

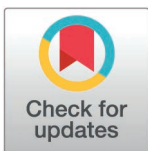
RESEARCH ARTICLE

Structural prediction of chimeric immunogen candidates to elicit targeted antibodies against betacoronaviruses

Jamel Simpson¹, Peter M. Kasson^{1,2,3*}

1 Program in Biophysics and Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, United States of America, **2** Departments of Chemistry and Biochemistry and Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States of America, **3** Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

* kassonlab@gmail.com



Abstract

Betacoronaviruses pose an ongoing pandemic threat. Antigenic evolution of the SARS-CoV-2 virus has shown that much of the spontaneous antibody response is narrowly focused rather than broadly neutralizing against even SARS-CoV-2 variants, let alone future threats. One way to overcome this is by focusing the antibody response against better-conserved regions of the viral spike protein. This has been demonstrated empirically in prior work, but we posit that systematic design tools will further potentiate antigenic focusing approaches. Here, we present a design approach to predict stable chimeras between SARS-CoV-2 and other coronaviruses, creating synthetic spike proteins that display a desired conserved region, in this case S2, and vary other regions. We leverage AlphaFold to predict chimeric structures and create a new metric for scoring chimera stability based on AlphaFold outputs. We evaluated 114 candidate spike chimeras using this approach. Top chimeras were further evaluated using molecular dynamics simulation as an intermediate validation technique, showing good stability compared to low-scoring controls. Experimental testing of five predicted-stable and two predicted-unstable chimeras confirmed 5/7 predictions, with one intermediate result. This demonstrates the feasibility of the underlying approach, which can be used to design custom immunogens to focus the immune response against a desired viral glycoprotein epitope.

OPEN ACCESS

Citation: Simpson J, Kasson PM (2025) Structural prediction of chimeric immunogen candidates to elicit targeted antibodies against betacoronaviruses. PLoS Comput Biol 21(2): e1012812. <https://doi.org/10.1371/journal.pcbi.1012812>

Editor: Rafael J. Najmanovich, Université de Montréal Faculté de médecine: Université de Montréal Faculté de médecine, CANADA

Received: October 1, 2024

Accepted: January 20, 2025

Published: February 5, 2025

Copyright: © 2025 Simpson, Kasson. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Code and scripts used to generate the chimeras, run AlphaFold, and analyze resulting scores are available from <https://github.com/kassonlab/coronavirus-chimera-prediction>. Multiple sequence alignments used are also in the above GitHub repository. Predicted structures simulated and analyzed, simulation trajectories, as well as primary experimental data are available

Author summary

Viral spike glycoproteins often show great antigenic variability in the most immunogenic regions, while broadly neutralizing antibodies frequently recognize secondary sites. This raises the question of how to elicit antibodies against less immunogenic but more conserved antigenic regions. Prior proof-of-concept studies have shown that a series of chimeric antigens can elicit such responses if the highly immunogenic regions are varied and the desired antigenic regions are held fixed. This introduces a protein design question: how to select viral spike chimeras that have comparable or increased stability compared to wild-type parent sequences. This study leverages structural modeling with

from the Zenodo repository doi: [10.5281/zenodo.13830739](https://doi.org/10.5281/zenodo.13830739).

Funding: This work was supported by NIH R01 GM138444, a Coulter Translational Research Award, and a Wallenberg Academy Fellowship 2020.0209 to P.M.K. J.S. was also supported by NSF 1810762. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

AlphaFold to predict the stability of chimeric coronavirus spike proteins, selecting a set of low-homology but high-stability chimeras. These chimeras were then computationally validated using molecular dynamics simulation and experimentally tested. The resulting chimeric spike proteins represent a starting point for further chimeric immunization strategies, and the underlying tools provide a general means to design chimeric proteins to assist future immunization strategies aimed at generating antigenically focused and potentially more broadly neutralizing immune responses.

Introduction

SARS-CoV-2 has caused a global pandemic that has resulted in at least 776 million diagnosed cases of Covid-19 and 7 million deaths as of September 2024 [1]. SARS-CoV-2 uses the receptor binding domain (RBD) of its S protein to bind the ACE-2 receptor on host cells, permitting subsequent fusion [2–4]. The S protein has been a focus for vaccine development because of its importance for infection, its display on the viral surface, and its greater immunogenicity compared to other SARS-CoV-2 proteins [5–8]. The RBD is the immunodominant domain of the S protein, eliciting a large majority of neutralizing antibodies [8–10]. However, it also displays the greatest propensity to mutate [11,12], and RBD mutations in major variants have reduced the effectiveness of current vaccination programs [11,13–17].

One strategy for overcoming variation in the RBD is to elicit neutralizing antibodies that target other parts of the S protein [18–21]. Two major domains that constitute other targets for neutralization are the N-terminal domain (NTD) and the stalk (S2 domain). Neutralizing antibodies against each have been recovered from patients, demonstrating the potential for SARS-CoV-2 neutralization via epitopes that may be better conserved across variants [20–22].

Immunization with “chimeric antigens” is one approach that has successfully elicited antibodies against such less immunogenic regions [23–27]. Chimeric antigens can be constructed by exchanging domains from homologous proteins encoded by different viruses to form a novel protein that maintains the overall approximate structure but varies large portions of the primary sequence. Serial immunization with a set of chimeric antigens that hold one domain constant while varying the others can thus amplify the immune response against the conserved domain.

Work by Krammer and colleagues [28] has tested serial immunization with chimeric antigens against influenza. In this approach, a series of chimeric influenza hemagglutinin proteins maintained the same stalk region in all the chimeras but replaced the head domain with different group 1 influenza virus head domains [28]. Ferrets vaccinated with this combination of chimeras had higher levels of anti-stalk antibody than those vaccinated with standard influenza vaccines [28]. More importantly, the chimeric immunization strategy elicited a broad neutralization response against different influenza subtypes, assessed both in ferret challenge studies [28] and in neutralization assays as part of human clinical trials [29] (NCT03300050).

This strategy of chimeric immunization has been extended to SARS-CoV-2 by Baric and co-workers [30]. In this work, mice were immunized with different chimeric sarbecovirus S proteins. The best-performing chimeras and immunization strategies were found to elicit antibodies with superior breadth of protection when compared to standard immunization regimens when assessed for protection against related sarbecoviruses in both neutralization assays and challenge studies. In addition, these approaches and S2-only vaccination approaches [31,32] have demonstrated the ability to elicit targeted antibody responses against less immunogenic regions of the S protein. In parallel, broadly-neutralizing antibodies against

S2 have been identified that confer protection against diverse betacoronaviruses [33]. Thus, the driving premise for this work is that an increased ability to design chimeric antigens will broaden the choice of parent sequences for chimeric antigens and facilitate better tuning of these targeted responses.

Based on these successes, we aim to develop a systematic approach for designing chimeric coronavirus antigens. To accomplish this, we leverage the recently-developed AlphaFold structure-prediction method and assess its use in stability prediction for chimeric protein design. AlphaFold2 is a powerful, neural-network-based protein structure prediction algorithm that substantially increased prediction accuracy over prior methods [34] and has been rapidly adopted for different protein design applications [35–37], also spawning a set of related deep-learning approaches for protein structure prediction [38,39]. AlphaFold 3 [40] of course offers additional algorithmic changes, but version 2 was used for this work in part due to the modeling throughput required. This approach complements prior methods for stability prediction, both those that utilize global structure and those that focus on point mutations [41–46]. One advantage of AlphaFold over prior methods is that it can in a single step produce predictions of both structure and global stability, particularly for large domain swaps.

We report both a pipeline and results for chimeric antigen design, where chimeras are formed by replacing the S1 domain of the SARS-CoV-2 S protein with the S1 domain from another coronavirus spike protein, leaving the S2 domain conserved across all chimeras (Fig 1). This approach complements computational designs focused on yielding stable S2-only immunogens [47]. We compare results against molecular dynamics simulation and experimental expression and stability testing and present a set of top-ranked antigens that will be used for immunological testing.

Results

To facilitate prediction of stable chimeras and evaluation of the results, we generated a curated set of 115 coronavirus spike protein sequences (S1 Table). These sequences are evolutionarily diverse and display particular variability in the receptor-binding and N-terminal domains,

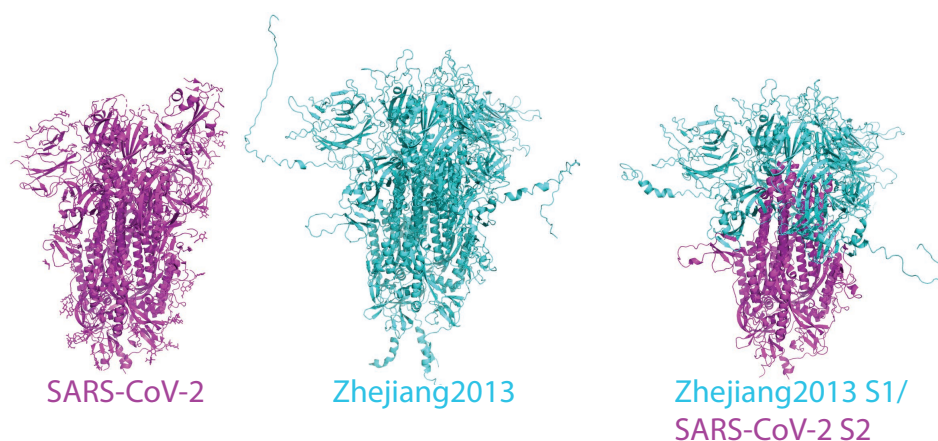


Fig 1. Design of chimeric spike proteins. The process of chimeric S1/S2 spike protein design is illustrated for a sample betacoronavirus sequence. The sequence of SARS-CoV-2 (structure rendered in magenta) is spliced with that of Bat Hp-betacoronavirus/Zhejiang2013 (predicted structure rendered in cyan) at the predicted S1/S2 junction. The predicted structure of the chimera is rendered in a mix of cyan and magenta representing the parental sequences for each domain. pLDDT (predicted local distance difference test) scores from AlphaFold for both the parental domains and chimeric sequences are calculated, and these are averaged per residue to yield the relative stability score.

<https://doi.org/10.1371/journal.pcbi.1012812.g001>

as demonstrated in an analysis of sequence entropy rendered on the SARS-CoV-2 S protein structure (Fig 2). Domain boundaries were identified via sequence alignment, and chimeras were generated computationally by splicing the [S1](#) domain from each of the 114 other sequences onto the [S2](#) domain of SARS-CoV-2 (Fig 1). For comparison, a similar operation

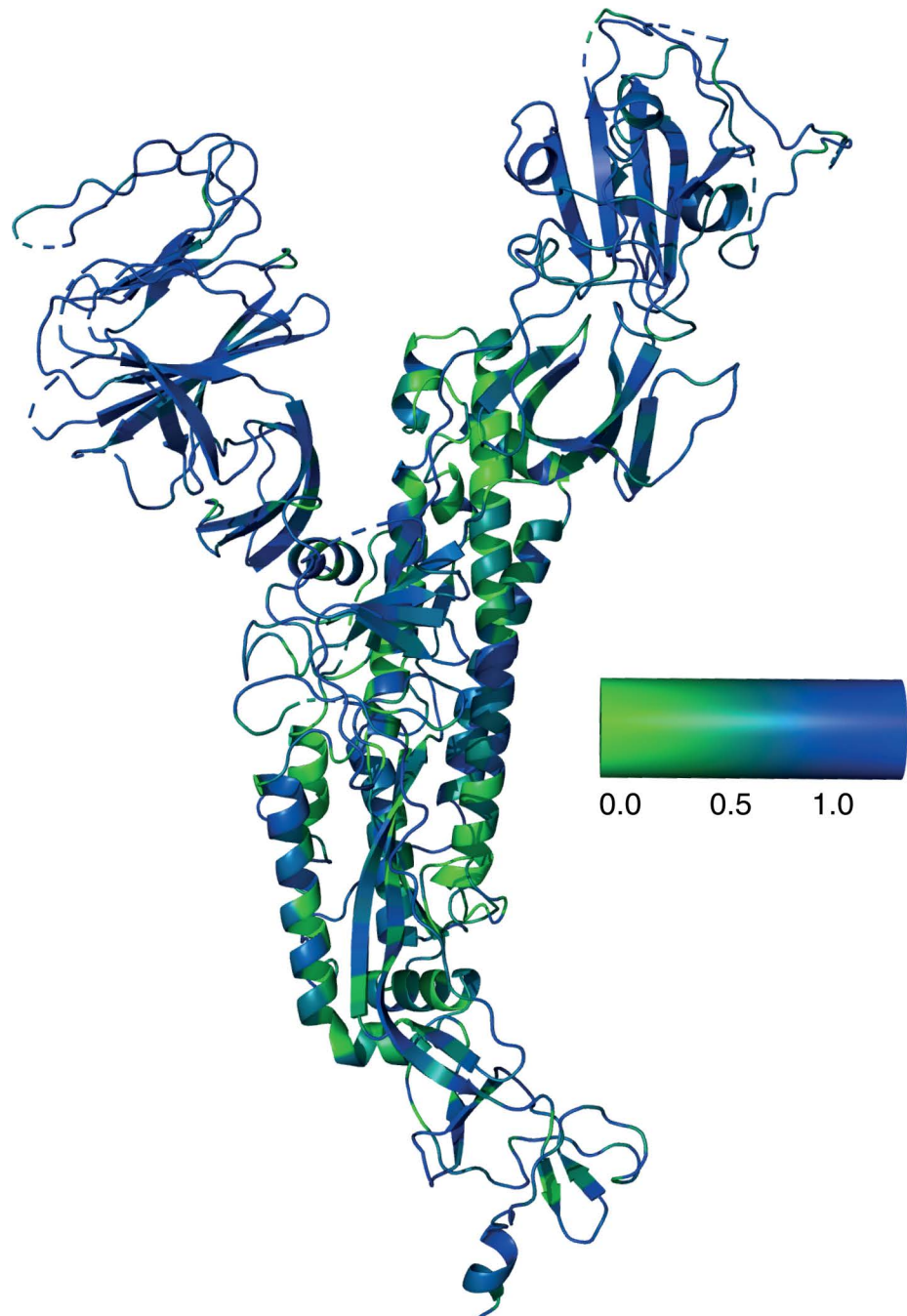


Fig 2. Relative sequence entropy among analyzed coronavirus spike proteins. Sequence entropy was calculated for a multiple sequence alignment of all coronavirus sequences used and rendered on the 6VSB structure, where green represents lowest sequence entropy and blue represents highest. As expected, the receptor-binding domain and N-terminal domain have the greatest sequence entropy, and loops tend to have higher entropy than adjacent structural elements.

<https://doi.org/10.1371/journal.pcbi.1012812.g002>

was performed to splice RBDs from each of the 114 other chimeras onto SARS-CoV-2; those results are presented in [S1 and S2 Figs](#) and [S3 Table](#). Structures for the spike protein trimer from each of these chimeras were then predicted using AlphaFold2, and the corresponding local confidence scores (pLDDT) recorded for use in stability predictions. Predictions were made for the trimeric spike structure because that is the stable folded unit on the surface of SARS-CoV-2 virions. In all cases including SARS-CoV-2 Wuhan, the top AlphaFold2 trimer model had all three receptor-binding domains in the “down” orientation.

Because each of the 114 spike protein sequences other than SARS-CoV-2 also corresponds to a native “wild-type” virus, it must represent a stable folded protein. We therefore evaluated the predicted gain or loss of stability for each of the 114 spliced [S1](#) domains in its chimera with SARS-CoV-2 versus in its full-length native sequence. The analogous comparison was also performed for the region of the chimera derived from SARS-CoV-2 (the [S2](#) domain). Stability scores were thus defined for each chimeric sequence i generated by splicing coronavirus i and SARS-CoV-2 as:

$$\text{Relative Stability} = \text{mean} \left(\frac{\left(p\text{LDDT}(\text{Chimera}_{i,j}) - p\text{LDDT}(\text{Native}_{i,j}) \right)}{p\text{LDDT}(\text{Native}_{i,j})} \right),$$

where pLDDT are predicted local distance difference test [\[48\]](#) scores emitted by AlphaFold, $\text{Native}_{i,j}$ was defined as the residue corresponding to j in either SARS-CoV-2 or coronavirus i , depending on which native protein the residue derived from.

Relative stabilities were calculated according to the above relationship and plotted versus sequence similarity in [Fig 3](#). As expected, coronavirus [S1](#) sequences highly similar to SARS-CoV-2 tended to yield stable chimeras, while sequences less similar to SARS-CoV-2 tended to yield less-stable chimeras. This was in line with our hypothesis that low-similarity coronavirus spike proteins would form poor [S1/S2](#) chimeras. As illustrated in [S3 Fig](#), the interface between [S1](#) and [S2](#) regions displays substantial interdigitation and a buried solvent-accessible surface area of 5154 Å² and thus is not expected to lend itself to modular swaps between dissimilar spikes. For comparison, the HexaPro stabilized SARS-CoV-2 spike trimer construct [\[50\]](#) was also modeled and yielded a relative stability of -8.0%. This makes sense because the protease-resistance mutations are not reflected in AlphaFold scoring, so the predicted stability increase mostly reflects stabilization of flexible regions. The Spearman rank correlation between sequence similarity and relative stability was calculated at 0.51 for the entire group of chimeras. The chimeras of greatest potential utility, however, were those that *violated* this sequence-stability correlation: where a sequence highly divergent from SARS-CoV-2 yielded a chimera with minimal decrease in stability. If correct, these would represent stable proteins that differ antigenically from SARS-CoV-2 in the exchanged region. For comparison, we also computed FoldX [\[41\]](#) stability scores for all chimeras; these are plotted in [S4 Fig](#) and show a 0.89 Spearman rank correlation with pLDDT and a 0.78 Spearman rank correlation with relative stability. Overall, the predicted ranking of chimeras is similar, but there are some substantial outliers.

The sequences in our analysis showed a bimodal similarity to SARS-CoV-2, with most sequences ≤40% similarity or ≥65% similarity in the [S1](#) domain. Because our goal was to identify antigenically different chimeras, we focused on the low-similarity group and selected the eight chimeras in this group with the highest relative stability for further analysis. Each of these chimeras was simulated using molecular dynamics for 100 ns as a computational proxy for physical stability. Structures were evaluated for unfolding as well as structural fluctuations, and the results are plotted in [Figs 4-6](#). Five low-similarity, low-stability sequences were also

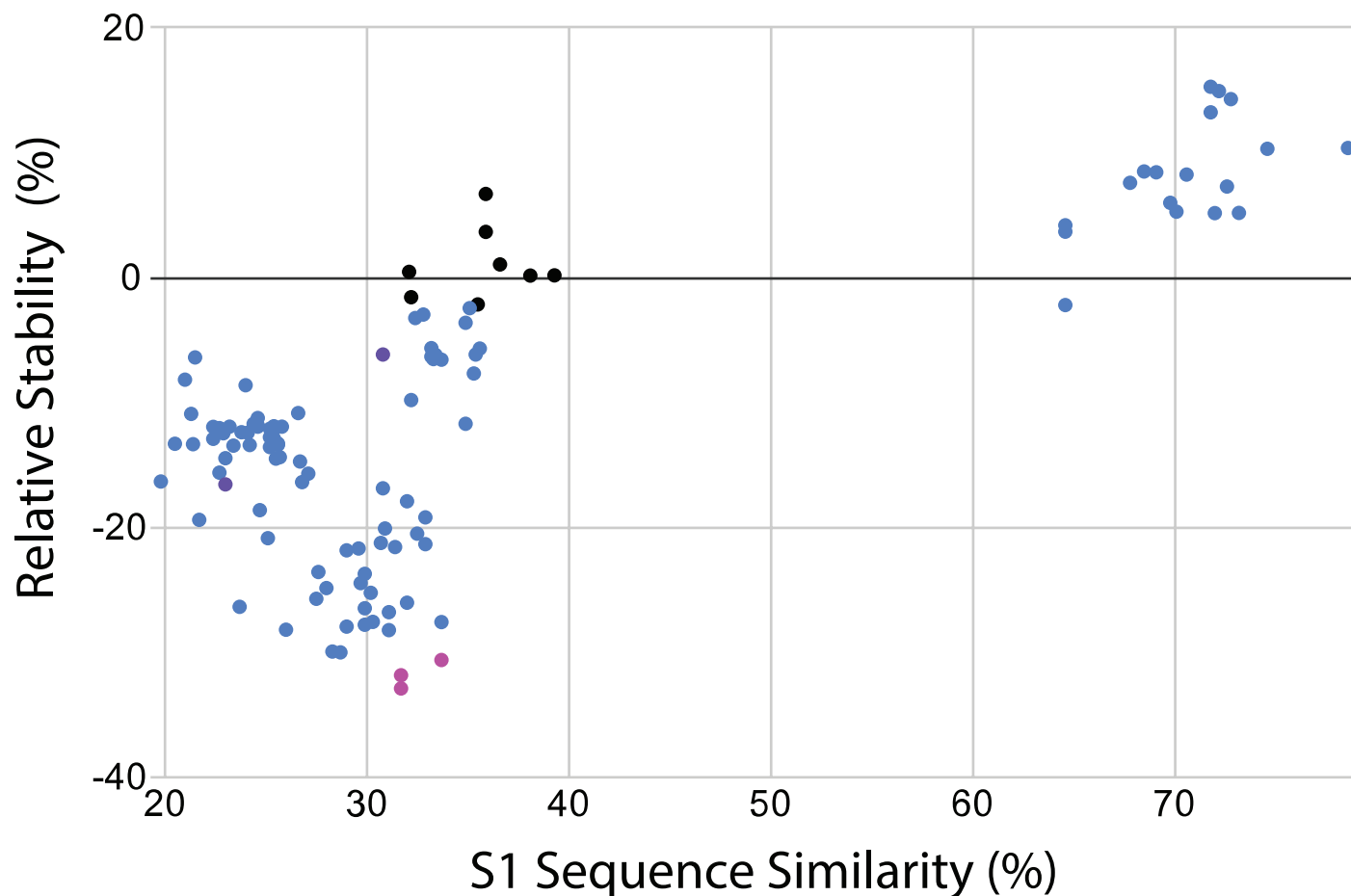


Fig 3. Predicted stability of coronavirus spike chimeras. The relative stability score for each [S1/S2](#) chimera is plotted against the sequence similarity between the parental [S1](#) sequence and SARS-CoV-2. The similarity between the two domain sequences was determined using EMBOSS [\[49\]](#). Sequences fall into two broad groups: one with high similarity and a gain in relative stability, and one with low similarity and a loss in relative stability. The outliers that are low similarity but relatively high relative stability are of particular interest for immunogen design. Predicted high-stability chimeras selected for simulation are plotted in black, and predicted low relative-stability chimeras used as controls are plotted in magenta. Additional low overall-stability controls are plotted in purple. Data are tabulated in [S4 Table](#).

<https://doi.org/10.1371/journal.pcbi.1012812.g003>

simulated as negative controls (three based on low relative stability and two on low mean pLDDT).

All predicted-stable chimeric proteins simulated showed minor structural fluctuations in the first 5 ns of simulation and then remained stable for the remainder of the simulation ([Fig 5](#)). Structural changes were limited; some loops and termini that were initially modeled as disordered became more compact and acquired secondary structure ([Fig 4f](#)). These changes for Zhejiang2013 caused the root-mean-squared deviation (RMSD) to be slightly higher than that of the lowest-RMSD control chimera, HKU15). Much of the conformational fluctuation observed in the simulations was concentrated near the furin cleavage site and similar loops; as well as the N-terminus and C-terminus ([Fig 6](#)). In part because the [S1/S2](#) loop was disordered in the original full-length SARS-CoV-2 structures and mutated away in others [\[6,51\]](#), it is unsurprising that AlphaFold modeled this loop in a physically less plausible state. In chimeras expressed as stable immunogens, we would expect to follow a strategy similar to that pursued successfully with SARS-CoV-2 [\[50\]](#) where either the furin site was mutated or the [S1/S2](#)

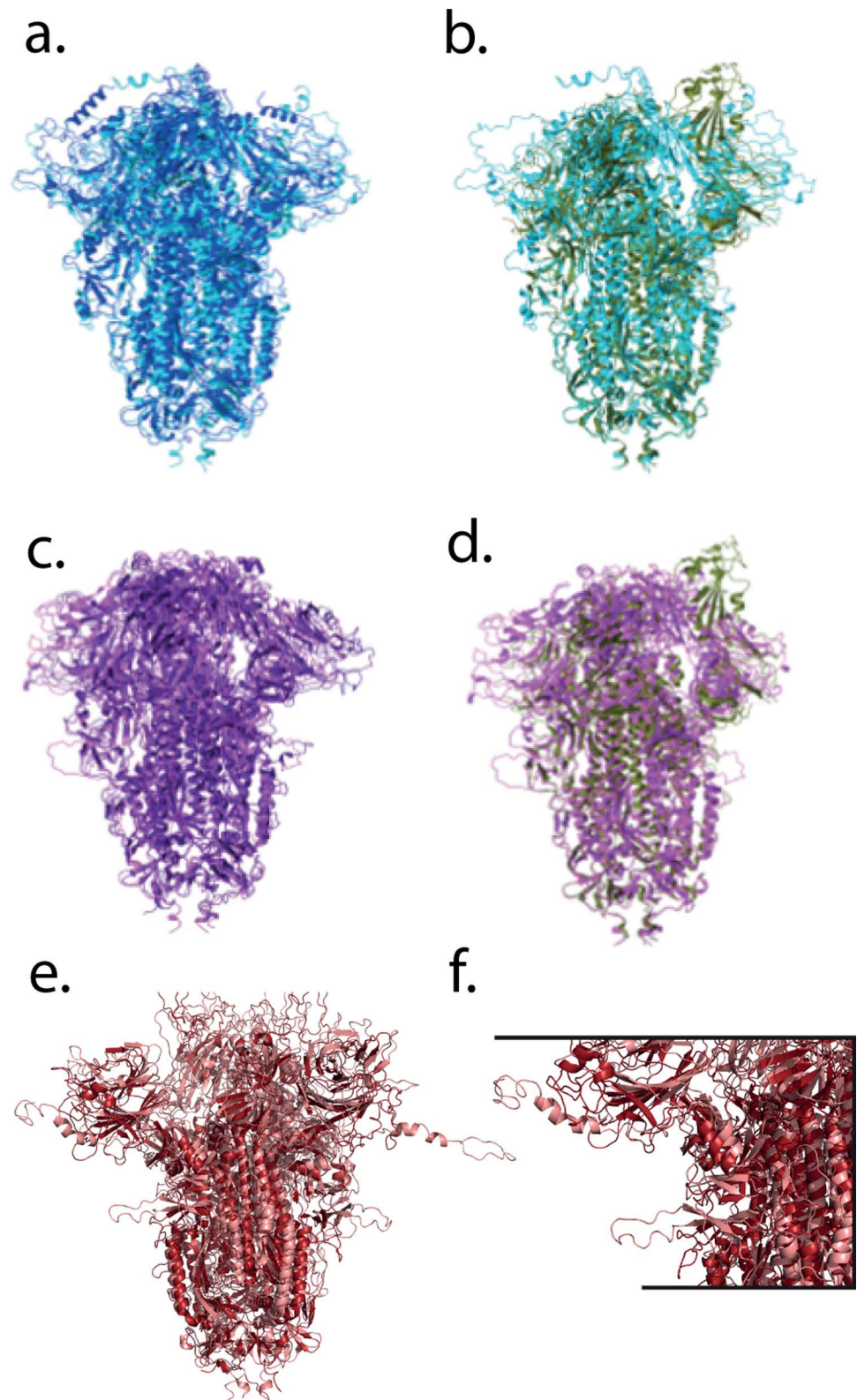


Fig 4. Predicted and simulated structures of top-scoring chimeras. Structures of two of the top-scoring chimeras are rendered as follows. Chimeras from Roussettus bat coronavirus GCCDC1 (panels (a) and (b)) and Hedgehog coronavirus 1 (panels (c) and (d)) are shown before and after 100 ns of molecular dynamics simulation (left column),

with the before and after structures also superimposed on renderings of the SARS-CoV-2 6VSB PDB structure (panels b and d, dark green). The chimera from Bat Hp-betacoronavirus Zhejiang2013 is rendered in panel (e), with a zoomed rendering in panel (f) showing the n-terminal and furin-cleavage loops that were initially modeled as extended becoming more compact over the simulation. For each chimera, the structure before simulation is rendered in a lighter shade and after simulation is rendered in a darker shade. The 6VSB SARS-CoV-2 structure has one receptor-binding domain in the “up” conformation, whereas both chimeras rendered are modeled as fully down and remain so throughout the simulation.

<https://doi.org/10.1371/journal.pcbi.1012812.g004>

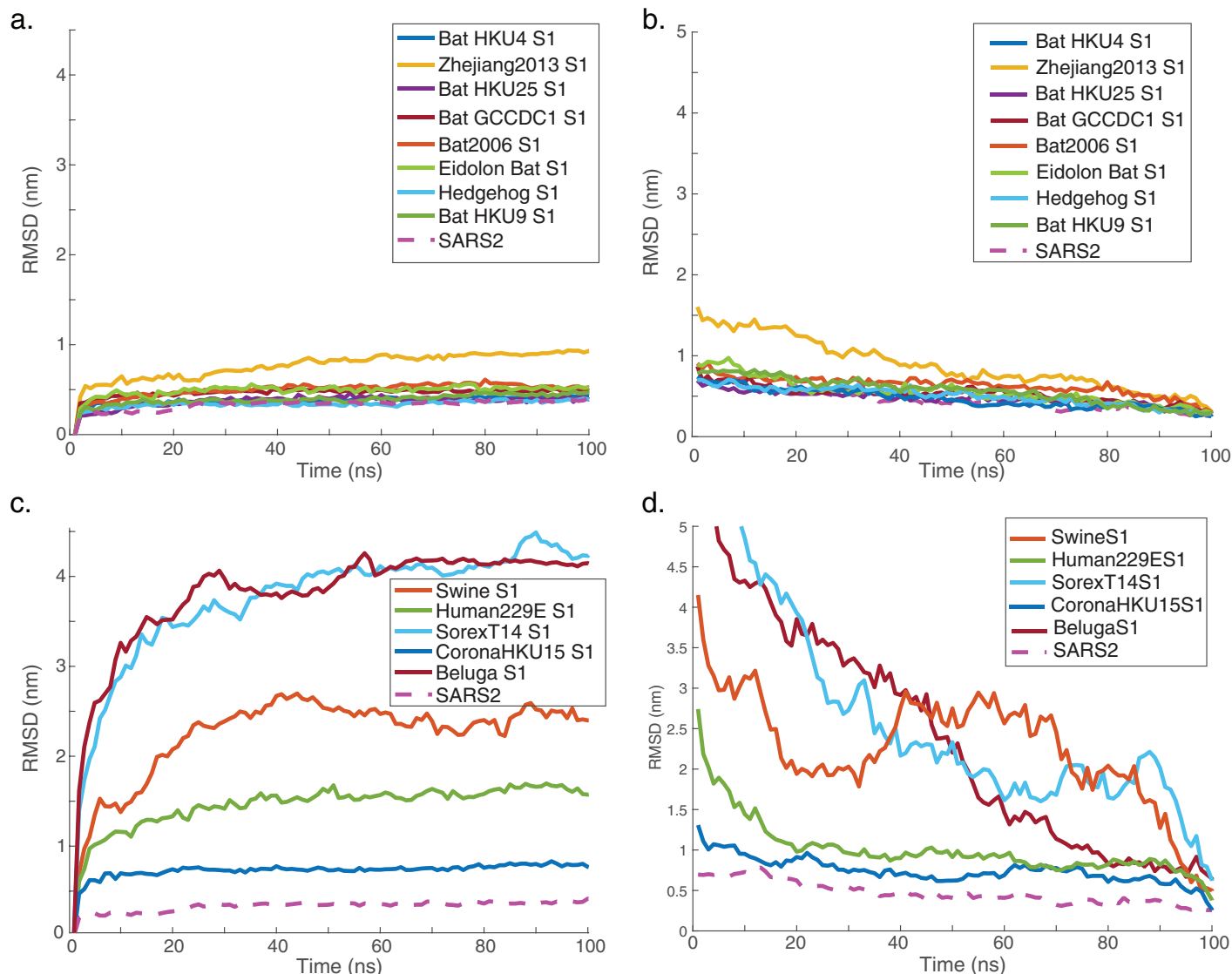


Fig 5. Structural stability of simulated chimeras. Molecular dynamics simulations were computed for predicted structures and root-mean-squared deviation (RMSD) from a reference structure is plotted versus time. Plots are given for (a-b) the eight chimeras in the low-sequence-similarity cluster with the highest predicted relative stability (c-d) five low-stability controls. SARS-Cov-2 is plotted in both panels as an additional control. Panels (a,c) plot RMSD from the starting structure, whereas panels (b,d) plot RMSD from the ending structure. As expected, the low-stability controls showed the greatest RMSD, with one showing gross structural changes. The eight predicted-stable chimeras and SARS-CoV-2 all showed an initial increase in RMSD followed by a stabilization over the course of the 100-ns simulation. This can also be seen in RMSD plots relative to the end of the simulation (S6 Fig).

<https://doi.org/10.1371/journal.pcbi.1012812.g005>

loop truncated altogether. Furthermore, the [S2](#) domains for the predicted-stable chimeras remained highly stable throughout the simulations ([Fig 6](#)), as would be hoped for constructs designed to help focus an immune response against conserved [S2](#) epitopes. Predicted-unstable chimeras showed much more structural fluctuation. This was due to a combination of two factors: regions that were intrinsically less stable (correct AlphaFold predictions; [S5 Fig](#)) and regions that became much more compact over the simulation (potentially incorrect AlphaFold predictions). This suggests that the scoring method used may have some false-negative results – chimeras where the predicted stability is much less than the simulated stability – but few or no false-positive results – predicted-stable chimeras were indeed highly stable on simulation.

Experimental testing was performed by expressing chimeric full-length spike proteins on pseudoviral particles in mammalian cell culture, purifying the particles, and assaying stability. Five predicted-stable chimeras were tested, along with two predicted-unstable chimeras and native SARS-CoV-2 as a comparator. Similar to previous work [\[50\]](#), expression of spike protein was assayed as a first indicator of stability. In this case, expression was measured via spike anti-[S2](#) immunofluorescence on individual pseudoviral particles. The results ([Fig 7](#)) show substantially stronger expression for Eidolon Bat [S1](#) and Bat GCCDC1 [S1](#) chimeras than for native SARS-CoV-2, while Bat2006 [S1](#) yielded similar expression on purified pseudoviral particles, and Zhejiang2013 [S1](#) yielded somewhat lower expression. The two predicted-unstable chimeras, Swine [S1](#) and Sorex T14 [S1](#), yielded 2.5-fold to 12-fold lower, along with one predicted-stable spike, Bat HKU25 [S1](#), at 4-fold lower expression, which we regard as a negative result. Of note, measuring expression via immunofluorescence also demonstrates that major [S2](#) epitopes are stably displayed on the surface of pseudoviral particles expressing the chimeras and capable of binding antibody.

Further testing of thermostability was performed using differential scanning fluorimetry of pseudoviral particles using particles lacking any spike construct as a negative control. All thermostability assays of course rely upon sufficient levels of expressed spike protein, so low-expressors are expected to behave more like “bald” pseudoviral particles in these assays. Thermal profiles are shown in [Fig 8](#). SARS-CoV-2, Eidolon Bat [S1](#), and Bat2006 [S1](#) all show a bimodal unfolding profile with inflection points in the 50–60 °C range. This bimodal profile is highly similar to that previously reported for purified SARS-CoV-2 spikes and variants thereof, including the HexaPro stabilized mutant [\[50\]](#). The profile likely reflects the multi-domain nature of the coronavirus spike trimer and incomplete cooperativity in the unfolding transition. Melting temperatures for full-length spike proteins on the surface of pseudovirus were somewhat higher than for the previously characterized soluble variants, as might be expected. Surprisingly for the degree of protein expression observed, Bat GCCDC1 [S1](#) did not show a high-magnitude unfolding peak over that temperature range. However, it did display substantial shoulders on the underlying pseudovirus particle unfolding peak that register as deviations from bald pseudovirus ([Fig 8b–c](#)) and may represent higher-temperature unfolding than native SARS-CoV-2, in the 70–80 °C range. Definitive assignment would require further confirmation, but the overall profile is consistent with a stable chimeric spike, in accordance with prediction. Overall, these sets of experiments suggest that 3/5 predicted-stable chimeras were indeed equivalently stable to native SARS-CoV-2, one was intermediate, and one was relatively unstable. 2/2 predicted-unstable chimeras were indeed unstable upon expression at 37 °C in mammalian cell culture.

Discussion

Here, we introduce a measure of relative stability to assess chimeric sequences: the predicted stability relative to the parent sequences for each portion of the chimera. As assessed by molecular dynamics simulations of the AlphaFold models, relative stability is a fairly effective metric for selecting stable chimeras. An alternate metric, however, would be average pLDDT, or absolute predicted stability. These two metrics rank the chimeras examined here very

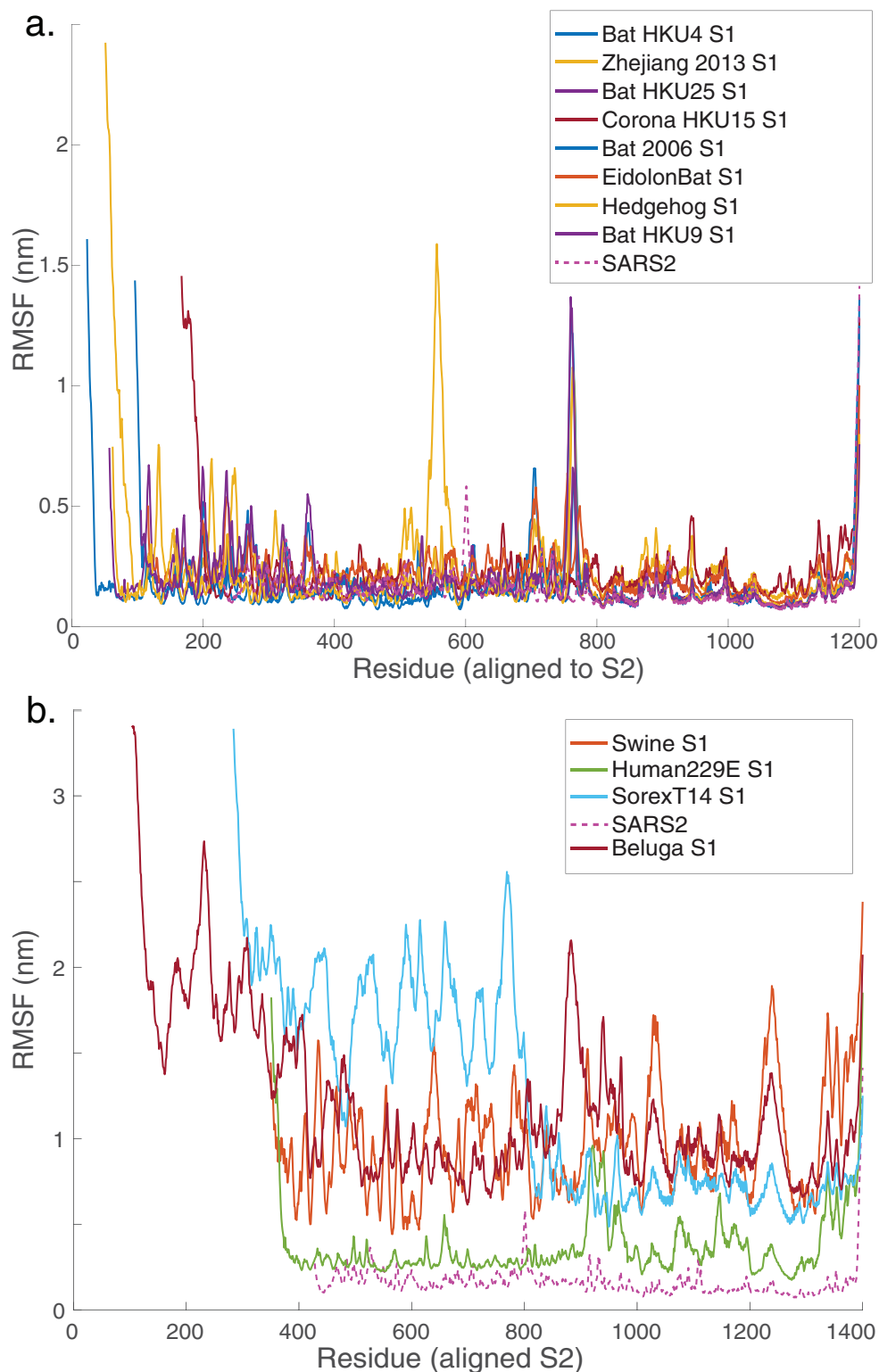


Fig 6. Per-residue structural stability of simulated chimeras. Root-mean-squared fluctuation (RMSF) values are plotted per residue for each of the simulated chimeras. Values were calculated on nanosecond intervals throughout the simulation trajectory. Panel (a) shows the predicted high-stability chimeras, and panel (b) shows the predicted low-stability chimeras. SARS-CoV-2 is included as a comparator. As expected, major loops as well as the C- and N-termini show the greatest fluctuation in the high-stability chimeras, and the low-stability chimeras have globally greater fluctuations.

<https://doi.org/10.1371/journal.pcbi.1012812.g006>

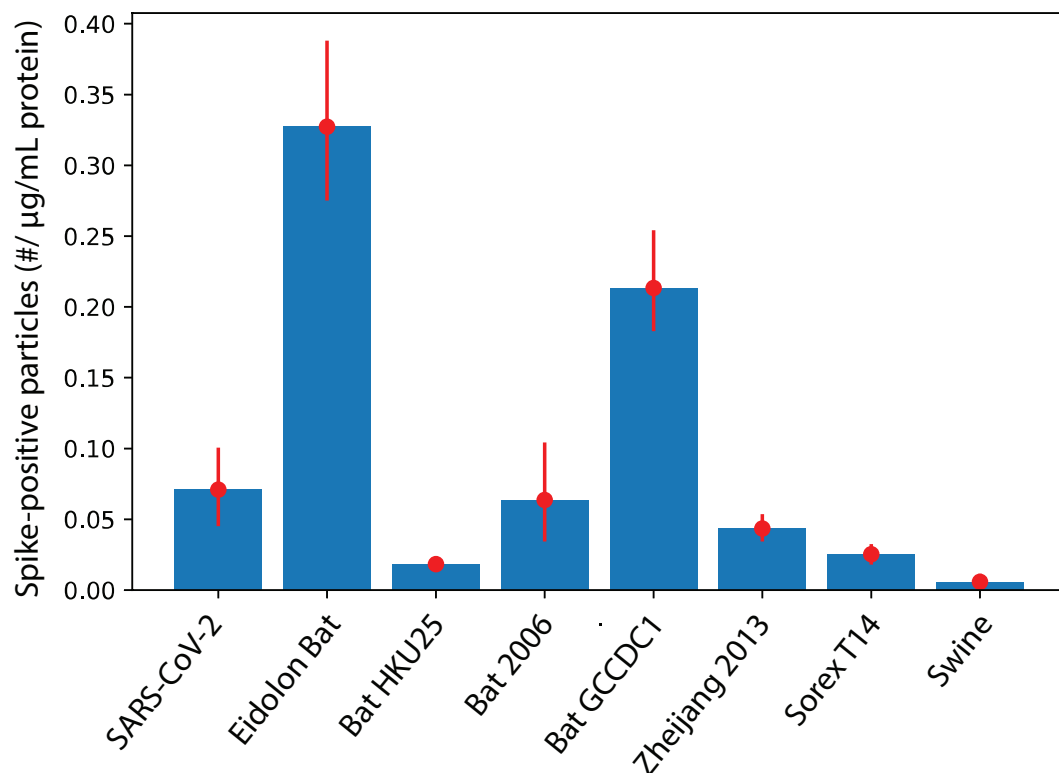


Fig 7. Expression of chimeric spike proteins on purified pseudoviral particles as assessed by single-particle immunostaining. Pseudoviral particles were purified, biotinylated, immobilized in a microfluidic flow cell via a biotin-neutravidin-biotin linkage, and then immunoassayed using an anti-S2 antibody. The number of spike-positive particles was then scaled by the total protein in the pseudoviral preparation to give a relative measure of spike expression per pseudovirus, which is plotted here. Eidolon Bat and Bat GCCDC1 showed substantially better expression than native SARS-CoV-2, Bat2006 was comparable to native SARS-CoV-2, while Zhejiang2013 was somewhat lower. BatHKU25 and the two predicted-unstable chimeras, Swine and Sorex T14 showed substantially worse expression, likely indicating spike instability at the stage of expression and pseudoviral budding. Error bars indicate 90% confidence intervals from bootstrap resampling across fields of view.

<https://doi.org/10.1371/journal.pcbi.1012812.g007>

similarly, with a Spearman rank correlation of 0.86. However, the ranking within the top 8 chimeras selected for simulation is essentially uncorrelated (Spearman rho 0.095). Relative stability is a much stronger predictor of stability in molecular dynamics simulation (Spearman rho 0.38 vs. -0.048 with average RMSD over each 100 ns simulation), although both of these are imprecise predictors of the molecular dynamics simulations. Our results thus suggest that relative stability is a better scoring metric for chimeric proteins but that a secondary screen with molecular dynamics simulation adds substantial further information. Experimental testing is largely concordant with the computational predictions, with the stable chimeric spikes showing long-term stability in molecular dynamics simulations, the predicted-unstable spikes being unstable throughout, and one of the two experimentally least-stable among the original positive predictions showing intermediate stability in molecular dynamics simulations. The only true outlier was Bat HKU25 [S1](#), which showed good stability in molecular dynamics simulations yet expressed poorly.

It is interesting to speculate as to why absolute stability may perform less well than relative stability. Relative stability scores take the native viral spike proteins as a baseline and measure predicted stability compared to those. They thus will correct for any AlphaFold scoring errors on the native sequence, but conversely they will fail to capture variation in stability among

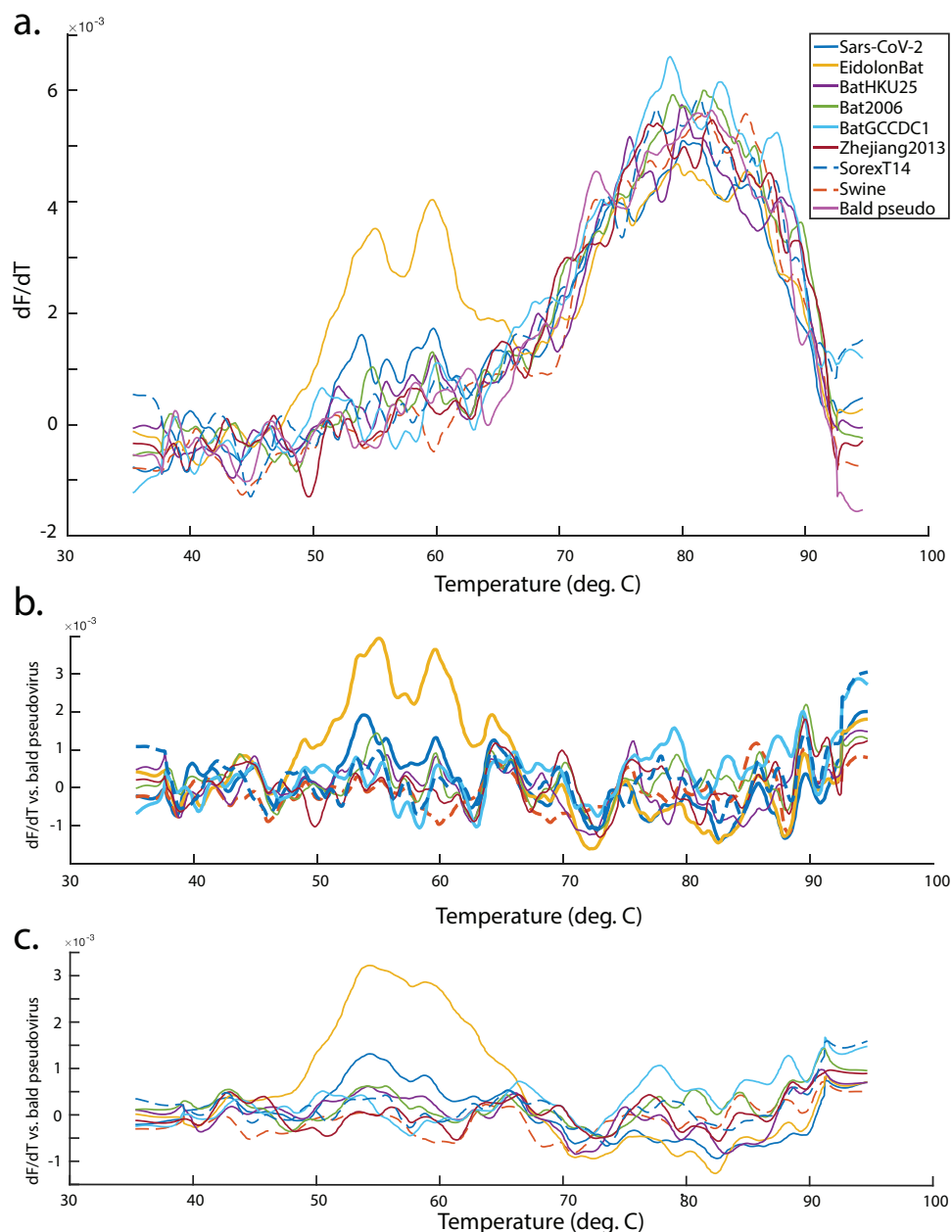


Fig 8. Thermal stability of pseudoviruses expressing chimeric spike proteins. Plotted are dF/dT values for pseudoviruses expressing SARS-CoV-2, 5 predicted-stable chimeric spikes, 2 predicted-unstable chimeric spikes, and bald pseudovirus produced with no spike. The dF/dT values (first derivative of the 350/330 nm fluorescence ratio with respect to temperature) are plotted in panel a, and the background-subtracted dF/dT are plotted in panels b-c, using bald pseudovirus as an estimator of non-spike background. Fluorescence ratios were smoothed using locally weighted scatterplot smoothing and showed peaks that were robust to choice of smoothing parameter over the range 40–80, with smoothing parameter of 50 shown in a–b and 80 shown in c. Eidolon bat chimeras showed stability comparable to native SARS-CoV-2, while Bat GCCDC1 chimeras were less well determined but consistent with thermal stabilization relative to SARS-CoV-2. Bat 2006 chimeras showed lower signal but were likely similar to SARS-CoV-2. Predicted-unstable chimeras had poor expression on pseudovirus produced in cell culture at 37 C (Fig 7) and were thus likely unstable at the point of expression and pseudoviral budding.

<https://doi.org/10.1371/journal.pcbi.1012812.g008>

native viral spike proteins. Thus, on the dataset examined here, AlphaFold prediction errors appear to be somewhat greater than variation in native viral spike stability, at least as scored by molecular dynamics simulation for the most stable predicted sequences.

The approach taken here did not explicitly consider open/closed transitions of the receptor-binding domain or glycosylation patterns of spike proteins. These limitations are reasonable considering our focus on conserved [S2](#) regions, which are indeed glycosylated [\[52\]](#) but potentially less heavily shielded than the [S1](#) region. Glycosylation has also been shown to play a role in RBD conformational transitions [\[53\]](#), which are of course key to antigenic changes within the RBD but have not been linked to any differences in neutralizing epitopes within [S2](#). Furthermore, RBD conformational state has been linked to antigenic exposure and antibody recognition for a number of RBD epitopes. We do not explicitly manipulate such factors in this approach because we are focused on conserved epitopes outside the RBD. However, both of these factors would be suitable subjects for future expansion of the method presented here.

Coronavirus spike immunogens are often expressed and used either as soluble constructs or as full-length proteins. This work analyzed predicted chimeric spikes in both constructs: soluble ectodomains in molecular dynamics simulations and full-length proteins presented in pseudoviral particles in expression and thermostability experiments. This choice was driven both by practical considerations, as full-length simulations in a membrane are more computationally expensive, and by a desire to cover both situations.

AlphaFold may have some generality as a predictor of protein stability for protein-design applications; here we use it specifically to predict chimeras of otherwise well-structured proteins that have a fair degree of overall homology. We use AlphaFold 2 because of its source-code availability (and it was the current version at time of project initiation), but it will be informative to systematically compare results with AlphaFold 3 [\[40\]](#). Within this more limited domain, the relative stability metric shows good success in guiding selection of chimeras that are stable as assessed by molecular dynamics simulation. Ultimately, experimental expression and testing of chimeric proteins will be required to test the best candidates. It is hoped that this approach of AlphaFold prediction, molecular dynamics simulation, followed by experimental testing and optimization as needed will assist in the design of chimeric immunogens against betacoronaviruses as well as other viral pathogens.

Methods

Sequence curation

Sequences were downloaded from the NCBI Virus portal on 10/1/2021 using the following search criteria: Viruses: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, Deltacoronavirus; Proteins: spike, spike protein, spike glycoprotein, surface glycoprotein, spike surface glycoprotein, membrane glycoprotein; Sequence length greater than 700 and less than 2000. All duplicate sequences and SARS-CoV-2 sequences were removed. The resulting sequences are given in [S1 Table](#). The reference sequence for SARS-CoV-2 was defined as the residues that were structurally resolved in chain B of the 6VSB PDB structure [\[54\]](#).

Sequence entropy calculation

All analyzed coronavirus spike sequences were aligned to the SARS-CoV-2 reference sequence, and substitution scores were calculated for each position using the BLOSUM62 matrix. Sequence entropy was then calculated according to the equation

$$H(x) = -\sum_{i=0}^n P(x=i) \log_n P(x=i)$$

where $P(x=i)$ is the probability of BLOSUM value i at position x , and n is the number of possible BLOSUM values.

Chimera Sequence Generation

The [S1](#) domain boundary was taken as previously defined [\[55\]](#) and applied to the SARS-CoV-2 reference sequence as defined above. Pairwise sequence alignments were created between SARS-CoV-2 and each other spike protein sequence retrieved above using Muscle [\[56\]](#) (v.3.8.31), and this alignment was used to define [S1](#) domain boundaries across the spike proteins examined. Sequence similarity between [S1](#) domains was scored using EMBOSS [\[49\]](#) (v6.6.0, alignment using the Needleman-Wunsch algorithm). Custom Python code (available from <https://github.com/kassonlab/coronavirus-chimera-prediction>) was used to generate 114 chimeras swapping the [S1](#) regions from each of the other coronaviruses with the [S1](#) of the SARS-CoV-2 reference sequence.

AlphaFold Modeling

The multimer prediction algorithm in AlphaFold [\[34,57\]](#) (v.2.2.2) was used to analyze the 115 wild-type spike proteins (SARS + 114 others) and 114 chimeric spike proteins. An additional 10 sequences did not yield AlphaFold predictions and are listed in [S2 Table](#); these were excluded from subsequent analysis. Databases passed to AlphaFold modules for generation of multiple sequence alignments were as follows: the Hhblits [\[58\]](#) module used reduced BFD [\[59\]](#), hmmsearch [\[60\]](#) used the PDB SEQRES database [\[61\]](#), and JackHMMer [\[60\]](#) used UniProt [\[62\]](#). Max template date was set to 09-12-2022, and 5 models were created per sequence with the top-scoring model selected after relaxation. Calculations were performed on NVIDIA A100 GPUs, enabling unified memory in Tensorflow to facilitate modeling of long trimeric sequences.

Relative Stability Scores

Relative stability scores were calculated as follows for each residue i :

$$Relative\ Stability_i = \frac{(pLDDT_{i, chimera} - pLDDT_{i, native})}{pLDDT_{i, native}},$$

where $pLDDT_{i, chimera}$ is the pLDDT calculated for residue i in the chimeric sequence and $pLDDT_{i, native}$ is defined as the pLDDT calculated for residue i in the context of the full sequence from which that amino acid derived (in the case of [S1/S2](#) chimeras, SARS-CoV-2 for [S2](#) and the variable coronavirus sequence for [S1](#)). Per-residue relative stability scores thus calculated were averaged across the full trimeric protein to yield a single score for each chimera.

Molecular Dynamics Simulations

Molecular dynamics simulations were run using GROMACS [\[63\]](#) (v2021.2) for each of the 8 chimeras with the greatest relative stability and sequence similarity <40%. Simulations were performed using the CHARMM36 force field [\[64\]](#). Each protein was placed in an octahedral periodic box with a minimum of 1 nm distance to the box boundary and solvated with TIP3P water and 150 mM NaCl. Simulations were run using hydrogen bonds constrained with LINCS and long-range electrostatics treated with Particle Mesh Ewald, temperature set to 310K with the velocity-rescaling thermostat [\[65\]](#), and pressure maintained at 1 bar with the Parrinello-Rahman barostat. Simulations were run for 100 ns using 2-fs timesteps.

Creation of chimeric protein expression vectors

[S1](#) domains corresponding to five predicted-stable chimeras and two predicted-unstable chimeras were synthesized from sequence by Twist Biosciences. Full-length SARS-CoV-2 Wuhan spike with a C-terminal ALAYT mutation was a kind gift of Jesse Bloom and used in his and our previous work [66,67]. Each [S1](#) was amplified via PCR and Dpn1 digested, as was the native SARS-CoV-2 expression plasmid minus SARS-CoV-2 [S1](#). The digested [S1](#) product (range 50–150 ng) was mixed with [S2](#)-vector product in a 1:2 ratio, added to 10 μ L of 2x Gibson Assembly master mix (New England Biolabs), volume-adjusted to 20 μ L with water, and incubated at 50 °C for 45 minutes. Transformation was performed as follows: 40 μ L of XL10 Gold high efficiency *E. coli* (Agilent) was thawed for 3 minutes and mixed with 2 μ L of Gibson Assembly product and incubated for 30 minutes on ice. Each tube was heated to 42 °C for 30 seconds and placed on ice for 2 minutes after adding 350 μ L of SOC media. Tubes were shaken at 190 rpm and 37°C for 1 hr, then plated on a 1 mg/ml Amp agar plate to incubate overnight at 37 °C. A single colony was selected for an overnight 1 mg/ml LB broth culture. The plasmids were extracted, verified via long read sequencing, and stored for pseudovirus expression.

Pseudovirus expression and purification

Chimeric spikes were purified and expressed in a variant of a previously published protocol ⁶⁸. The following plasmids were mixed in a tube: 0.85 μ g chimeric spike plasmid, 2.5 μ g luciferase-IRES-ZsGreen (BEI NR-52516), 0.55 μ g HDM-Hgpm2 (BEI NR-52517), 1.62 μ g pRC-CMV-Rev1b (BEI NR-52519), and 0.55 μ g HDM-tat1b (BEI NR-52518). Opti-MEM media was added for a total volume of 236 μ L, and 14 μ L of lipofectamine was added to 236 μ L of OMEM and incubated for 5 minutes. This mixture was used to transfect 293T cells in a 75-cm² flask; media was replaced after 6 hours, and supernatant harvested after an additional 48 hours. Supernatant was centrifuged at 800 rcf, 4 °C for 8 min and passed through a 0.45 μ m filter to remove cell debris. Pseudoviral particles were crudely purified by ultracentrifugation at 140,000 rcf, 4 °C for 2 hours on a 2.5 ml 25% sucrose/HEPES cushion. Pellets were resuspended in HEPES-MES buffer and rocked gently overnight at 4 °C.

Immunofluorescence

Pseudoviral particles were fluorescently labeled as previously described [66] by incubation with 1% Texas Red-DHPE in HEPES-MES 2 hours at room temperature with gentle rocking, followed by overnight incubation with biotin-PE (83 nM final concentration). A 2- μ L aliquot was diluted 1:3 in HEPES-MES and added to a microfluidic flow cell functionalized with PLL-PEG-Biotin-neutravidin as previously described [66]. The mixture was incubated for one hour at room temperature, washed twice with 500 μ L of HEPES-MES, and blocked with 10 μ L of 2% normal goat serum, and washed once more. Primary immunolabeling was performed with 10 μ L of anti-[S2](#) monoclonal antibody (Invitrogen MA5-35946) at 1 μ g/ml, incubated overnight at 4°C, then washed twice. Secondary immunolabeling was performed using an Alexa 488-labeled goat anti-mouse antibody (Invitrogen) at 1 μ g/ml with a 45-minute incubation followed by a 45-minute incubation and two washes. Imaging was performed on a Zeiss AxioObserver A2 microscope with a Lumencor Spectra-X light engine and an Andor Zyla 4.2plus sCMOS camera, with measurements recorded on five fields of view per condition. Background-subtracted fluorescence intensity was used as a measure of relative spike concentration, controlled against total pseudoviral protein from a microBCA assay.

Nanoscale differential scanning fluorimetry

Pseudoviral particles were diluted to 30 µg/ml of total protein as per microBCA performed above. Fluorescence was measured in a Nanotemper Tycho NT.6, using a scanning rate of 30 °C/minute from 35 °C to 95 °C. The ratio of fluorescence between 350 and 330 nm was measured over this temperature range and smoothed using locally weighted scatterplot smoothing. To measure robustness to smoothing, the smoothing parameter was varied between 40 and 80, and the conclusions drawn remained unchanged over this range. The derivative of fluorescence ratio with respect to temperature (dF/dT) was measured and used to assess thermal unfolding. Particle size distributions were additionally characterized by nanoparticle tracking analysis at both room temperature and after heating to 55 °C. The primary populations were all consistent with monomeric pseudovirus, and there was not a major difference in aggregate formation between SARS-CoV-2 expressing pseudoviruses and bald pseudoviruses after heating, suggesting that, in these expression regimes, spike unfolding is not a major driver of pseudovirus aggregation.

Supporting information

S1 Fig. Predicted Stability of Coronavirus RBD Chimeras. The relative stability score for each receptor-binding domain (RBD) chimera is plotted against the sequence similarity between the parental RBD sequence and SARS-CoV-2. There is less of a clear low-similarity, high-stability cluster than for the S1/S2 chimeras plotted in Fig 3. However, the 9 sequences with similarity <45% and relative stability >-10% were selected for further analysis by molecular dynamics simulation.

(PDF)

S2 Fig. Per-residue structural stability of simulated RBD chimeras. Root-mean-squared fluctuation (RMSF) values are plotted per residue for each of the simulated chimeras. Values were calculated on nanosecond intervals throughout the simulation trajectory.

(PDF)

S3 Fig. Interdigitation of S1 and S2 domains. The model for SARS-CoV-2 Wuhan spike trimer from PDB 6VSB is rendered here with S1 in magenta and S2 in transparent green. Substantial interdigitation can be seen between the domains, with a large contact interface. The buried solvent-accessible surface area at the S1/S2 interface was calculated at 5154 Å². This led us to the hypothesis that chimeras between low-similarity coronavirus spikes may frequently be unstable.

(PDF)

S4 Fig. Comparison of FoldX Scores with AlphaFold pLDDT and Relative Stability Scores. FoldX stability scores were calculated for each chimera and compared both to the relative stability metric (a) and the average pLDDT output from AlphaFold (b). Overall trends were similar, with a Spearman rank correlation of -0.89 between FoldX stability score and AlphaFold pLDDT and -0.78 between FoldX stability score and relative stability. Both plots show outliers with substantially higher FoldX stability.

(PDF)

S5 Fig. Plots of pLDDT versus RMSF over all simulated S1 chimeras. Panels (a-n) show the per-residue pLDDT from AlphaFold plotted against the RMSF from molecular dynamics simulations across the entire dataset of simulated S1 chimeras. Predicted high-stability chimeras are plotted in red, and predicted low-stability chimeras are plotted in magenta. The Spearman correlation coefficient between pLDDT and RMSF over the dataset is -0.69.

(PDF)

S6 Fig. Structural stability of predicted chimeras relative to the end of the simulation.

Root-mean-squared deviation values are plotted relative to simulated structures after 100 ns of simulation. Panel (a) shows the predicted-stable chimeras, and panel (b) shows the predicted-unstable chimeras.

(PDF)

S1 Table. Sequences used in this study.

(PDF)

S2 Table. Sequences Excluded due to AlphaFold Failures.

(PDF)

S3 Table. Relative Stability Scores and Sequence Similarity for RBD Chimeras. Names for chimeras selected for MD simulation are colored in red.

(PDF)

S4 Table. Relative Stability Scores and Sequence Similarity for S1 Chimeras. Names for predicted high-stability chimeras selected for MD simulation are colored in red, and names for predicted low-stability chimeras selected for MD simulation are colored in magenta.

(PDF)

Acknowledgments

Computational resources were provided by Research Computing at the University of Virginia and the PDC Center for High-Performance Computing supported by the National Academic Infrastructure for Supercomputing in Sweden.

Author contributions

Conceptualization: Peter M Kasson.

Formal analysis: Jamel Simpson, Peter M Kasson.

Investigation: Jamel Simpson, Peter M Kasson.

Methodology: Peter M Kasson.

Software: Jamel Simpson.

Writing – original draft: Peter M Kasson.

Writing – review & editing: Jamel Simpson, Peter M Kasson.

References

1. Organization WH. WHO Coronavirus Dashboard. WHO Coronavirus (COVID-19) Dashboard. 2024.
2. Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol.* 2022;23(1):3–20. <https://doi.org/10.1038/s41580-021-00418-x> PMID: [34611326](https://pubmed.ncbi.nlm.nih.gov/34611326/)
3. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* 2020;181(2):271–280.e8. <https://doi.org/10.1016/j.cell.2020.02.052> PMID: [32142651](https://pubmed.ncbi.nlm.nih.gov/32142651/)
4. Letko M, Marzi A, Munster V (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* 5: 562–9. <https://doi.org/10.1038/s41564-020-0688-y> PMID: [32094589](https://pubmed.ncbi.nlm.nih.gov/32094589/)
5. Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg Microbes Infect.* 2020;9(1):382–5. <https://doi.org/10.1080/22221751.2020.1729069> PMID: [32065055](https://pubmed.ncbi.nlm.nih.gov/32065055/)
6. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell.* 2020;181(2):281–292.e6. <https://doi.org/10.1016/j.cell.2020.02.058> PMID: [32155444](https://pubmed.ncbi.nlm.nih.gov/32155444/)

7. Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature*. 2020;584(7821):450–6. <https://doi.org/10.1038/s41586-020-2571-7> PMID: [32698192](https://pubmed.ncbi.nlm.nih.gov/32698192/)
8. Robbani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature*. 2020.
9. Piccoli L, Park Y-J, Tortorici MA, Czudnochowski N, Walls AC, Beltramello M, et al. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding Domain by structure-guided high-resolution serology. *Cell*. 2020;183(4):1024–1042.e21. <https://doi.org/10.1016/j.cell.2020.09.037> PMID: [32991844](https://pubmed.ncbi.nlm.nih.gov/32991844/)
10. Barnes C, West A, Huey-Tubman K, Hoffmann M, Sharaf N. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *bioRxiv*. 2020.
11. Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe*. 2021;29(1):44–57.e9. <https://doi.org/10.1016/j.chom.2020.11.007> PMID: [33259788](https://pubmed.ncbi.nlm.nih.gov/33259788/)
12. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*. 2020;182(5):1295–1310.e20. <https://doi.org/10.1016/j.cell.2020.08.012> PMID: [32841599](https://pubmed.ncbi.nlm.nih.gov/32841599/)
13. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife*. 2020;9:e61312. <https://doi.org/10.7554/eLife.61312> PMID: [33112236](https://pubmed.ncbi.nlm.nih.gov/33112236/)
14. Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature*. 2021.
15. Liu Z, VanBlargan L, Bloyet L-M, Rothlauf P, Chen R, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe*. 2021.
16. Willett BJ, Grove J, MacLean OA, Wilkie C, Logan N, et al. The hyper-transmissible SARS-CoV-2 Omicron variant exhibits significant antigenic change, vaccine escape and a switch in cell entry mechanism. *medRxiv*. 2022. <https://doi.org/10.1101/2022.01.03.21268111>
17. Liu L, Iketani S, Guo Y, Chan J, Wang M, et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature*. 2022;602:676–81. <https://doi.org/10.1038/s41586-021-04388-0> PMID: [35016198](https://pubmed.ncbi.nlm.nih.gov/35016198/)
18. Amanat F, Thapa M, Lei T, Ahmed SMS, Adelsberg DC, Carreño JM, et al. SARS-CoV-2 mRNA vaccination induces functionally diverse antibodies to NTD, RBD, and S2. *Cell*. 2021;184(15):3936–3948.e10. <https://doi.org/10.1016/j.cell.2021.06.005> PMID: [34192529](https://pubmed.ncbi.nlm.nih.gov/34192529/)
19. Voss W, Hou Y, Johnson N, Kim J, Delidakis G, et al. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes in COVID-19 convalescent plasma. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.12.20.423708> PMID: [33398269](https://pubmed.ncbi.nlm.nih.gov/33398269/)
20. Wang Z, Muecksch F, Cho A, Gaebler C, Hoffmann H-H, Ramos V, et al. Analysis of memory B cells identifies conserved neutralizing epitopes on the N-terminal domain of variant SARS-Cov-2 spike proteins. *Immunity*. 2022;55(6):998–1012.e8. <https://doi.org/10.1016/j.immuni.2022.04.003> PMID: [35447092](https://pubmed.ncbi.nlm.nih.gov/35447092/)
21. Changrob S, Fu Y, Guthmiller JJ, Halfmann PJ, Li L, Stamper CT, et al. Cross-neutralization of emerging SARS-CoV-2 variants of concern by antibodies targeting distinct epitopes on spike. *mBio*. 2021;12(6):e0297521. <https://doi.org/10.1128/mBio.02975-21> PMID: [34781736](https://pubmed.ncbi.nlm.nih.gov/34781736/)
22. Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science*. 2020;369(6504):650–5. <https://doi.org/10.1126/science.abc6952> PMID: [32571838](https://pubmed.ncbi.nlm.nih.gov/32571838/)
23. London SD, Schmaljohn AL, Dalrymple JM, Rice CM. Infectious enveloped RNA virus antigenic chimeras. *Proc Natl Acad Sci U S A*. 1992;89(1):207–11. <https://doi.org/10.1073/pnas.89.1.207> PMID: [1370348](https://pubmed.ncbi.nlm.nih.gov/1370348/)
24. Major ME, Vitvitski L, Mink MA, Schleef M, Whalen RG, et al. DNA-based immunization with chimeric vectors for the induction of immune responses against the hepatitis C virus nucleocapsid. *J Virol*. 1995;69: 5798–805.
25. Muster T, Ferko B, Klima A, Purtscher M, Trkola A, Schulz P, et al. Mucosal model of immunization against human immunodeficiency virus type 1 with a chimeric influenza virus. *J Virol*. 1995;69(11):6678–86. <https://doi.org/10.1128/JVI.69.11.6678-6686.1995> PMID: [7474077](https://pubmed.ncbi.nlm.nih.gov/7474077/)
26. Kitson JD, Burke KL, Pullen LA, Belsham GJ, Almond JW. Chimeric polioviruses that include sequences derived from two independent antigenic sites of foot-and-mouth disease virus (FMDV)

- induce neutralizing antibodies against FMDV in guinea pigs. *J Virol.* 1991;65:3068–75. <https://doi.org/10.1128/JVI.65.6.3068-3075.1991> PMID: 1709696
27. Li S, Polonis V, Isobe H, Zaghoulani H, Guinea R, Moran T, et al. Chimeric influenza virus induces neutralizing antibodies and cytotoxic T cells against human immunodeficiency virus type 1. *J Virol.* 1993;67(11):6659–66. <https://doi.org/10.1128/JVI.67.11.6659-6666.1993> PMID: 7692083
 28. Nachbagauer R, Liu W-C, Choi A, Wohlbold TJ, Atlas T, Rajendran M, et al. A universal influenza virus vaccine candidate confers protection against pandemic H1N1 infection in preclinical ferret studies. *NPJ Vaccines.* 2017;2:26. <https://doi.org/10.1038/s41541-017-0026-4> PMID: 29263881
 29. Nachbagauer R, Feser J, Naficy A, Bernstein D, Guptill J. A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial. *Nature Medicine.* 2021;27:106–14. <https://doi.org/10.1038/s41591-020-1118-7> PMID: 33288923
 30. Martinez DR, Schäfer A, Leist SR, De la Cruz G, West A, Atochina-Vasserman EN, et al. Chimeric spike mRNA vaccines protect against Sarbecovirus challenge in mice. *Science.* 2021;373(6558):991–8. <https://doi.org/10.1126/science.abi4506> PMID: 34214046
 31. Ng KW, Faulkner N, Finsterbusch K, Wu M, Harvey R, Hussain S, et al. SARS-CoV-2 S2-targeted vaccination elicits broadly neutralizing antibodies. *Sci Transl Med.* 2022;14(655):eabn3715. <https://doi.org/10.1126/scitranslmed.abn3715> PMID: 35895836
 32. Halfmann PJ, Frey SJ, Loeffler K, Kuroda M, Maemura T, Armbrust T, et al. Multivalent S2-based vaccines provide broad protection against SARS-CoV-2 variants of concern and pangolin coronaviruses. *EBioMedicine.* 2022;86:104341. <https://doi.org/10.1016/j.ebiom.2022.104341> PMID: 36375316
 33. Zhou P, Song G, Liu H, Yuan M, He W-T, Beutler N, et al. Broadly neutralizing anti-S2 antibodies protect against all three human betacoronaviruses that cause deadly disease. *Immunity.* 2023;56(3):669–86.e7. <https://doi.org/10.1016/j.immuni.2023.02.005> PMID: 36889306
 34. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
 35. Jendrusch M, Korbel JO, Sadiq SK. (2021) AlphaDesign: a de novo protein design framework based on AlphaFold. *bioRxiv.* 2021.2010.2011.463937.
 36. Moffat L, Greener JG, Jones DT. (2021) Using AlphaFold for rapid and accurate fixed backbone protein design. *bioRxiv.* 2021.2008.2024.457549.
 37. Anishchenko I, Pellock SJ, Chidyausiku TM, Ramelot TA, Ovchinnikov S, Hao J, et al. De novo protein design by deep network hallucination. *Nature.* 2021;600(7889):547–52. <https://doi.org/10.1038/s41586-021-04184-w> PMID: 34853475
 38. Lin Z, Akin H, Rao R, Hie B, Zhu Z. Evolutionary-scale prediction of atomic level protein structure with a language model. *bioRxiv.* 2022. <https://doi.org/10.1101/2022.07.20.500902>
 39. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 2021;373(6557):871–6. <https://doi.org/10.1126/science.abj8754> PMID: 34282049
 40. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature.* 2024;630(8016):493–500. <https://doi.org/10.1038/s41586-024-07487-w> PMID: 38718835
 41. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res.* 2005;33(Web Server issue):W382–8. <https://doi.org/10.1093/nar/gki387> PMID: 15980494
 42. Yin S, Ding F, Dokholyan NV. Eris: an automated estimator of protein stability. *Nat Methods.* 2007;4(6):466–7. <https://doi.org/10.1038/nmeth0607-466> PMID: 17538626
 43. Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* 2018;46(W1):W350–5. <https://doi.org/10.1093/nar/gky300> PMID: 29718330
 44. Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics.* 2016;32(19):2936–46. <https://doi.org/10.1093/bioinformatics/btw361> PMID: 27318206
 45. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.* 2006;34(Web Server issue):W239–42. <https://doi.org/10.1093/nar/gkl190> PMID: 16845001
 46. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins.* 2006;62(4):1125–32. <https://doi.org/10.1002/prot.20810> PMID: 16372356

47. Nuqui X, Casalino L, Zhou L, Shehata M, Wang A, Tse AL, et al. Simulation-driven design of stabilized SARS-CoV-2 spike S2 immunogens. *Nat Commun.* 2024;15(1):7370. <https://doi.org/10.1038/s41467-024-50976-9> PMID: 39191724
48. Mariani V, Biasini M, Barbato A, Schwede T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics.* 2013;29(21):2722–8. <https://doi.org/10.1093/bioinformatics/btt473> PMID: 23986568
49. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 2000;16(6):276–7. [https://doi.org/10.1016/s0168-9525\(00\)00204-2](https://doi.org/10.1016/s0168-9525(00)00204-2) PMID: 10827456
50. Hsieh C-L, Goldsmith JA, Schaub JM, DiVenere AM, Kuo H-C, Javanmardi K, et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science.* 2020;369(6510):1501–5. <https://doi.org/10.1126/science.abd0826> PMID: 32703906
51. Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM Jr, et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science.* 2020;369(6511):1586–92. <https://doi.org/10.1126/science.abd4251> PMID: 32694201
52. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science.* 2020;369(6501):330–3. <https://doi.org/10.1126/science.abb9983> PMID: 32366695
53. Sztain T, Ahn S-H, Bogetti A, Casalino L, Goldsmith JA, et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat Chem.* 2021;13(10):963–8. <https://doi.org/10.10101/2021.02.15.431212> PMID: 33619492
54. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* 2020;367(6483):1260–3. <https://doi.org/10.1126/science.abb2507> PMID: 32075877
55. Huang Y, Yang C, Xu X-F, Xu W, Liu S-W. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacol Sin.* 2020;41(9):1141–9. <https://doi.org/10.1038/s41401-020-0485-4> PMID: 32747721
56. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
57. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, et al. (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv.* 2004;463034.
58. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics.* 2019;20(1):473. <https://doi.org/10.1186/s12859-019-3019-7> PMID: 31521110
59. Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. *Nat Commun.* 2018;9(1):2542. <https://doi.org/10.1038/s41467-018-04964-5> PMID: 29959318
60. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46(W1):W200–4. <https://doi.org/10.1093/nar/gky448> PMID: 29905871
61. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN. The protein data bank. *Nucleic Acids Res.* 2000;28:235–42. <https://doi.org/10.1093/nar/28.1.235> PMID: 10592235
62. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47(D1):D506–15. <https://doi.org/10.1093/nar/gky1049> PMID: 30395287
63. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 2013;29(7):845–54. <https://doi.org/10.1093/bioinformatics/btt055> PMID: 23407358
64. Huang J, MacKerell AD Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem.* 2013;34(25):2135–45. <https://doi.org/10.1002/jcc.23354> PMID: 23832629
65. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys.* 2007;126(1):014101. <https://doi.org/10.1063/1.2408420> PMID: 17212484
66. Sengar A, Cervantes M, Bondalapati S, Hess T, Kasson P. Single-virus fusion measurements reveal multiple mechanistically equivalent pathways for SARS-CoV-2 entry. *J Virol.* 2023;97(5):e01992. <https://doi.org/10.1128/jvi.01992-22> PMID: 37133381
67. Crawford KHD, Eguia R, Dingens AS, Loes AN, Malone KD, Wolf CR, et al. Protocol and reagents for pseudotyping lentiviral particles with SARS-CoV-2 spike protein for neutralization assays. *Viruses.* 2020;12(5):513. <https://doi.org/10.3390/v12050513> PMID: 32384820