\left(\frac{F_{mut,mNG2^+}}{F_{mut,mNG2^-}} \right) \quad (6)$$

Fusion scores were normalized (U_{mut}^{norm}) such that the U_{mut}^{avg} of silent mutations equals 1, and the U_{mut}^{avg} of nonsense mutations equals 0 using Eq. (7):

$$U_{mut}^{norm} = \frac{U_{mut} - U_{nonsense}^{avg}}{U_{WT}^{avg} - U_{nonsense}^{avg}} \quad (7)$$

Then, the final average score (U_{mut}^{avg}) for each mutant was calculated using Eq. (8):

$$U_{mut}^{avg} = \frac{1}{2} \times \left(U_{mut}^{norm,rep1} + U_{mut}^{norm,rep2} \right) \quad (8)$$

Adjusted fusion score of each mutant is equal to the residual of that mutant in a linear regression model of fusion score against expression score. The linear regression model and residuals were calculated using the 'lm' and 'resid' functions in RStudio v2022.12.0+353.

Sequence conservation analysis

Sequences were obtained from GenBank or GISAID (Tables S1, S7). A BLAST database was created, and the reference sequence of the DMS region (residues 883-1034) was used to run tblastn to generate BlastXML files. The reference sequence used was the founder strain of SARS-CoV-2 (GenBank accession number: MN908947.3)[68]. Extracted information was obtained by running 'XML_Extraction.py'[73]. Multiple alignment using MAFFT was then performed[74]. Sequence conservation was calculated based on the residue conservation at each position relative to the reference sequence. Mean expression score and mean fusion score were calculated by taking the average of the expression scores and fusion scores of all mutants, respectively, at that position.

Fluorescence microscopy

Images were captured with an ECHO Revolve epifluorescence microscope (ECHO) with a UPLANFL N 10×/0.30 NA objective (Olympus) using the FITC channel for mNG2 fluorescence. Brightfield images were also obtained using white light. Fluorescent and brightfield images were then overlaid. Identical exposure and intensity settings were used to capture images. Scale bars correspond to 100 µm for all micrographs.

Cryogenic electron microscopy

To prepare cryoEM grid, an aliquot of 3.5 μL purified protein at ~1 mg/mL concentration was applied to a 300-mesh Quantifoil R1.2/1.3 Cu grid pre-treated with glow-discharge, blotted in a Vitrobot Mark IV machine (force -5, time 3 s), and plunge-frozen in liquid ethane. The grid was loaded in a Titan Krios microscope equipped with Gatan BioQuantum K3 imaging filter and camera. A 10-eV slit was used for the filter. Data collection was done with serialEM v4.0[75]. Images were recorded at 130,000× magnification, corresponding to a pixel size of 0.33 Å/pix at super-resolution mode of the camera. A defocus range of -0.8 μm to -1.5 μm was set. A total dose of 50 $e^-/Å^2$ of each exposure was fractionated into 50 frames. The first two frames of the movie stacks were not included in motion-correction. CryoEM data processing was performed on the fly with cryoSPARC Live v3.3.2 (Structura Biotechnology)[76] following regular single-particle procedures. Statistics are provided in Table S8. Structures were visualized using UCSF ChimeraX v1.5 (UCSF).

Rosetta-based mutagenesis

The structure of S was obtained from the Protein Data Bank (PDB ID: 6ZGE). N-acetyl-D-glucosamine and water molecules were removed using PyMOL v2.4.0 (Schrödinger), and amino acids were renumbered using pdb-tools[77]. The 'fixbb' application in Rosetta v3.11 (RosettaCommons) was used to generate the D994Q mutation in all protomers. One-hundred poses were obtained, and the lowest scoring pose was used for further processing. A constraint file was generated using the lowest-scoring pose from fixed backbone mutagenesis as input, and the 'minimize_with_cst' application in Rosetta. Fast relax was subsequently performed using the 'relax' application[78] with the constraint file. The lowest scoring pose out of thirty was used for structural analysis.

Antibody expression and purification

Codon-optimized oligonucleotides encoding the heavy chain and light chain of the indicated antibodies were cloned into phCMV3 plasmids in an IgG1 Fc format with a mouse immunoglobulin kappa signal peptide. Plasmids encoding the heavy chain and light chain of antibodies were transfected into Expi293F cells using an Expifectamine 293 transfection kit (Gibco) in a 2:1 mass ratio following the manufacturer's protocol. Supernatant was harvested 6 days post-transfection and centrifuged at 4000 × *g* for 30 min at 4 °C to remove cells and debris. The supernatant was subsequently clarified using a polyethersulfone membrane filter with a 0.22 μm pore size (Millipore).

CaptureSelect CH1-XL beads (Thermo Scientific) were washed with MilliQ H_2O thrice, and resuspended in 1× PBS. The clarified supernatant was incubated with washed beads overnight at 4 °C with gentle rocking. Then, flowthrough was collected, and beads washed once with 1× PBS. Beads were incubated in 60 mM sodium acetate, pH 3.7 for 10 min at 4 °C. The eluate containing antibody was buffer-exchanged into 1× PBS using a centrifugal filter unit with a 30 kDa molecular weight cut-off (Millipore) four times. Antibodies were stored at 4 °C.

Soluble S protein expression and purification

SARS-CoV-2 S ectodomain (residues 1-1213, which includes the native signal peptide) with the PRRA motif in the furin cleavage site deleted, C-terminal SGGGG linker, biotinylation site, thrombin cleavage site, Foldon trimerization sequence, and 6×His-tag were all cloned in-frame into a phCMV3 vector via PCR. Site-directed mutagenesis via PCR was performed to generate the indicated mutants of soluble S protein.

Expi293F cells were transfected with vectors encoding the indicated soluble spike protein mutant using an Expifectamine 293 transfection kit following the manufacturer's protocol. Cells were harvested six days post-transfection. The supernatant was collected via centrifugation at 4000 × *g* for 30 min at 4 °C, and further clarified using a polyethersulfone membrane with a 0.22 μm pore size (Millipore). The clarified supernatant was incubated with washed Ni sepharose excel His-tagged protein purification resin (Cytiva) with gentle rocking overnight at 4 °C. Flow-through was collected. Beads were washed once with 20 mM imidazole in 1× PBS, then washed once with 40 mM imidazole in 1× PBS, and finally eluted with 300 mM imidazole in 1× PBS thrice. Wash and elution fractions were subjected to denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. S7a). All elution fractions were combined and concentrated using a centrifugal filter unit with a 30 kDa molecular weight cut-off (Millipore) via centrifugation at 4000 × *g* and 4 °C for 15 min. The concentrated protein mixture was passed through a Superdex 200 XK 16/100 column in 20 mM Tris-HCl pH 8.0 and 150 mM NaCl for size-exclusion chromatography (Fig. S7b, c). Data was collected using ChromLab software v6.1 (Bio-Rad). Fractions corresponding to ~540 kDa were pooled and concentrated using a centrifugal filter unit with a 30 kDa molecular weight cut-off (Millipore) via centrifugation at 4000 × *g* and 4 °C for 15 min.

Differential scanning fluorimetry

200 ng/μL of purified S protein and 5× SYPRO orange (Thermo Fisher Scientific) were added into 20 mM Tris-HCl pH 8.0, 150 mM NaCl in optically clear tubes. SYPRO orange fluorescence intensity in relative fluorescence units (RFU) was measured over temperatures ranging from 10 °C to 95 °C using a CFX Connect Real-Time PCR Detection System (Bio-Rad). Melting temperature (T_m) was calculated as the temperature at which the first derivative of fluorescence intensity with respect to temperature, $-\frac{d(RFU)}{dT}$, was minimum.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Structures from the following identifiers from the Protein Data Bank (PDB) were used in this study: 6VXX and 6VYB. The cryoEM map of 2PQ spike can be accessed at the Electron Microscopy Data Bank (EMDB) using accession code EMD-29374. Raw deep sequencing data generated in this study have been submitted to the NIH Sequence Read Archive under accession number: PRJNA826665. Source data are provided with this paper.

Code availability

Custom codes to analyze deep mutational scanning, thermal stability, and flow cytometry data have been deposited to https://doi.org/10.5281/zenodo.7742830[79].

References

1. Cai, Y. et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586–1592 (2020).
2. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
3. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).
4. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
5. Chiliveri, S. C., Louis, J. M., Ghirlando, R. & Bax, A. Transient lipid-bound states of spike protein heptad repeats provide insights into SARS-CoV-2 membrane fusion. *Sci. Adv.* **7**, eabk2226 (2021).
6. Marcink, T. C. et al. Intermediates in SARS-CoV-2 spike–mediated cell entry. *Sci. Adv.* **8**, eabo3153 (2022).

7. Dodero-Rojas, E., Onuchic, J. N. & Whitford, P. C. Sterically confined rearrangements of SARS-CoV-2 Spike protein control cell invasion. *eLife* **10**, e70362 (2021).
8. Walls, A. C. et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. USA* **114**, 11157–11162 (2017).
9. Gorgun, D., Lihan, M., Kapoor, K. & Tajkhorshid, E. Binding mode of SARS-CoV-2 fusion peptide to human cellular membrane. *Biophys. J.* **120**, 2914–2926 (2021).
10. Koppisetti, R. K., Fulcher, Y. G. & Van Doren, S. R. Fusion peptide of SARS-CoV-2 spike rearranges into a wedge inserted in bilayered micelles. *J. Am. Chem. Soc.* **143**, 13205–13211 (2021).
11. Cerutti, G. et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **29**, 819–833.e7 (2021).
12. Qing, E. et al. Inter-domain communication in SARS-CoV-2 spike proteins controls protease-triggered cell entry. *Cell Rep.* **39**, 110786 (2022).
13. Barnes, C. O. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).
14. Liu, L. et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020).
15. Yuan, M., Liu, H., Wu, N. C. & Wilson, I. A. Recognition of the SARS-CoV-2 receptor binding domain by neutralizing antibodies. *Biochem. Biophys. Res. Commun.* **538**, 192–203 (2021).
16. Bowen, J. E. et al. SARS-CoV-2 spike conformation determines plasma neutralizing activity. *bioRxiv*, https://doi.org/10.1101/2021.12.19.473391 (2021).
17. Pallesen, J. et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. USA* **114**, E7348–E7357 (2017).
18. Yuan, Y. et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092 (2017).
19. Kirchdoerfer, R. N. et al. Pre-fusion structure of a human coronavirus spike protein. *Nature* **531**, 118–121 (2016).
20. Gilbert, P. B. et al. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science* **375**, 43–50 (2022).
21. Skowronski, D. M. & De Serres, G. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New Engl. J. Med.* **384**, 1576–1577 (2021).
22. Sadoff, J. et al. Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *New Engl. J. Med.* **384**, 2187–2201 (2021).
23. Heath, P. T. et al. Safety and efficacy of NVX-CoV2373 Covid-19 vaccine. *New Engl. J. Med.* **385**, 1172–1183 (2021).
24. Hsieh, C.-L. et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020).
25. Juraszek, J. et al. Stabilizing the closed SARS-CoV-2 spike trimer. *Nat. Commun.* **12**, 244 (2021).
26. Riley, T. P. et al. Enhancing the prefusion conformational stability of SARS-CoV-2 spike protein through structure-guided design. *Front. Immunol.* **12**, 660198 (2021).
27. Olmedillas, E. et al. Structure-based design of a highly stable, covalently-linked SARS-CoV-2 spike trimer with improved structural properties and immunogenicity. *bioRxiv*, https://doi.org/10.1101/2021.05.06.441046 (2021).
28. Sanders, R. W. & Moore, J. P. Virus vaccines: proteins prefer prolines. *Cell Host Microbe* **29**, 327–333 (2021).
29. Matreyek, K. A. et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).
30. Suiter, C. C. et al. Massively parallel variant characterization identifies NUDT15 alleles associated with thiopurine toxicity. *Proc. Natl. Acad. Sci. USA* **117**, 5394–5401 (2020).
31. Chiasson, M. A. et al. Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact. *eLife* **9**, e58026 (2020).
32. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res.* **48**, e1 (2020).
33. Matreyek, K. A., Stephany, J. J. & Fowler, D. M. A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res.* **45**, e102 (2017).
34. Ouyang, W. O. et al. Probing the biophysical constraints of SARS-CoV-2 spike N-terminal domain using deep mutational scanning. *bioRxiv*, https://doi.org/10.1101/2022.06.20.496903 (2022).
35. Kondo, N., Miyauchi, K., Meng, F., Iwamoto, A. & Matsuda, Z. Conformational changes of the HIV-1 envelope protein during membrane fusion are inhibited by the replacement of its membrane-spanning domain. *J. Biol. Chem.* **285**, 14681–14688 (2010).
36. Baviskar, P. S., Hotard, A. L., Moore, M. L. & Oomens, A. G. The respiratory syncytial virus fusion protein targets to the perimeter of inclusion bodies and facilitates filament formation by a cytoplasmic tail-dependent mechanism. *J. Virol.* **87**, 10730–10741 (2013).
37. Atanasiu, D. et al. Dual split protein-based fusion assay reveals that mutations to herpes simplex virus (HSV) glycoprotein gB alter the kinetics of cell-cell fusion induced by HSV entry glycoproteins. *J. Virol.* **87**, 11332–11345 (2013).
38. Meng, B. et al. SARS-CoV-2 Spike N-Terminal Domain modulates TMPRSS2-dependent viral entry and fusogenicity. *Cell Rep.* **40**, 111220 (2022).
39. Meng, B. et al. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* **603**, 706–714 (2022).
40. Bradel-Tretheway, B. G. et al. Nipah and Hendra virus glycoproteins induce comparable homologous but distinct heterologous fusion phenotypes. *J. Virol.* **93**, e00577–19 (2019).
41. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23**, 102–107 (2005).
42. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
43. Feng, S. et al. Improved split fluorescent proteins for endogenous protein labeling. *Nat. Commun.* **8**, 370 (2017).
44. Yuan, M. et al. Structural basis of a shared antibody response to SARS-CoV-2. *Science* **369**, 1119–1123 (2020).
45. Zhou, P. et al. A human antibody reveals a conserved site on beta-coronavirus spike proteins and confers protection against SARS-CoV-2 infection. *Sci. Transl. Med.* **14**, eabi9215 (2022).
46. Krammer, F. SARS-CoV-2 vaccines in development. *Nature* **586**, 516–527 (2020).
47. Klasse, P. J., Nixon, D. F. & Moore, J. P. Immunogenicity of clinically relevant SARS-CoV-2 vaccines in nonhuman primates and humans. *Sci. Adv.* **7**, eabe8065 (2021).
48. McCallum, M. et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332–2347.e16 (2021).
49. Claireaux, M. et al. A public antibody class recognizes an S2 epitope exposed on open conformations of SARS-CoV-2 spike. *Nat. Commun.* **13**, 4539 (2022).
50. Henderson, R. et al. Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020).
51. Binley, J. M. et al. A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J. Virol.* **74**, 627–643 (2000).

52. Sanders, R. W. et al. Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *J. Virol.* **76**, 8875–8889 (2002).
53. Kong, L. et al. Uncleaved prefusion-optimized gp140 trimers derived from analysis of HIV-1 envelope metastability. *Nat. Commun.* **7**, 12040 (2016).
54. Sanders, R. W. et al. A next-generation cleaved, soluble HIV-1 Env Trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Path* **9**, e1003618 (2013).
55. McLellan, J. S. et al. Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113–1117 (2013).
56. Loomis, R. J. et al. Structure-based design of Nipah virus vaccines: a generalizable approach to paramyxovirus immunogen development. *Front. Immunol.* **11**, 842 (2020).
57. Hastie, K. M. et al. Structural basis for antibody-mediated neutralization of Lassa virus. *Science* **356**, 923–928 (2017).
58. Rutten, L. et al. Structure-based design of prefusion-stabilized filovirus glycoprotein trimers. *Cell Rep.* **30**, 4540–4550.e3 (2020).
59. Caradonna, T. M. & Schmidt, A. G. Protein engineering strategies for rational immunogen design. *npj Vaccines* **6**, 154 (2021).
60. Heinz, F. X. & Stiasny, K. Distinguishing features of current COVID-19 vaccines: knowns and unknowns of antigen presentation and modes of action. *npj Vaccines* **6**, 104 (2021).
61. Premkumar, L. et al. The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci. Immunol.* **5**, eabc8413 (2020).
62. Wang, C. et al. A conserved immunogenic and vulnerable site on the coronavirus spike protein delineated by cross-reactive monoclonal antibodies. *Nat. Commun.* **12**, 1715 (2021).
63. Pinto, D. et al. Broad betacoronavirus neutralization by a stem helix-specific human antibody. *Science* **373**, 1109–1116 (2021).
64. Hurlburt, N. K. et al. Structural definition of a pan-sarbecovirus neutralizing epitope on the spike S2 subunit. *Commun. Biol.* **5**, 342 (2022).
65. Zhou, P. et al. Broadly neutralizing anti-S2 antibodies protect against all three human betacoronaviruses that cause severe disease. *Immunity* **56**, 669–686 (2023).
66. Kirchdoerfer, R. N. et al. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **8**, 15701 (2018).
67. Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D. mRNA vaccines —a new era in vaccinology. *Nat. Rev. Drug Discov.* **17**, 261–279 (2018).
68. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
69. Hoffmann, M., Kleine-Weber, H. & Pöhlmann, S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol. Cell* **78**, 779–784.e5 (2020).
70. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643–2651 (2014).
71. Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
72. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
73. Ream, D. & Kiss, A. J. NCBI/GenBank BLAST Output XML Parser Tool. https://www.semanticscholar.org/paper/NCBI-GenBank-BLAST-Output-XML-Parser-Tool-Ream-Kiss/3ead0ae31b91d3096369de11f3488024f752bdc5 (2013).
74. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
75. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
76. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
77. Rodrigues, J. P., Teixeira, J. M., Trellet, M. & Bonvin, A. M. pdb-tools: a swiss army knife for molecular structures. *F1000Res.* **7**, 1961 (2018).
78. Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E. & Baker, D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci.* **23**, 47–55 (2014).
79. Tan, T. J. C. et al. High-throughput identification of prefusion-stabilizing mutations in SARS-CoV-2 spike (this paper). SARS2_S_fusogenicity_DMS-main. https://doi.org/10.5281/zenodo.7742830 (2023).

Acknowledgements

We thank the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign for assistance with fluorescence-activated cell sorting and deep sequencing. We thank the cryogenic-electron microscopy core facility at the Case Western Reserve University School of Medicine. This work was supported by National Institutes of Health (NIH) R01 AI167910 (N.C.W.), DP2 AT011966 (N.C.W.), R35 GM142886 (K.A.M.), the Michelson Prizes for Human Immunology and Vaccine Research (N.C.W.), the Searle Scholars Program (N.C.W.), and the Bill and Melinda Gates Foundation INV-004923 (I.A.W.).

Author contributions

T.J.C.T. and N.C.W. conceived and designed the study. T.J.C.T. established the fusion assay and performed the deep mutational scanning experiments. T.J.C.T. and N.C.W. analyzed the deep mutational scanning data. T.J.C.T., R.L. and W.O.O. expressed and purified recombinant proteins. Z.M. and X.D. performed cryo-EM analysis. K.A.M. provided the landing pad cells and helped establish the fusion assay. M.Y. and I.A.W. provided the CC12.3 antibody; G.S. and R.A. provided the CC40.8 antibody. T.J.C.T. and C.K. performed the microscopy analysis. T.J.C.T. and N.C.W. wrote the paper and all authors reviewed and/or edited the paper.

Competing interests

N.C.W., K.A.M. and T.J.C.T. have filed a provisional patent application (IL0042US.L) with the University of Illinois covering the deep mutational scanning-based method to identify prefusion-stabilizing mutations for vaccine design and SARS-CoV-2 spike with the K986P/V987P/D994Q mutations described in this article. N.C.W. serves as a consultant for HeliXon. The remaining authors declare no other competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-37786-1.

Correspondence and requests for materials should be addressed to Nicholas C. Wu.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023