Trivalent cocktail of *de novo* designed immunogens enables the robust induction and focusing of functional antibodies *in vivo*

Authors

Sesterhenn F[1,2*], Yang C[1,2*], Cramer JT[3], Bonet J[1,2], Wen X[4], Abriata LA[1,2], Kucharska I[5,6], Chiang CI[7], Wang Y[7], Castoro G[3], Vollers SS[1,2], Galloux M[8], Richard CA[8], Rosset S[1,2], Corthésy P[1,2], Georgeon S[1,2], Villard M[1,2], Descamps D[8], Delgado T[9], Rameix-Welti MA[10], Más V[9], Ervin S[11], Eléouët JF[8], Riffault S[8], Bates JT[12], Julien JP[5,6], Li Y[7], Jardetzky T[4], Krey T[3,13] & Correia BE[1,2].

***These authors contributed equally.**

1 Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland. 2 Swiss Institute of Bioinformatics (SIB), Lausanne CH-1015, Switzerland. 3 Institute of Virology, Hannover Medical School, Germany. 4 Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305, USA; 5 Program in Molecular Medicine, Hospital for Sick Children Research Institute, Toronto, ON, M5G 0A4, Canada. 6 Departments of Biochemistry and Immunology, University of Toronto, Toronto, ON M5S 1A8, Canada. 7 Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD 20850, USA. 8 Unité de Virologie et Immunologie Moléculaires (UR892), INRA, Université Paris-Saclay, 78352, Jouy-en-Josas, France. 9 Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain. 10 UMR1173, INSERM, Université de Versailles St. Quentin, 78180 Montigny le Bretonneux, France. 11 Wake Forest Baptist Medical Center, Winston Salem NC 27157, USA. 12 University of Mississippi Medical Center, Mississippi 39216, USA. 13 German Center for Infection Research (DZIF), Hannover, Germany.

Abstract

De novo protein design has been increasingly successful in expanding beyond nature's sequence and structural space. However, most *de novo* designed proteins lack biological function, in part due to the structural complexity required for functional purposes. An important domain where protein design has raised expectations was on the induction of precise antibody responses that may lead to improved vaccines. Here, we showcase two computational design approaches to stabilize irregular and discontinuous binding motifs in *de novo* designed immunogens and tested them for the induction of respiratory syncytial virus neutralizing antibodies *in vivo*. The designs mimic the native conformations of the neutralization epitopes with sub-angstrom accuracy. *In vivo*, cocktail formulations of the immunogens induce robust neutralizing serum responses targeting three epitopes, and re-focus pre-existing antibody responses towards *bona fide* neutralization epitopes. Our work provides a blueprint for epitope-centric vaccine design for pathogens that have frustrated traditional vaccine development efforts, and a general methodological pipeline to create novel proteins with functional sites within tailored protein topologies.

Introduction

Efforts to design novel proteins from first principles have revealed a variety of rules to control the structural features of *de novo* proteins (*1-4*). However, the *de novo* design of functional proteins has been far more challenging (*5*). A commonly used strategy to design *de novo* functional proteins is to transplant the binding motifs found in existing protein structures to pre-existing or *de novo* templates. In nearly all cases, the binding motifs transplanted were commonly found in existing protein structures, such as linear helical segments, allowing the grafting of such motifs without extensive backbone adjustments (*6-8*). Most protein functional sites, however, are not contained within a single, regular segment in protein structures but arise from the three-dimensional arrangement of several, often irregular, structural elements that are supported by defined topological features of the overall structure (*9-11*). As such, it is of utmost importance for the field to develop computational approaches to endow *de novo* designed proteins with irregular and multi-segment complex structural motifs that can perform the desired functions.

Functional protein design has raised expectations in the domain of immune response modulation; in particular, on the induction of neutralizing antibodies (nAbs) *in vivo* (*12*). Inducing nAbs targeting defined epitopes remains an overarching challenge for vaccine development (*13*). Our increasing structural understanding of many nAb-antigen interactions has provided templates for the rational design of immunogens for respiratory syncytial virus (RSV), influenza, HIV, dengue and others (*14-16*). Despite this extensive structural knowledge, these and other pathogens are still lacking efficacious vaccines, highlighting the need for next-generation vaccines to efficiently guide antibody responses towards key neutralization epitopes in both naïve and pre-exposed immune systems. The elicitation of antibody-responses with defined epitope specificities has been an enduring challenge for immunogens derived from modified viral proteins (*17*).

Recently, Correia and colleagues (*12*) have shown that computationally designed immunogens could elicit epitope-specific responses. The RSVF antigenic site II, a linear helix-turn-helix motif, was transplanted onto a heterologous protein scaffold, which elicited nAbs in non-human-primates (NHPs) after repeated boosting immunizations. Despite being a proof-of-principle for the induction of functional antibodies using a computationally designed immunogen, several major caveats emerged; namely, the lack of applicability of the computational approach to structurally complex epitopes, and the inconsistent neutralization titers observed in the immunogenicity studies.

To address these limitations, here we designed epitope-focused immunogens mimicking irregular and discontinuous RSV neutralization epitopes (site 0 (*18*) and IV (*19*), Fig 1) and showcase two computational design methodologies that enable the presentation of these structurally challenging motifs in *de novo* designed proteins. *In vivo*, cocktail formulations including an optimized site II immunogen (*20*) yielded consistent neutralization levels above the protective threshold directed against all three epitopes. The design strategies presented provide a blueprint to engineer proteins stabilizing irregular and discontinuous binding sites, applicable to vaccine design for pathogens that require fine control over the antibody specificities induced, and more generally for the design of *de novo* proteins displaying complex functional motifs.

Results

De novo design of immunogens with structurally complex epitopes

Designing proteins with structurally complex functional sites has remained a largely unmet challenge in the field of computational protein design (*5*). We sought to design accurate mimetics of RSV neutralization epitopes, which have been structurally well characterized, and evaluate their functionality in immunization studies. We chose antigenic sites 0 and IV (Fig 1), which are both targeted by potent nAbs, and are structurally distinct from functional motifs that have previously been handled by computational protein design algorithms. The antigenic site 0 presents a structurally complex and discontinuous epitope consisting of a kinked 17-residue alpha helix and a disordered loop of 7 residues, targeted by nAbs D25 and 5C4 (*18, 21*), while site IV presents an irregular 6-residue bulged beta-strand and is targeted by nAb 101F (*19*).

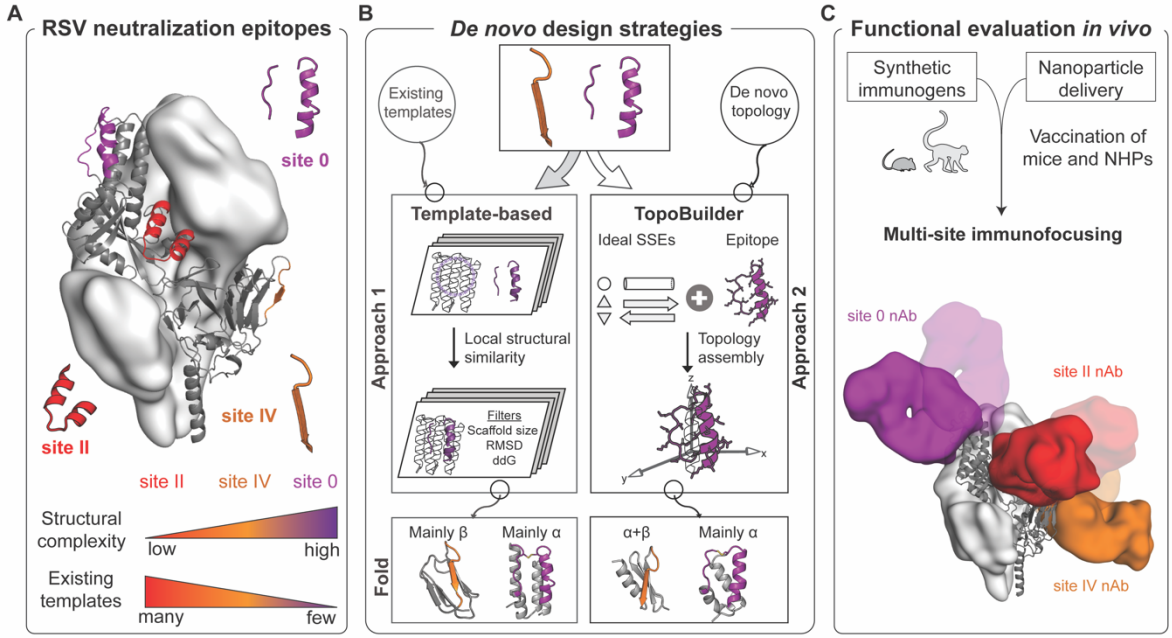


Fig 1
Conceptual overview of the computational design of immunogens to elicit RSV neutralizing antibodies focused on three distal epitopes. (A) Prefusion RSVF structure (PDB 4JHW) with sites 0, II and IV highlighted. An immunogen for site II was previously reported (*12*). (B) Computational protein design strategies. Approach 1: Design templates were identified in the PDB based on loose structural similarity to site 0/IV, followed by *in silico* folding and design, and sequence optimization through directed evolution. Approach 2: A motif-centric *de novo* design approach was developed ("TopoBuilder") to tailor the protein topology to the motif's structural constraints. Bottom: Computational models of designed immunogens using different approaches. (C) Cocktail formulations of three synthetic immunogen nanoparticles to elicit nAbs focused on three non-overlapping epitopes.

The computational design of proteins mimicking structural motifs has previously been performed by first identifying compatible protein scaffolds, either from naturally occurring structures or built *de novo*, which then serve as design templates to graft the motif (*6, 7, 22-24*). Given the structural complexity of sites 0 and IV, this approach did not provide any promising matches, even using loose structural criteria (Fig S1).

Thus, for site IV, we noticed that a small structural domain that resembles an immunoglobulin fold containing the epitope could be excised from the prefusion RSVF (preRSVF) structure. We hypothesized this would be a conservative approach to maintain its native, distorted epitope structure (Fig S2). We optimized the sequence for stability and epitope mimicry using Rosetta FunFolDes (*25*) (Fig 2a), and our best computational design (S4_1.1) bound with a K_D > 85 μM to the 101F target antibody. To improve binding affinity, we performed deep mutational scanning followed by next-generation sequencing, as previously described (*26*)

(Fig S2). We tested combinations of enriched positions in recombinantly expressed proteins for antibody binding, obtaining a double mutant (S4_1.5) that bound with a K_D of 35 nM to the 101F target antibody, showed a circular dichroism (CD) spectrum of a folded protein, and was thermostable up to 65 °C (Fig 2b-d and Fig S3).

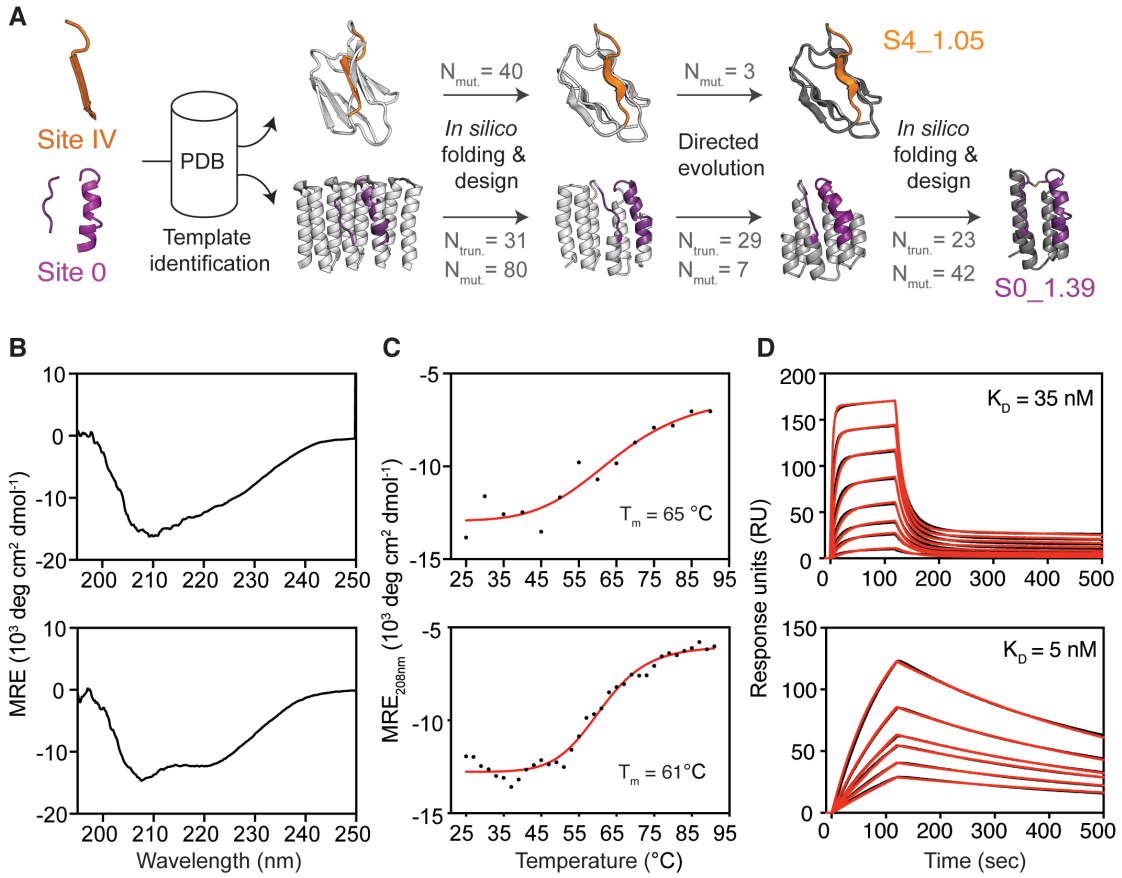


Fig 2
Templated computational design and biophysical characterization of synthetic immunogens. (A) Protein design strategy - templates with structural similarity to sites IV and 0 were identified by native domain excision or loose structural matching, followed by *in silico* folding, design and directed evolution. An additional *in silico* folding and design step was necessary to install site 0 on a truncated template sequence revealed by directed evolution. Computational models of intermediates and final designs (S4_1.5 and S0_1.39) are shown, and the number of mutations (N_{mut}) and truncated residues (N_{trun}) are indicated for each step. (B) CD spectra measured at 20 °C of S4_1.5 (top) and S0_1.39 (bottom), are in agreement with the expected secondary structure content of the design model. (C) Thermal melting curves measured by CD at 208 nm in presence of 5 mM TCEP reducing agent. (D) Binding affinity measured by SPR against target antibodies 101F (top) and D25 (bottom). Sensorgrams are shown in black and fits in red lines. CD - Circular dichroism, T_m - melting temperature, SPR - Surface plasmon resonance.

The discontinuous structure of site 0 was not amenable for domain excision and stabilization. Thus, we searched for template structures that mimicked the helical segment of the epitope,

and simultaneously allowed grafting the loop segment, ultimately selecting a designed helical repeat protein as template (PDB 5cwj) (Fig 2a and Fig S4) (*27*). In order to avoid steric clashes with the target antibody D25, we truncated the N-terminal 31 residues of the 5cwj template, and performed *in silico* folding and design simulations to sample local and global changes on the scaffold to allow the presentation of the site 0 epitope (Fig 2a). Out of 9 sequences tested, the best design (S0_1.1) bound with a K_D of 1.4 μM to the D25 antibody (Fig S5), which is four orders of magnitude lower than the affinity of D25 to preRSVF (*28*). Following multiple rounds of directed evolution using yeast display, we found an enriched sequence that was C-terminally truncated by 29 residues, and showed a ~5-fold increased affinity towards D25 (Fig S4-S5). We used the truncated structure as a new template for *in silico* folding and design. Ultimately, this multi-stage process yielded S0_1.39, a design further truncated by additional 23 residues, which bound with 5 nM to D25 (Fig 2d). S0_1.39 was also recognized by the 5C4 antibody (Fig S5), which has been shown to engage site 0 in a different orientation compared to D25 (*21*), with an affinity of 5 nM, identical to that of the 5C4-preRSVF interaction (*28*).

The primary goals for the designs were achieved in terms of the stabilization of irregular and complex binding motifs in a conformation relevant for antibody binding, however, the overall strategy presented important limitations with respect to its general utility. Despite the large number of structures available to serve as design templates, the fraction of those that are practically useful for the design of functional proteins becomes increasingly limited with the structural complexity of the motif (Fig S1). As described above, suboptimal design templates require extensive backbone flexibility on the design process and multiple rounds of directed evolution until a sequence with high-affinity binding is identified. Additionally, the starting topology determines the overall shape of the designed protein, which may be suboptimal for the accurate stabilization of the motif, and may impose unwanted tertiary steric constraints that interfere with the designed function. In particular, for immunogen design it is desired to preserve native-like accessibility of the epitope in the context of the designed immunogen, thereby maximizing the induction of antibodies that can cross-react with the native antigen presented by the pathogen. An illustrative example on how a template-based design approach can fail to fulfil these criteria is the comparison between the quaternary environment of the site 0 epitope in preRSVF and S0_1.39 showing that this topology does not mimic such environment, albeit allowing the binding of several monoclonal antibodies (Fig S6).

To overcome these limitations, we developed a template-free design protocol - the TopoBuilder - that generates tailor-made topologies to stabilize complex functional motifs. Within the TopoBuilder, we parametrically sample the placement of idealized secondary structure elements which are then connected by loop segments, to assemble topologies that

can stabilize the desired conformation of the structural motif. Next, these topologies are diversified to enhance structural and sequence diversity with a folding and design stage using Rosetta FunFolDes (see Fig S7 and methods for details). For this approach, we defined two new design objectives which were unmet by our previous template-based designs: (1) building stable *de novo* topologies that stabilize the epitope, while mimicking their native quaternary environment; (2) fine-tuning the topology's secondary structure arrangements to maximize the fold stability and optimize epitope presentation for high affinity antibody binding.

To present antigenic site IV, we designed a fold composed of a ß-sheet with 4 antiparallel strands and one helix (Fig 3a), referred to as S4_2 fold. Within the S4_2 topology, we generated three structural variants (S4_2_bb1-3), by parametrically sampling three distinct helical secondary structural elements, varying both orientations and lengths to maximize the packing interaction against the ß-sheet. Sequences generated from two structural variants (S4_2_bb2 and S4_2_bb3) showed a strong propensity to recover the designed structures in Rosetta *abinitio* simulations (Fig 3a and Fig S8).

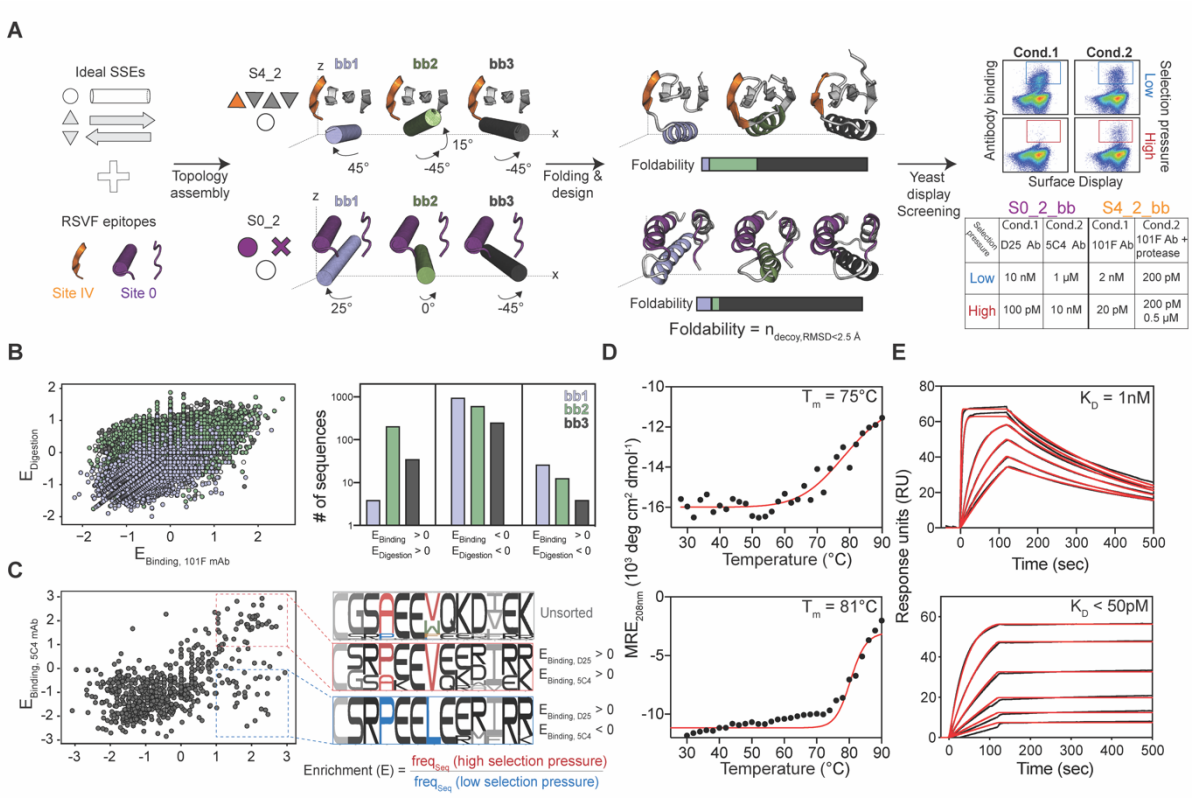


Fig 3
Motif-centric *de novo* design of epitope-focused immunogens. (A) Ideal secondary structure elements (SSEs) are assembled around RSVF epitopes, sampling different orientations within the same topology, followed by a single round of *in silico* folding and design. Rosetta *abinitio* simulations are performed for designs of each topology to assess its propensity to fold into the designed structures, returning a foldability score. Selected designs are then displayed on yeast surface and sorted under two different selection pressures for

subsequent deep sequencing. (B) All three designed topological variants were screened for high affinity binding and resistance to chymotrypsin to select stably folded proteins. Enrichment analysis revealed a strong preference for one of the designed helix orientations (S4_2_bb2, green) to resist protease digestion and bind with high affinity to 101F. (C) Enrichment analysis of sorted populations under high and low selective pressures. Sequences highly enriched for both D25 and 5C4 binding show convergent sequence features in critical core positions of the site 0 scaffold. (D) Thermal melting curves measured by CD for best designs (S4_2.45 (top) and S0_2.126 (bottom)) showing high thermostability. (E) Dissociation constants (K_D) of S4_2.45 to 101F (top) and S0_2.126 to D25 (bottom) antibodies measured by SPR.

We screened a defined set of computationally designed sequences using yeast display and applied two selective pressures – binding to 101F and resistance to the nonspecific protease chymotrypsin, an effective method to digest partially unfolded proteins (*6, 29, 30*). Deep sequencing of populations sorted under different conditions revealed that S4_2_bb2-based designs were strongly enriched under stringent selection conditions for folding and 101F binding, showing that subtle topological differences in the design template can have substantial impact on function and stability. We expressed 15 S4_2_bb2 design variants and successfully purified and biochemically characterized 14. The designs showed mixed alpha/beta CD spectra and bound to 101F with affinities ranging from 1 nM to 200 nM (Fig S9). The best variant, S4_2.45 (K_D = 1 nM), was well folded and thermostable according to CD and NMR with a T_m of 75 °C (Fig 3d and Fig S10).

Similarly, we built a minimal *de novo* topology to present the tertiary structure of the site 0 epitope. The choice for this topology was motivated by the fact that site 0, in its native preRSVF environment, is accessible for antibody binding from diverse angles (*21*), in contrast to the S0_39 natural template which topologically constrained site 0 accessibility (Fig S6). By building a template *de novo,* we attempted to respect site 0's native quaternary constraints, while stabilizing both irregular epitope segments with high accuracy.

We explored the topological space within the shape constraints of preRSVF and built three different helical orientations (S0_2_bb1-3) that support both epitope segments. Evaluation of the designed sequences with Rosetta *abinitio* showed that only sequences generated based on one topology (S0_2_bb3) presented a funnel-shaped energy landscape (Fig S11). A set of computationally designed sequences based on S0_2_bb3 was screened in yeast under the selective pressure of two site 0-specific antibodies (D25 and 5C4) to ensure the presentation of the native epitope's conformation. Deep sequencing of the double-enriched clones and subsequent sequence analysis revealed that a valine at position 28 is critical to retain a cavity formed between the two epitope segments, ensuring binding to both antibodies (Fig 3b). We selected five sequences, differing from 3 to 21 mutations, for further biochemical

characterization (Fig S12). The design with best solution behaviour (S0_02.126) showed a CD spectrum of a mostly helical protein, with extremely high thermostability even under reducing conditions (T_m = 81 °C, Fig 3d) and a well-dispersed HSQC NMR spectrum (Fig S10). Strikingly, S0_2.126 bound with ~50 pM affinity to D25, similar to that of the preRSVF-D25 interaction (~150 pM), and with a K_D = 4 nM to 5C4 (Fig 3e and Fig S13).

Overall, the properties of the designs generated by topological assembly with the TopoBuilder showed improved binding affinities and thermal stabilities as compared to those using available structural templates. To investigate whether this design and screening procedure yielded scaffolds that better mimicked the viral epitope presented, or rather revealed sequences with a highly optimized interface towards the antibodies used during the selection, we determined the affinities of S4_2.45 and S0_2.126 against a panel of human site-specific antibodies. Compared to the first-generation designs, S4_2.45 and S0_2.126 showed large affinity improvements across the antibody panels, exhibiting a geometric mean affinity closely resembling that of the antibodies to preRSVF (Fig S13). In the light of such results, we concluded that the topologically designed immunogens were superior mimetics of sites IV and 0 relative to the template-based designs.

De novo designed topologies adopt the predicted structures with high accuracy

To evaluate the structural accuracy of the computational design approach, we solved the crystal structure of S4_2.45 in complex with 101F at 2.6 Å resolution. The structure closely matched our design model, with a full-atom RMSD of 1.5 Å (Fig 4a). The epitope was mimicked with an RMSD of 0.135 Å, and retained all essential interactions with 101F (Fig 4d,e). Importantly, the structural data confirmed that we presented an irregular beta strand, a common motif found in many protein-protein interactions (*31*), in a fully *de novo* designed protein with sub-angstrom accuracy.

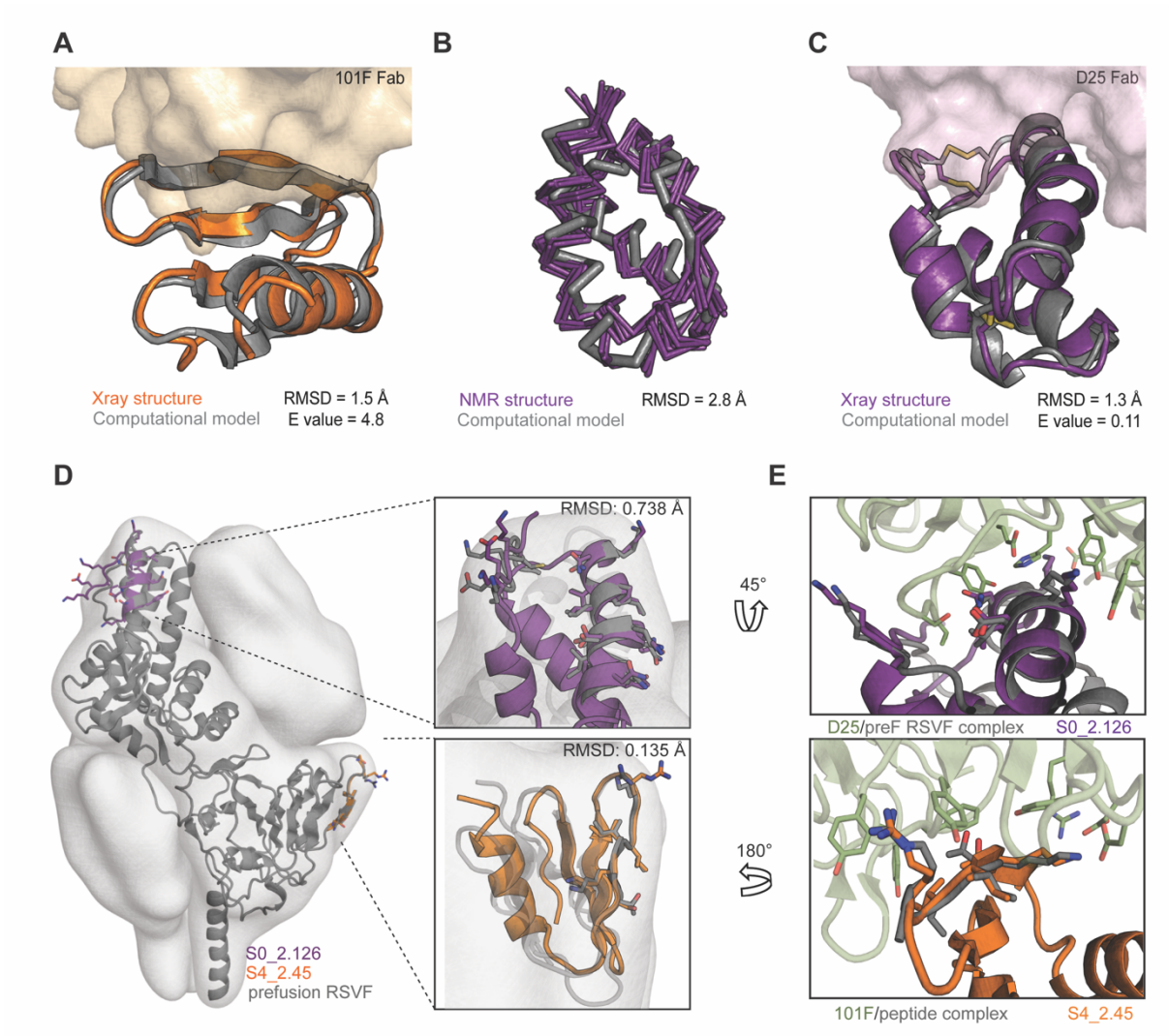


Fig 4
Structural characterization of *de novo* designed immunogens. (A) Crystal structure of S4_2.45 (orange) bound to 101F Fab closely matches the design model (grey, RMSD = 1.5 Å). (B) NMR structural ensemble of S0_2.126 (purple) superimposed to the computational model (grey). The NMR structure is in agreement with the design model (backbone RMSD of 2.8 Å). (C) Crystal structure of S0_2.126 (purple) bound to D25 Fab closely resembles the design model (grey, RMSD = 1.3 Å). (D) Superposition of the preRSVF sites 0/IV and designed immunogens shows sub-angstrom mimicry of the epitopes. Designed scaffolds are compatible with the shape constraints of preRSVF (surface representation). (E) Close-up view of the interfacial side-chain interactions between D25 (top) and 101F (bottom) with designed immunogens as compared to the starting epitope structures.

Next, we solved an unbound structure of S0_2.126 by NMR, confirming the accuracy of the designed fold with a backbone RMSD between the average structure and the model of 2.8 Å (Fig 4b). Additionally, we solved a crystal structure of S0_2.126 bound to D25 at a resolution of 3.0 Å. The structure showed an overall RMSD of 1.5 Å to the design model, and an RMSD of 0.9 Å over the discontinuous epitope compared to preRSVF (Fig 4c-e). To the best of our knowledge, this is the first computationally *de novo* designed protein that presents a two-segment, structurally irregular, binding motif with atomic-level accuracy. In comparison with

native proteins, S0_2.126 showed exceptionally low packing due to a large core cavity (Fig S14), but retained a very high thermal stability. The core cavity was essential for antibody binding and highlights the potential of *de novo* approaches to design small proteins hosting structurally challenging motifs and preserving cavities required for function (*2*). Notably, due to the level of control and precision of the TopoBuilder, both designed antigens respected the shape constraints of the respective epitope in their native environment within preRSVF, a structural feature that may be important for the improved elicitation of functional antibodies (Fig S6).

Cocktails of designed immunogens elicit neutralizing antibodies *in vivo*

Lastly, we sought to evaluate the designed antigens for their ability to elicit focused nAb responses *in vivo*. Our rationale for combining site 0, II and IV immunogens in a cocktail formulation is that all three sites are non-overlapping, as verified by electron microscopy analysis (Fig S15), and thus could induce a more potent antibody response *in vivo*. To increase immunogenicity, each immunogen was multimerized on self-assembling protein nanoparticles. We chose the RSV nucleoprotein (RSVN), a self-assembling ring-like structure of 10-11 subunits, previously shown to be an effective carrier for the site II immunogen (*20*), and formulated a trivalent immunogen cocktail containing equimolar amounts of S0_1.39, S4_1.05 and S2_1.2 immunogen nanoparticles ("Trivax1", Fig S16). The fusion of S0_2.126 and S4_2.45 to RSVN yielded poorly soluble nanoparticles, prompting us to use ferritin particles for multimerization, with a 50% occupancy (~12 copies), creating a second cocktail comprising S2_1.2 in RSVN and the remaining immunogens in ferritin ("Trivax2", Fig S17).

In mice, Trivax1 elicited low levels of RSVF cross-reactive antibodies, and sera did not show RSV neutralizing activity in most animals (Fig S18). In contrast, Trivax2 induced robust levels of RSVF cross-reactive serum levels, and the response was balanced against all three epitopes (Fig 5a,b). Strikingly, Trivax2 immunization yielded RSV neutralizing activity above the protective threshold in 6/10 mice (Fig 5c). These results show that vaccine candidates composed of *de novo* designed proteins mimicking viral neutralization epitopes can induce robust antibody responses *in vivo*, targeting multiple specificities. This is an important finding given that mice have been a traditionally difficult model to induce neutralizing antibodies with scaffold-based immunogens (*12, 23*).

In parallel, we sought to test the potential of a trivalent immunogen cocktail in NHPs. The previously designed site II immunogen showed promise in NHPs, but the induced neutralizing titers were low and inconsistent across animals, requiring up to five immunizations to elicit neutralizing antibodies in 2/4 animals (*12*). We immunized seven RSV-naïve NHPs with

Trivax1, as detailed in Fig 5d. In contrast to mice, NHPs developed robust levels of RSVF cross-reactive serum titer in all animals (Fig 5e), and antibodies induced were directed against all three epitopes (Fig 5f). Strikingly, we found that 6/7 NHPs showed RSV neutralizing serum levels above the protective threshold after a single boosting immunization (median IC_{50} = 312) (Fig 5g). Neutralization titers were maximal at day 84 (median IC_{50} = 408), four-fold above the protective threshold (*28*), and measurements were confirmed by an independent laboratory (Fig S19).

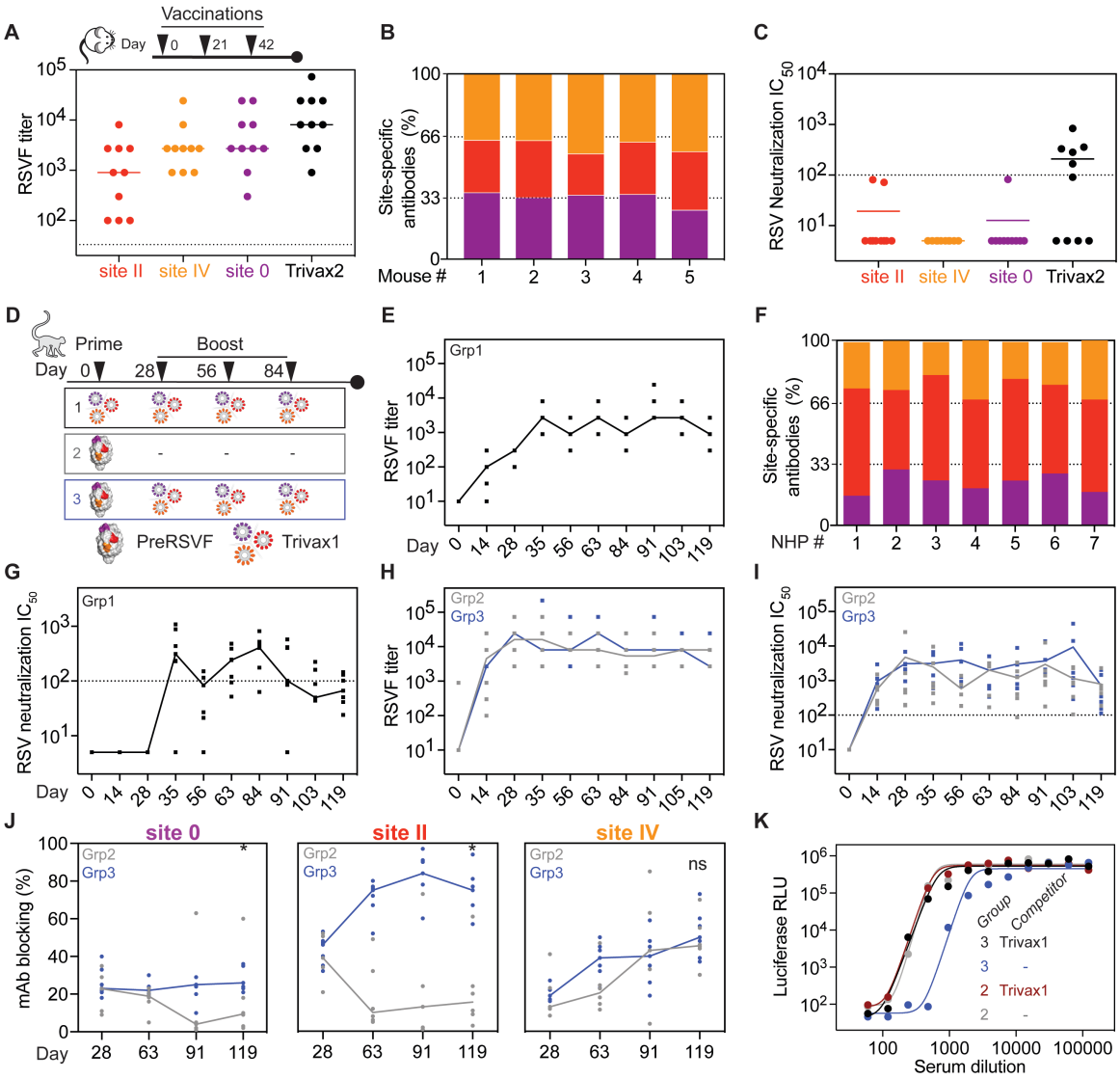


Fig 5

Synthetic immunogens elicit neutralizing serum responses in mice and NHPs, and focus pre-existing immunity on sites 0 and II. (A-C) Trivax2 immunization study in mice. (A) PreRSVF cross-reactive serum levels following three immunizations with single immunogens or Trivax2 cocktail (day 56). (B) Serum specificity shown for 5 representative mice immunized with Trivax2, as measured by an SPR competition assay with D25, Motavizumab and 101F IgGs as competitors, exhibiting an equally balanced response towards all sites. (C) RSV neutralization titer of mice at day 56, immunized with Trivax2 components individually and as cocktail. Dotted line (IC_{50} = 100) indicates protective threshold as defined

by protective level of Palivizumab. (D-K) Trivax1 immunization study in NHPs. (D) NHP immunization scheme. (E) PreRSVF cross-reactive serum levels for group 1. (F) Serum antibodies target all three antigenic sites in all 7 animals as measured by an SPR competition assay. (G) RSV neutralization titers of group 1. (H) PreRSVF titer in group 2 (grey) and 3 (blue). (I) RSV neutralization titer of group 2 and 3. (J) Site-specific antibody levels measured by SPR competition assay. Site 0 and site II-specific titers were significantly higher in group 3 compared to 2 following Trivax1 boosting ($p < 0.05$, Mann-Whitney U test). (K) RSV neutralization curves upon depletion of day 91 sera with site 0, II, IV-specific scaffolds. 60% of the neutralizing activity is competed in group 3, whereas no significant decrease is observed in the control group 2.

Immunization studies in naïve animals are essential to test the capability of the designed immunogens to induce functional antibodies. However, an overarching challenge for vaccine development to target pathogens such as RSV, influenza, dengue and others is to focus or reshape pre-existing immunity of broad specificity on defined neutralizing epitopes that can confer long-lasting protection (*20, 32, 33*). To mimic a serum response of broad specificity towards RSV, we immunized 13 NHPs with preRSVF. All animals developed strong preRSVF-specific titers and cross-reactivity with all the epitope-focused immunogens, indicating that epitope-specific antibodies were primed and recognize the designed immunogens (Fig S20). Group 2 (6 animals) subsequently served as control group to follow the dynamics of epitope-specific antibodies over time, and group 3 (7 animals) was boosted three times with Trivax1 (Fig 5d). PreRSVF-specific antibody and neutralization titers maximized at day 28 and were maintained up to day 119 in both groups (Fig 5h,i). Analysis of the site-specific antibody levels showed that site 0, II and IV responses were dynamic in the control group, with site II dropping from 37% to 13% and site 0 from 17% to 4% at day 28 and 91, respectively (Fig 5j). In contrast, site IV specific responses increased from 13% to 43% over the same time span. Although Trivax1 boosting immunizations did not significantly change the magnitude of the preRSVF-specific serum response, they reshaped the serum specificities in primed animals. Site II specific titers were 6.5-fold higher (day 91) compared to the non-boosted control group (84% vs 13%, $p = 0.02$, Mann-Whitney), and unlike the rapid drop of site 0-specific antibodies in the non-boosted group, these antibodies were maintained upon Trivax1 boosting (25% vs 4%, $p=0.02$, Mann-Whitney) (Fig 5j). In contrast, site IV specific responses increased to similar levels in both groups, 43% and 40% in group 2 and 3, respectively. Strikingly, upon depletion of site 0, II and IV specific antibodies from pooled sera, we observed a 60% drop in neutralizing activity in group 3 as compared to only a 7% drop in the non-boosted control group, indicating that Trivax1 boosting reshaped a serum response of broad specificity towards a more focused response that predominantly relies on site 0, II and IV-specific antibodies for RSV neutralization (Fig 5k).

Altogether, the design strategies utilized, yielded antigens presenting structurally complex neutralization epitopes that induce neutralizing antibodies upon cocktail formulation, providing a strong rationale for including multiple, ideally non-overlapping epitopes in an epitope-focused vaccination strategy. While the first-generation immunogens were inferior according to biophysical parameters and failed to induce neutralization in mice, they were successful under two different immunological scenarios in NHPs, and we show that a second generation can now induce neutralizing antibodies in mice. This is an important step to optimize and test different nanoparticles, formulations and delivery routes in a small animal model, and we foresee that the second-generation immunogens will prove superior in inducing neutralizing serum responses in NHPs.

Discussion & Conclusions

Our work showcased two computational protein design strategies to design immunogens which present structurally complex epitopes with atomic accuracy, and validated their functionality to elicit nAb responses in cocktail formulations both in mice and NHPs. We have shown that through computational design of pre-existing templates with full backbone flexibility, irregular and discontinuous epitopes were successfully stabilized in heterologous scaffolds. However, this design strategy required extensive *in vitro* evolution optimization and the resulting scaffolds remained suboptimal regarding their biochemical and biophysical properties. Moreover, the lack of precise topological control of the designed proteins is a major limitation for the design of functional proteins that require defined topological similarity in addition to local mimicry of the transplanted site. For instance, the design template of the site 0 immunogen did not mimic the quaternary environment of the epitope of interest, which may have contributed to the low levels of functional antibodies induced in mice.

To overcome these limitations, we developed the TopoBuilder, a motif-centric design approach that tailors a protein fold directly to the functional site of interest. Compared to previously employed functional *de novo* design protocols, in which a stable scaffold topology was constructed first and endowed with binding motifs in a second step (*6*), our method has significant advantages for structurally complex motifs. First, it tailors the topology to the structural requirements of the functional motif from the start of the design process, rather than through the adaptation (and often destabilization) of a stable protein to accommodate the functional site. Second, the topological assembly and fine-tuning allowed to select for optimal backbone orientations and sequences that stably folded and bound with high affinity in a single screening round, without further optimization through directed evolution, as often necessary in computational protein design efforts (*6, 7, 34, 35*). Together, our approach enabled the computational design of *de novo* proteins presenting irregular and discontinuous structural

motifs that are typically required to endow proteins with diverse biochemical functions (e.g. binding or catalysis), thus providing a new means for the *de novo* design of functional proteins.

As to the functional aspect of our design work, we showed *in vivo* that these immunogens consistently elicited neutralizing serum levels in mice and NHPs as cocktail formulations. The elicitation of focused neutralizing antibody responses by vaccination remains the central goal for vaccines against pathogens that have frustrated conventional vaccine development efforts (*13*). Using RSV as a model system, we have shown that cocktails of computationally designed antigens can robustly elicit neutralizing serum levels in naïve animals. These neutralization levels were much superior to any previous report on epitope-focused immunogens (*12*) and provide a strong rationale for an epitope-focused vaccination strategy involving multiple, non-overlapping epitopes. Also, their capability to dramatically reshape the nature of non-naïve repertoires in NHPs addresses an important challenge for many next-generation vaccines to target pathogens for which efficacious vaccines are needed (*20, 33, 36*). An important pathogen from this category is influenza, where the challenge is to overcome immunodominance hierarchies (*37*), which have been established during repeated natural infections, and that favour strain-specific antibody specificities, rather than cross-protecting nAbs found in the hemagglutinin stem region (*38*). The ability to selectively boost subdominant nAbs targeting defined, broadly protective epitopes that are surrounded by strain-specific epitopes could overcome long-standing challenges in vaccine development, given that cross-neutralizing antibodies can persist for years once elicited (*39*). A tantalizing future application for epitope-focused immunogens could marry this technology with engineered components of the immune system, which could be used to stimulate antibody production of adoptively transferred, engineered B-cells that express monoclonal therapeutic antibodies *in vivo* (*40*).

Altogether, this study provides a blueprint for the design of epitope-focused vaccines against pathogens that have eluded traditional vaccine development approaches. Beyond immunogens, our approach to design *de novo* proteins presenting complex binding sites will be broadly applicable to engineer novel functional proteins with defined structural properties.

Contributions: FS, CY and BEC conceived the work and designed the experiments. FS and CY performed computational design and experimental characterization. JTC, GC, TK, XW and TJ solved x-ray structures. JB developed the TopoBuilder protocol. LAA performed NMR characterization and solved NMR structure. IK and JPJ performed and analysed samples by electron microscopy. CIC, YW, SSV, MG, SR, PC, SG, MV and MAR performed experiments and analysed data. JTB contributed to the design and planning of animal studies. FS, CY and BEC wrote the manuscript, with input from all authors.

Data availability: All code used for this study is available through a public github repository: https://github.com/lpdi-epfl/trivalent_cocktail. It contains the TopoBuilder source code, RosettaScripts used for the design, analysis scripts and detailed information on how designs were selected. Structures have been deposited in the Protein Data Bank under accession codes 6S28 (solution NMR structure of S0_2.126).

Funding: This work was supported by the Swiss initiative for systems biology (SystemsX.ch), the European Research Council (Starting grant - 716058), Swiss National Science Foundation (Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung; 310030_163139) and the EPFL's Catalyze4Life initiative. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program (J.P.J.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements: We thank James Crowe for providing RSVF site IV antibodies used in this study. We thank Aline E. Christine and Florence Pojer in PTPSP facility at EPFL for crystallography support. We thank Andrew McCarthy at ESRF for beam line support, as well as Heather Hotchin, Amy Beierschmitt, and Roberta Palmour from the Behavioural Science Foundation for the NHP immunization and PBMC isolation. We thank the EPFL phenogenomics center (Céline Waldvogel, Raphaël Doenlen) for help with the animal experiments and the protein expression core facility (David Hacker, Laurence Durrer, Soraya Quinche) for help with mammalian protein expression. We thank Davide Demurtas (CIME) and Sergey Nazarov PTBIOEM (EPFL, Lausanne, Switzerland) for electron microscopy support. We thank the Flow Cytometry facility (FCCF) at the EPFL for technical support. These computational simulations were facilitated by SCITAS at EPFL.

References

1. N. Koga *et al.*, Principles for designing ideal protein structures. *Nature* **491**, 222-227 (2012).
2. E. Marcos *et al.*, Principles for designing proteins with cavities formed by curved beta sheets. *Science* **355**, 201-206 (2017).
3. P. S. Huang *et al.*, De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat Chem Biol* **12**, 29-34 (2016).
4. M. Mravic *et al.*, Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science* **363**, 1418-1423 (2019).
5. P. S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature* **537**, 320-327 (2016).
6. A. Chevalier *et al.*, Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74-79 (2017).
7. E. Procko *et al.*, A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* **157**, 1644-1656 (2014).
8. S. Berger *et al.*, Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. *Elife* **5**, (2016).
9. S. Jones, J. M. Thornton, Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93**, 13-20 (1996).
10. N. D. Rubinstein *et al.*, Computational characterization of B-cell epitopes. *Mol Immunol* **45**, 3477-3489 (2008).
11. G. L. Holliday, J. D. Fischer, J. B. Mitchell, J. M. Thornton, Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. *FEBS J* **278**, 3835-3845 (2011).
12. B. E. Correia *et al.*, Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201-206 (2014).
13. R. Rappuoli, M. J. Bottomley, U. D'Oro, O. Finco, E. De Gregorio, Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *J Exp Med* **213**, 469-481 (2016).
14. J. S. McLellan, Neutralizing epitopes on the respiratory syncytial virus fusion glycoprotein. *Curr Opin Virol* **11**, 70-75 (2015).
15. N. S. Laursen, I. A. Wilson, Broadly neutralizing antibodies against influenza viruses. *Antiviral Res* **98**, 476-483 (2013).
16. D. Sok, D. R. Burton, Recent progress in broadly neutralizing antibodies to HIV. *Nat Immunol* **19**, 1179-1188 (2018).
17. F. Sesterhenn, J. Bonet, B. E. Correia, Structure-based immunogen design-leading the way to the new age of precision vaccines. *Curr Opin Struct Biol* **51**, 163-169 (2018).
18. J. S. McLellan *et al.*, Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113-1117 (2013).
19. J. S. McLellan *et al.*, Structure of a major antigenic site on the respiratory syncytial virus fusion glycoprotein in complex with neutralizing antibody 101F. *J Virol* **84**, 12236-12244 (2010).
20. F. Sesterhenn *et al.*, Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen. *PLoS Biol* **17**, e3000164 (2019).
21. D. Tian *et al.*, Structural basis of respiratory syncytial virus subtype-dependent neutralization by an antibody targeting the fusion glycoprotein. *Nat Commun* **8**, 1877 (2017).

22. M. L. Azoitei *et al.*, Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science* **334**, 373-376 (2011).
23. J. S. McLellan *et al.*, Design and characterization of epitope-scaffold immunogens that present the motavizumab epitope from respiratory syncytial virus. *J Mol Biol* **409**, 853-866 (2011).
24. S. J. Fleishman *et al.*, Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816-821 (2011).
25. J. Bonet *et al.*, Rosetta FunFolDes - A general framework for the computational design of functional proteins. *PLoS Comput Biol* **14**, e1006623 (2018).
26. T. A. Whitehead *et al.*, Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* **30**, 543-548 (2012).
27. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580-584 (2015).
28. J. S. McLellan *et al.*, Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* **342**, 592-598 (2013).
29. P. Kristensen, G. Winter, Proteolytic selection for protein folding using filamentous bacteriophages. *Fold Des* **3**, 321-328 (1998).
30. M. D. Finucane, M. Tuna, J. H. Lees, D. N. Woolfson, Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry* **38**, 11604-11612 (1999).
31. A. M. Watkins, P. S. Arora, Anatomy of beta-strands at protein-protein interfaces. *ACS Chem Biol* **9**, 1747-1754 (2014).
32. S. F. Andrews *et al.*, Immune history profoundly affects broadly protective B cell responses to influenza. *Sci Transl Med* **7**, 316ra192 (2015).
33. G. Barba-Spaeth *et al.*, Structural basis of potent Zika-dengue virus antibody cross-neutralization. *Nature* **536**, 48-53 (2016).
34. D. A. Silva *et al.*, De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186-191 (2019).
35. E. M. Strauch *et al.*, Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site. *Nat Biotechnol* **35**, 667-671 (2017).
36. S. F. Andrews *et al.*, High preexisting serological antibody levels correlate with diversification of the influenza vaccine response. *J Virol* **89**, 3308-3317 (2015).
37. D. Angeletti *et al.*, Defining B cell immunodominance to viruses. *Nat Immunol* **18**, 456-463 (2017).
38. D. Corti *et al.*, A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* **333**, 850-856 (2011).
39. J. Lee *et al.*, Persistent Antibody Clonotypes Dominate the Serum Response to Influenza over Multiple Years and Repeated Vaccinations. *Cell Host Microbe* **25**, 367-376 e365 (2019).
40. H. F. Moffett *et al.*, B cells engineered to express pathogen-specific antibodies protect against infection. *Sci Immunol* **4**, (2019).

Computational design of template-based epitope-focused immunogens

Site 0

The structural segments entailing the antigenic site 0 were extracted from the prefusion stabilized RSVF Ds-Cav1 crystal structure, bound to the antibody D25 (PDB ID: 4JHW) (*1*). The epitope consists of two segments: a kinked helical segment (residues 196-212) and a 7-residue loop (residues 63-69).

The MASTER software (*2*) was used to perform structural searches over the Protein Data Bank (PDB, from August 2018), containing 141,920 protein structures, to select template scaffolds with local structural similarities to the site 0 motif. A first search with a Cα RMSD threshold below 2.5 Å did not produce any usable structural matches both in terms of local mimicry as well as global topology features. A second search was performed, where extra structural elements that support the epitope in its native environment were included as part of the query motif to bias the search towards matches that favoured motif-compatible topologies rather than those with close local similarities. The extra structural elements included were the two buried helices that directly contact the site 0 in the preRSVF structure (4JHW residues 70-88 and 212-229). The search yielded initially 7,600 matches under 5 Å of backbone RMSD, which were subsequently filtered for proteins with a length between 50 and 160 residues, high secondary structure content, as well as for accessibility of the epitope for antibody binding. Remaining matches were manually inspected to select template-scaffolds suitable to present the native conformation of antigenic site 0. Subsequently, we selected a computationally designed, highly stable, helical repeat protein (*3*) consisting of 8 regular helices (PDB ID: 5CWJ) with an RMSD of 4.4 Å to the query (2.82 Å for site 0 segments only). To avoid steric clashes with the D25 antibody, we truncated the 5CWJ template structure at the N-terminus by 31 residues, resulting in a structural topology composed of 7 helices.

Using Rosetta FunFolDes (*4*) the truncated 5CWJ topology was folded and designed to stabilized the grafted site 0 epitope recognized by D25. We generated 25,000 designs and selected the top 300 by Rosetta energy score (RE), designed backbones that presented obvious flaws, such as low packing scores, distorted secondary structural elements and buried unsatisfied atoms were discarded. From the top 300 designs, 3 were retained for follow-up iterative cycles of structural relaxation and design using Rosetta FastDesign (*5*), generating a total of 100 designed sequences.

The best 9 designs by Rosetta energy score were recombinantly expressed in E. coli. 2 designed sequences derived from the same backbone, were successfully expressed and purified. The best variant was named S0_1.1, and subjected to experimental optimization using yeast surface display (Fig S4-S5). In one of the libraries, we found a truncated sequence (S0_1.17) enriched for expression and binding, which served as template for a second round of computational design (Fig S4-S5). We performed 25,000 folding and design simulations using Rosetta FunFolDes (*4*). The best 300 decoys by total Rosetta energy score were extracted, and relaxed using the Rosetta Relax application (*6*). We computed the mean total RE, and selected designs that showed a lower energy score than the mean of the design population (RE = -155.2), RMSD drift of the epitope after relaxing of less than 0.7 Å, and a cavity volume <60 $Å^3$. We selected one of the best 5 scoring decoys, truncated the C-terminal 29 and N-terminal 23 residues which did not contribute to epitope stabilization, and introduced a disulfide bond between residue 1 and 43. Four sequences were experimentally tested (S0_1.37-40). The best variant according to binding, S0_1.39, bound with 5 nM affinity to antibody D25, and, importantly, also gained binding to the 5C4 antibody (K_D = 5 nM).

Site IV

When the design simulations were carried out, there was no structure available of the full RSVF protein in complex with a site IV-specific nAb, nevertheless a peptide epitope of this site recognized by the 101F nAb had been previously reported (PDB ID: 3O41) (*7*).

The crystallized peptide-epitope corresponds to the residues 429-434 of the RSVF protein. Structurally the 101F-bound peptide-epitope adopts a bulged strand and several studies suggest that 101F recognition extends beyond the linear β-strand, contacting other residues located in antigenic site IV (*8*). Despite the apparent structural simplicity of the epitope, structural searches for designable scaffolds failed to yield promising starting templates. However, we noticed that the antigenic site IV of RSVF is self-contained within an individual domain that could potentially be excised and designed as a soluble folded protein. To maximize these contacts, we first truncated the seemingly self-contained region from RSVF pre-fusion structure (PDB

ID: 4JHW, residue: 402-459) forming a β-sandwich and containing site IV. We used Rosetta FastDesign to optimize the core positions of this minimal topology, obtaining our initial design: S4_wt. However, S4_wt did not show a funnel-shaped energy landscape in Rosetta *ab initio* simulations, and we were unable to obtain expression in *E.coli*.

In an attempt to improve the conformation and stabilization of S4_wt, we used Rosetta FunFolDes to fold and design this topology, while keeping the conformation of the site IV epitope fixed. Out of 25,000 simulations, the top 1 % decoys according to RE score and overall RMSD were selected for manual inspection, and 12 designed sequences were selected for recombinant expression in *E.coli*.

The Rosetta scripts and analysis scripts to perform the designs are available in the GitHub repository accompanying this paper, together with detailed instructions on how to run the code.

TopoBuilder - Motif-centric *de novo* design

Given the limited availability of suitable starting templates to host structurally complex motifs such as site 0 and site IV, we developed a template-free design protocol, which we named TopoBuilder (see Fig S7). In contrast to adapting an existing topology to accommodate the epitope, the design goal is to build protein scaffolds around the epitope from scratch, using idealized secondary structures (beta strands and alpha helices). The length, orientation and 3D-positioning are defined by the user for each secondary structure with respect to the epitope, which is extracted from its native environment. The topologies built were designed to meet the following criteria: (1) Small, globular proteins with a high contact order between secondary structures and the epitope, to allow for stable folding and accurate stabilization of the epitope in its native conformation; (2) Context mimicry, i.e. respecting shape constraints of the epitope in its native context (Fig S6). For assembling the topology, the default distances between alpha helices was set to 11 Å and for adjacent beta-strands was 5 Å. For each discontinuous structural sketch, a connectivity between the secondary structural elements was defined and loop lengths were selected to connect the secondary structure elements with the minimal number of residues that can cover a given distance, while maintaining proper backbone geometries.

For site 0, the short helix of S0_1.39 preceding the epitope loop segment was kept, and a third helix was placed on the backside of the epitope to: (1) provide a core to the protein and (2) allow for the proper connectivity between the secondary structures. A total of three different orientations (45°, 0° and -45° degrees to the plane formed by site 0) where tested for the designed supporting alpha helix (Fig. 3).

In the case of site IV, the known binding region to 101F (residues 428F-434F) was extracted from prefusion RSVF (PDB 4JWH). Three antiparallel beta strands, pairing with the epitope, plus an alpha helix on the buried side, were assembled around the 101F epitope. Three different configurations (45°, (-45°,0°,10°) and -45° degrees with respect to the β-sheet) were sampled parametrically for the alpha helix (Fig. 3 and Fig S8).

The structural sketches were used to generate Cα distance constraints to guide Rosetta FunFolDes (*4*) folding trajectories. Around 25,000 trajectories were generated for each sketch. The newly generated backbones were further subjected to layer-based FastDesign (*5*), meaning that each amino acid position was assigned a layer (combining 'core', 'boundary', 'surface' and 'sheet' or 'helix') on the basis of its exposure and secondary structure type, that dictated the allowed amino acid types at that position.

After iterative cycles of sequence design, unconstraint FastRelax (*9*) (i.e sidechain repacking and backbone minimization) was applied over the designs to evaluate their conformational stability of the epitope region. After each relax cycle, structural changes of the epitope region were evaluated (epitope RMSD drift). Designs with epitope RMSD drifts higher than 1.2 Å were discarded. Designs were also ranked and selected according hydrophobic core packing (packstat score), with a cutoff of 0.5 for site 0 and 0.6 for the site IV design series, and a cavity volume of < 50 $Å^3$. Between 1,000 and 10,000 of the designed sequences were generated from this computational protocol. We evaluated sequence profiles for the designs, and encoded the critical positions combinatorially by assembling overlapping oligos. Upon PCR assembly, libraries were transformed in yeast and screened for antibody binding and stability as assessed by protease digestion assays (*10-12*).

The Rosetta scripts and analysis scripts to perform the designs are available in the GitHub repository accompanying this paper, together with detailed instructions on how to run the code.

Mouse immunizations

All animal experiments were approved by the Veterinary Authority of the Canton of Vaud (Switzerland) according to Swiss regulations of animal welfare (animal protocol number 3074). Female Balb/c mice (6-week old) were purchased from Janvier labs. Immunogens were thawed on ice, mixed with equal volumes of adjuvant (2% Alhydrogel, Invivogen or Sigma Adjuvant System, Sigma) and incubated for 30 minutes. Mice were injected subcutaneously with 100 μl vaccine formulation, containing in total 10 μg of immunogen (equimolar ratios of each immunogen for Trivax immunizations). Immunizations were performed on day 0, 21 and 42. 100-200 μl blood were drawn on day 0, 14 and 35. Mice were euthanized at day 56 and blood was taken by cardiac puncture.

NHP immunizations

Twenty-one african green monkeys (AGM, 3-4 years) were divided into three experimental groups with at least two animals of each sex. AGMs were pre-screened as seronegative against prefusion RSVF (preRSVF) by ELISA. Vaccines were prepared 1 hour before injection, containing 50 µg preRSVF or 300 µg Trivax1 in 0.5 ml PBS, mixed with 0.5 ml alum adjuvant (Alhydrogel, Invivogen) for each animal. AGMs were immunized intramuscularly at day 0, 28, 56, and 84. Blood was drawn at days 14, 28, 35, 56, 63, 84, 91, 105 and 119.

RSV neutralization assay

The RSV neutralization assay was performed as described previously (*13*). Briefly, Hep2 cells were seeded in Corning 96-well tissue culture plates (Sigma) at a density of 40,000 cells/well in 100 μl of Minimum Essential Medium (MEM, Gibco) supplemented with 10% FBS (Gibco), L-glutamine 2 mM (Gibco) and penicillin-streptomycin (Gibco), and grown overnight at 37 °C with 5% CO2. Serial dilutions of heat-inactivated sera were prepared in MEM without phenol red (M0, Life Technologies, supplemented with 2 mM L-glutamine and penicillin/streptomycin) and were incubated with 800 pfu/well (final MOI 0.01) RSV-Luc (A2 strain carrying a luciferase gene) for 1 hour at 37 °C. Serum-virus mixture was added to Hep2 cell layer, and incubated for 48 hours. Cells were lysed in lysis buffer supplemented with 1 µg/ml

luciferin (Sigma) and 2 mM ATP (Sigma), and luminescence signal was read on a Tecan Infinite 500 plate reader. The neutralization curve was plotted and fitted using the GraphPad variable slope fitting model, weighted by $1/Y^2$.

Serum fractionation

Monomeric Trivax1 immunogens (S2_1, S0_1.39 and S4_1.5) were used to deplete the site 0, II and IV specific antibodies in immunized sera. HisPurTM Ni-NTA resin slurry (Thermo Scientific) was washed with PBS containing 10 mM imidazole. Approximately 1 mg of each immunogen was immobilized on Ni-NTA resin, followed by two wash steps to remove unbound scaffold. 60 μl of sera pooled from all animals within the same group were diluted to a final volume of 600 μl in wash buffer, and incubated overnight at 4 °C with 500 μl Ni-NTA resin slurry. As control, the same amount of sera was incubated with Ni-NTA resin that did not contain scaffolds. Resin was pelleted down at 13,000 rpm for 5 minutes, and the supernatant (depleted sera) was collected and then used for neutralization assays.

Site saturation mutagenesis library (SSM)

A SSM library was assembled by overhang PCR, in which 11 selected positions surrounding the epitope in the S4_1.1 design model were allowed to mutate to all 20 amino acids, with one mutation allowed at a time. Each of the 11 libraries was assembled by primers (Table S1) containing the degenerate codon 'NNK' at the selected position. All 11 libraries were pooled, and transformed into EBY-100 yeast strain with a transformation efficiency of $1x10^6$ transformants.

Combinatorial library

Combinatorial sequence libraries were constructed by assembling multiple overlapping primers (Table S2) containing degenerate codons at selected positions for combinatorial sampling of hydrophobic amino acids in the protein core. The theoretical diversity was between $1x10^6$ and $5x10^6$. Primers were mixed (10 μM each), and assembled in a PCR reaction (55 °C annealing for 30 sec, 72 °C extension time for 1 min, 25 cycles). To amplify full-length assembled products, a second PCR reaction was performed, with forward and reverse primers specific for the full-length

product. The PCR product was desalted, and transformed into EBY-100 yeast strain with a transformation efficiency of at least $1x10^7$ transformants (*14*).

Protein expression and purification

Designed scaffolds

All genes of designed proteins were purchased as DNA fragments from Twist Bioscience, and cloned via Gibson assembly into either pET11b or pET21b bacterial expression vectors. Plasmids were transformed into *E.coli* BL21 (DE3) (Merck) and grown overnight in LB media. For protein expression, precultures were diluted 1:100 and grown at 37 °C until the OD_{600} reached 0.6, followed by the addition of 1 mM IPTG to induce expression. Cultures were harvested after 12-16 hours at 22 °C. Pellets were resuspended in lysis buffer (50 mM Tris, pH 7.5, 500 mM NaCl, 5% Glycerol, 1 mg/ml lysozyme, 1 mM PMSF, 1 µg/ml DNase) and sonicated on ice for a total of 12 minutes, in intervals of 15 seconds sonication followed by 45 seconds pause. Lysates were clarified by centrifugation (20,000 rpm, 20 minutes) and purified via Ni-NTA affinity chromatography followed by size exclusion on a HiLoad 16/600 Superdex 75 column (GE Healthcare) in PBS buffer.

Antibodies - IgG and Fab constructs

Plasmids encoding cDNAs for the heavy chain of IgG were purchased from Genscript and cloned into the pFUSE-CHIg-hG1 vector (Invivogen). Heavy and light chain DNA sequences of antibody fragments (Fab) were purchased from Twist Bioscience and cloned separately into the pHLsec mammalian expression vector (Addgene, #99845) via Gibson assembly. HEK293-F cells were transfected in a 1:1 ratio with heavy and light chains, and cultured in FreeStyle medium (Gibco) for 7 days. Supernatants were collected by centrifugation and purified using a 1 ml HiTrap Protein A HP column (GE Healthcare) for IgG expression and 5 ml kappa-select column (GE Healthcare) for Fab purification. Bound antibodies/Fabs were eluted with 0.1 M glycine buffer (pH 2.7), immediately neutralized by 1 M Tris ethylamine buffer (pH 9), and buffer exchanged to PBS.

Prefusion stabilized RSVF

The construct encoding the thermostabilized the preRSVF protein corresponds to the sc9-10 DS-Cav1 A149C Y458C S46G E92D S215P K465Q variant (referred to as DS2) reported by Joyce and colleagues (*15*). The sequence was codon-optimized for mammalian cell expression and cloned into the pHCMV-1 vector flanked with two C-terminal Strep-Tag II and one 8x His tag. Expression and purification were performed as described previously (*13*).

Nanoring-based immunogens

The full-length N gene derived from the human RSV strain Long, ATCC VR-26 (GenBank accession number AY911262.1) was PCR amplified and cloned into pET28a+ at NcoI-XhoI sites to obtain the pET-N plasmid. Immunogens presenting sites 0, II and IV epitopes were cloned into the pET-N plasmid at NcoI site as pET-S0_1.39-N, pET-S2_1.2-N and pET-S4_1.5-N, respectively. Expression and purification of the nanoring fusion proteins was performed as described previously (*13*).

Ferritin-based immunogens

The gene encoding Helicobacter pylori ferritin (GenBank ID: QAB33511.1) was cloned into the pHLsec vector for mammalian expression, with an N-terminal 6x His Tag. The sequence of the designed immunogens (S0_2.126 and S4_2.45) were cloned upstream of the ferritin gene, spaced by a GGGGS linker. Ferritin particulate immunogens were produced by co-transfecting a 1:1 stochiometric ratio of “naked” ferritin and immunogen-ferritin in HEK-293F cells, as previously described for other immunogen-nanoparticle fusion constructs (*16*). The supernatant was collected 7-days post transfection and purified via Ni-NTA affinity chromatography and size exclusion on a Superose 6 increase 10/300 GL column (GE).

Negative-stain transmission electron microscopy

Sample preparation

RSVN and Ferritin- based nanoparticles were diluted to a concentration of 0.015 mg/ml. The samples were absorbed on carbon-coated copper grid (EMS, Hatfield, PA, United States) for 3 mins, washed with deionized water and stained with freshly prepared 0.75 % uranyl formate.

Data acquisition

The samples were viewed under an F20 electron microscope (Thermo Fisher) operated at 200 kV. Digital images were collected using a direct detector camera Falcon III (Thermo Fisher) with the set-up of 4098 X 4098 pixels. The homogeneity and coverage of staining samples on the grid was first visualized at low magnification mode before automatic data collection. Automatic data collection was performed using EPU software (Thermo Fisher) at a nominal magnification of 50,000X, corresponding to pixel size of 2 Å, and defocus range from -1 µm to -2 µm.

Image processing

CTFFIND4 program (*17*) was used to estimate the contrast transfer function for each collected image. Around 1000 particles were manually selected using the installed package XMIPP within SCIPION framework (*18*). Manually picked particles were served as input for XMIPP auto-picking utility, resulting in at least 10,000 particles. Selected particles were extracted with the box size of 100 pixels and subjected for three rounds of reference-free 2D classification without CTF correction using RELION-3.0 Beta suite (*19*).

RSVF-Fabs complex formation and negative stain EM

20 µg of RSVF trimer was incubated overnight at 4°C with 80 µg of Fabs (Motavizumab, D25 or 101F). For complex formation with all three monoclonal Fabs, 80 µg of each Fab was used. Complexes were purified on a Superose 6 Increase 10/300 column using an Äkta Pure system (GE Healthcare) in TBS buffer. The main fraction containing the complex was directly used for negative stain EM.

Purified complexes of RSVF and Fabs were deposited at approximately 0.02 mg/ml onto carbon-coated copper grids and stained with 2% uranyl formate. Images were collected with a field-emission FEI Tecnai F20 electron microscope operating at 200 kV. Images were acquired with an Orius charge-coupled device (CCD) camera (Gatan Inc.) at a calibrated magnification of ×34,483, resulting in a pixel size of 2.71 Å. For the complexes of RSVF with a single Fab, approximately 2,000 particles were manually selected with Cryosparc2 (*20*). Two rounds of 2D classification of particle images were performed with 20 classes allowed. For the complexes of RSVF with D25, Motavizumab and 101F Fabs, approximately 330,000 particles were picked using Relion 3.0 (*19*) and subsequently imported to Cryosparc2 for two rounds of 2D classification with 50 classes allowed.

Determining binding affinities by Surface plasmon resonance (SPR)

SPR measurements were performed on a Biacore 8K (GE Healthcare) with HBS-EP+ as running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20, GE Healthcare). Ligands were immobilized on a CM5 chip (GE Healthcare # 29104988) via amine coupling. Approximately 2000 response units (RU) of IgG were immobilized, and designed monomeric proteins were injected as analyte in two-fold serial dilutions. The flow rate was 30 µl/min for a contact time of 120 seconds followed by 400 seconds dissociation time. After each injection, surface was regenerated using 3 M magnesium chloride (101F as immobilized ligand) or 0.1 M Glycine at pH 4.0 (Motavizumab and D25 IgG as an immobilized ligand). Data were fitted using 1:1 Langmuir binding model within the Biacore 8K analysis software (GE Healthcare #29310604).

Dissection of serum antibody specificities by SPR

To quantify the epitope-specific antibody responses in bulk serum from immunized animals, we performed an SPR competition assay with the monoclonal antibodies (D25, Motavizumab and 101F) as described previously (*13*). Briefly, approximately 400 RU of prefusion RSVF were immobilized on a CM5 chip via amine coupling, and serum diluted 1:10 in running buffer was injected to measure the total response ($RU_{non\text{-}blocked\ surface}$). After chip regeneration using 50 mM NaOH, the site 0/II/IV epitopes were blocked by injecting saturating amounts of either D25, Motavizumab, or 101F IgG, and serum was injected again to quantify residual response ($RU_{blocked\ surface}$). The delta serum response (ΔSR) was calculated as follows:

$$\Delta \mathrm{SR} = \mathrm{RU}_{\mathrm{(non-)blocked\ surface}} - \mathrm{RU}_{\mathrm{Baseline}}$$

Percent blocking was calculated for each site as:

$$\%\ \text{blocking} = \left(1 - \left(\frac{\Delta \mathrm{SR}_{\mathrm{blocked\ surface}}}{\Delta \mathrm{SR}_{\mathrm{non-blocked\ surface}}}\right)\right) * 100$$

SEC-MALS

Size exclusion chromatography with an online multi-angle light scattering (MALS) device (miniDAWN TREOS, Wyatt) was used to determine the oligomeric state and

molecular weight for the protein in solution. Purified proteins were concentrated to 1 mg/ml in PBS (pH 7.4), and 100 µl of sample was injected into a Superdex 75 300/10 GL column (GE Healthcare) with a flow rate of 0.5 ml/min, and UV_{280} and light scattering signals were recorded. Molecular weight was determined using the ASTRA software (version 6.1, Wyatt).

Circular Dichroism

Far-UV circular dichroism spectra were measured using a Jasco-815 spectrometer in a 1 mm path-length cuvette. The protein samples were prepared in 10 mM sodium phosphate buffer at a protein concentration of 30 µM. Wavelengths between 190 nm and 250 nm were recorded with a scanning speed of 20 nm min^{-1} and a response time of 0.125 sec. All spectra were averaged 2 times and corrected for buffer absorption. Temperature ramping melts were performed from 25 to 90 °C with an increment of 2 °C/min in presence or absence of 2.5 mM TCEP reducing agent. Thermal denaturation curves were plotted by the change of ellipticity at the global curve minimum to calculate the melting temperature (T_m).

Yeast surface display

Libraries of linear DNA fragments encoding variants of the designed proteins were transformed together with linearized pCTcon2 vector (Addgene #41843) based on the protocol previously described by Chao and colleagues (*14*). Transformation procedures generally yielded ~10^7 transformants. The transformed cells were passaged twice in SDCAA medium before induction. To induce cell surface expression, cells were centrifuged at 7,000 r.p.m. for 1 min, washed with induction media (SGCAA) and resuspended in 100 ml SGCAA with a cell density of 1 x 10^7 cells/ml SGCAA. Cells were grown overnight at 30 °C in SGCAA medium. Induced cells were washed in cold wash buffer (PBS + 0.05% BSA) and labelled with various concentration of target IgG or Fab (101F, D25, and 5C4) at 4°C. After one hour of incubation, cells were washed twice with wash buffer and then incubated with FITC-conjugated anti-cMyc antibody and PE-conjugated anti-human Fc (BioLegend, #342303) or PE-conjugated anti-Fab (Thermo Scientific, #MA1-10377) for an additional 30 min. Cells were washed and sorted using a SONY SH800 flow cytometer in 'ultra-purity' mode. The sorted cells were recovered in SDCAA medium, and grown for 1-2 days at 30 °C.

In order to select stably folded proteins, we washed the induced cells with TBS buffer (20 mM Tris, 100 mM NaCl, pH 8.0) three times and resuspended in 0.5 ml of TBS buffer containing 1 μM of chymotrypsin. After incubating five-minutes at 30°C, the reaction was quenched by adding 1 ml of wash buffer, followed by five wash steps. Cells were then labelled with primary and secondary antibodies as described above.

ELISA

96-well plates (Nunc MediSorp plates; Thermo Scientific) were coated overnight at 4°C with 50 ng/well of purified antigen (recombinant RSVF or designed immunogen) in coating buffer (100 mM sodium bicarbonate, pH 9) in 100 μl total volume. Following overnight incubation, wells were blocked with blocking buffer (PBS + 0.05% Tween 20 (PBST) containing 5% skim milk (Sigma)) for 2 hours at room temperature. Plates were washed five times with PBST. 3-fold serial dilutions were prepared and added to the plates in duplicates, and incubated at room temperature for 1 hour. After washing, anti-mouse (abcam, #99617) or anti-monkey (abcam, #112767) HRP-conjugated secondary antibody were diluted 1:1,500 or 1:10,000, respectively, in blocking buffer and incubated for 1 hour. An additional five wash steps were performed before adding 100 μl/well Pierce TMB substrate (Thermo Scientific). The reaction was terminated by adding an equal volume of 2 M sulfuric acid. The absorbance at 450 nm was measured on a Tecan Safire 2 plate reader, and the antigen specific titers were determined as the reciprocal of the serum dilution yielding a signal two-fold above the background.

NMR

Protein samples for NMR were prepared in 10 mM sodium phosphate buffer, 50 mM sodium chloride at pH 7.4 with the protein concentration of 500 μM. All NMR experiments were carried out in a 18.8T (800 MHz proton Larmor frequency) Bruker spectrometer equipped with a CPTC ^{1}H,^{13}C,^{15}N 5 mm cryoprobe and an Avance III console. Experiments for backbone resonance assignment consisted in standard triple resonance spectra HNCA, HN(CO)CA, HNCO, HN(CO)CA, CBCA(CO)NH and HNCACB acquired on a 0.5 mM sample doubly labelled with ^{13}C and ^{15}N (*21*). Sidechain assignments were obtained from HCCH-TOCSY experiments acquired on the same sample plus HNHA, NOESY-^{15}N-HSQC and TOCSY-^{15}N-HSQC acquired on a ^{15}N-labeled sample. The NOESY-^{15}N-HSQC was used together with a 2D

NOESY collected on an unlabelled sample for structure calculations. Spectra for backbone assignments were acquired with 40 increments in the ^{15}N dimension and 128 increments in the ^{13}C dimension, and processed with 128 and 256 points by using linear prediction. HCCH-TOCSY were recorded with 64-128 increments in the ^{13}C dimensions and processed with twice the number of points. ^{15}N-resolved NOESY and TOCSY spectra were acquired with 64 increments in ^{15}N dimension and 128 in the indirect ^{1}H dimension, and processed with twice the number of points. ^{1}H-^{1}H 2D-NOESY and 2D TOCSY spectra were acquired with 256 increments in the indirect dimension, processed with 512 points. Mixing times for NOESY spectra were 100 ms and TOCSY spin locks were 60 ms. Heteronuclear ^{1}H-^{15}N NOE was measured with 128 ^{15}N increments processed with 256 points, using 64 scans and a saturation time of 6 seconds. All samples were prepared in 20 mM phosphate buffer pH 7, with 10% $^{2}H_2O$ and 0.2% sodium azide to prevent sample degradation.

All spectra were acquired and processed with Bruker's TopSpin 3.0 (acquisition with standard pulse programs) and analyzed manually with the program CARA (http://cara.nmr.ch/doku.php/home) to obtain backbone and sidechain resonance assignments. Peak picking and assignment of NOESY spectra (a ^{15}N-resolved NOESY and a 2D NOESY) were performed automatically with the program UNIO-ATNOS/CANDID (*22, 23*) coupled to Cyana 2.1 (*24*), using standard settings in both programs. The run was complemented with dihedral angles derived from chemical shifts with Talos-n (*25*).

X-ray crystallization and structural determination

Co-crystallization of complex D25 Fab with S0_2.126

After overnight incubation at 4°C, the S0_2.126/D25 Fab complex was purified by size exclusion chromatography using a Superdex200 26 600 (GE Healthcare) equilibrated in 10 mM Tris pH 8, 100 mM NaCl and subsequently concentrated to ~10 mg/ml (Amicon Ultra-15, MWCO 3,000). Crystals were grown at 291K using the sitting-drop vapor-diffusion method in drops containing 1 µl purified protein mixed with 1 µl reservoir solution containing 10% PEG 8000, 100 mM HEPES pH 7.5, and 200 mM calcium acetate.

For cryo protection, crystals were briefly swished through mother liquor containing 20% ethylene glycol.

Data collection and structural determination of the S0_2.126/D25 Fab complex

Diffraction data was recorded at ESRF beamline ID30B. Data integration was performed by XDS (*26*) and a high-resolution cut at I/σ=1 was applied. The dataset contained a strong off-origin peak in the Patterson function (88% height rel. to origin) corresponding to a pseudo translational symmetry of ½, 0, ½. The structure was determined by the molecular replacement method using PHASER (*27*) using the D25 structure (*1*) (PDB ID 4JHW) as a search model. Manual model building was performed using Coot (*28*), and automated refinement in Phenix (*29*). After several rounds of automated refinement and manual building, paired refinement (*30*) determined the resolution cut-off for final refinement.

<u>Co-crystallization of complex 101F Fab with S4_2.45</u>

The complex of S4_2.45 with the F101 Fab was prepared by mixing two proteins in 2:1 molar ratio for 1 hour at 4 °C, followed by size exclusion chromatography using a Superdex-75 column (Fig S21). Complexes of S4_2.45 with the 101F Fab were verified by SDS–PAGE (Fig S21). Complexes were subsequently concentrated to 6–8 mg/ml. Crystals were grown using hanging drops vapor-diffusion method at 20 °C. The S4_2.45/101F protein complex was mixed with equal volume of a well solution containing 0.2 M Magnesium acetate, 0.1 M Sodium cacodylate pH 6.5, 20 %(w/v) PEG 8000. Native crystals were transferred to a cryoprotectant solution of 0.2 M Magnesium acetate, 0.1 M Sodium cacodylate pH 6.5, 20 % (w/v) PEG 8000 and 15% glycerol, followed by flash-cooling in liquid nitrogen.

Data collection and structural determination of the S4_2.45/101F Fab complex

Diffraction data were collected at SSRL facility, BL9-2 beamline at the SLAC National Accelerator Laboratory. The crystals belonged to space group P3221. The diffraction data were initially processed to 2.6 Å with X-ray Detector Software (XDS) (Table S9). Molecular replacement searches were conducted with the program PHENIX PHASER using 101F Fab model (PDB ID: 3O41) and S4_2.45/101F Fab computational model generated from superimposing epitope region of S4_2.45 with the peptide-bound

structure (PDB ID: 3O41), and yielded clear molecular replacement solutions. Initial refinement provided a R_{free} of 42.43% and R_{work} of 32.25% and a complex structure was refined using Phenix Refine, followed by manual rebuilding with the program COOT. The final refinement statistics, native data and phasing statistics are summarized in Table S9.

Next-generation sequencing of design pools

After sorting, yeast cells were grown overnight, pelleted and plasmid DNA was extracted using Zymoprep Yeast Plasmid Miniprep II (Zymo Research) following the manufacturer's instructions. The coding sequence of the designed variants was amplified using vector-specific primer pairs, Illumina sequencing adapters were attached using overhang PCR, and PCR products were desalted (Qiaquick PCR purification kit, Qiagen). Next generation sequencing was performed using an Illumina MiSeq 2x150bp paired end sequencing (300 cycles), yielding between 0.45-0.58 million reads/sample.

For bioinformatic analysis, sequences were translated in the correct reading frame, and enrichment values were computed for each sequence. We defined the enrichment value E as follows:

$$E_{Seq} = \frac{count_{Seq}(high\ selective\ pressure)}{count_{Seq}(low\ selective\ pressure)}$$

The high selective pressure corresponds to low labelling concentration of the respective target antibodies (100 pM D25, 10 nM 5C4 or 20 pM 101F, as shown in Fig. 3), or a higher concentration of chymotrypsin protease (0.5 μM). The low selective pressure corresponds to a high labelling concentration with antibodies (10 nM D25, 1 μM 5C4 or 2 nM 101F), or no protease digestion, as indicated in Fig. 3. Only sequences that had at least one count in both sorting conditions were included in the analysis.

Tables.

Table S1.
Primers used for constructing single-site saturation mutagenesis library for S4_1 design.

S4_1_SSM_fw	CAGGCTAGTGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGC
S4_1_SSM_rw	CTGTTGTTATCAGATCTCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCC
S4_1_#18_rw	TTTCGGGCATTTGACTTTGATACCATTGCTGT
S4_1_#18_fw	CAATGGTATCAAAGTCAAATGCCCGAAANNKGGTGAATGTACGATTAAAGACAGTCAACG
S4_1_#20_rw	CTTTCGGGCATTTGACTTTGATACCATTGCTGT
S4_1_#20_fw	GCAATGGTATCAAAGTCAAATGCCCGAAAGGCGGTNNKTGTACGATTAAAGACAGTCAACGTGG
S4_1_#22_rw	CCTTTCGGGCATTTGACTTTGATACCATTGCTGT
S4_1_#22_fw	GCAATGGTATCAAAGTCAAATGCCCGAAAGGCGGTGAATGTNNKATTAAAGACAGTCAACGTGGCATTATC
S4_1_#25_rw	TTTAATCGTACATTCACCGCCTTTCG
S4_1_#25_fw	CGAAAGGCGGTGAATGTACGATTAAANNKAGTCAACGTGGCATTATCAAAACC
S4_1_#36_rw	GCTAAAGGTTTTGATAATGCCACGTTGAC
S4_1_#36_fw	CAACGTGGCATTATCAAAACCTTTAGCNNKGGTACGGAAGAAGTTCGCAGTC
S4_1_#39_rw	CGTACCAGAGCTAAAGGTTTTGATAATGCCA
S4_1_#39_fw	GCATTATCAAAACCTTTAGCTCTGGTACGNNKGAAGTTCGCAGTCCGTCCCTG
S4_1_#43_rw	GCGAACTTCTTCCGTACCAGAGCTAAAG
S4_1_#43_fw	GCTCTGGTACGGAAGAAGTTCGCNNKCCGTCCCTGGGCAAAGTGACCGT
S4_1_#45_fw	GCTCTGGTACGGAAGAAGTTCGCAGTCCGNNKCTGGGCAAAGTGACCGTTGGTGATAAC
S4_1_#48_rw	GCCCAGGGACGGACTGCGAACTTC
S4_1_#48_fw	GTTCGCAGTCCGTCCCTGGGCNNKGTGACCGTTGGTGATAACACGTTC
S4_1_#50_fw	GTTCGCAGTCCGTCCCTGGGCAAAGTGNNKGTTGGTGATAACACGTTCGAAGCG

Table S2.
Primers used for encoding computationally designed sequences of S4_2 design series.

S4_2_uni_O1	GACAATAGCTCGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAGGCTAGTGGTGGAGGAGG
S4_2_uni_O2	CCCTCCGCCTCCGCTACCGCCTCCACCAGAGCCTCCTCCACCACTAGCCTG
S4_2_bb1_O3.1	GTAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCGTCCATCYACTCAKWCGTTSYTGGTGGGAACATCAAGGTGAAGTGC
S4_2_bb1_O3.2	GTAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCGTCCATCYACTCAKWCGTTSYTGGGAACATCAAGGTGAAGTGC

S4_2_bb1_O3.3	GTAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCGTCCATCYACTCAKWCGTTSYTAACGGGAACATCAAGGTGAAGTGC
S4_2_bb1_O4.1	GGTCTTGATGATGCCACGCTGGCTATCCTCGATGGTACATTTGTCACCAGTGCACTTCACCTTGATGTTCCC
S4_2_bb1_O4.2	GGTCTTGATGATGCCACGATTCTTGTTCTCGATGGTACATTTGTCACCAGTGCACTTCACCTTGATGTTCCC
S4_2_bb1_O4.3	GGTCTTGATGATGCCACGCTGGCTATCCTCGATGGTACACTTGCCCTCGTGGCACTTCACCTTGATGTTCCC
S4_2_bb1_O5.1	GCGTGGCATCATCAAGACCACGAATGTTGATATTGCTGAGGAGRYGYRGAAGCAGSYTCAAGAGBYTBWGGAAGMGAAACGTAAGGGCTCGTGGGGCTCG
S4_2_bb1_O5.2	GCGTGGCATCATCAAGACCTTCACGGGGTTCGAGCCCGAGGAGRYGYRGAAGCAGSYTCAAGAGBYTBWGGAAGMGAAACGTAAGGGCTCGTGGGGCTCG
S4_2_bb1_O5.3	GCGTGGCATCATCAAGACCGTCCCGATGATCGAGACAGGGGAGGAGRYGYRGAAGCAGSYTCAAGAGBYTBWGGAAGMGAAACGTGGCTCGTGGGGCTCG
S4_2_uni_O6	CAGAAATAAGCTTTTGTTCGGATCCGGGCTCAGCCTATTAGTGGTGGTGGTGGTGGTGCGAGCCCCACGAGCC
S4_2_uni_O7	GGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGCTCGAGATCTGATAAC
S4_2_uni_O8	GTACGAGCTAAAAGTACAGTGGGAACAAAGTCGATTTTGTTACATCTACACTGTTGTTATCAGATCTCGAGCTATTACAAGTCC
S4_2_bb2_O3.1	TAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCAAAWACCHACGTAWTTGAAGCAGGCDTCAGCTTCACCTGCTTAGGTGAGAAGTGCACCATCGAGGAC
S4_2_bb2_O3.2	TAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCAAAWACCHACGTAWTTCCCTCGDTCAGCTTCACCTGCTTAGGTGAGAAGTGCACCATCGAGGAC
S4_2_bb2_O3.3	TAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCAAAWACCHACGTAWTTCCCTCGDTCAGCTTCACCTGCCCTAAGGGGGGGAAGTGCACCATCGAGGAC
S4_2_bb2_O4.1	CGGTCTTGATGATCCCACGTTGTGAGTCCTCGATGGTGCACTTC
S4_2_bb2_O4.2	CGGTCTTGATGATCCCACGATCGTCCTCGATGGTGCACTTC
S4_2_bb2_O4.3	CGGTCTTGATGATCCCACGCGAGCGGTCCTCGATGGTGCACTTC
S4_2_bb2_O5.1	CGTGGGATCATCAAGACCGGCAAAAATGCCGAGGAGKYCDKGGAAGATBTCGAGAAGVRGGHGCGTGCCCAGGGCTCGTGGGGCTCGCAC
S4_2_bb2_O5.2	CGTGGGATCATCAAGACCGGCACACATCCAGAGGAGKYCDKGGAAGATBTCGAGAAGVRGGHGCGTGCCCAGGGCTCGTGGGGCTCGCAC
S4_2_bb2_O5.3	CGTGGGATCATCAAGACCGGCAAAAATAAGGAGGAGKYCDKGGAAGATBTCGAGAAGVRGGHGCGTGCCCAGGGCTCGTGGGGCTCGCAC
S4_2_bb3_O3.1	TAGCGGAGGCGGAGGGTCGGCTAGCCATATGGTCTKSAGTKKTGTAGYTGGGGAGAACTATTCARYTAAGTGTACTGGCGACAAGTGCACCATCGAGGAC
S4_2_bb3_O3.2	TAGCGGAGGCGGAGGGTCGGCTAGCCATATGGTCTKSAGTKKTGTAGYTACCCCGACATTTTCARYTAAGTGTACTGGCGACAAGTGCACCATCGAGGAC

S4_2_bb3_O3.3	TAGCGGAGGCGGAGGGTCGGCTAGCCATATGTKSAGTKKTGTAGY TGGGGAGAACTATTCARYTAAGTGTCCTAAGGGGGGCAAGTGCAC CATCGAGGAC
S4_2_bb3_O4.1	GGTCTTGATGATCCCGCGCTGTGAGTCCTCGATGGTGCACTTG
S4_2_bb3_O4.2	GGTCTTGATGATCCCGCGATTCTTGTCCTCGATGGTGCACTTG
S4_2_bb3_O4.3	GGTCTTGATGATCCCGCGCCCGCCATAGTCCTCGATGGTGCACTTG
S4_2_bb3_O5.1	CGCGGGATCATCAAGACCACGATTGGAGATACATGTGAGSHGKYG KMTAAGGCGGYTCAAAAGGCTSVGAAAGGCTCGTGGGGCTCG
S4_2_bb3_O5.2	CGCGGGATCATCAAGACCGTTACTGGCAGTCGCTGTGAGSHGKYG KMTAAGGCGGYTCAAAAGGCTSVGAAAGGCTCGTGGGGCTCG

Table S3.
Computationally designed protein sequences for S4_1 design series.

Design name	Sequence	Expression vector
S4_1.1	MDGTLQINSNGIKVKCPKGGECTIKDSQRGIIKTFSSGTEEVRSPS LGKVTVGDNTFEASNGSWLEHHHHHH	pET21b
S4_1.2	MHHHHHHWGSPGTVTLNSNGLTVTGNDNYNLTVTGNDRGIIKT FSPSTETTNDDGMSTITVGNLTVTLGN	pET11b
S4_1.3	MHHHHHHWGSQSTVNVQDKNIRIEVDDKNSVQVNGSNRGIIKTF SPGTVQISSKNGDTVTVGNVRVNMGG	pET11b
S4_1.4	MHHHHHHWGSQSTVNVQDKNIRIECDDNCGVQVNGSNRGIIKT FSPGTVQISSKNGDTVTVGNVRVNMGG	pET11b
S4_1.5	MHHHHHHWGSDGTLQINSHGVKVKAPPGSGATVKDSQRGIIKTF SSGYEEVRSPSLGKVTVGDNTFEVSN	pET11b
S4_1.6	MHHHHHHWGSDGTLQINSHGVKVKCPKGSECTVKDSQRGIIKTF SSGYEEVRSPSLGKVTVGDNTFEVSN	pET11b
S4_1.7	MHHHHHHGSKVTFRQDKNGIKIRVNGNKGLVIRTNDRGIIKTFS NGTYDIPNSGYNRFTVGGTQFDWNE	pET11b
S4_1.8	MHHHHHHGSKVTFRQDKNGIKFRVNGNKGAVIRTNDRGIIKTFS NGTYDIPNSGYNRFTVGGNTFDWNE	pET11b
S4_1.9	MDGTLQINSNGVKVKCPKGVECTVKDSQRGIIKTFSSGTEEVRSP SLGKVTVGDNTFEVSNGSWLEHHHHHH	pET21b
S4_1.10	MDGTLQINSNGVKVKCPKGAECTVKVSQRGIIKTFSSGTEEVRSP SLGKVTVGDNTTEVSNGSWLEHHHHHH	pET21b
S4_1.11	MHHHHHHWGSPGTVKLNSNGLTVRGNDSYGLTVRGNDRGIIKT FSPSTEVVQSKGMSTITVGNLDVRLGE	pET11b

Table S4.
Computationally designed protein sequences for S0_1 design series.

Design name	Sequence	Expression vector
S0_1.1	PEDAQKEASKGSEVRELKNIIDKQLLPIVNKTSCSGAEQAAEAL REALEGAGSCDAVEQLLGNIKEIKCGTDAGRALIRILAEVAREI GCPRAIDQVAEWVRRIAKAVGGEEAKKQVKEVEKEIRREKG	pET21
S0_1.17	PEDAQKEASKGSEVRELKNIIDKQLLPILNKASCSGAEQLLEAL REALEGAGSCDAVEQLLGNIKEIKCGTDAGRALKRILEEVQREI GCGSW	pET21
S0_1.37	CDQLKNYIDKQLLPIVNKQSCANGEEALKDIEKALRGAGSKDC WKELLSNIKEIKCGKEFAEKLKKEWERIKKEAGD	pET21

S0_1.38	CDQIKNYIDKQLLPIVNKAGCGSAEEALKDIEKALRLAGSKDCL KEIFSNIKEIKCGKEFAEKLKKEWERIKKEAGD	pET21
S0_1.39	CDQIKNYIDKQLLPIVNKAGCGSAEEVLKDIEKALRNAGSKDCL KEIFSGIKEIKCGKEQAEKLKKEWERIKKEAGDG	pET21
S0_1.40	ADQIKNYIDKQLLPIVNKAGCGSAEEVLKDIEKALRNAGSKDA LKEIFSGIKEIKCGKEQAEKLKKEWERIKKEAGDG	pET21

Table S5.
Protein sequences for S4_2 design series.

Design name	Sequence	Successful expression
S4_2.07	MCSVVVGENYSIKCNPDGKCTIEDKNRGIIKTVTGSRCELLYKAVQ KAQKGSWGSHHHHHH	yes
S4_2.19	MPNTNVFPSFSFTCLPDGKCIIEDSQRGIIKTGKNKEEFMEDFEKQV RAQGSWGSHHHHHH	yes
S4_2.20	MPSIYSDVPGGNIKVKCHEGKCTIEDSQRGIIKTVPMIETGEEMWK QVQEVLEEKRGSWGSHHHHHH	yes
S4_2.21	MPKTNVIPSFSFTCLGEKCTIEDSQRGIIKTGKNKEEVLEDFEKEER AQGSWSHHHHHH	yes
S4_2.22	MPSIYSDVPGNIKVKCHEGKCTIEDSQRGIIKTVPMIETGEEMWKQ PQELLEEKRGSWGSHHHHHH	yes
S4_2.35	MPNTNVFPSFSFTCLPDGKCIIEDSQRGIIKTGKNKEEFMEDFEKKV RAQGSWGSHHHHHH	yes
S4_2.45	MVCSVVVGENYSIKCDATKCTIEDKNRGIIKTVTGSRCEELAKAV QKAQKGSWGSHHHHHH	yes
S4_2.60	MPSIYSDVPGGNVKVKCHEGKCTIEDSQRGIIKTVPMIETGEEMWK QVQEVVEEKRGSWGSHHHHHH	yes
S4_2.68	MPSIHSVVVGGNIKVKCHEGKCTIEDSQRGIIKTVPMIETGEEMQK QVQEFLEAKRGSWGSHHHHHH	yes
S4_2.73	MVCSVVVGENYSIKCDATKCTIEDSQRGIIKTGTHPEEFLEDLEKK ARAQGSWGSHHHHHH	yes
S4_2.74	MVFSCVVGENYSIKCDATKCTIEDSQRGIIKTGTHPEEFLEDLEKK ARAQGSWGSHHHHHH	yes
S4_2.84	MPSIHSVVPGGNIKVKCHEGKCTIEDSQRGIIKTVPMIETGEEMWK QPQELLEEKRGSWGSHHHHHH	yes
S4_2.85	MPNTNVFPSFSFTCLPDGKCIIEDSQRGIIKTGKNKEEFMEDFEKQV RAQGSWGSHHHHHH	yes
S4_2.94	HMPSIHSVVAGGNIKVKCHEGKCTIEDSQRGIIKTFTGFEPEEVWK QAQEFLEEKRGSWGSHHHHH	yes

Table S6.
Protein sequences for S0_2 design series.

Design name	Sequence	Successful expression
S0_2.37	MSCDQIKNYIDKQLLPIVNKAGCSRPEELEERIRRALKKFGDT DCLKDILLGIKEWKCGGSLEHHHHHH	no
S0_2.79	MPCDKQKNYIDKQLLPIVNKAGCSRPEEVEEMVRRALKKLGE TPCLEDILRGIKEIKCGGSLEHHHHHH	yes
S0_2.10	MPCDDAKNYIDKQLLPIVNKAGCSRPEEVERAVRKMLKKMG NTDCLEDILRGIKEIKCGGSLEHHHHHH	yes

S0_2.102	MSCDQIKNYIDKQLLPIVNKAGCGSAKEVQKDIEKALRNAGV KDCLEDILRGIKEWKCGGSLEHHHHHH	no
S0_2.31	MSCDESKNYIDKQLLPIVNKAGCDRPEDVERWIRKALKKMG DTSCFDEILKGLKEIKCGGSLEHHHHHH	no
S0_2.197	MSCDQIKNYIDKQLLPIVNKAGCSRPEEVEERIRRALKKMGDT SCFDEIMKGLKEIKCGGSLEHHHHHH	no
S0_2.57516	MSCDQIKNYIDKQLLPIVNKAGCNRPEEFEEWIKRALKKLGDT SCLEDILRGIKEIKCGGSLEHHHHHH	no
S0_2.57575	MSCDQIKNYIDKQLLPIVNKAGCSRPEEVEEMVRRALKKLGE TPCLEDILRGIKEWKCGGSLEHHHHHH	no
S0_2.57588	MSCDQIKNYIDKQLLPIVNKAGCSRPEEVERAVRKMLKKMG NTDCLEDILRGIKEIKCGGSLEHHHHHH	no
S0_2.57855	MSCDQIKNYIDKQLLPIVNKAGCGSAKEVQKDIEKALRNAGV KDCLKEIFSGIKEIKCGGSLEHHHHHH	no
S0_2.57910	MSCDQIKNYIDKQLLPIVNKAGCGSAKEVQKDIEKALRNAGV KDCLEDILRGIKEIKCGGSLEHHHHHH	no
S0_2.57911	MSCEEAKNYIDKQLLPIVNKAGCGSAEEVQKDIEKALRNAGV KDCLEDILRGIKEWKCGGSLEHHHHHH	no
S0_2.57	MPCDDAKNYIDKQLLPIVNKAGCSRPEEVEERIRRALKKMGD TSCFDEIMKGLKEIKCGGSLEHHHHHH	yes
S0_2.58980	MSCEEAKNYIDKQLLPIVNKAGCSRPEELEEMIRRALKKMGD TSCFDEIMKGLKEIKCGGSLEHHHHHH	no
S0_2.611	MPCDKQKNYIDKQLLPIVNKAGCGSAKEVQKDIEKALRNAG VKDCLEDILRGIKEWKCGGSLEHHHHHH	yes
S0_2.126	MPCDKQKNYIDKQLLPIVNKAGCSRPEEVEERIRRALKKMGD TSCFDEILKGLKEIKCGGSLEHHHHHH	yes

Table S7.
Refinement statistics of the S0_2.126 NMR structure.

NMR restraints	
Total NOEs from Unio [a]	*306*
Intraresidual	124
Interresidual	182
Sequential (i – j = 1)	112
Medium-range (1 < i – j < 5)	47
Long-range (i – j ≥ 5)	23
Dihedral Angles from Talos-n [b]	*88*
ϕ	43
ψ	45
Structural statistics	
Violations[c]	
Distance restraints (Å)	0.0254 ± 0.009
Dihedral angle constraints (°)	6.8 ± 0.12
Ramachandran plot (all residues/ordered residues) [d]	
Most favored (%)	84.7 / 95.8
Additionally allowed (%)	14.3 / 4.5
Generously allowed (%)	0.98 / 0.1
Disallowed (%)	0 / 0
Average pairwise RMSD (Å) [e]	

Heavy	3.3 / 1.8
Backbone	2.8 / 1.2
Structure Quality Factors (raw score/z-score) [e]	
Procheck G-factor (phi/psi)	0.15 / 0.9
Procheck G-factor (all)	-0.48 / -2.84

[a] From UNIO-ATNOS/CANDID's last cycle (cycle 7)
[b] Obtained from chemical shifts with Talos-N server
[c] From Cyana in Unio's last cycle
[d] All residues from Cyana un Unio's last cycle; ordered residues (5-22,26-57) from the Protein Structure Validation Suite at http://psvs-1_5-dev.nesg.org/results/testbc/OUTPUT.html
[e] From the Protein Structure Validation Suite

Table S8.
X-ray data collection and refinement statistics of S0_2.126 crystal structure.

D25 S0_2.126	
Wavelength	0.9763
Resolution range	49.09-3.0 (3.107-3.0)
Space group	P 21 21 21
Unit cell	126.3 127.0 156.1 90 90 90
Total reflections	700184 (72248)
Unique reflections	50740 (5000)
Multiplicity	13.8 (14.4)
Completeness (%)	98.76 (99.22)
Mean I/sigma(I)	12.63 (2.00)
Wilson B-factor	74.78
R-merge	0.1622 (1.484)
R-meas	0.1684 (1.538)
R-pim	0.04506 (0.4019)
CC1/2	0.999 (0.893)
CC*	1 (0.971)
Reflections used in refinement	50284 (4971)
Reflections used for R-free	2519 (249)
R-work	0.2699 (0.3677)
R-free	0.2936 (0.3972)
CC(work)	0.949 (0.817)
CC(free)	0.958 (0.793)
Number of non-hydrogen atoms	14453
macromolecules	14452
Protein residues	1921
RMS(bonds)	0.004
RMS(angles)	1.02
Ramachandran favored (%)	94.45
Ramachandran allowed (%)	5.07
Ramachandran outliers (%)	0.48

Rotamer outliers (%)	0.00
Clashscore	7.35
Average B-factor	97.74
macromolecules	97.74
solvent	59.33
Number of TLS groups	12

Table S9.
X-ray data collection and refinement statistics of S4_2.45 crystal structure.

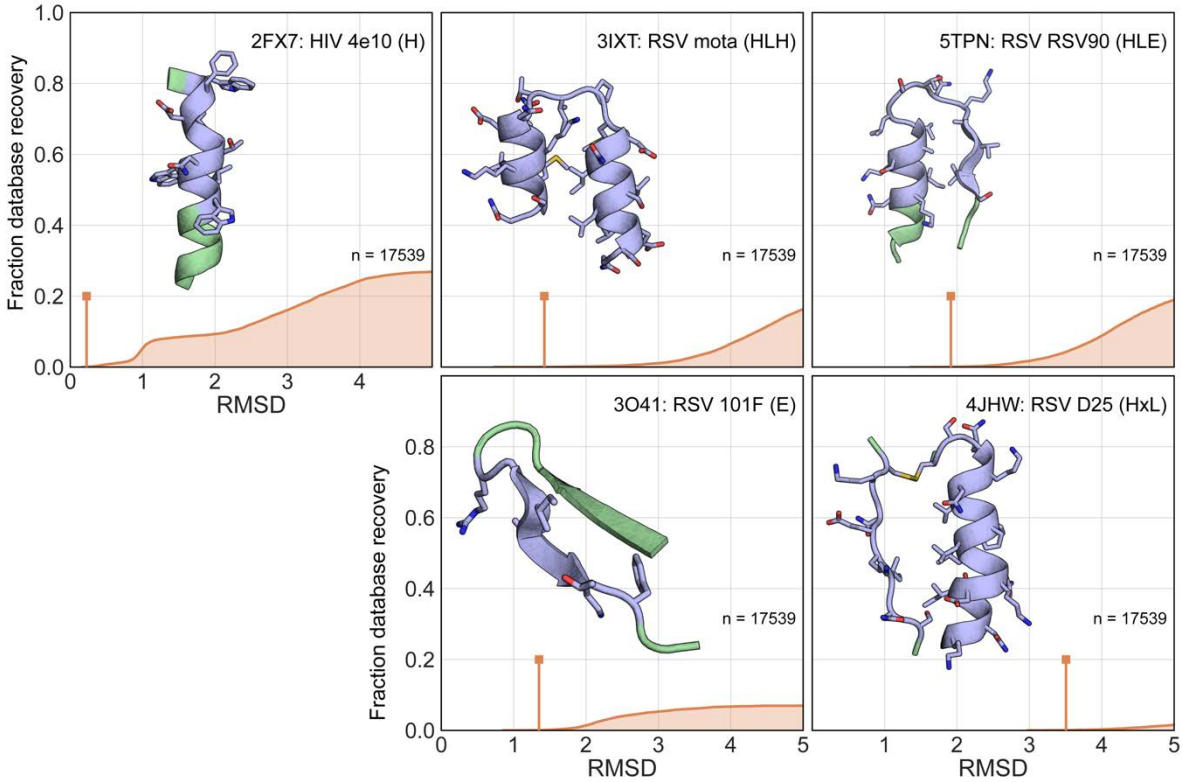
101F S4_2.45	
Wavelength	0.98
Resolution range	38.49 - 2.6 (2.693 - 2.6)
Space group	P 32 2 1
Unit cell	148.224 148.224 45.046 90 90 120
Total reflections	113069 (7302)
Unique reflections	17464 (1567)
Multiplicity	6.5 (4.7)
Completeness (%)	98.57 (89.58)
Mean I/sigma(I)	17.03 (1.66)
Wilson B-factor	56.09
R-merge	0.06712 (0.8361)
R-meas	0.07282 (0.9424)
R-pim	0.02776 (0.4231)
CC1/2	0.999 (0.635)
CC*	1 (0.881)
Reflections used in refinement	17455 (1565)
Reflections used for R-free	1748 (166)
R-work	0.2298 (0.3682)
R-free	0.2736 (0.3503)
CC(work)	0.462 (0.203)
CC(free)	0.353 (0.190)
Number of non-hydrogen atoms	3794
macromolecules	3686
solvent	108
Protein residues	485
RMS(bonds)	0.010
RMS(angles)	1.46
Ramachandran favored (%)	93.53
Ramachandran allowed (%)	5.64
Ramachandran outliers (%)	0.84
Rotamer outliers (%)	0.96
Clashscore	2.19
Average B-factor	38.90

macromolecules	38.37
solvent	56.78
Number of TLS groups	3

References

1. J. S. McLellan *et al.*, Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113-1117 (2013).
2. J. Zhou, G. Grigoryan, Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci* **24**, 508-524 (2015).
3. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580-584 (2015).
4. J. Bonet *et al.*, Rosetta FunFolDes - A general framework for the computational design of functional proteins. *PLoS Comput Biol* **14**, e1006623 (2018).
5. X. Hu, H. Wang, H. Ke, B. Kuhlman, High-resolution design of a protein loop. *Proc Natl Acad Sci U S A* **104**, 17668-17673 (2007).
6. P. Conway, M. D. Tyka, F. DiMaio, D. E. Konerding, D. Baker, Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci* **23**, 47-55 (2014).
7. J. S. McLellan *et al.*, Structure of a major antigenic site on the respiratory syncytial virus fusion glycoprotein in complex with neutralizing antibody 101F. *J Virol* **84**, 12236-12244 (2010).
8. V. Mas *et al.*, Engineering, Structure and Immunogenicity of the Human Metapneumovirus F Protein in the Postfusion Conformation. *PLoS Pathog* **12**, e1005859 (2016).
9. M. D. Tyka *et al.*, Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* **405**, 607-618 (2011).
10. M. D. Finucane, M. Tuna, J. H. Lees, D. N. Woolfson, Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry* **38**, 11604-11612 (1999).
11. P. Kristensen, G. Winter, Proteolytic selection for protein folding using filamentous bacteriophages. *Fold Des* **3**, 321-328 (1998).
12. A. Chevalier *et al.*, Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74-79 (2017).
13. F. Sesterhenn *et al.*, Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen. *PLoS Biol* **17**, e3000164 (2019).
14. G. Chao *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* **1**, 755-768 (2006).
15. M. G. Joyce *et al.*, Iterative structure-based improvement of a fusion-glycoprotein vaccine against RSV. *Nat Struct Mol Biol* **23**, 811-820 (2016).
16. B. Briney *et al.*, Tailored Immunogens Direct Affinity Maturation toward HIV Neutralizing Antibodies. *Cell* **166**, 1459-1470 e1411 (2016).
17. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).
18. J. M. de la Rosa-Trevin *et al.*, Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* **195**, 93-99 (2016).

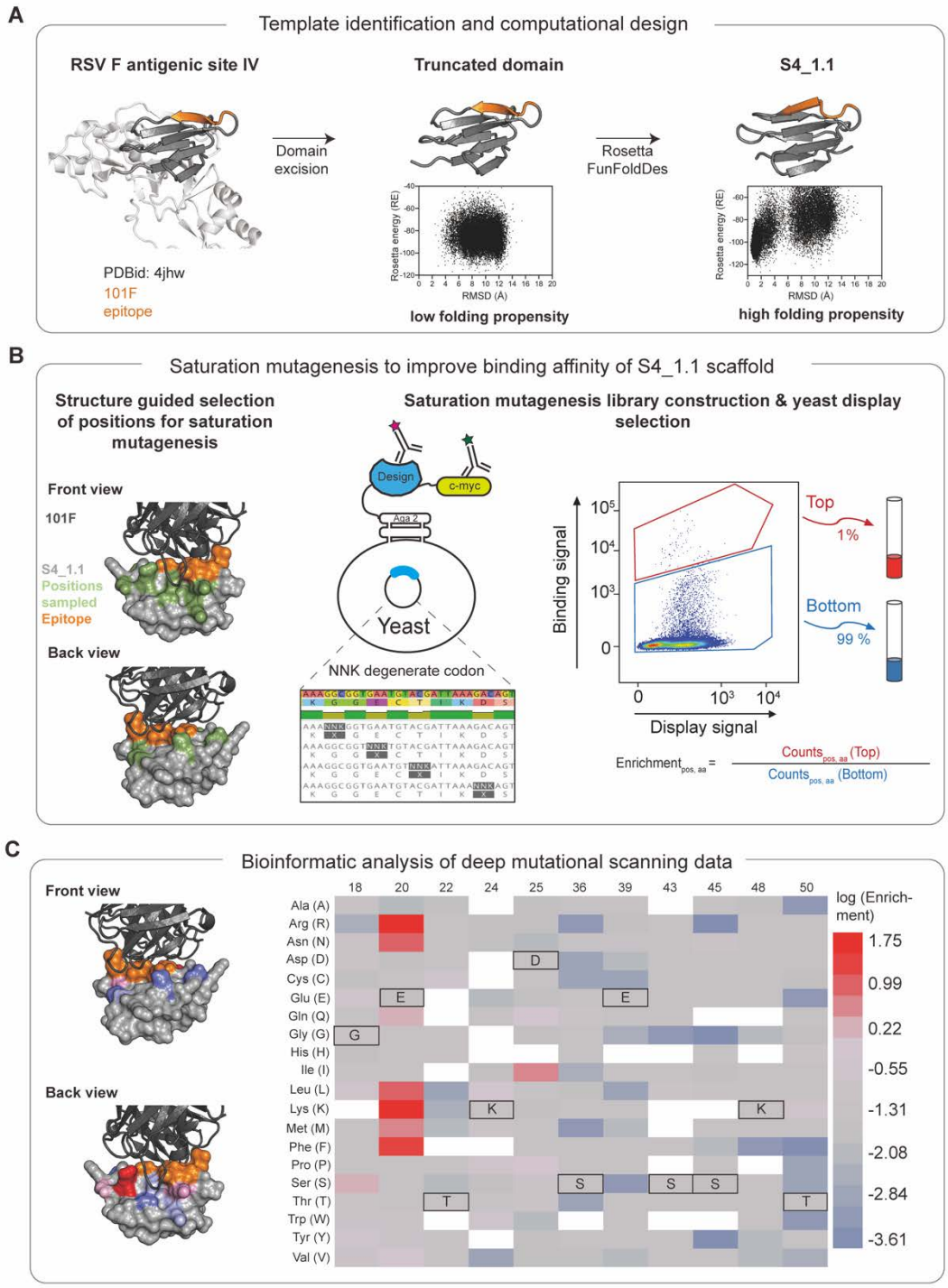
19. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).
20. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
21. M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* **34**, 93-158 (1999).
22. T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *Journal of Biomolecular Nmr* **24**, 171-189 (2002).
23. T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology* **319**, 209-227 (2002).
24. D. Gottstein, D. K. Kirchner, P. Guntert, Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space. *J Biomol NMR* **52**, 351-364 (2012).
25. Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular Nmr* **56**, 227-241 (2013).
26. W. Kabsch, Xds. *Acta Crystallogr D* **66**, 125-132 (2010).
27. A. J. Mccoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
28. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D* **66**, 486-501 (2010).
29. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213-221 (2010).
30. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).



Supplementary Figure 1

The increase in structural complexity of the functional motifs determine the number of designable templates that are found in known structures. A MASTER search (*1*) was performed over the nrPDB30 database (nonredundant subset of the PDB with a 30% sequence identity cutoff) containing a total of 17539 structures, querying the number of matches for different neutralization epitopes (colored in blue in the structures) of increasing structural complexity. The fraction of the database recovered is plotted on the y-axis. Matches were filtered for protein size <180 residues. The vertical line (orange) indicates the RMSD cutoff in Å

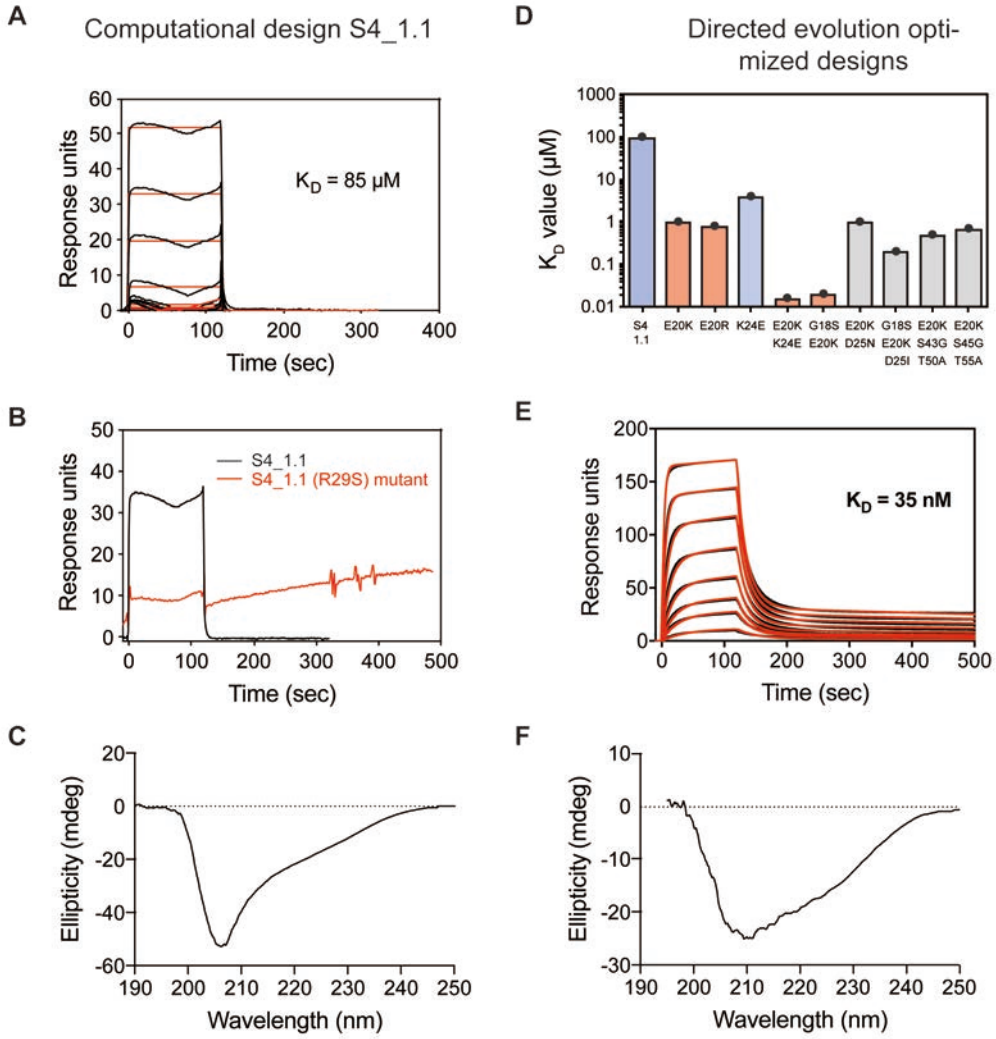
for the first 10 scaffold identified. Secondary structure composition of the motifs is represented by: E - strand; L – Loop; H – helix; x – chain break.



Supplementary Figure 2

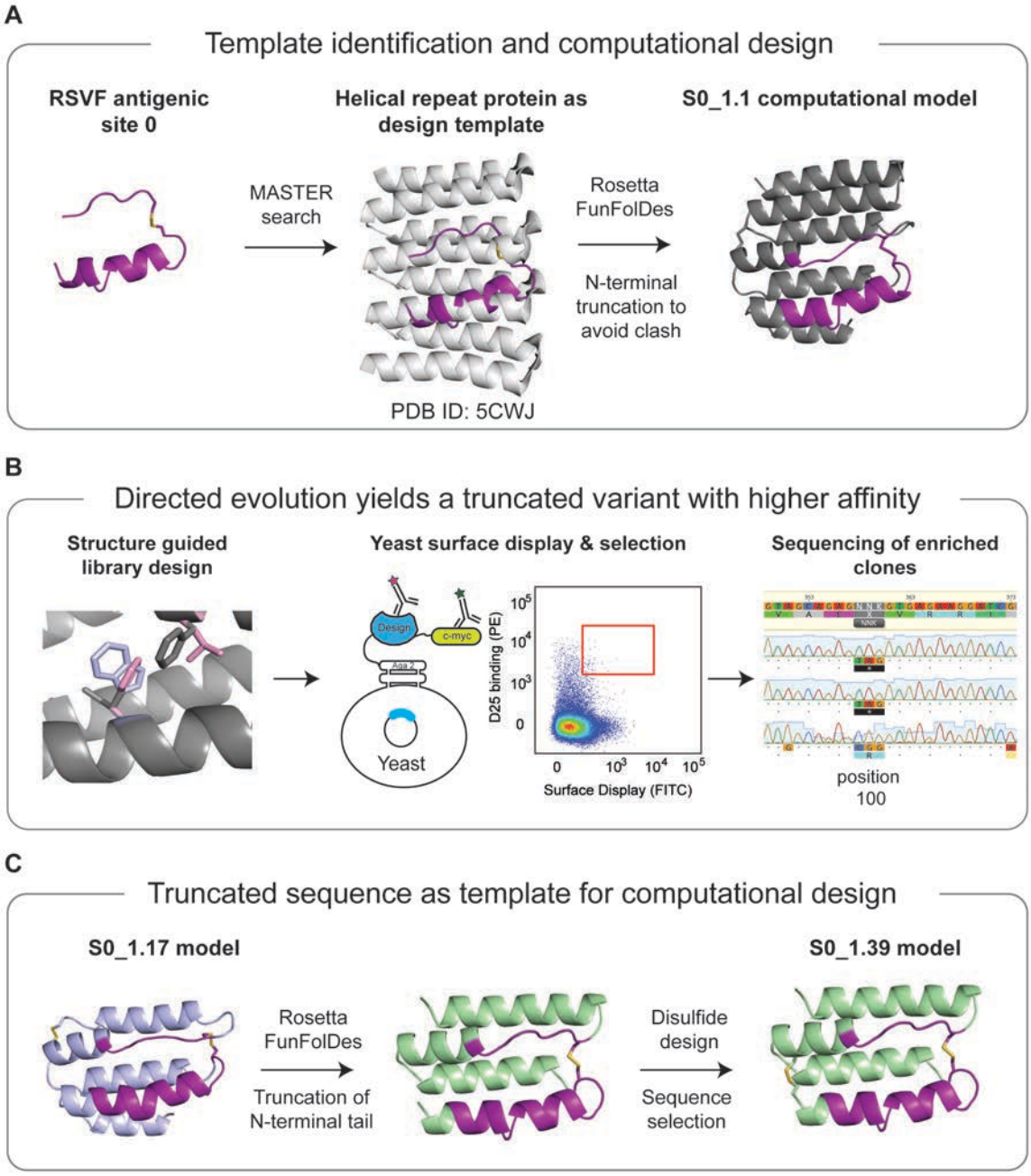
Computational design and experimental optimization of S4_1 design series. A) Template identification and computational design of S4_1.1. RSVF antigenic site IV is located in a small contained domain of preRSVF. This excised domain failed to show a folding funnel in Rosetta *abinitio* predictions, and failed to express recombinantly in *E.coli*. Using the excised domain as template, we folded and sequence-designed this topology using Rosetta FunFolDes, yielding design S4_1.1 which showed a strong funnel-shape energy landscape in *abinitio* folding simulation. B) Experimental optimization of S4_1.1 through saturation mutagenesis. A saturation mutagenesis library was constructed using overhang PCR for 11 positions proximal (green) to the

site IV epitope (orange), allowing one position at a time to mutate to any of the 20 amino acids, encoded by the degenerate codon 'NNK'. The library (size 11 positions x 32 codons = 352) was transformed in yeast, and designs were displayed on the cell surface. The selection was done by labeling the cells with 125 nM of 101F antibody. The top 1 % of clones binding with high affinity to 101F antibody were then sorted, as well as the bottom 99 % as shown. Following next-generation sequencing of the two populations, the enrichment values were computed for each sequence variant, corresponding to the relative abundance of each variant in the top versus bottom gate. C) Bioinformatic analysis of deep mutational scanning data. The log(enrichment) is shown as heatmap (right) for each sequence variant, and mapped to the structure (left). White indicates missing data. Position 20 showed the highest enrichment for arginine and lysine, together with other less pronounced enrichments seen for other positions.



Supplementary Figure 3

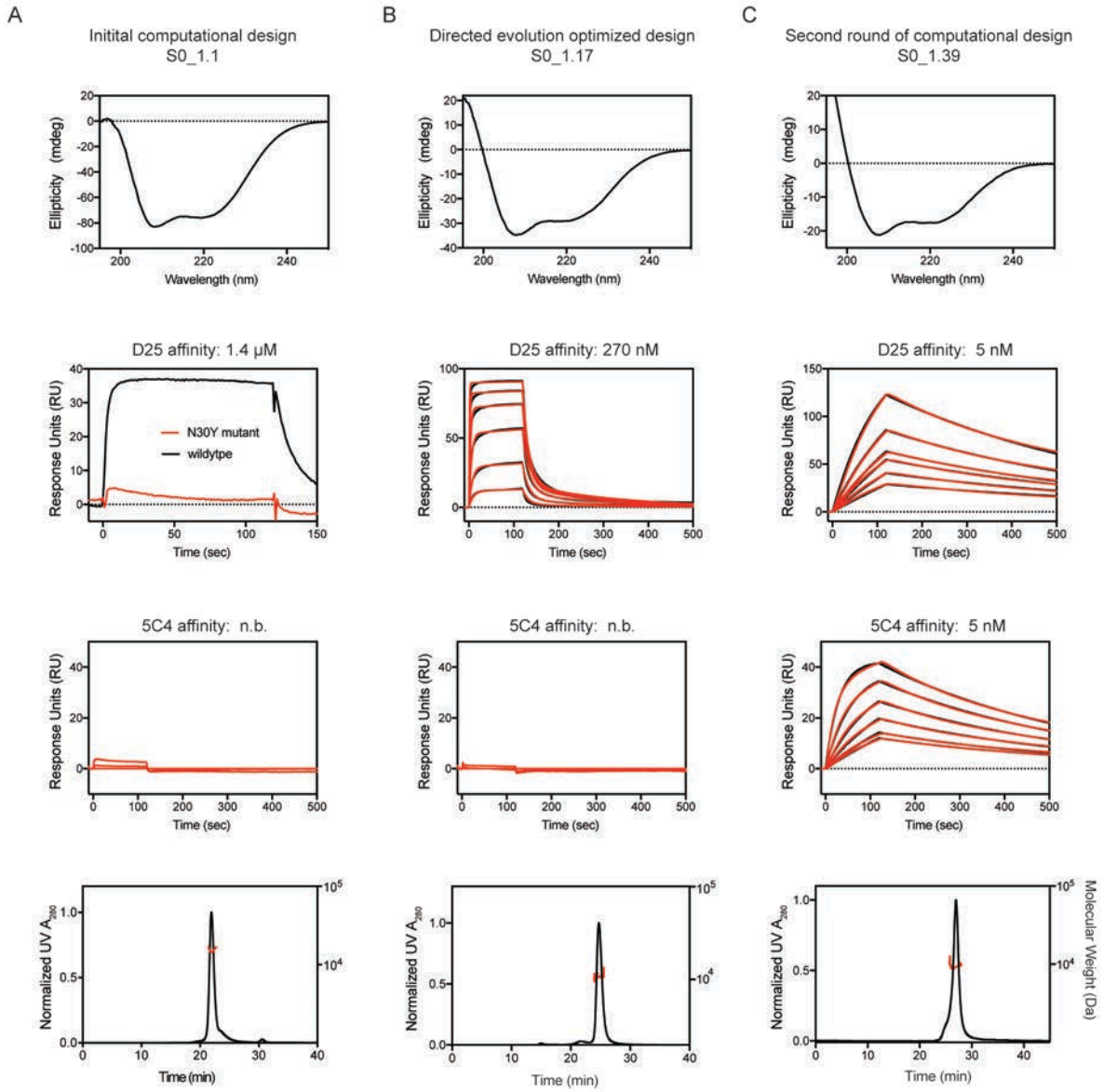
Experimental characterization of S4_1 design series. A) Surface plasmon resonance measurement for the initial computational design S4_1.1 against 101F antibody revealed a dissociation constant of > 85 μM. B) Despite low affinity, an R29S mutant revealed that binding was specific to the epitope of interest. C) Circular dichroism spectrum of S4_1.1 at 20 ºC. D) Dissociation constants (K_D) for single and combined mutations of S4_1.1 that were identified in the deep mutational scanning screen. E20K/K24E double mutant (named S4_1.5) showed a binding affinity of 35 nM. E) SPR sensorgram of S4_1.5 against 101F. F) Circular dichroism spectrum of S4_1.5 at 20 ºC.



Supplementary Figure 4

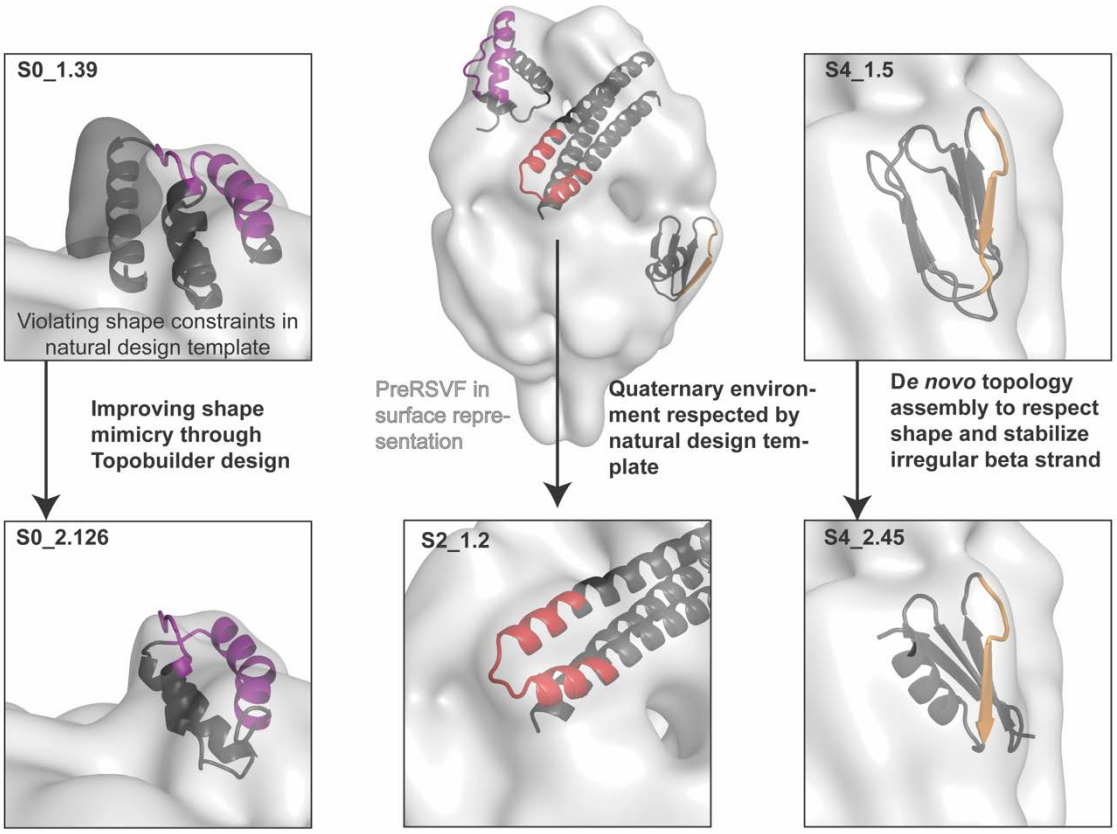
Computational design and experimental optimization of S0_1 design series. A) Template identification and design. Using MASTER, we identified a designed helical repeat protein (PDB ID: 5CWJ) to serve as design template to present and stabilize antigenic site 0 (see methods for details). The N-terminal 29 residues were truncated to avoid clashing with the D25 antibody, and Rosetta FunFolDes was used to design S0_1.1. See methods for details on the design process. B) Based on S0_1.1, a combinatorial sequence library was constructed and screened using yeast surface display. After three consecutive sorts of high-affinity binding clones, individual colonies were sequenced. Position 100 was frequently found to be mutated to a stop codon, leading to a truncated variant with increased expression yield, and a ~5-fold improved binding affinity to D25 (Fig S5). C) A model of the truncated variant served as template for a second round *of in silico* folding and design. We truncated the template further

by the N terminal 14 residues, and introduced a disulfide bond between residues 1 and 43, leading to S0_1.39. See methods for full details on the design selection process.



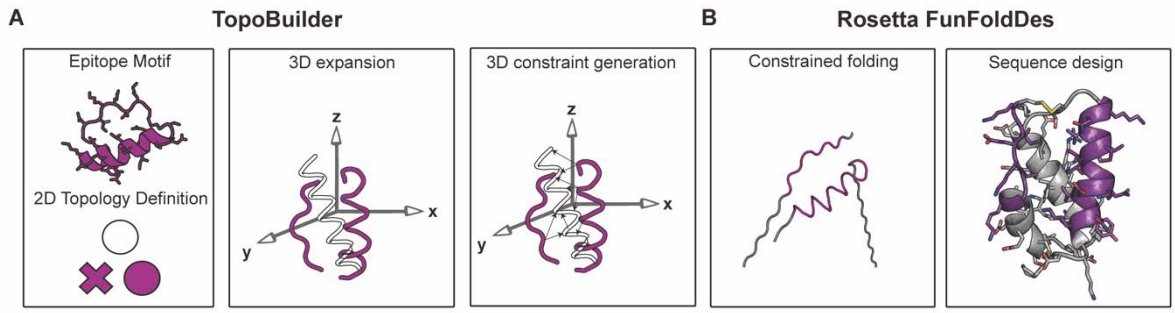
Supplementary Figure 5
Biophysical characterization of the S0_1 design series.
Top: Circular dichroism spectra at 20 ºC. Middle: Surface plasmon resonance measurements against D25 and 5C4. Bottom: Multi-angle light scattering coupled to size exclusion chromatography. A) S0_1.1 bound with a K_D of 1.4 μM to D25 and no detectable binding to 5C4. To verify that the binding interaction was specific to the epitope we generated a knockout mutant (N30Y) and observed that the binding interaction was absent. B) S0_1.17 showed a K_D of 270 nM to D25 and no binding to 5C4. C) SPR sensorgrams of S0_1.39 binding to D25 and 5C4 antibodies. D25 or 5C4 IgG was immobilized as ligand on the sensor chip surface, and S0_1.39 was flown as

analyte. All designs showed CD spectra typical of helical proteins and behaved as monomers is solution (Top and bottom rows).

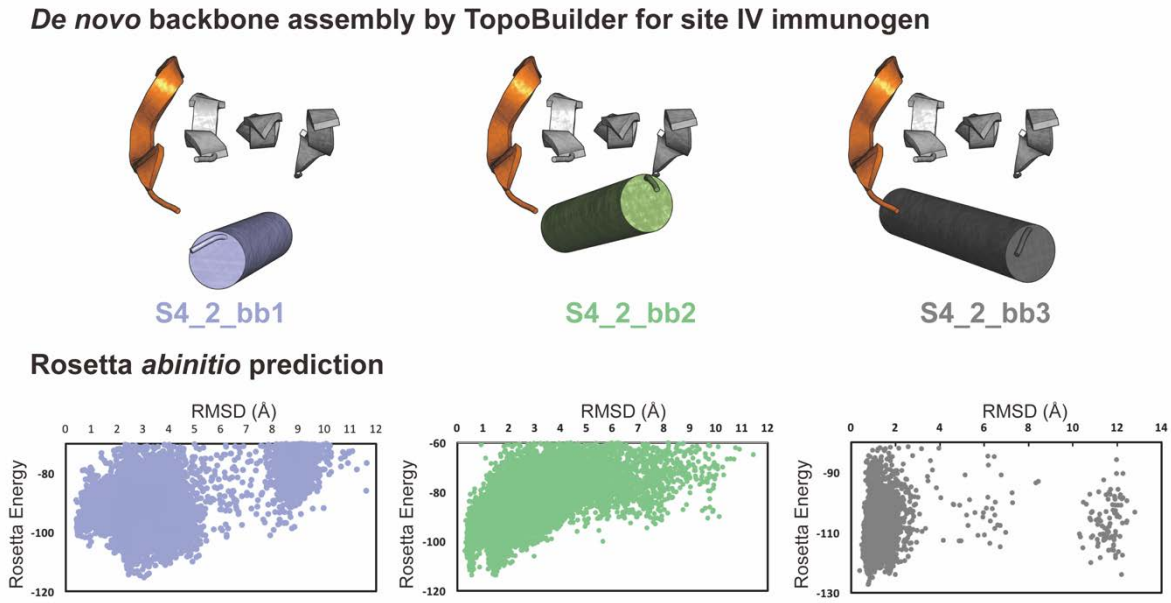


Supplementary Figure 6

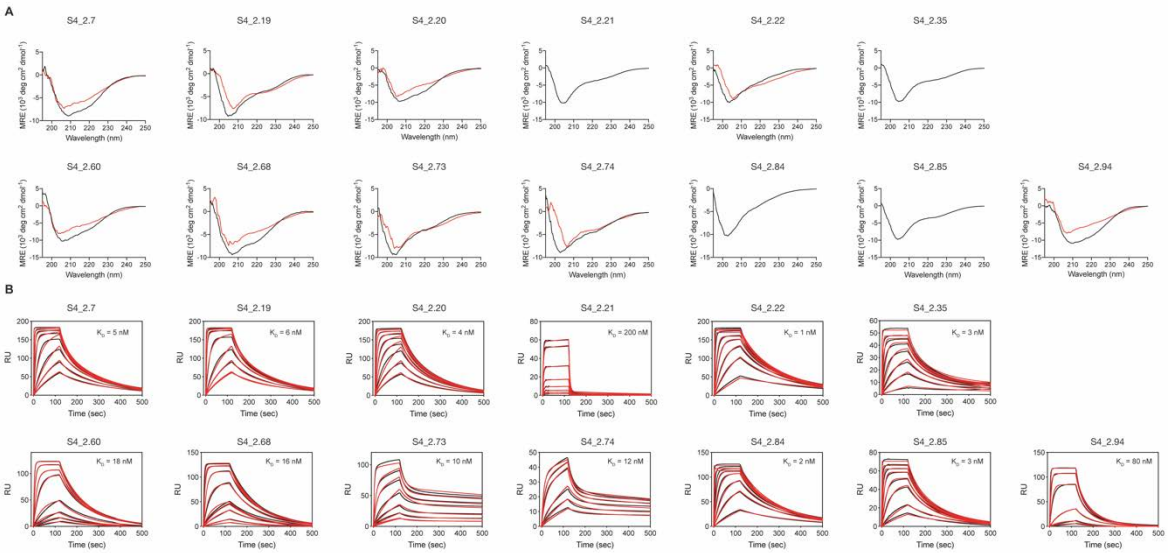
Shape mimicry of computationally designed immunogens compared to prefusion RSVF. Prefusion RSVF is shown in surface representation (light grey), with designed immunogens superimposed. Close-up views are shown for template-based designs (S0_1.39 and S4_1.5, top row). While site 0 is freely accessible for antibody binding in preRSVF, the C-terminal helix of S0_1.39 constrains its accessibility (dark grey surface). Through defined backbone assembly using TopoBuilder, S0_2.126 was designed, mimicking the native quaternary environment of site 0 (bottom left). RSVF antigenic site II, which is a structurally simple helix-turn-helix motif frequently found in natural proteins, was previously designed based on a design template that respects the quaternary constraints of site II in its native environment (S2_1.2, bottom middle). For site IV, a topology was assembled (S4_2.45) that respects the shape constraints while improving the stabilization of the irregular, bulged beta strand compared to the S4_1.5 design (right).



Supplementary Figure 7
TopoBuilder design strategy. A) The motif of interest is extracted from its native environment, and a 2D form is generated that allows to connect the discontinuous epitope segments. The 2D form is then expanded to the 3D space, applying user defined rotations and translations along x,y and z coordinates. From the 3D sketch, Cα constraints are generated to guide the folding process. B) Rosetta FunFolDes (*2*) is used to fold the idealized 3D sketch (using fragment insertions of sizes 3 and 9), and to build connecting loops between the secondary structures. A sequence that stabilizes the folded pose is designed in a last step using Rosetta FastDesign. Further details on the design process, the TopoBuilder code and scripts used for folding and design are available in the online repository.

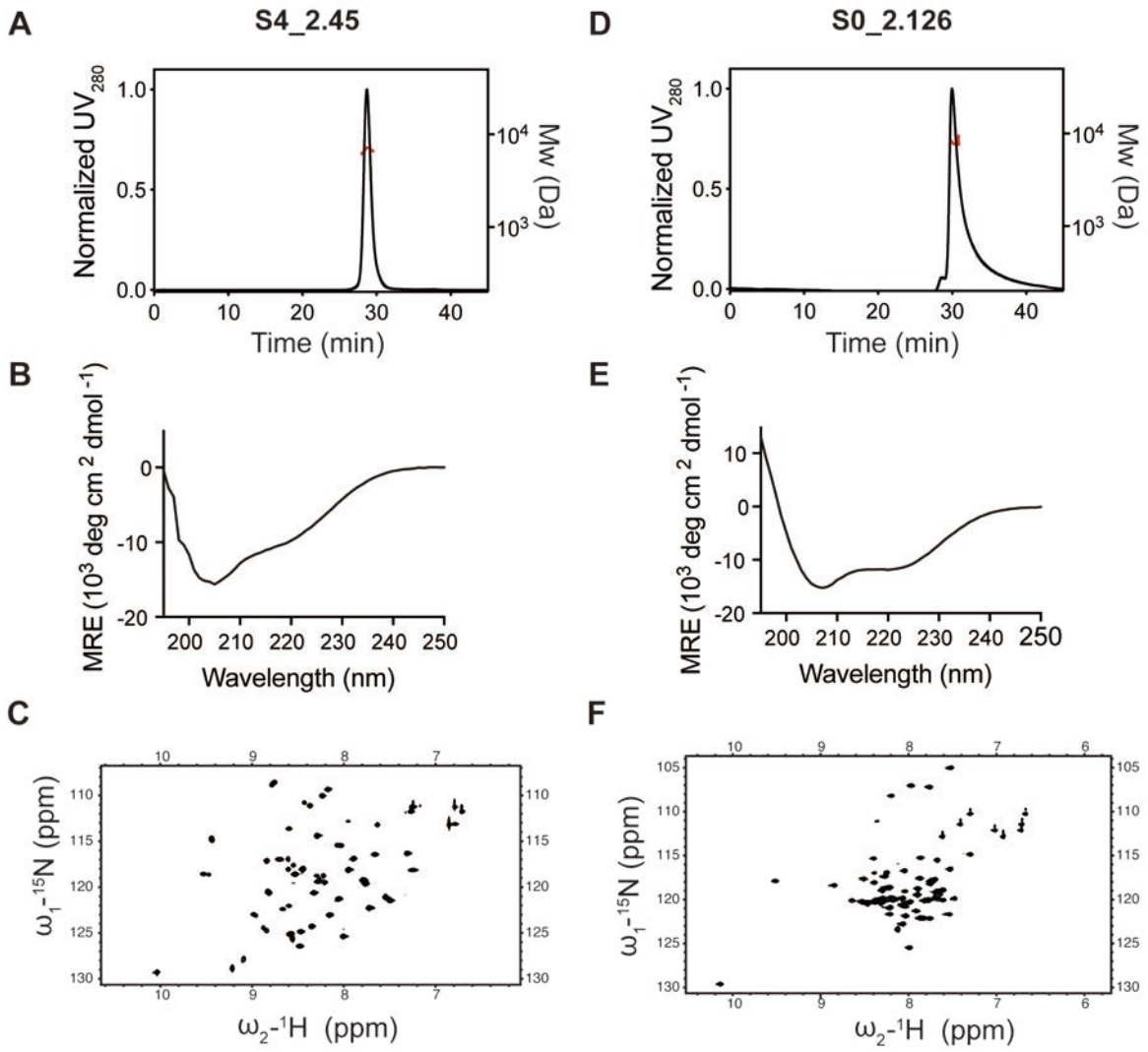


Supplementary Figure 8
***De novo* backbone assembly for site IV immunogen.** The site IV epitope was stabilized with three antiparallel beta strands built *de novo*, and a helix packing in various orientations against this beta sheet (bb1-bb3). Each backbone was simulated in Rosetta *abinitio* simulations for its ability to fold into a low energy state that is close to the design model, indicating that S4_2_bb2 and bb3 have a stronger tendency to converge into the designed fold.

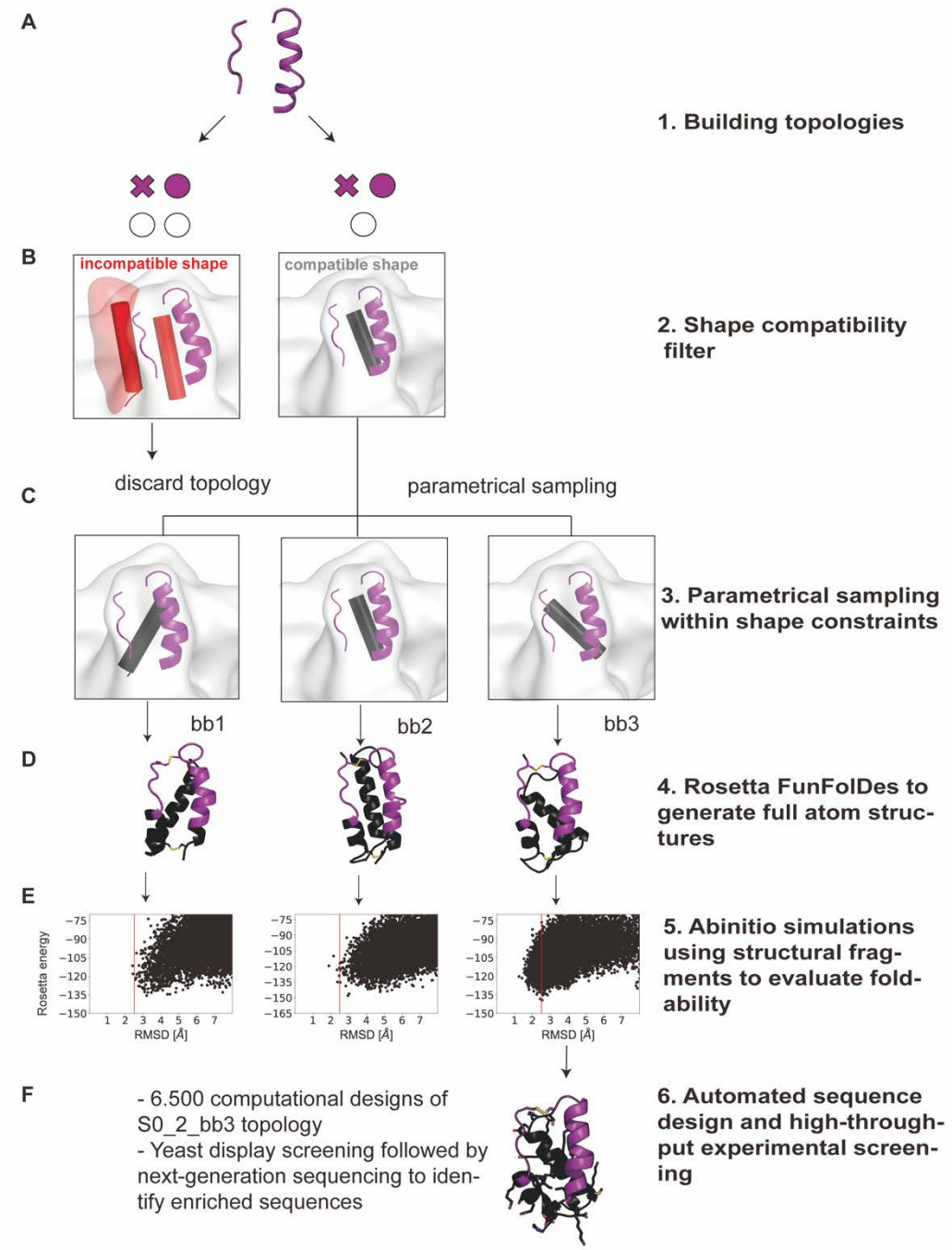


Supplementary Figure 9

Biophysical characterization of S4_2 design series. A) Circular dichroism spectra for 13 designs of the S4_2 design series that were enriched for protease resistance and binding to 101F in the yeast display selection assay. Black: spectrum at 20 ºC. Red: spectrum at 90 ºC. B) SPR sensorgrams for binding to 101F for the same designs. 101F IgG was immobilized on the sensor chip surface, and the designs were flown as analyte.



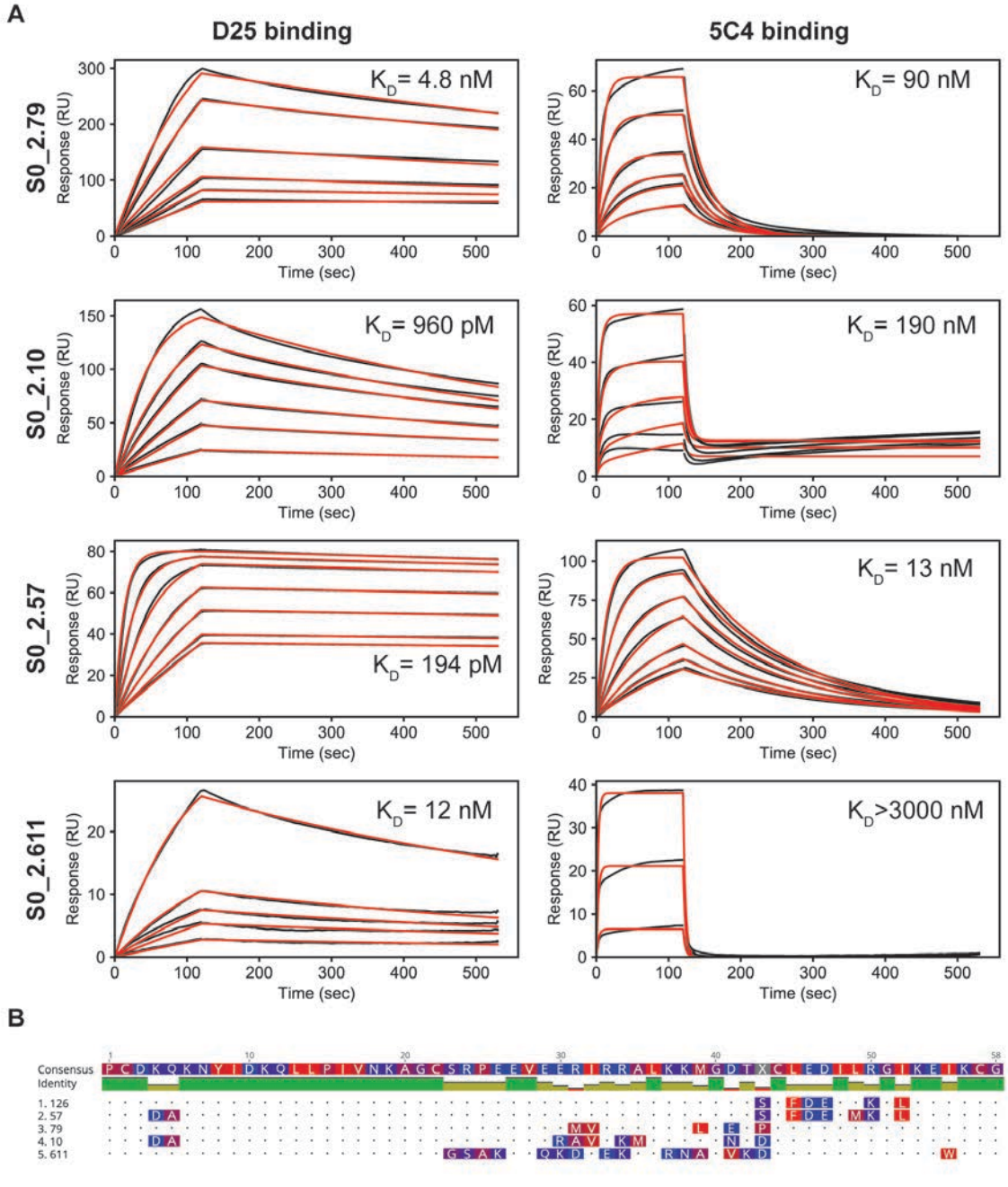
Supplementary Figure 10
Biophysical characterization of S4_2.45 (A-C) and S0_2.126 (D-F). A,D: S4_2.45 (A) and S0_2.126 (D) are monomeric in solution as shown by SEC-MALS profile. B,E: Circular dichroism spectra at 25 ºC. C,F: 2D NMR of ^{15}N HSQC spectra for S4_2.45 (C) and S0_2.126 (F) are well dispersed, confirming that the designs are well folded in solution. Mw: Molecular Weight.



Supplementary Figure 11

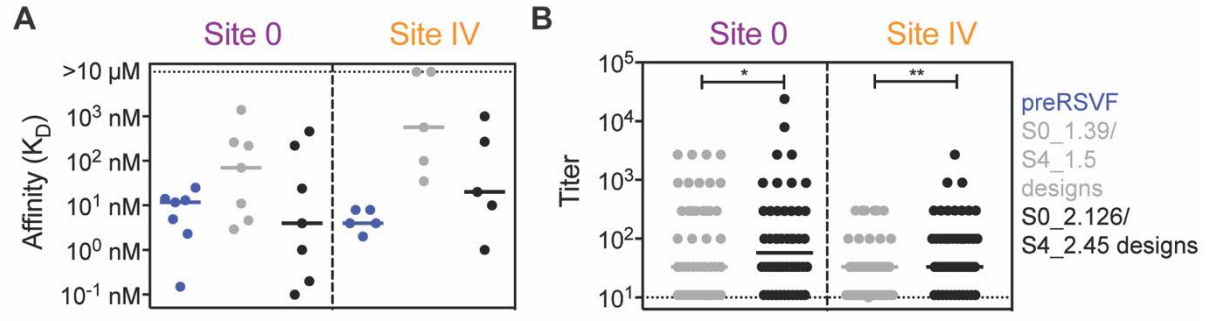
***De novo* topology assembly to stabilize site 0 using TopoBuilder.** A) Topologies that allow to connect the discontinuous site 0 motif are built as 2D form, and then translated into a 3D sketch. B) Generated sketches are evaluated for their compatibility with the shape constraints of prefusion RSVF. A compatible shape (right) respects the quaternary constraints of the motif in its native environment, whereas an incompatible shape (left) is one that sterically constrains the motif's accessibility in a non-native manner (red surface). C) Once a compatible topology is identified, the secondary structure is parametrically sampled (3 different helical orientations, S0_2_bb1-bb3). D) Three customized helical orientations were assembled (S0_2_bb1-bb3) to support site 0 epitope, and evaluated for their ability to fold into the designed topology in Rosetta *abinitio* simulations (E). S0_2_bb3 showed a funnel-shaped energy

landscape, and was selected for subsequent automated sequence design and experimental screening (F).

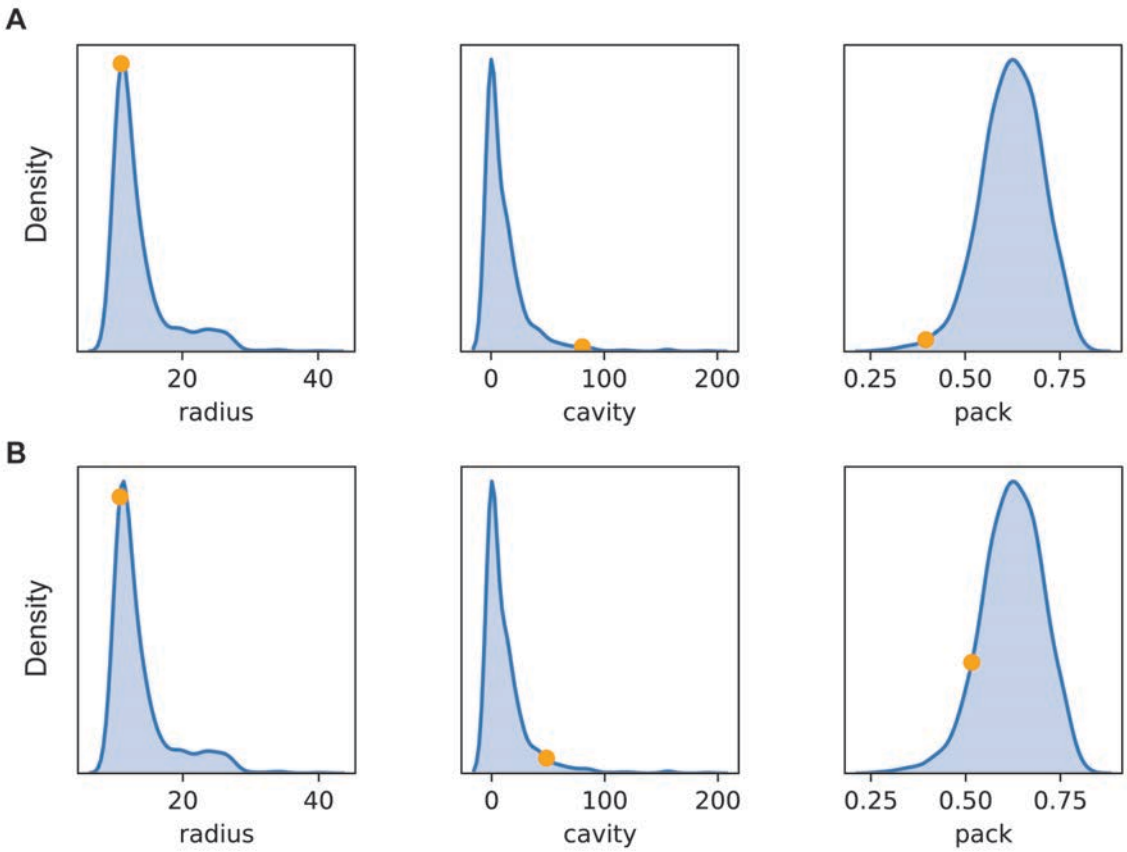


Supplementary Figure 12

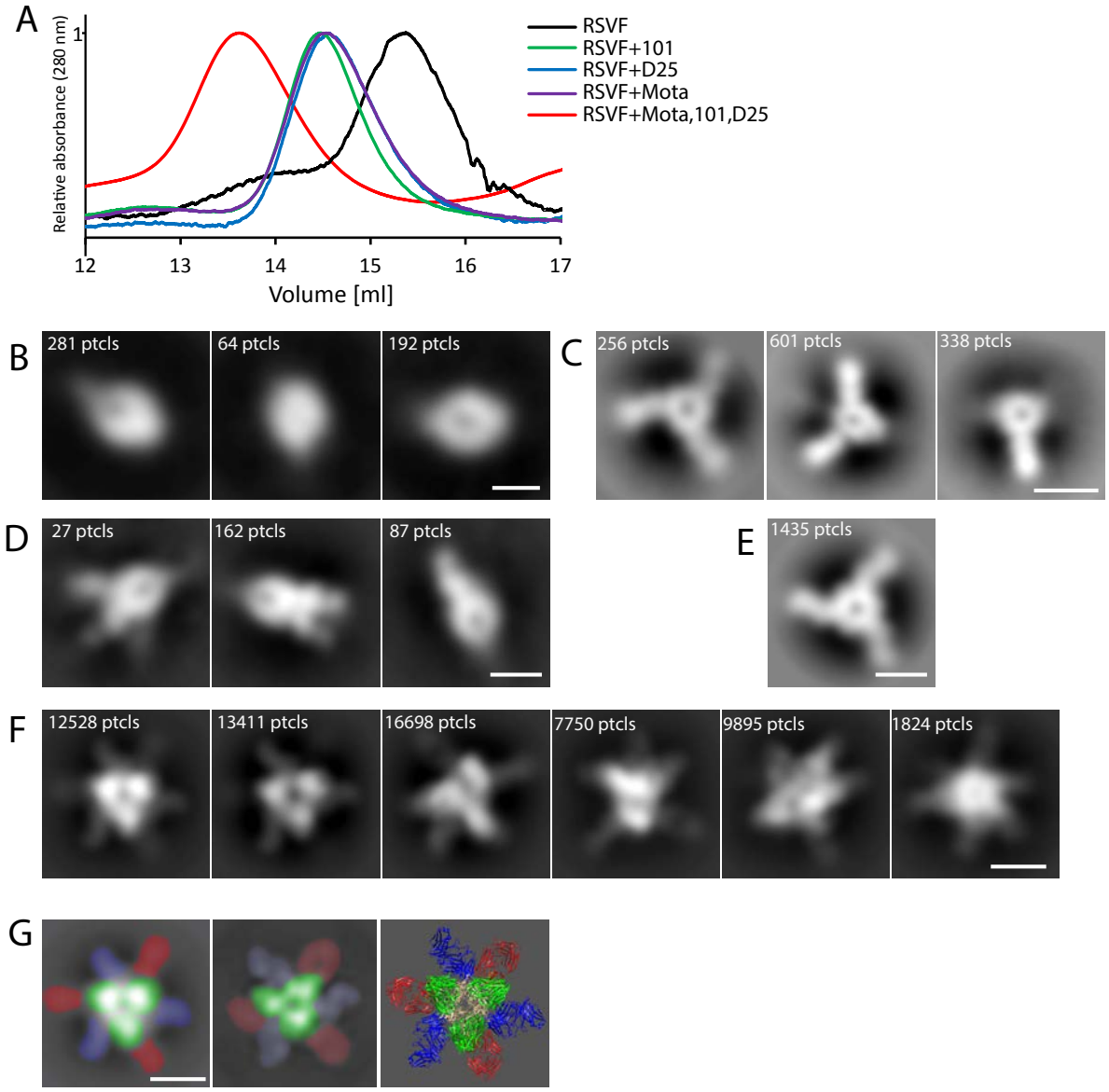
Biophysical characterization of S0_2 design series. (A) Binding affinity measurement for D25 and 5C4 binding of *de novo* site 0 scaffolds. Shown are the SPR sensorgrams of enriched designs that were successfully expressed and purified after the yeast display selection. D25 or 5C4 IgG were immobilized on the sensor chip surface, and scaffolds were injected as analyte. Dissociation constants shown are kinetic fits using a 1:1 Langmuir model. (B) Sequence alignment of experimentally characterized sequences, in comparison to S0_2.126. The closest sequence homolog to S0_2.126 is S0_2.57, differing in 3 amino acids.



Supplementary Figure 13
Binding affinity of designed immunogens towards panels of site-specific, human neutralizing antibodies and human sera.
A) Binding affinity (K_D, determined by SPR flowing Fabs as analyte) of S0_1.39 (grey) and S0_2.126 (black) towards a diverse panel of site-specific neutralizing antibodies, in comparison to prefusion RSVF (blue). Antibodies shown for site 0 are 5C4, D25 (*3*), ADI-14496, ADI-18916, ADI-15602, ADI-18900 and ADI-19009 (*4*). For site IV, the binding affinity was tested against 101F (*5*), ADI-15600 (*4*), 17E10, 6F18 and 2N6 (*6*), comparing S4_1.5 (grey) and S4_2.45 (black) to prefusion RSVF. The higher binding affinity of the second-generation designs (S0_2.126 and S4_2.45) compared to the first-generation and to prefusion RSVF indicates a greatly improved, near-native epitope mimicry of the respective antigenic sites in the designed immunogens.
B) ELISA reactivity of designed immunogens with sera obtained from 50 healthy human adults that were seropositive for prefusion RSVF. Both S0_2.126 and S4_2.45 showed significantly increased reactivity compared to the first-generation designs, confirming an improved epitope-mimicry on the serum level (* $p< 0.05$ and ** $p< 0.01$, Wilcoxon test). Data are representative from one out of two independent experiments.

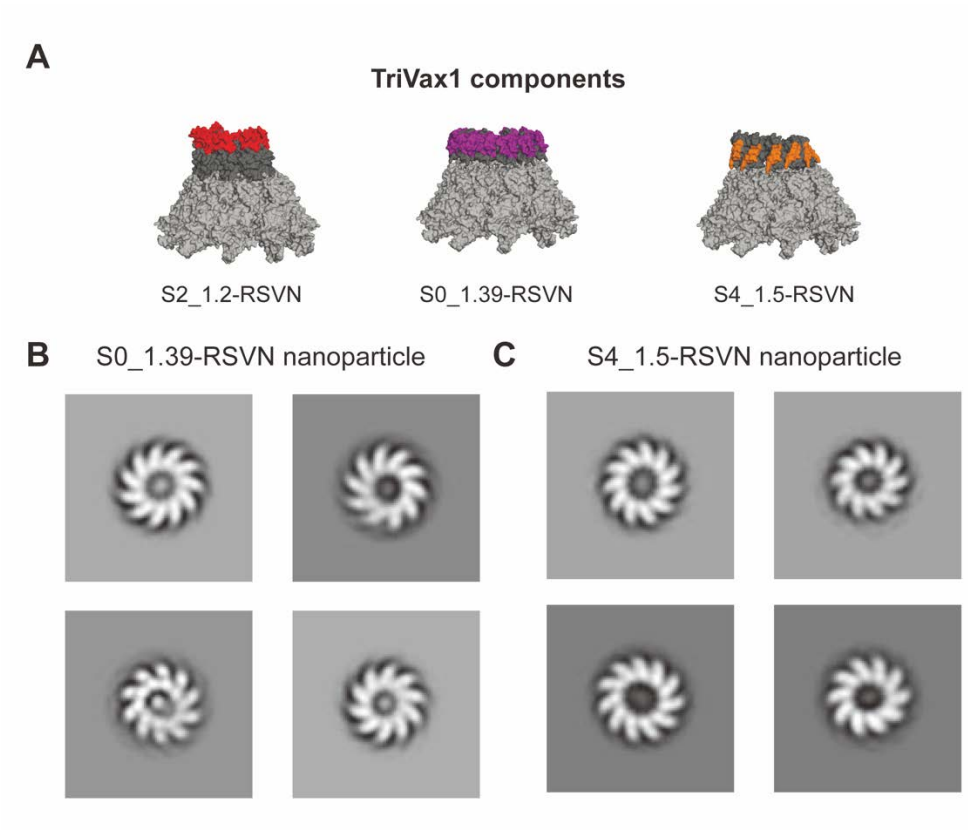


Supplementary Figure 14
Comparison of S0_2.126 Rosetta scores against natural proteins of similar size. Protein structures within the same size as S0_2.126 (57 +/- 5 residues) were downloaded from the CATH database and filtered by 70 % sequence homology, yielding a representative database of natural proteins with similar size as S0_2.126 (n = 1,013 structures). Proteins were then minimized and scored by Rosetta to compute their radius of gyration, intra-protein cavities (cavity) and core packing (packstat). Plotted is the distribution for these score terms in 1,013 natural proteins (blue density plot), and the same scores for S0_2.126 are shown in orange. The NMR structure of S0_2.126 is shown in (A), the computational model of S0_2.126 is shown in (B), indicating that, despite similar radius of gyration, S0_2.126 shows a substantial cavity volume as well as a very low core packing compared to natural proteins of similar size. CATH database and scores were pre-calculated, loaded and visualized using the rstoolbox python library (*7*).

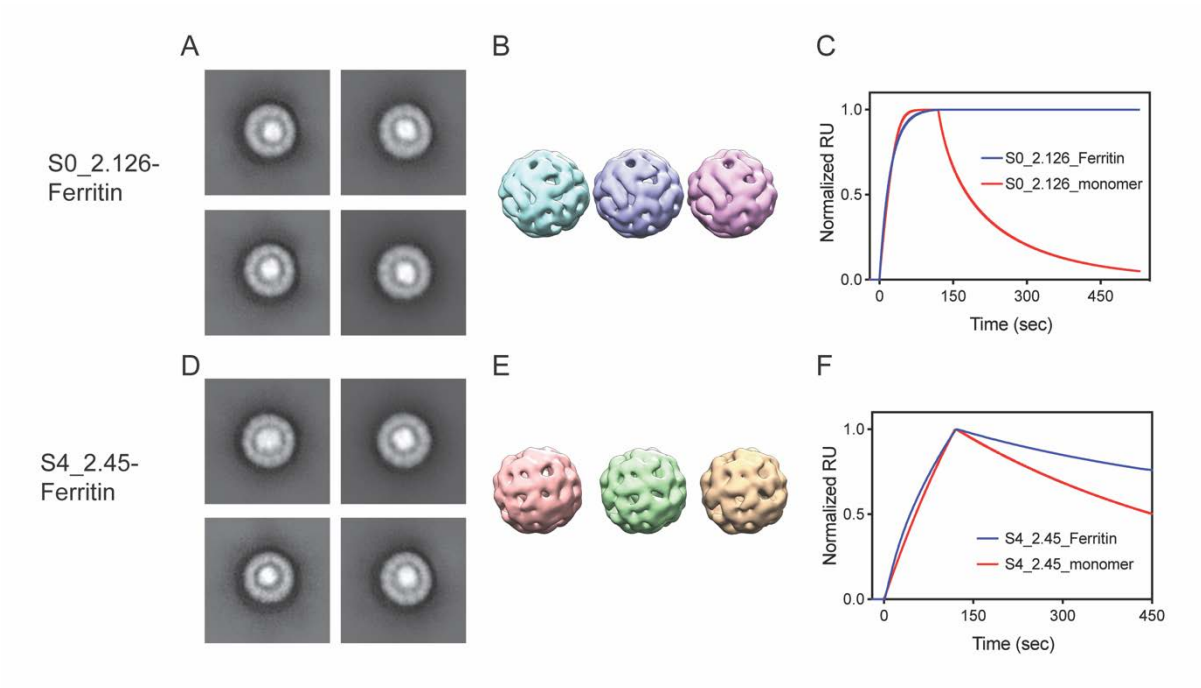


Supplementary Figure 15

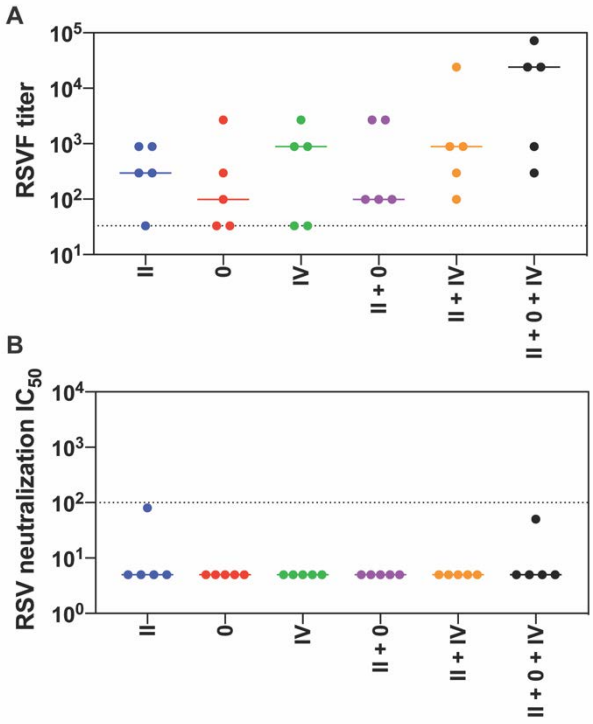
Electron microscopy analysis of site-specific antibodies in complex with RSVF trimer. (A) Superposed size-exclusion profiles of unliganded RSVF (black line) and RSVF in complex with 101F (green line), D25 (blue line), Mota (purple line) and all three (101F, D25, Mota - red line) Fabs. (B-F) Representative reference-free 2D class averages of the unliganded RSVF trimer (B) and RSVF in complex with 101F (C), D25 (D), Mota (E) or all three (101F, D25, Mota (F)) Fabs. Fully-saturated RSVF trimers bound by Fabs are observed, as well as sub-stoichiometric classes. (G) Left panel: reference-free 2D class average of RSVF trimer with three copies of 101F, D25 and Mota Fabs visibly bound. The predicted structure of RSVF in complex with 101F, D25 and Mota was used to simulate 2D class averages in Cryosparc2, and simulated 2D class average with all three types of Fabs is shown in the middle panel. Right panel: predicted structure of RSVF trimer with bound 101F, D25 and Mota Fabs based on the existing structures of RSVF with individual Fabs (PDB ID 4JHW, 3QWO and 3O45). Fabs are colored as follow: red - 101F; blue - Mota; green - D25. Scale bar - 100 Å.



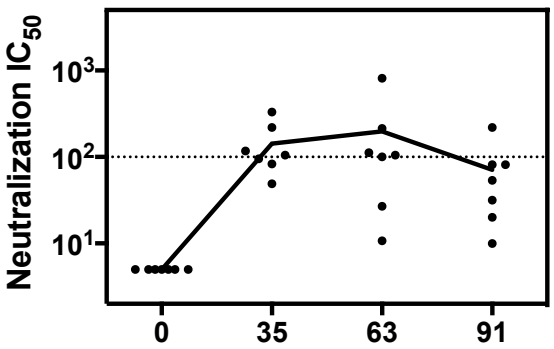
Supplementary Figure 16
Composition and EM analysis of Trivax1 RSVN nanoparticles.
A) Trivax1 contains equimolar amounts of site II, 0 and IV epitope focused immunogens fused to the self-assembling RSVN nanoparticle with a ring-like structure (n = 10-11 subunits). The site II-RSVN nanoparticle has been described previously (*8*). Shown are the computational models for the nanoparticles-immunogen fusion proteins.
B,C) Negative stain electron microscopy for S0_1.39-RSVN and S4_1.5-RSVN nanoparticles confirms that the ring-like structure is maintained upon fusion of the designed immunogens.



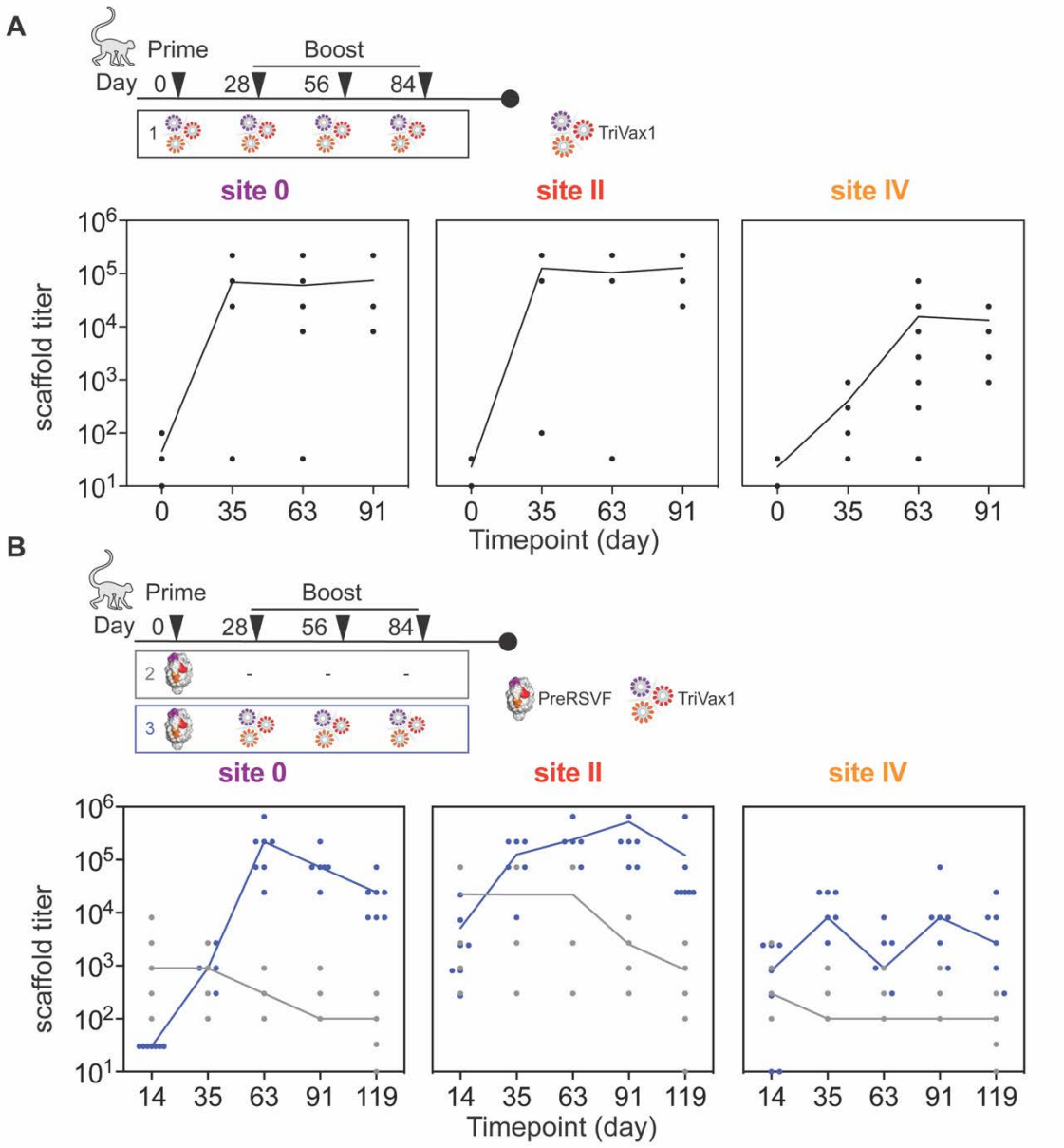
Supplementary Figure 17
EM analysis of Trivax2 ferritin nanoparticles.
A,B,D,E) Negative stain electron microscopy (A,D) and 3D reconstruction (B,E) for S0_2.126 and S4_2.45 fused to ferritin nanoparticles. C) Binding affinity of S0_2.126 nanoparticle (blue) to 5C4 antibody in comparison to S0_2.126 monomer (red), showing that S0_2.126 has been successfully multimerized and antibody binding sites are accessible. F) Binding of S4_2.45 to 101F antibody when multimerized on ferritin nanoparticle (blue) compared to monomeric S4_2.45 (red), indicating that the scaffold is multimerized and the epitope is accessible for antibody binding.



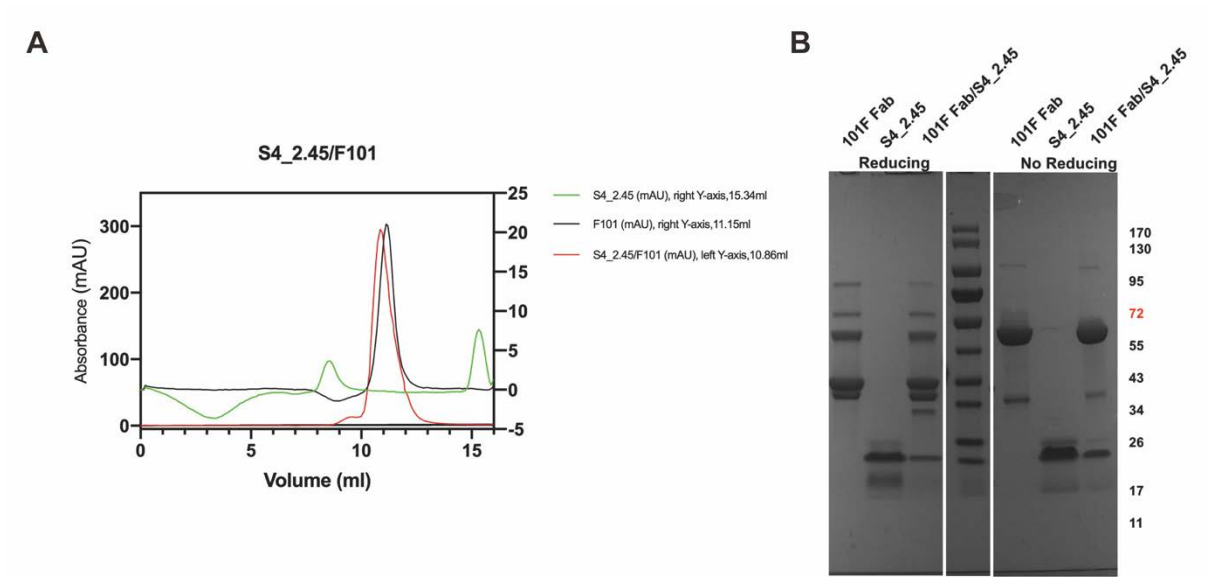
Supplementary Figure 18
Mouse immunization studies with Trivax1.
A) RSVF cross-reactivity of epitope-focused immunogens formulated individually, as cocktail of two, and three (Trivax1).
B) RSV neutralizing serum titer of mice immunized with designed immunogens and combinations thereof.



Supplementary Figure 19
Confirmation of NHP neutralization titer by an independent laboratory.
Sera from indicated timepoints were tested for RSV neutralization by an independent laboratory in a different RSV neutralization assay, using a Vero-118 cell line and a GFP readout. See (*9*) for method details.



Supplementary Figure 20
NHP serum reactivity with designed immunogens.
A) ELISA titer of NHP group 1 (immunized with Trivax1) measured at different timepoints. All animals responded to Trivax1 immunogens at day 91, with site IV immunogen reactivity lower compared to site 0 and site II reactivity.
B) ELISA titer of NHP group 2 (grey, RSVF prime) and 3 (blue, RSVF prime, Trivax1 boost). Following the priming immunization, all animals developed detectable cross-reactivity with the designed immunogens, indicating that the designed scaffolds recognized relevant antibodies primed by RSVF.



Supplementary Figure 21
Purification of protein complex of S4_2.45/101F Fab
A) The size exclusion chromatograph of individual components (Fab and scaffold) and the protein complex.
B) SDS gel of purified protein complex shows the collected fractions contain both scaffold and Fab.

1. J. Zhou, G. Grigoryan, Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci* **24**, 508-524 (2015).
2. J. Bonet *et al.*, Rosetta FunFolDes - A general framework for the computational design of functional proteins. *PLoS Comput Biol* **14**, e1006623 (2018).
3. J. S. McLellan *et al.*, Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113-1117 (2013).
4. M. S. Gilman *et al.*, Rapid profiling of RSV antibody repertoires from the memory B cells of naturally infected adult donors. *Sci Immunol* **1**, (2016).
5. J. S. McLellan *et al.*, Structure of a major antigenic site on the respiratory syncytial virus fusion glycoprotein in complex with neutralizing antibody 101F. *J Virol* **84**, 12236-12244 (2010).
6. J. J. Mousa *et al.*, Human antibody recognition of antigenic site IV on Pneumovirus fusion proteins. *PLoS Pathog* **14**, e1006837 (2018).
7. J. Bonet, Z. Harteveld, F. Sesterhenn, A. Scheck, B. E. Correia, rstoolbox - a Python library for large-scale analysis of computational protein design data and structural bioinformatics. *BMC Bioinformatics* **20**, 240 (2019).
8. F. Sesterhenn *et al.*, Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen. *PLoS Biol* **17**, e3000164 (2019).
9. E. Olmedillas *et al.*, Chimeric Pneumoviridae fusion proteins as immunogens to induce cross-neutralizing antibody responses. *EMBO Mol Med* **10**, 175-187 (2018).