RESEARCH ARTICLE

PROTEIN ENGINEERING

Unsupervised evolution of protein and antibody complexes with a structure-informed language model

Varun R. Shanker[1,2,3], Theodora U. J. Bruun[2,3,4], Brian L. Hie[3,4]†*, Peter S. Kim[3,4,5]*

Large language models trained on sequence information alone can learn high-level principles of protein design. However, beyond sequence, the three-dimensional structures of proteins determine their specific function, activity, and evolvability. Here, we show that a general protein language model augmented with protein structure backbone coordinates can guide evolution for diverse proteins without the need to model individual functional tasks. We also demonstrate that ESM-IF1, which was only trained on single-chain structures, can be extended to engineer protein complexes. Using this approach, we screened about 30 variants of two therapeutic clinical antibodies used to treat severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. We achieved up to 25-fold improvement in neutralization and 37-fold improvement in affinity against antibody-escaped viral variants of concern BQ.1.1 and XBB.1.5, respectively. These findings highlight the advantage of integrating structural information to identify efficient protein evolution trajectories without requiring any task-specific training data.

Evolution generates diverse proteins at the level of biological sequences by exploring a vast search space of potential mutations and acquiring those that improve fitness. However, it is the three-dimensional structure encoded by these sequences that ultimately determines the function and activity of a protein. Therefore, as proteins accumulate mutations, they undergo corresponding structural changes, which in turn facilitate functional adaptations (*1*).

In the laboratory, this tendency for greater sequence change to cause structural divergence poses a major challenge for engineering better proteins through a stepwise evolutionary process. Mutations added in sequential rounds of artificial evolution are increasingly likely to destabilize the structure and therefore diminish the protein's evolvability (*2*). Identifying functionally beneficial mutations is also challenged by the fact that almost all mutations to a prototypical protein are deleterious, or at best neutral, and only a rare subset are beneficial on its fitness landscape (*3*–*8*). In total, these phenomena can often reduce the evolutionarily accessible paths and make evolution more susceptible to local fitness optima (*9*, *10*), further complicating attempts to increase fitness.

To address both the structural constraints of protein design and the high dimensionality of the mutational search space, we used a general protein language model augmented with structural information and trained across millions of nonredundant, single sequence-structure pairs on the sequence recovery task, ESM-IF1 (*11*). Most simply, the model considers the inverse task of that performed by many of the recent powerful structure-prediction tools, including AlphaFold and ESMFold (*12*, *13*): prediction of a sequence that will adopt the fold of a desired target structure (Fig. 1A). This is accomplished by predicting the identity of an amino acid given both the preceding amino acid sequence (referred to as autoregressive modeling) and the entire structure's backbone coordinates (see the materials and methods). Thus, sequences assigned high likelihood scores by the structure-informed language model are expected to fold into the backbone of the input structure with high confidence (Fig. 1B).

The problem of designing a sequence for a desired target structure, such as in inverse folding, is considered only in terms of protein folding (*14*–*16*) and thus does not guarantee a functional protein (*17*). A key barrier to finding an optimal solution for this sequence design problem is that many sequences often fold into a given backbone conformation (*18*). Our framework for protein design does not model an explicit protein function or definition of protein fitness. Rather, using a structure-guided paradigm, we leverage this sequence-structure degeneracy to indirectly explore the underlying fitness landscape by focusing exploration to regions where the backbone fold of the protein is preserved. We hypothesize that constraining evolution to regions of high sequence likelihood can serve as an effective prior for high-fitness variants and thereby improve the efficiency of evolution (Fig. 1C).

We reasoned that this approach may be particularly valuable for the evolution of human antibodies, which are used clinically to treat a broad range of diseases (*19*). Antibodies offer protection by selectively binding to a target antigen involved in pathogenesis and modifying or disrupting its function (*20*). An important optimization step in the development of most therapeutic human antibodies involves an intensive process to identify amino acid substitutions that further enhance potency and efficacy. Here, we demonstrate that this task can be accomplished efficiently with machine learning using an inverse folding model. A central concept of this study is to use the complete structure of the antibody-antigen complex to guide evolution. By conditioning the structure-informed language model on the entire antibody-antigen complex, we sought to enable the discovery of mutations that preserve or enhance the stability of the entire complex and thus improve antibody function.

Results

Enriching sequence exploration for high-function protein variants across diverse tasks with a structure-informed language model

We evaluated whether adding structural information to a language model could be used to guide protein evolution by predicting mutations that improve a protein's activity for a specified property without training on or explicitly modeling the task itself. Accordingly, for 10 proteins from diverse families among four organisms and with functions ranging from enzyme catalysis (TPMT) to oncogenesis (HRAS) to transcriptional regulation (GAL4), we scored variants profiled in large deep mutational scanning experiments (*21*–*30*) against the target backbone of the wild-type protein (*31*–*40*) to compute sequence log likelihoods (see the materials and methods and table S1). These predictions were made in a completely unsupervised setting, with the model never having been trained on any experimental data. To demonstrate the utility for a practical user who wishes to find the most beneficial mutations, we assessed prediction precision by comparing the top-scoring variants against their experimentally determined functional activity in the relative context of the entire sequence-fitness landscape.

Within just the set of top 10 predictions, we identified numerous high-fitness protein variants out of the thousands of tested for each protein, with experimentally determined activities ranking in the top percentiles of the entire deep mutational scanning screen (Fig. 1D). Our analysis demonstrates that conditioning on structural information serves to

[1]Stanford Biophysics Program, Stanford University School of Medicine, Stanford, CA 94305, USA. [2]Stanford Medical Scientist Training Program, Stanford University School of Medicine, Stanford, CA 94305, USA. [3]Sarafan ChEM-H, Stanford University, Stanford, CA 94305, USA. [4]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA. [5]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.
*Corresponding author. Email: brianhie@stanford.edu (B.L.H.); kimpeter@stanford.edu (P.S.K.)
†Present address: Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA; Stanford Data Science, Stanford University, Stanford, CA 94305, USA; and Arc Institute, Palo Alto, CA 94304, USA.

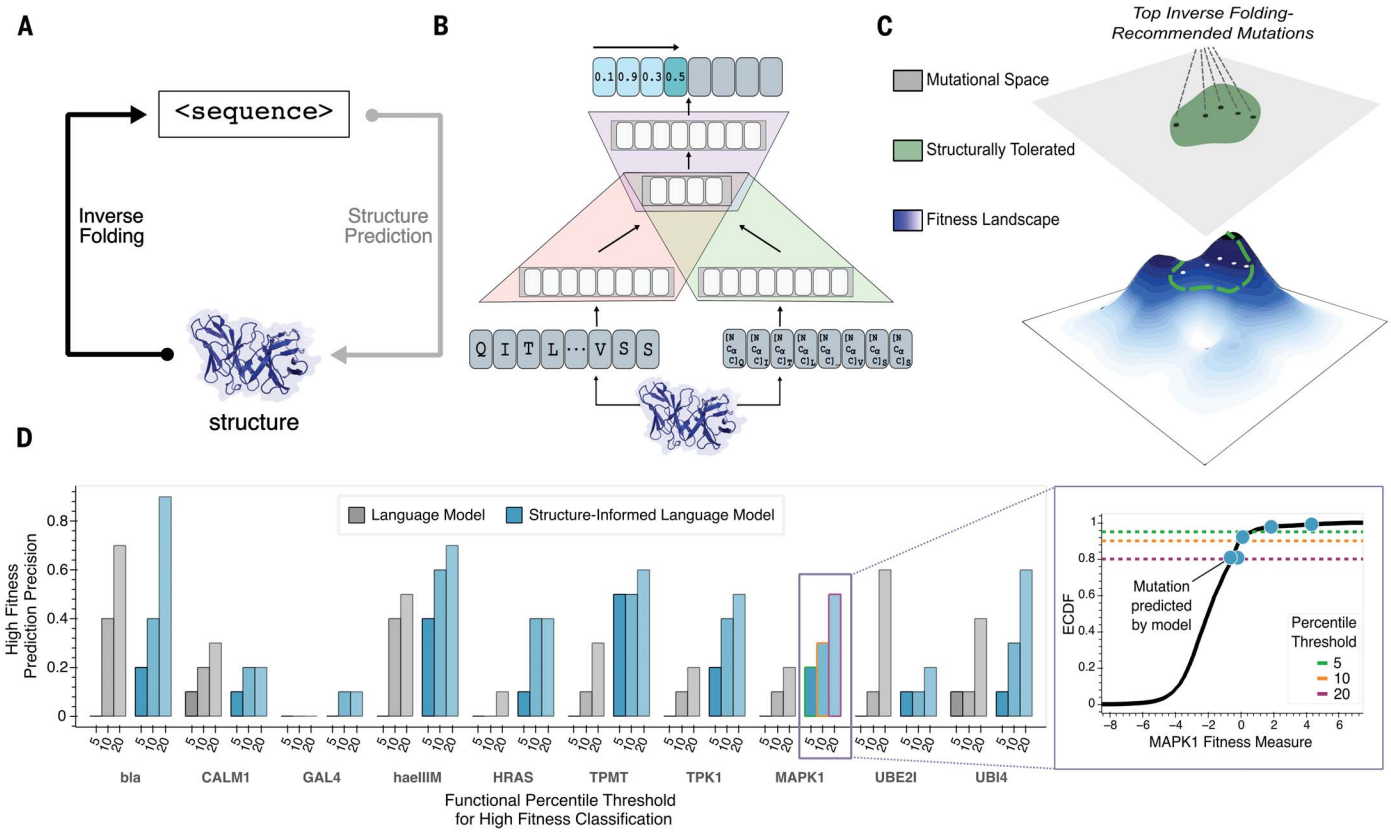

Fig. 1. Guiding evolution of diverse proteins with a structure-guided language model. (**A**) The sequence design problem refers to the prediction of a protein amino acid sequence that will adopt the fold of a given three-dimensional backbone structure, which is conceptually analogous to the inverse problem solved by structure prediction tools such as AlphaFold (*12*). (**B**) A hybrid autoregressive model (*11*) integrates amino acid values and backbone structural information to evaluate the joint likelihood over all positions in a sequence. Amino acids from the protein sequence are tokenized (red), combined with geometric features extracted from a structural encoder (green), and modeled with an encoder-decoder transformer (purple). (**C**) Our structure-guided framework for protein design indirectly explores the underlying fitness landscape without modeling a specific definition of fitness or requiring any task-specific training data by constraining the search space to regions where the backbone fold preserved. (**D**) High-fitness sensitivity analysis reveals that multimodal input improves language model performance compared with sequence-only input across 10 proteins from diverse protein families (left). "High-fitness prediction precision" refers to the fraction of the top 10 single–amino acid substitution predictions that are experimentally determined to confer high protein fitness, defined as having an activity level above the specified percentile threshold among all experimentally screened variants. A representative plot (right) demonstrates this metric for assessing the enrichment of high-fitness MAPK1 mutations. Given the vastness of the search space, finding any function-enhancing variant is valuable for most practical settings, so only successfully predicted mutations are highlighted (blue) on the empirical cumulative density function (ECDF) of the experimental data (black). The three different thresholds, as defined by percentiles, are also shown as dashed lines. Structure-informed language model predictions were more enriched, on average, for high-fitness variants across the various tested thresholds for high-fitness classification. bla, β-lactamase TEM; CALM1, Calmodulin-1; haeIIIM, type II methyltransferase M.HaeIII; HRAS, GTPase HRas; TMPT, thiopurine S-methyltransferase; TPK1, thiamin pyrophosphokinase 1; UBI4, polyubiquitin; UBE2I, SUMO-conjugating enzyme UBC9.

improve predictive capabilities of protein language models by having successfully recovered mutations in the top fifth percentile for nine of 10 proteins compared with just two proteins using a state-of-the-art general protein language model trained only on sequence information and specifically for variant prediction (ESM-1v) (*41*) (Fig. 1D). This improvement in prediction also holds with increasingly relaxed thresholds for classification as high-fitness variants.

On the basis of this experiment, we conclude that structure-based sequence design offers a promising alternative to brute force experimental searches for functionally beneficial mutations. Some of the top mutations predicted are also the same ones discovered from exhaustive experimental exploration. For example, for restriction enzyme haeIIIM, variant Q18E is recommended as one of the top five single amino acid predictions and experimentally ranks as the second-best substitution (and >5 SDs above the mean) of the nearly 2000 substitutions screened across the endonuclease (*30*). Another key advantage of our task-independent framework, in addition to being broadly applicable across diverse proteins, is its ability to improve a single protein for multiple desired properties without the need to develop specialized high-throughput assays to screen each independently. From just the top 10 predictions for mitogen-activated protein kinase 1 (MAPK1), we identified substitutions Q105M and Y64D, which are experimentally shown to confer resistance to two different oncogenic-targeting MAPK1 kinase inhibitors (*24*).

Structural information enables state-of-the-art zero-shot antibody mutational effect prediction for language models

To analyze the effectiveness of augmenting a general protein language model with structural information specifically for antibody variant prediction, we compared the likelihoods of sequences for three antibodies across entire mutational landscapes against corresponding experimental fitness values from a total of five

existing mutagenesis datasets. The mutational landscapes of the first two antibodies were evaluated by measuring the scFv equilibrium dissociation constants (K_D) of all possible evolutionary intermediates between the inferred germline and the somatic sequence of naturally affinity-matured influenza broadly neutralizing antibodies (bnAbs) CR9114 and CR6261, which bind the conserved stem epitope of the influenza surface protein hemagglutinin (HA) (*42*). For both bnAbs, only mutations in the heavy chain, which is responsible for antigen binding, were characterized. The profiled mutational landscape of CR9114 includes all possible combinations of 16 substitutions, whereas that of CR6261 includes all possible combinations of 11 substitutions, totaling 2^{16} = 65,536 and 2^{11} = 2048 variant antibody sequences, respectively. Each of these libraries was screened for binding against two distinct influenza HA subtypes: H1 and H3 for CR9114 and H1 and H9 for CR6261. The fifth dataset assesses the effects of all possible single–amino acid substitutions with a deep mutational scan profiling 4275 mutations in the variable regions for both heavy chain (VH) and light chain (VL) of antibody G6.31 to binding with its ligand, vascular endothelial growth factor A (VEGF-A) (*43*).

For each dataset, we computed the Spearman correlation between the log likelihood estimated by the structure-informed language model and the experimentally determined binding measure for a given antigen across all sequences in the mutational library. We scored the likelihood of each candidate sequence in the library using the backbone coordinates of a structure with the mature antibody bound to its target antigen (*44*–*46*).

Across all five experimental binding datasets, we found that the structure-informed language model performed better than three other sequence-based methods: (i) a general protein language model trained across diverse protein sequences, ESM-1v (*41*); (ii) a specialized antibody language model trained exclusively on sequences sampled from the Observed Antibody Space (OAS) database, AbLang (*47*); and (iii) a site-independent model of mutational frequency curated with extensive antibody sequence alignments, abYsis (*48*). In nearly all experimental scenarios, supplementing sequence information with the backbone coordinates of the antibody alone without providing antigen information as input was sufficient to outperform other sequence-only methods. A notable feature of the autoregressive architecture is that it computes the joint likelihood over all positions in a sequence, making it well suited to score combinatorial sequence changes. We found that this method could capture complex epistatic interactions, or potential interdependence among individual amino acids, because it performs well on the CR9114 and CR6261 libraries composed of sequences with multiple mutations (Fig. 2, A and B).

We achieved the greatest improvement in performance on all five experimental screens by incorporating the structure of both the antibody and antigen (Fig. 2A), indicating that the

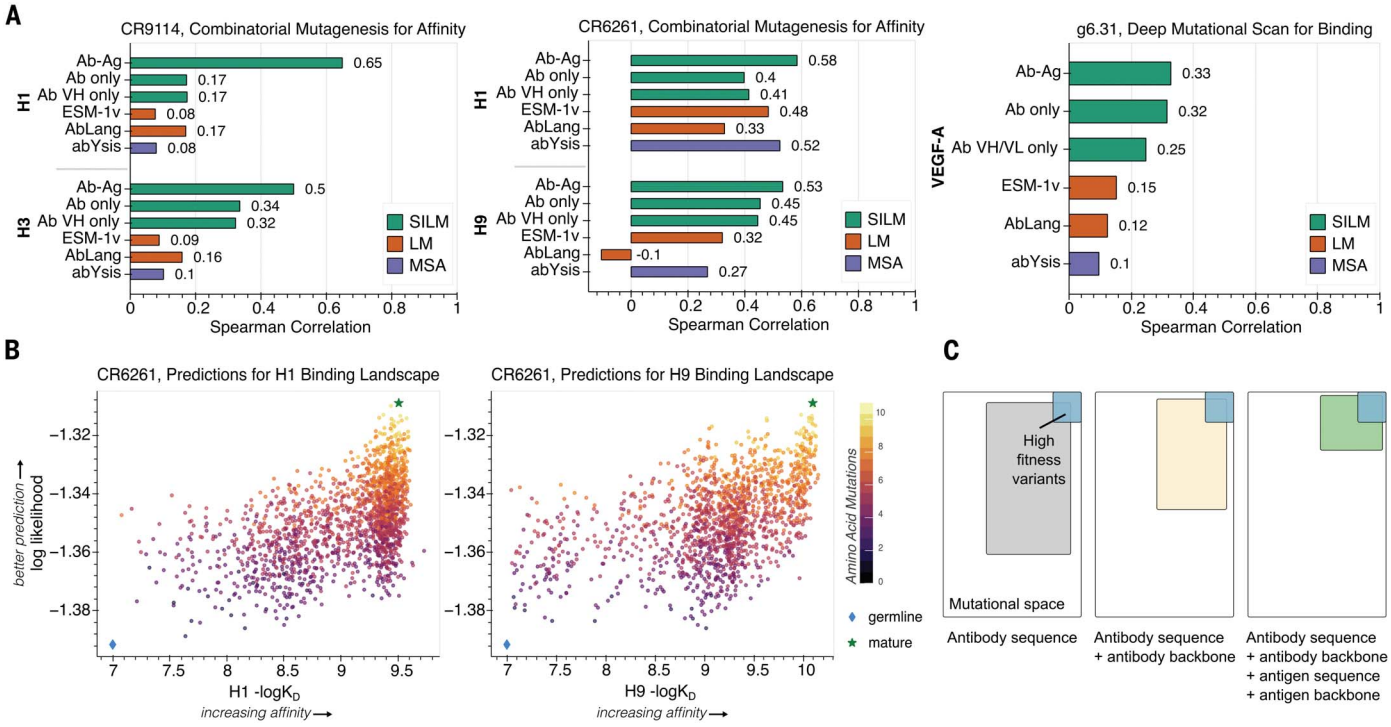

Fig. 2. Prediction of antibody-antigen complexes resolves mutational landscapes by implicitly learning features of binding and protein epistasis. (**A**) Spearman correlation using the structure-informed language model and the sequence-based modeling approaches ESM-1v (*41*), AbLang (*47*), and abYsis (*48*) reported for three antibodies screened with corresponding antigens. Bars are colored by the type of model used: SILM, structure-informed language model (green); LM, language model (orange); and MSA, multiple sequence alignment (purple). The structure-informed language model was evaluated in three different settings: (i) by providing the entire antibody variable region and antigen complex (Ab-Ag); (ii) by providing only the antibody-variable region (Ab only); and (iii) by providing only the single-antibody variable region of the chain responsible for binding or being mutated (Ab VH only or Ab VH/VL only). Antibody sequences scored by the structure-informed language model with antigen information were computed using input complexes of CR9114 with H5 HA [PDB 4FQI (*44*)], CR6261 with H1 HA [PDB 3GBN (*45*)], and g6.31 with VEGF-A [PDB 2FJG (*46*)]. (**B**) Scatter plots showing predictions against experimentally determined dissociation constants of CR6261 against HA-H1(left) and HA to H9 (right). The germline and mature sequences are highlighted on all plots, as indicated on the right. For visualization, all scatter plots omit points on the lower limit of quantitation. (**C**) Conceptual illustration of protein language model performance with improved priors. Providing sequence and structural information for both the antibody and antigen enables the structure-informed language model to most efficiently enrich for high-fitness antibody variants (top right, blue square) by identifying and guiding focused sequence exploration (green square) away from regimes of mutations destabilizing to the complex.

structure-informed model can implicitly learn features of binding (Fig. 2C). This result is notable given that the model is only trained on single-chain protein structures, whereas the antibody-antigen complexes that we used as inputs are composed of either three (G6.31) or four (CR9114 and CR6261) protein chains. The most substantial contribution of antigen information was observed in the case of CR9114-H1, for which the correlation increased from 0.17 with only antibody information to 0.65 with sequence and backbone coordinates of the entire complex. By contrast, this same performance improvement was not observed when the sequence-only general protein language model was provided with additional context of the paired antibody chain or antigen sequence (fig. S4). We found that extending our model beyond the monomeric structures seen during training to protein complexes also performed better for antibody prediction than ProteinMPNN (*49*), an alternate structure-based deep learning method that was trained on a dataset that includes multichain protein structures (fig. S5).

These results showed that we could even predict the effects of mutations to a cross-reactive antibody on binding to a strain of influenza different from the one used as input to the model (Fig. 2, A and B). Despite using a target antibody structure in complex with HA from H5N1 influenza to score CR9114 variants, we obtained correlations of 0.65 and 0.50 with experimental binding data for H1 and H3, respectively. This is particularly important because the antibody epitope, which spans both HA subunits, only has 67% sequence identity between the H5 strain of the structure used to make predictions (A/Vietnam/1203/2004) and the H1 strain experimentally tested against (A/New Caledonia/20/99) (fig. S6 and table S2). This same cross-reactive predictive capability was observed with CR6261 (Fig. 2A), for which the experimentally tested H9 (A/Hong Kong/1073/1999) differs at more than one-third of the residues in the epitope from the 1918 H1N1 influenza strain used in the structure (A/Brevig Mission/1/1918). Although the structure-informed language model cannot learn explicit chemical rules of binding (e.g., hydrogen bonding or disulfide bridge formation) because it does not have access to amino acid side chain atomic coordinates, these results suggest that structural principles such as interface packing or potential steric interference are not only implicitly accessible from residue identities, but are also informative for binding prediction.

Our model's top recommended mutations were made independently of a specific definition of fitness; they simply represent a set of variants with a high likelihood of folding into the input backbone structure. Therefore, our model's recommendations may also help to identify mutations that improve other useful biochemical properties beyond affinity. For example, the top recommended mutation to the VL of G6.31 is F83A, which was identified in the original screening study to be particularly interesting because it confers a threefold increase in VEGF-A binding affinity and a 5°C improvement in melting temperature despite being 25 Å from the antigen and in the antibody framework region. It was determined that the VL F83A substitution induces more compact packing, and the site serves as a conformational switch that affects biological activity at the antibody-antigen interface by modulating both interdomain and elbow angle dynamics (*43*). However, whereas our model successfully enriches for high-fitness variants across many settings, an associated consequence of this structure-based framework is the limited ability to identify mutations that impart beneficial effects by modifying the backbone of the mature antibody.

Engineering therapeutic antibodies for increased potency and resilience

Finally, we investigated whether the structure-augmented language model's predictive capabilities could not only resolve trends on large sets of experimental data, but also enable efficient and successful directed evolution campaigns while testing only a small number (on the order of tens) of variants. To do so, we considered the task of improving the potency and resilience (effectiveness against a virus as it mutates over time) of two mature, clinical monoclonal antibody therapies.

LY-CoV1404 (bebtelovimab), isolated from a COVID-19 convalescent donor, binds to the receptor-binding domain (RBD) of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike protein (*50*). It was approved by the US Food and Drug Administration (FDA) on 11 February 2022 given its activity against both the original Wuhan strain and the Omicron SARS-CoV-2 variants, and it was the last remaining approved monoclonal antibody therapy withstanding viral evolution (*51*) until its discontinuation on 30 November 2022 due to antibody evasion by VOC BQ.1.1 (*52*).

SA58 (BD55-5840) was isolated from a vaccinated individual and is one of two RBD-targeting neutralizing antibodies in a rationally developed antibody cocktail. SA58 alone retained efficacy against all Omicron subvariants, including in vivo protection against BA.5 (*53*, *54*), and was shown to be effective as a postexposure prophylaxis in a clinical study (*55*).

For both antibody engineering campaigns, we used the structure-informed language model to compute likelihoods of all ~4300 possible single-residue substitutions in the VH or VL regions of the antibody. In the first round of evolution, we selected only the top 10 predictions at distinct residues in each chain for experimental validation. An important practical benefit of our method is the ability to optimize against measures of fitness most relevant to the protein's downstream function, such as viral neutralization or receptor agonism, rather than being limited to indirecting surrogate measures such as affinity that are more amenable to high-throughput screening (*4*, *56*). We leveraged this advantage to directly evolve these antibodies for their ability to more potently neutralize SARS-CoV-2–pseudotyped lentivirus.

Variants recommended by the structure-informed language model were assessed by comparing the half-maximal inhibitory concentration (IC_{50}) relative to the wild-type antibody. Although we chose to only test 20 single-site substitutions for each of the two clinical monoclonal antibody therapies, approximately one-third of them improved neutralizing potency. Several of these antibody variants improved neutralization IC_{50} by more than threefold with just a single amino acid change (Fig. 3A and data S1). We also observed greater variance in changes to neutralization for SA58 than Ly-1404, which may be reflective of intrinsic differences in the number of residues critical for and participating in neutralization and binding even beyond the antigen interface.

Prompted by recent evidence showing that conservation of the overall RBD structure is robust to SARS-CoV-2 evolution (*57*), we next sought to determine whether we could also evolve the previously mature antibodies against SARS-CoV-2 BQ.1.1, the variant responsible for diminished therapeutic efficacy. Although the antibodies were previously effective, a change in antigen conceptually represents a fundamental shift in the underlying fitness landscape (Fig. 3B). Accordingly, the baseline neutralization activity against BQ.1.1 for LY-1404 and SA58 dropped to IC_{50} values in the nanomolar range. However, from the same set of 20 single–amino acid substitutions to LY-CoV1404, we found that nearly half improved neutralization of variant BQ.1.1. In addition to a high success rate, we also found that many of these mutations provided a large magnitude of improvement. Several single–amino acid substitutions to LY-CoV1404 individually resulted in a more than fourfold improvement, whereas the most beneficial mutation to SA58 resulted in a nearly fivefold improvement (Fig. 3C).

Approximately two-thirds and one-half of tested single amino acid substitutions to LY-CoV1404 and SA58, respectively, were beneficial for neutralization of either the original strain or BQ.1.1. For both antibodies, the most potent single–amino acid mutations were distinct to the two different strains tested (fig. S7). These results reinforce the idea that despite all of them being predicted to have the same backbone fold, the top set of designed variants feature functional diversity and can be used for distinct notions of protein fitness.

A common challenge in directed evolution is contending with the combinatorial explosion of possible sequences that emerges from trying to combine a set of individually beneficial mutations. In the second round of evolution, we simply used the structure-informed model again to acquire up to five top-scoring combinations of mutations to each antibody chain (see the materials and methods). Across both evolutionary trajectories, all 11 LY-CoV1404 and five SA58 antibody designs with multiple mutations had IC_{50} values better than wild type, with many designs showing synergistic effects upon combination. For example, just a single–amino acid mutation in each of the two chains of SA58 led to a more than 14-fold improvement (Fig. 3, C and D). Similarly, the most potent evolved design of LY-CoV1404 is a combination of seven of the eight beneficial single amino acid substitution to the VH and improved neutralization 25-fold (Fig. 3D). These improvements to neutralizing potency against BQ.1.1 do not sacrifice potency against the original strains. We found that the top SA58

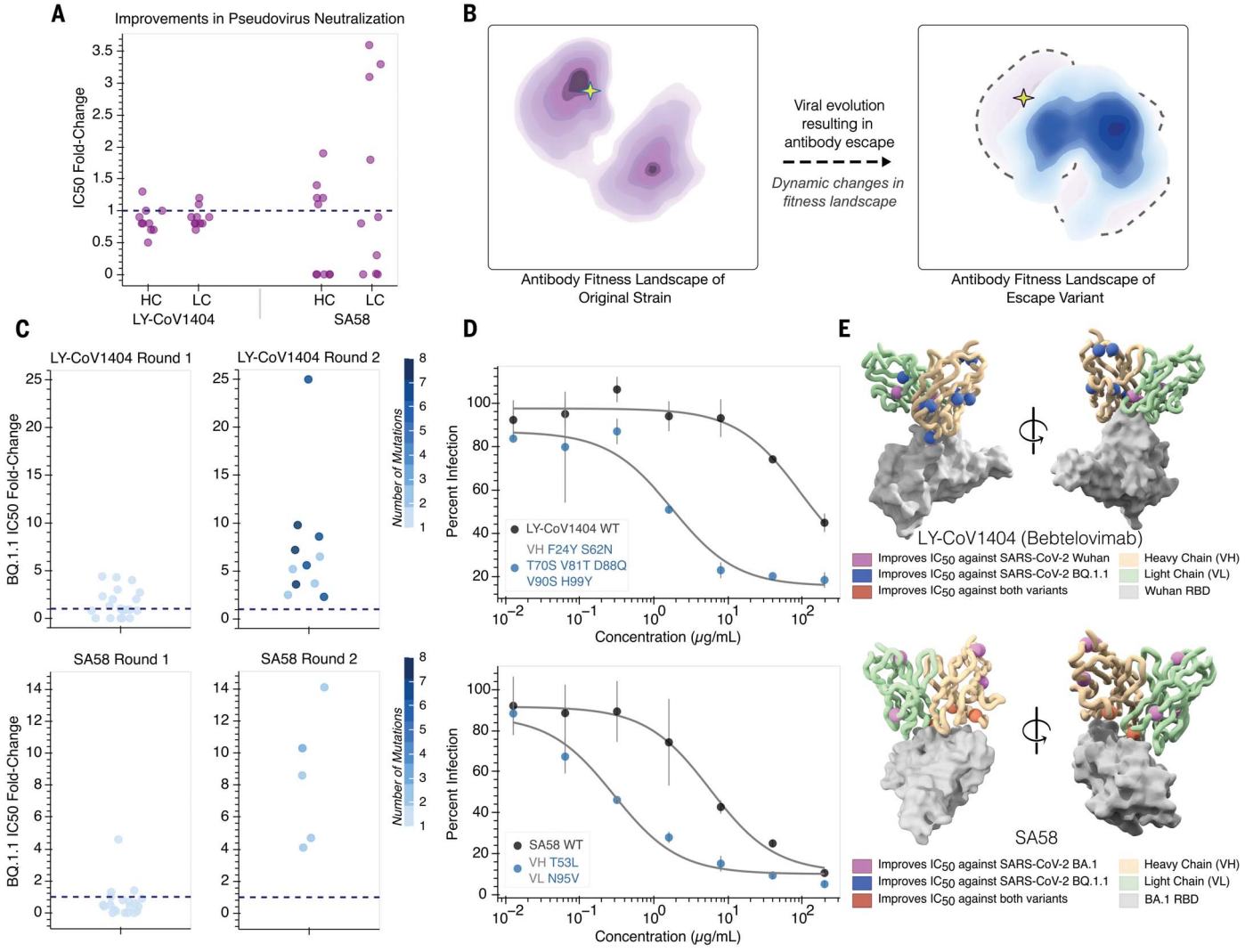

Fig. 3. Evolution of antibodies with a structure-informed language model improves neutralization potency and resilience. (**A**) Each point represents the fold change in IC_{50} of pseudovirus neutralization for antibody variants with single–amino acid mutations. Antibodies were tested against the viral strain represented in the input structure (Ly1404, Wuhan; SA58-BA.1, Omicron). A dashed line is shown at a fold change of 1, corresponding to no change. Improved antibody potency is defined as a 1.1-fold or higher improvement in IC_{50} compared with wild type. (**B**) Conceptual representation of viral evolution. Selection for immune evasion drives antibody escape, which fundamentally represents a dynamic change in the underlying fitness landscape for the antibody. This antigenic drift displaces a potent antibody from a peak on the previous fitness landscape (left) to a new starting point at lower activity (right). (**C**) Strip plots visualizing antibody evolution across two rounds. Each point shows the corresponding fold change in IC_{50} of pseudovirus neutralization for a designed variant and is colored according to the number of mutations it has (1 to 8). Consistent with preserving backbone fold, all 55 designed variants across both antibody evolutionary campaigns could be expressed. All round 1 variants are only composed of only single–amino acid changes, whereas beneficial mutations are combined in round 2. All round 2 variants have improved neutralization activity compared with their respective wild-type antibody (dotted line). (**D**) Pseudovirus neutralization curves for the most potent evolved antibody variant consisting of mutations annotated on the left. The top LY-CoV1404 variant, bearing seven amino acid substitutions in VH, achieves a 25-fold improvement in neutralization against BQ.1.1 (top). The top SA58 variant, bearing single–amino acid mutations in both VH and VL, achieves a 14-fold improvement in neutralization against BQ.1.1 (bottom). (**E**) Residues at which mutations improve neutralization against either the structure-encoded strain, BQ.1.1, or both viral strains are highlighted with spheres for antibodies LY-CoV1404 [PDB 7MMO (*50*)] and SA58 [PDB 7Y0W (*54*)]. Beneficial mutations are identified both within the binding interface as well as distal to the antigen. Neutralization enhancing mutations are labeled in fig. S10.

design against BQ.1.1 after the second round of evolution also improved BA.1 neutralization by more than threefold (data S1).

To rigorously evaluate the benefit of adding structural information, we also performed identical evolutionary campaigns for both LY-CoV1404 and SA58 using an ensemble of general sequence-only protein language models to recommend variants (*58*) (see the materials and methods), an approach that has previously been experimentally validated in applications for antibody engineering and serves as a competitive unsupervised baseline. Consistent with the computational results, we found that the structure-informed language model led to final antibody designs with substantially greater overall magnitudes in improvement (25-fold versus twofold for LY-CoV1404 and 14-fold versus fourfold for SA58) (fig. S8). Compared with our structurally informed evolution campaigns, combinations of language model–recommended beneficial mutations have limited additive effects. These results further underscore the value of selecting mutations from the outset that are known to be structurally compatible and thereby enable a more efficient ascent up the fitness landscape.

Additional characterization of evolved antibodies

To further characterize the basis for enhanced neutralization of SARS-CoV-2 VOC BQ.1.1, we tested the binding affinity of all variant antibodies to RBD as bivalent immunoglobulin G (IgG) using biolayer interferometry (BLI) to obtain the apparent dissociation constant ($K_{D,app}$). For LY-CoV1404, a total of 23 designs across both rounds of evolution exhibited improved viral neutralization, and each of these improved antibodies was confirmed to have increased apparent binding affinities of up to ~27-fold. However, we found improved apparent affinity to not be a sufficient condition for improved neutralization potency, because four additional model-recommended mutations that were neutral or deleterious to neutralization actually improved binding. Across all variants, there was a Spearman correlation of 0.45 between the fold change in IC_{50} and the fold change in $K_{D,app}$ (Fig. 4, A and B).

We similarly screened the SA58 variants for binding to the RBD of BQ.1.1. However, because the $K_{D,app}$ of the wild-type antibody as IgG was already subpicomolar, further improvements to binding were below the limit of quantitation and indistinguishable using this measure. Given this strong binding affinity of wild-type SA58 to BQ.1.1 RBD, we also screened this same set of variants against emerging VOC XBB.1.5 and observed improvements in $K_{D,app}$ of up to 37-fold (Fig. 4, C and D).

By testing several top affinity-matured designs in a polyspecificity assay, we also confirmed that improvements in binding are not mediated by generalized enhancements of nonspecific interactions (fig. S9A). In this assay, we observed no substantial changes in off-target binding of the evolved antibodies to membrane-soluble proteins, particularly within a therapeutically viable range (as defined by controls of clinically approved antibodies with recorded high and low polyspecificity). Furthermore, we found no correlation between the fold change in polyspecificity and the fold change in affinity (fig. S9B).

Analysis of evolutionary exploration

Confronted by the large number of possible mutations, traditional experiment-based methods for antibody affinity maturation often restrict the mutational search space to only a few regions of the antibody. Specifically, binding optimization efforts are typically focused within the complementarity-determining regions, which are hotspots for natural somatic hypermutation. However, using our unbiased approach to consider all regions of the variable domain allows for many discoveries that may be less intuitive to a rational designer. For example, the most beneficial substitutions to LY-CoV1404, VH F24Y, and VH V90S are located within framework regions and positioned distally from the binding interface (fig. S10 and table S3). They both improve neutralization of BQ.1.1 by more than fourfold and are not deleterious to Wuhan neutralization. In other cases, the structure-informed language model also successfully predicts beneficial substitutions using residues rarely observed among human antibody sequences. Substitution VL N95V in SA58, which improves neutralization by about fivefold against BQ.1.1, is mediated by the incorporation of a valine observed in only 0.7% of human antibody sequences at that position and enhances antibody-antigen contact. Although the model is capable of successfully making new predictions, in some instances it also suggests reverting residues to ones frequently selected for in natural somatic hypermutation. Mutation VL F51Y in LY-CoV1404 changes a phenylalanine observed in just 5%

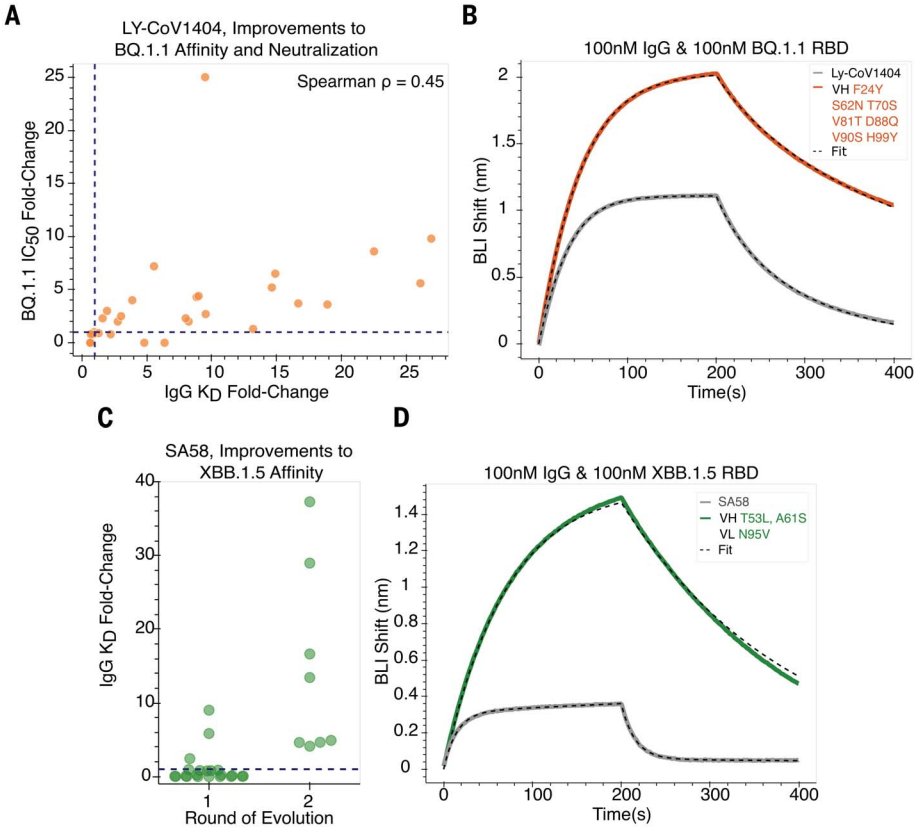

Fig. 4. Antibodies evolved for high potency also exhibit improved affinity. (**A**) LY-CoV1404 antibody variants showing a Spearman correlation of 0.45 between apparent affinity fold change and potency fold change. Improved affinity is observed to be necessary but not sufficient for improved neutralization activity. Four variants exhibit improved affinity but do not enhance neutralization. All variants with improved neutralization also display improved affinity. The top LY-CoV1404 design with a 25-fold improvement in neutralization has a 9.5-fold improvement in affinity to BQ.1.1 RBD, as measured using BLI. (**C**) SA58 antibodies evolved for improved potency against BQ.1.1 also exhibit improved affinity against VOC XBB.1.5 up to 37-fold. (**B** and **D**) Representative traces of BLI-binding assays for LY-CoV1404 and SA58 variants with improved affinity, respectively.

of sequences to a tyrosine observed in 86% of sequences. However, this variant results in no change to Wuhan neutralization. Overall, these results highlight the value in augmenting a language model with structural information to evolve antibodies and protein complexes.

Discussion

The discovery of mutations that improve protein function is inherently challenging because of the large sequence search space and complex rules that govern the relationship between sequence and function, such as stability or environmental selection pressures. We have shown that an inverse folding protein language model informed with the sequence and backbone structural coordinates of a protein can considerably improve directed evolution efforts by serving as an improved prior compared with sequence-only deep learning methods. A structure-guided approach can investigate protein fitness landscapes indirectly without needing to explicitly model individual functional tasks or properties, making it broadly applicable to proteins across diverse settings ranging from enzyme catalysis to antibiotic and chemotherapy resistance (Fig. 1D). We also demonstrate that the structure-informed language model generalizes to multimeric proteins despite being trained only on single-chain proteins through its ability to implicitly learn features of binding. This result is particularly noteworthy considering that the model has no access to amino acid side chain atoms, coordinates, or bond information.

Equipped with these capabilities, we evolved clinical therapeutic antibodies and identified several mutations that act synergistically to improve antibody potency and resilience against emerging variants of concern. In the context of pandemics and emergency-use situations, where monoclonal antibody therapies are limited in supply and vulnerable to resistance from viral evolution, the ability to rapidly make improvements in potency with a general method could have major clinical and economic implications.

Machine learning has transformed protein engineering across several design objectives. Methods that design sequences for de novo proteins with specified folds have enabled entirely new capabilities to address previously intractable problems in many settings (*49*, *59*–*62*). Here, we consider the directed evolution problem of improving a desired function of an existing protein. Compared with 14 other promising machine learning–guided protein evolution methods used to experimentally guide directed evolution campaigns on various proteins (*8*, *56*, *58*, *63*–*73*), our success rates of generating designs with functional activity better than the wild-type protein compare favorably while not requiring any assay-labeled fitness data to use for training or task-specific model supervision (fig. S11 and data S5). These results support the notion that the protein's structure itself may be used in lieu of learned surrogate functions of fitness. By eliminating the reliance on any initial data collection or prior knowledge of the protein, we show our structure-informed method has the potential to accelerate entire evolutionary campaigns.

Computational methods such as the one we propose here have the opportunity to democratize protein engineering efforts. Not only is our approach more efficient than conventional resource-intensive techniques that experimentally test the effects of all single-residue changes on biochemical functions such as binding affinity, but it also enables directed evolution based on properties that are not easily measured at scale or are incompatible with high-throughput screening. By overcoming these limitations, we anticipate that our structure-based paradigm will be useful for evolving proteins across many domains.

REFERENCES AND NOTES

1. C. Chothia, A. M. Lesk, *EMBO J.* **5**, 823–826 (1986).
2. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5869–5874 (2006).
3. D. D. Axe, N. W. Foster, A. R. Fersht, *Biochemistry* **37**, 7157–7166 (1998).
4. P. A. Romero, F. H. Arnold, *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
5. S. Shafikhani, R. A. Siegel, E. Ferrari, V. Schellenberger, *Biotechniques* **23**, 304–310 (1997).
6. H. H. Guo, J. Choe, L. A. Loeb, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9205–9210 (2004).
7. D. Rennell, S. E. Bouvier, L. W. Hardy, A. R. Poteete, *J. Mol. Biol.* **222**, 67–88 (1991).
8. P. J. Ogden, E. D. Kelsic, S. Sinai, G. M. Church, *Science* **366**, 1139–1143 (2019).
9. F. J. Poelwijk, D. J. Kiviet, D. M. Weinreich, S. J. Tans, *Nature* **445**, 383–386 (2007).
10. B. J. Wittmann, Y. Yue, F. H. Arnold, *Cell Syst.* **12**, 1026–1045.e7 (2021).
11. C. Hsu *et al*., "Learning inverse folding from millions of predicted structures" in *Proceedings of the 39th International Conference on Machine Learning* (PMLR, 2022); pp. 8946–8970; https://proceedings.mlr.press/v162/hsu22a.html.
12. J. Jumper *et al*., *Nature* **596**, 583–589 (2021).
13. Z. Lin *et al*., *Science* **379**, 1123–1130 (2023).
14. K. Yue, K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4163–4167 (1992).
15. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **282**, 1462–1467 (1998).
16. B. I. Dahiyat, S. L. Mayo, *Science* **278**, 82–87 (1997).
17. W. A. Lim, R. T. Sauer, *Nature* **339**, 31–36 (1989).
18. K. F. Lau, K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 638–642 (1990).
19. P. J. Carter, G. A. Lazar, *Nat. Rev. Drug Discov.* **17**, 197–223 (2018).
20. H. W. Schroeder Jr., L. Cavacini, *J. Allergy Clin. Immunol.* **125** (Suppl 2), S41–S52 (2010).
21. B. J. Livesey, J. A. Marsh, *Mol. Syst. Biol.* **16**, e9380 (2020).
22. J. Weile *et al*., *Mol. Syst. Biol.* **13**, 957 (2017).
23. P. Bandaru *et al*., *eLife* **6**, e27810 (2017).
24. L. Brenan *et al*., *Cell Rep.* **17**, 1171–1183 (2016).
25. K. A. Matreyek *et al*., *Nat. Genet.* **50**, 874–882 (2018).
26. P. Mishra, J. M. Flynn, T. N. Starr, D. N. A. Bolon, *Cell Rep.* **15**, 588–598 (2016).
27. B. P. Roscoe, D. N. A. Bolon, *J. Mol. Biol.* **426**, 2854–2870 (2014).
28. J. O. Kitzman, L. M. Starita, R. S. Lo, S. Fields, J. Shendure, *Nat. Methods* **12**, 203–206, 4, 206 (2015).
29. M. A. Stiffler, D. R. Hekstra, R. Ranganathan, *Cell* **160**, 882–892 (2015).
30. L. Rockah-Shmuel, Á. Tóth-Petróczy, D. S. Tawfik, *PLOS Comput. Biol.* **11**, e1004421 (2015).
31. W. M. Hewitt *et al*., *Angew. Chem. Int. Ed.* **55**, 5703–5707 (2016).
32. L. T.-Y. Cho *et al*., *Structure* **26**, 533–544.e3 (2018).
33. B. U. Klink, R. S. Goody, A. J. Scheidig, *Biophys. J.* **91**, 981–992 (2006).
34. R. A. Ward *et al*., *J. Med. Chem.* **58**, 4790–4801 (2015).
35. H. Wu *et al*., *Proteins* **67**, 198–208 (2007).
36. P. Meyer *et al*., *EMBO J.* **23**, 511–519 (2004).
37. A. M. Grishin *et al*., *Structure* **22**, 878–888 (2014).
38. M. Hong *et al*., *Structure* **16**, 1019–1026 (2008).
39. G. Minasov, X. Wang, B. K. Shoichet, *J. Am. Chem. Soc.* **124**, 5333–5340 (2002).
40. A. Didovyk, G. L. Verdine, *J. Biol. Chem.* **287**, 40099–40105 (2012).
41. J. Meier *et al*., Language models enable zero-shot prediction of the effects of mutations on protein function. bioRxiv 450648 [Preprint] (2021); https://doi.org/10.1101/2021.07.09.450648.
42. A. M. Phillips *et al*., *eLife* **10**, e71393 (2021).
43. P. Koenig *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, E486–E495 (2017).
44. C. Dreyfus *et al*., *Science* **337**, 1343–1348 (2012).
45. D. C. Ekiert *et al*., *Science* **324**, 246–251 (2009).
46. G. Fuh *et al*., *J. Biol. Chem.* **281**, 6625–6631 (2006).
47. T. H. Olsen, I. H. Moal, C. M. Deane, *Bioinform. Adv.* **2**, vbac046 (2022).
48. M. B. Swindells *et al*., *J. Mol. Biol.* **429**, 356–364 (2017).
49. J. Dauparas *et al*., *Science* **378**, 49–56 (2022).
50. K. Westendorf *et al*., *Cell Rep.* **39**, 110812 (2022).
51. E. Takashita *et al*., *N. Engl. J. Med.* **387**, 468–470 (2022).
52. US Food and Drug Administration, Center for Drug Evaluation and Research, "FDA announces bebtelovimab is not currently authorized in any US region" (FDA, 2022); https://www.fda.gov/drugs/drug-safety-and-availability/fda-announces-bebtelovimab-not-currently-authorized-any-us-region.
53. Y. Cao *et al*., *Nature* **608**, 593–602 (2022).
54. Y. Cao *et al*., *Cell Rep.* **41**, 111845 (2022).
55. R. Song *et al*., *Emerg. Microbes Infect.* **12**, 2212806 (2023).
56. E. K. Makowski *et al*., *Nat. Commun.* **13**, 3788 (2022).
57. T. N. Starr *et al*., *Science* **377**, 420–424 (2022).
58. B. L. Hie *et al*., *Nat. Biotechnol.* **42**, 275–283 (2023).
59. J. B. Ingraham *et al*., *Nature* **623**, 1070–1078 (2023).
60. J. L. Watson *et al*., *Nature* **620**, 1089–1100 (2023).
61. K. H. Sumida *et al*., *J. Am. Chem. Soc.* **146**, 2054–2061 (2024).
62. O. J. Goudy, A. Nallathambi, T. Kinjo, N. Z. Randolph, B. Kuhlman, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2307371120 (2023).
63. S. Biswas, G. Khimulya, E. C. Alley, K. M. Esvelt, G. M. Church, *Nat. Methods* **18**, 389–396 (2021).
64. C. N. Bedbrook *et al*., *Nat. Methods* **16**, 1176–1184 (2019).
65. Z. Wu, S. B. J. Kan, R. D. Lewis, B. J. Wittmann, F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8852–8858 (2019).
66. F. Cadet *et al*., *Sci. Rep.* **8**, 16757 (2018).
67. Y. Saito *et al*., *ACS Synth. Biol.* **7**, 2014–2022 (2018).
68. S. Biswas *et al*., Toward machine-guided design of proteins. bioRxiv 337154 [Preprint] (2018); https://doi.org/10.1101/337154.
69. P. A. Romero, A. Krause, F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E193–E201 (2013).
70. C. N. Bedbrook, K. K. Yang, A. J. Rice, V. Gradinaru, F. H. Arnold, *PLOS Comput. Biol.* **13**, e1005786 (2017).
71. J. Liao *et al*., *BMC Biotechnol.* **7**, 16 (2007).
72. D. Repecka *et al*., *Nat. Mach. Intell.* **3**, 324–333 (2021).
73. J.-E. Shin *et al*., *Nat. Commun.* **12**, 2403 (2021).
74. Code and scripts for: V. R. Shanker, T. U. J. Bruun, B. L. Hie, P. S. Kim, Unsupervised evolution of protein and antibody complexes with a structure-informed language model, Zenodo (2024); https://doi.org/10.5281/zenodo.11260509.
75. Raw data for: V. R. Shanker, T. U. J. Bruun, B. L. Hie, P. S. Kim, Unsupervised evolution of protein and antibody complexes with a structure-informed language model, Zenodo (2024); https://doi.org/10.5281/zenodo.11260318.

ACKNOWLEDGMENTS

We thank D. Xu, S. Kim, and the members of the Kim lab for helpful discussions on this project and D. Xu for assistance with protein graphics. **Funding:** This work was supported by the Virginia & D.K. Ludwig Fund for Cancer Research and the Chan Zuckerberg Biohub. V.R.S. acknowledges the support of the Stanford University Medical Scientist Training Program (grants T32-GM007365 and T32-GM145402). V.R.S. and T.U.J.B. are both supported by the Sarafan ChEM-H Chemistry/Biology Interface Training Program. T.U.J.B. is also supported by the Knight-Hennessy Graduate Scholarship Fund and a CIHR Doctoral Foreign Study Award (FRN:170770). B.L.H. is supported by the Stanford Science Fellows Program. **Author contributions:** Conceptualization, methodology,

interpretation: V.R.S., B.L.H., P.S.K.; Computational experiments and software development: V.R.S.; Antibody and antigen cloning, expression, and purification: V.R.S., T.U.J.B.; Lentivirus production and pseudovirus neutralization: T.U.J.B; Binding assays: V.R.S.; Writing – original draft: V.R.S with assistance from B.L.H. and P.S.K.; Writing – final draft: all authors. **Competing interests:** V.R.S., B.L.H., and P.S.K. are named as inventors on a patent application applied for by Stanford University and the Chan Zuckerberg Biohub entitled "Antibody compositions and optimization methods." B.L.H. acknowledges an outside interest in Prox Biosciences as a scientific cofounder. T.U.J.B. declares no competing interests. **Data, code, and materials availability:** Open-source code is provided that allows a user to easily evaluate the structure-informed language model on protein complexes of interest and predict beneficial amino acid substitutions at https://github.com/varun-shanker/structural-evolution. This is implemented as a simple call to a Python script with a pdb/cif file and the target chain to be mutated as the main arguments. Code and scripts (*74*) and raw data and files used in this study (*75*) have been deposited to Zenodo.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk8946
Materials and Methods
Figs. S1 to S11
Tables S1 to S3
References (*76*–*80*)
Data S1 to S6
MDAR Checklist

Submitted 18 September 2023; resubmitted 15 March 2024
Accepted 29 May 2024
10.1126/science.adk8946