Structure based stabilization of native-like antigens with deep learning

Naveen Jasti

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Neil King, Chair

Mary-Claire King

Andrew McGuire

Kelly Lee

Program Authorized to Offer Degree:

Molecular Engineering

University of Washington

Abstract

Structure based stabilization of native-like antigens with deep learning

Naveen Jasti

Chair of the Supervisory Committee:

Neil King

Department of Biochemistry

Effective vaccines prevent illness and death by stimulating protective immune responses against infectious pathogens. Protective immune responses can recognize and neutralize antigens that are important proteins used by pathogens to infect and replicate. However, these key proteins often evolve multiple conformations or instability to infect and evade the host immune system. Structure-guided design has advanced vaccine development by introducing mutations that stabilize these proteins to elicit stronger immune responses. Deep learning based methods have transformed our ability to predict and design protein structures. In this work, I investigated how to best apply these novel protein design methods to the stabilization of native-like antigens. I focused on four pathogens that cause significant morbidity and mortality: human rotavirus,

Group A *Streptococcus*, *Mycobacterium tuberculosis*, and Rabies virus. For each pathogen, I chose key proteins that are compelling vaccine targets and present distinct challenges for antigen stabilization. By pairing existing methods with novel deep learning based tools, I identified mutations that improve antigen stability and immunogenicity. In the process, I have identified principles which may be applicable for structure guided design across a wide range of antigens.

Introduction

Vaccines remain one of the most effective public health measures for reducing morbidity and mortality[1,2]. Vaccines are estimated to avert 4.8 million deaths per year, and since the rollout of the WHO's Expanded Program on Immunization in 1974 are thought to have been responsible for 40% of the reduction in infant mortality observed over the last 50 years[3,4]. Vaccines offer protection by eliciting an adaptive immune response to prevent infection or illness during subsequent encounters with the target pathogen[5,6]. While immune correlates of protection are specific to each pathogen, serum antibody titers are a clear and consistent correlate across most vaccines[7]. Historically, many vaccines have been derived from attenuated versions of live pathogens, which in some cases provide lifelong protection, but may present risks associated with reassortment or reversion to virulence[8,9]. Many modern vaccines consist of non-infectious protein or carbohydrate subunits of pathogens, due to their improved safety profiles and the ability to precisely engineer the subunits to elicit protective immune responses[6,10].

A central requirement of eliciting functional antibodies with subunit vaccines is the ability to manufacture an antigen that is stable in a biologically and immunologically relevant conformation[11,12]. However, pathogen proteins have been selected for pathogen fitness, not optimal stability, and many cannot be produced recombinantly or are unstable when taken out of their native context. Introducing targeted mutations that stabilize these proteins in native-like conformations can improve manufacturability, shelf-life stability, and the potency of vaccine-elicited antibody responses[13–24]. For example, stabilization of the class I fusion proteins from RSV and SARS-CoV-2 in the prefusion conformation significantly improved the titers of neutralizing antibodies, a main correlate of protection,

elicited by vaccination and enabled the development of clinically approved vaccines[25–36]. Improved potency of neutralizing antibody responses have also been observed with stabilized fusion proteins of HIV and influenza[36–39]. Mutations that improve the biophysical stability of an antigen can increase the yield of recombinantly produced antigens for subunit vaccines. Stabilized variants of the malarial antigens Rh5 and Pfs48/45 display more than 50- and 30-fold improvements in yield, respectively[40,41]. In addition, stabilization might extend the time that antigen remains intact in germinal centers, which can improve the quality and magnitude of antibody responses[42,43]. Vaccine development against a number of persistent and emerging pathogens would be facilitated by antigen design, but the complexity of protein structures and energetics make it challenging to identify mutations that stabilize antigens in native-like conformations[16,44].

Numerous structure-based vaccine design efforts provide examples of manual and computational approaches to stabilize native-like antigens. Many efforts have involved identifying potentially stabilizing mutations from manual inspection of high-resolution structures[13,26,41]. These mutations are individually tested and then iteratively combined to select a stabilized antigen[15,27,45]. This approach has identified prolines that disfavor alternate conformations, disulfides that rigidify conformationally dynamic regions, and mutations that create additional noncovalent packing or electrostatic interactions as successful and generalizable strategies for the prefusion stabilization of class I viral fusion proteins[46]. Meanwhile, a computational approach pairing consensus mutations from multiple sequence alignments (MSAs) with previously identified mutations

successfully stabilized fusion proteins from diverse strains of HIV and influenza[13,47,48]. Finally, computational protein design with tools such as Rosetta has also been used to improve the stability and immunogenicity of a number of antigens, either in isolation or paired with information from MSAs[49–52]. While these successes highlight the utility of structure-guided design, many antigens from devastating pathogens have not yet been stabilized, precluding the development of safe and effective subunit vaccines.

Considerable advances over the last four years in modeling protein sequence-structure relationships provide new opportunities for designing efficacious vaccine candidates[53–57]. For example, AlphaFold (AF) and other methods have enabled significant advances in protein structure prediction[58–60], yielding high-confidence models for 500,000 sequences which lack experimental structures and capturing multiple conformations of metastable proteins with subsampled MSAs or custom structural templates[61–64]. Sequence design networks such as ProteinMPNN have successfully designed amino acid sequences encoding soluble and highly thermostable de novo proteins with sub-Ångstrom accuracy[65,66]. Meanwhile, redesign of native proteins has enhanced the soluble yield and thermostability of transporter proteins, enzymes, and membrane protein analogs[67–69]. Finally, generative methods have enabled the design of novel protein backbones for protein binders, oligomers, and scaffolds that present epitopes from target antigens[70–73]. Recent efforts have begun to employ such deep learning methods to stabilize class I fusion proteins, building on extensive prior work[74,75]. Similarly, stabilization of the Langya virus G protein and the prefusion conformation of class III viral fusion proteins from multiple herpesviruses have also been reported[76,77].

However, the potential of deep learning methods to stabilize antigens and improve the immunogenicity of subunit vaccines has not yet been surveyed across a range of vaccine design challenges.

We set out to assess how recent advances in deep learning methods for protein design could be effective across diverse antigens and structure-based design challenges that have to date precluded the development of subunit vaccine candidates. We focused on antigens from human rotavirus (HRV), Group A Strep (GAS), *Mycobacterium tuberculosis* (Mtb), and Rabies virus (RABV), because these pathogens continue to cause a substantial burden of disease globally. Our efforts resulted in six stabilized antigens with a range of improved properties from these four pathogens, and uncovered design principles that may be broadly useful in structure-based antigen design.

References

1. Rodrigues, C. M. C. & Plotkin, S. A. Impact of Vaccines; Health, Economic and Social Perspectives. *Front. Microbiol.* **11**, 1526 (2020).
2. Plotkin, S. A., Orenstein, W. A. & Mortimer, E. A. *Vaccines*. (Saunders, 1999).
3. Shattock, A. J. *et al.* Contribution of vaccination to improved survival and health: modelling 50 years of the Expanded Programme on Immunization. *Lancet* **403**, 2307–2316 (2024).
4. Carter, A. *et al.* Modeling the impact of vaccination for the immunization Agenda 2030: Deaths averted due to vaccination against 14 pathogens in 194 countries from 2021 to 2030. *Vaccine* (2023) doi:10.1016/j.vaccine.2023.07.033.
5. Leo, O., Cunningham, A. & Stern, P. L. Vaccine immunology. *Perspect. Vaccinol.* **1**, 25–59 (2011).
6. Pollard, A. J. & Bijker, E. M. A guide to vaccinology: from basic principles to new developments. *Nat. Rev. Immunol.* **21**, 83–100 (2021).
7. Burton, D. R. Antibodies, viruses and vaccines. *Nat. Rev. Immunol.* **2**, 706–713 (2002).
8. Ghattas, M., Dwivedi, G., Lavertu, M. & Alameh, M.-G. Vaccine Technologies and Platforms for Infectious Diseases: Current Progress, Challenges, and Opportunities. *Vaccines (Basel)* **9**, (2021).
9. Gotuzzo, E., Yactayo, S. & Córdova, E. Efficacy and duration of immunity after yellow fever vaccination: systematic review on the need for a booster every 10 years. *Am. J. Trop. Med. Hyg.* **89**, 434–444 (2013).
10. Moyle, P. M. & Toth, I. Modern subunit vaccines: development, components, and

research opportunities. *ChemMedChem* **8**, 360–376 (2013).

11. Abbas, A. K., Lichtman, A. H. & Pillai, S. *Cellular and Molecular Immunology E-Book: Cellular and Molecular Immunology E-Book*. (Elsevier Health Sciences, 2014).
12. Park, J. & Champion, J. A. Effect of antigen structure in subunit vaccine nanoparticles on humoral immune responses. *ACS Biomater. Sci. Eng.* **9**, 1296–1306 (2023).
13. Milder, F. J. *et al.* Universal stabilization of the influenza hemagglutinin by structure-based redesign of the pH switch regions. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2115379119 (2022).
14. Rutten, L. *et al.* Structure-based design of prefusion-stabilized filovirus glycoprotein trimers. *Cell Rep.* **30**, 4540–4550.e3 (2020).
15. Sponholtz, M. R. *et al.* Structure-based design of a soluble human cytomegalovirus glycoprotein B antigen stabilized in a prefusion-like conformation. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2404250121 (2024).
16. Graham, B. S., Gilman, M. S. A. & McLellan, J. S. Structure-Based Vaccine Antigen Design. *Annu. Rev. Med.* **70**, 91–104 (2019).
17. Moon-Walker, A. *et al.* Structural basis for antibody-mediated neutralization of lymphocytic choriomeningitis virus. *Cell Chem Biol* **30**, 403–411.e4 (2023).
18. Hsieh, C.-L. *et al.* Structure-based design of prefusion-stabilized human metapneumovirus fusion proteins. *Nat. Commun.* **13**, 1299 (2022).
19. Mishra, A. K. *et al.* Structural basis of synergistic neutralization of Crimean-Congo hemorrhagic fever virus by human antibodies. *Science* **375**, 104–109 (2022).

20. Hastie, K. M. *et al.* Structural basis for antibody-mediated neutralization of Lassa virus. *Science* **356**, 923–928 (2017).
21. Chan, Y.-P. *et al.* Biochemical, conformational, and immunogenic analysis of soluble trimeric forms of henipavirus fusion glycoproteins. *J. Virol.* **86**, 11457–11471 (2012).
22. Wong, J. J. W., Paterson, R. G., Lamb, R. A. & Jardetzky, T. S. Structure and stabilization of the Hendra virus F glycoprotein in its prefusion form. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1056–1061 (2016).
23. Loomis, R. J. *et al.* Structure-Based Design of Nipah Virus Vaccines: A Generalizable Approach to Paramyxovirus Immunogen Development. *Front. Immunol.* **11**, 842 (2020).
24. Langedijk, J. P. M. *et al.* Universal paramyxovirus vaccine design by stabilizing regions involved in structural transformation of the fusion protein. *Nat. Commun.* **15**, 4629 (2024).
25. McLellan, J. S. *et al.* Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113–1117 (2013).
26. McLellan, J. S. *et al.* Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* **342**, 592–598 (2013).
27. Krarup, A. *et al.* A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism. *Nat. Commun.* **6**, 8143 (2015).
28. Joyce, M. G. *et al.* Iterative structure-based improvement of a fusion-glycoprotein vaccine against RSV. *Nat. Struct. Mol. Biol.* **23**, 811–820 (2016).
29. Killikelly, A. M., Kanekiyo, M. & Graham, B. S. Pre-fusion F is absent on the surface

of formalin-inactivated respiratory syncytial virus. *Sci. Rep.* **6**, 34108 (2016).

30. Crank, M. C. *et al.* A proof of concept for structure-based vaccine design targeting RSV in humans. *Science* **365**, 505–509 (2019).
31. Chang, L. A. *et al.* A prefusion-stabilized RSV F subunit vaccine elicits B cell responses with greater breadth and potency than a postfusion F vaccine. *Sci. Transl. Med.* **14**, eade0424 (2022).
32. Papi, A. *et al.* Respiratory Syncytial Virus Prefusion F Protein Vaccine in Older Adults. *N. Engl. J. Med.* **388**, 595–608 (2023).
33. Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7348–E7357 (2017).
34. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
35. Corbett, K. S. *et al.* SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* **586**, 567–571 (2020).
36. Bong, Y.-S. *et al.* S6P mutation in Delta and Omicron variant spike protein significantly enhances the efficacy of mRNA COVID-19 vaccines. *Front. Immunol.* **15**, 1495561 (2024).
37. Yang, X., Wyatt, R. & Sodroski, J. Improved elicitation of neutralizing antibodies against primary human immunodeficiency viruses by soluble stabilized envelope glycoprotein trimers. *J. Virol.* **75**, 1165–1171 (2001).
38. Sanders, R. W. *et al.* A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not

non-neutralizing antibodies. *PLoS Pathog.* **9**, e1003618 (2013).

39. Sanders, R. W. *et al.* HIV-1 VACCINES. HIV-1 neutralizing antibodies induced by native-like envelope trimers. *Science* **349**, aac4223 (2015).
40. Weldon, W. C. *et al.* Enhanced immunogenicity of stabilized trimeric soluble influenza hemagglutinin. *PLoS One* **5**, (2010).
41. Campeotto, I. *et al.* One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 998–1002 (2017).
42. McLeod, B. *et al.* Vaccination with a structure-based stabilized version of malarial antigen Pfs48/45 elicits ultra-potent transmission-blocking antibody responses. *Immunity* **55**, 1680–1692.e8 (2022).
43. Tam, H. H. *et al.* Sustained antigen availability during germinal center initiation enhances antibody responses to vaccination. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6639–E6648 (2016).
44. Cirelli, K. M. *et al.* Slow Delivery Immunization Enhances HIV Neutralizing Antibody and Germinal Center Responses via Modulation of Immunodominance. *Cell* **177**, 1153–1171.e28 (2019).
45. Li, B., Fooksa, M., Heinze, S. & Meiler, J. Finding the needle in the haystack: towards solving the protein-folding problem computationally. *Crit. Rev. Biochem. Mol. Biol.* **53**, 1–28 (2018).
46. Stewart-Jones, G. B. E. *et al.* Structure-based design of a quadrivalent fusion glycoprotein vaccine for human parainfluenza virus types 1–4. *Proceedings of the National Academy of Sciences* **115**, 12265–12270 (2018).

47. Byrne, P. O. & McLellan, J. S. Principles and practical applications of structure-based vaccine design. *Curr. Opin. Immunol.* **77**, 102209 (2022).
48. Rutten, L. *et al.* A universal approach to optimize the folding and stability of prefusion-closed HIV-1 envelope trimers. *Cell Rep.* **23**, 584–595 (2018).
49. Rawi, R. *et al.* Automated design by structure-based stabilization and consensus repair to achieve prefusion-closed envelope trimers in a wide variety of HIV strains. *Cell Rep.* **33**, 108432 (2020).
50. Kudlacek, S. T. *et al.* Designed, highly expressing, thermostable dengue virus 2 envelope protein dimers elicit quaternary epitope antibodies. *Sci Adv* **7**, eabg4084 (2021).
51. Phan, T. T. N. *et al.* A conserved set of mutations for stabilizing soluble envelope protein dimers from dengue and Zika viruses to advance the development of subunit vaccines. *J. Biol. Chem.* **298**, 102079 (2022).
52. Gonzalez, K. J. *et al.* A general computational design strategy for stabilizing viral class I fusion proteins. *Nat. Commun.* **15**, 1–13 (2024).
53. Ellis, D. *et al.* Structure-based design of stabilized recombinant influenza neuraminidase tetramers. *Nat. Commun.* **13**, 1825 (2022).
54. Ovchinnikov, S. & Huang, P.-S. Structure-based protein design with deep learning. *Curr. Opin. Chem. Biol.* **65**, 136–144 (2021).
55. Khakzad, H. *et al.* A new age in protein design empowered by deep learning. *Cell Syst* **14**, 925–939 (2023).
56. Kortemme, T. De novo protein design-From new structures to programmable functions. *Cell* **187**, 526–544 (2024).

57. Chu, A. E., Lu, T. & Huang, P.-S. Sparks of function by de novo protein design. *Nat. Biotechnol.* **42**, 203–215 (2024).
58. Notin, P., Rollins, N., Gal, Y., Sander, C. & Marks, D. Machine learning for functional protein design. *Nat. Biotechnol.* **42**, 216–228 (2024).
59. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
60. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
61. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
62. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2021).
63. Del Alamo, D., Sala, D., Mchaourab, H. S. & Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *Elife* **11**, (2022).
64. Monteiro da Silva, G., Cui, J. Y., Dalgarno, D. C., Lisi, G. P. & Rubenstein, B. M. High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nat. Commun.* **15**, 2464 (2024).
65. Wayment-Steele, H. K. *et al.* Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 832–839 (2024).
66. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).

67. Akpinaroglu, D. *et al.* Structure-conditioned masked language models for protein sequence design generalize beyond the native sequence space. *bioRxiv* 2023.12.15.571823 (2023) doi:10.1101/2023.12.15.571823.
68. Goverde, C. A. *et al.* Computational design of soluble and functional membrane protein analogues. *Nature* **631**, 449–458 (2024).
69. Shroff, R. *et al.* Discovery of novel gain-of-function mutations guided by structure-based deep learning. *ACS Synth. Biol.* **9**, 2927–2935 (2020).
70. Sumida, K. H. *et al.* Improving Protein Expression, Stability, and Function with ProteinMPNN. *J. Am. Chem. Soc.* **146**, 2054–2061 (2024).
71. Anishchenko, I. *et al.* De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
72. Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).
73. Lisanza, S. L. *et al.* Multistate and functional protein design using RoseTTAFold sequence space diffusion. *Nat. Biotechnol.* (2024) doi:10.1038/s41587-024-02395-w.
74. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
75. Bakkers, M. J. G. *et al.* Efficacious human metapneumovirus vaccine based on AI-guided engineering of a closed prefusion trimer. *Nat. Commun.* **15**, 6270 (2024).
76. Nitesh Mishra, Gabriel Avillion, Sean Callaghan, et al. and Bryan Briney. Conformational ensemble-based framework enables rapid development of Lassa virus vaccine candidates. *bioRxivorg* (2024) doi:10.1101/2024.11.21.624760v1.

77. Wang, Z. *et al.* Structure and design of Langya virus glycoprotein antigens. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2314990121 (2024).
78. McCallum, M. & Veesler, D. Computational design of prefusion-stabilized Herpesvirus gB trimers. *bioRxivorg* (2024) doi:10.1101/2024.10.23.619923.

Rotavirus

Despite the availability of licensed vaccines, rotavirus, a double-stranded RNA virus in the family *Reoviridae*, remains the leading cause of diarrheal death among children under five years of age[1–3]. Oral live-attenuated vaccines for rotavirus display variable efficacy, from 80-90% in high-income countries to 40-60% in low-and-middle-income countries[4–6]. Although clear reasons for this difference in efficacy have not been demonstrated, multiple factors may contribute, including maternal antibodies in breast milk, enteric pathogen load and inflammation, and malnutrition[5,7]. A parenteral rotavirus vaccine could potentially circumvent these factors, complementing existing vaccines to improve vaccine efficacy in low- and middle-income countries.

Rotavirus possesses a non-enveloped, triple-layered capsid comprising the spike protein VP4, outer capsid protein VP7, middle capsid protein VP6, and inner capsid protein VP2[2,8]. VP4 is processed by host trypsin in the small intestine into the attachment protein VP8* and the cell entry protein VP5*, which perforates host membranes and mediates endosomal escape through a conformational change from an upright, asymmetric trimer to a reversed, symmetric trimer[9–14]. VP7 forms Ca^{2+}-dependent trimers which dissociate upon endosomal perforation to expose the double-layered capsid[15,16]. Both VP4 and VP7 are the targets of neutralizing antibodies[17–21], and passive transfer of serum IgG antibodies has been shown to protect non-human primates from rotavirus challenge[22]. Characterization of antibodies isolated from human intestinal B cells revealed that all targeting VP5* and most targeting VP7 neutralize multiple strains of rotavirus[23]. By contrast, antibodies targeting VP8* were predominantly non-neutralizing or neutralized only a single tested strain[23]. Efforts to

develop a parenteral, non-replicating rotavirus vaccine (NRRV) have focused on VP8* because it can be readily produced recombinantly, but the resulting NRRV vaccine does not provide sufficient protection in infants[24,25]. The non-neutralizing or strain-specific activity of antibodies targeting VP8* might contribute to NRRV's inability to elicit a protective immune response. While antibodies targeting VP5* neutralize multiple strains of rotavirus, the recombinant production of VP5* has been challenging, hindering its inclusion in a parenteral vaccine[24].

We sought to stabilize the symmetric dimer within the upright form of VP5*, as we reasoned that— analogous to the prefusion conformation of enveloped virus fusion proteins— the upright form would be the predominant conformation on virions prior to cell entry[26,27]. We focused on the antigenic fragment from amino acids 247-478 of the human rotavirus VP4 protein and subsequently use VP5* to refer to this portion of the protein[11](Fig 1-1A). As VP4 lacks glycans or disulfide bonds, we set out to express stabilized VP5* and VP4 antigens in *Escherichia coli* cells and purify them by immobilized metal affinity chromatography (IMAC; see Methods). We tested sequences from the attenuated G1P[8] rotavirus strain CDC-9, which has been shown to have a greater proportion of VP4 in the upright conformation on virions during *in vitro* culturing[27], and the attenuated G1P[8] Rotarix® strain, as the Rotarix® vaccine containing a single rotavirus strain displays comparable efficacy to Rotateq®, which contains five strains[4,28,29]. As we obtained higher yield with VP5* sequence from the Rotarix® strain and it contains fewer attenuating mutations, we maintained that as our base sequence in all further work (designated VP5*-base). However, despite optimizing

expression conditions and lysis buffer composition (see Methods), we obtained very low soluble yield of VP5*-base, consistent with previous work on recombinant expression of human rotavirus proteins in *E. coli*[24].

We reasoned that membrane interactions of the VP5* protein, and not simply instability of the protein, might cause its association with the insoluble fraction of *E. coli* during purification. If that interaction affected soluble yield, redesign to stabilize the protein would not resolve the issue. To abrogate these membrane interactions, we took two approaches: either directly redesigning the hydrophobic apex of VP5* or instead redesigning the VP8* protein to serve as a chaperone during expression and purification. The latter approach was motivated by the native role of VP8* in occluding the hydrophobic apex of VP5* until attachment to host cells[26]. To redesign the apex of VP5*, we selected four hydrophobic, solvent-exposed residues (284, 287, 333, and 391) and redesigned them with ProteinMPNN. We chose the most frequent amino acid identified across 200 ProteinMPNN-redesigned sequences, which was a serine at each position, and visually examined them to ensure they did not introduce potential steric clashes. All four mutations were incorporated into a single design (VP5*-apex). To redesign VP8* we kept the native interface between VP8* and VP5* and allowed ProteinMPNN to design the remaining positions. 200 designs were filtered by AlphaFold 2 (AF2) predictions, with the four sequences with the highest model confidence (pLDDT > 90) and structure most similar to native VP8* (backbone RMSD < 2 Å) chosen to test experimentally alongside native VP8*. Both approaches improved the soluble yield of VP5*, but redesign of VP8* as a chaperone allowed us to avoid mutations at an epitope

in VP5* that may be targeted by neutralizing antibodies. When analyzed by ELISA with a small panel of neutralizing antibodies (nAbs; VP5* specific mAbs #41, 2, 33, VP4 specific mAb 30, and VP8* specific mAbs 9 and 47)[23], VP5*-apex displayed reduced binding relative to VP5*-base. SDS-PAGE analysis of bicistronic constructs co-expressing native VP8* or several redesigned versions as chaperones indicated that a variant of VP8* with only 27% identity to the native sequence led to the highest purified yield of VP5* (Supp Fig 1-5). Importantly, both SDS-PAGE and ELISA analysis indicated that none of the VP8* remained after IMAC purification (Supp Fig 1-6). All subsequent purifications of VP5* involved use of the redesigned VP8*, named cVP8*-M5, as a chaperone.

Able to obtain enough protein to support more detailed characterization, we then set out to redesign VP5* to stabilize the upright dimer. We initially used the 3.5 Å cryo-EM structure of the human rotavirus CDC-9 virion with VP5* in the upright conformation, in complex with mAb #41, as our input structure (PDB ID 7UMS)[27]. The structure was aggressively redesigned with ProteinMPNN. As only the epitope of mAb #41 has been structurally characterized, with data on the positions of escape mutations for a limited number of other mAbs, we kept the apex, known epitopes, and neighboring residues intact (see Methods)(Fig 1-1B). Both VP5*-base and the redesigned sequences were analyzed by AF2 prediction using 7UMS as a structural template. AF2 predicted VP5*-base to fold into a dimer that shows deviation from the upright dimer within 7UMS (pLDDT = 84.5, backbone RMSD > 5 Å). Many of the ProteinMPNN-redesigned sequences were also predicted as dimers that aligned well with the cryo-EM structure,

but with higher model confidence (Fig 1-1C). We selected the 48 top-ranking designs across multiple AF2 models (pLDDT > 88, PAE <8.0, and backbone RMSD to 7UMS < 2.0 Å) for experimental characterization. These designs comprised an average of 45 mutations, which were a mix of mutations that predominantly alter packing in the core of VP5* and reduce solvent-exposed hydrophobic residues.

Following expression in *E. coli* and purification by IMAC, VP5*-base and the 48 designed antigens were screened by ELISA using mAbs 2, 33, and 41 to capture total reactivity, a measure of both soluble yield and antigenicity. Although VP5*-base bound all mAbs, none of the designed antigens displayed reactivity, indicating that the designed mutations likely resulted in misfolding. With the hypothesis that a single deleterious mutation introduced into the core of VP5* could disrupt folding, we next adopted a two-step design approach. In the first step, we determined the positions at which ProteinMPNN suggested mutations that appeared deleterious by manual inspection of the structure. In the second step, we fixed the native residues at these positions to preclude mutation and only allowed ProteinMPNN to redesign positions where it appeared to suggest beneficial or neutral mutations. The same filtering and experimental screening processes were used as before for this set of designs. Once again, none of 48 designed antigens displayed reactivity during ELISA.

Further inspection of available VP5* structures revealed a crystal structure of the simian rotavirus VP5* dimer, which has 83% amino acid identity to human G1P[8] strains, at 1.6 Å resolution (PDB ID 2B4H)[11]. The high resolution of the 2B4H structure provided a

potential explanation by revealing that an extensive network of ordered water molecules make hydrogen bonds to backbone polar atoms of VP5* (Fig 1-1D). Current deep learning methods for structure prediction and sequence design, including AF2 and ProteinMPNN, lack representations for solvent molecules[30,31]. Therefore, these methods do not account for solvent-mediated interactions that are likely key to the proper folding of VP5*, potentially explaining why our first two rounds of design were unsuccessful.

We therefore took a more conservative approach to stabilizing VP5*, utilizing multiple sequence alignments (MSAs) to identify positions where the amino acid in VP5*-base differs from the consensus. We constructed MSAs with 95% or higher identity to VP5*-base (MSA1) or 70% or higher identity to VP5*-base (MSA2). While all sequences in MSA1 were from human rotavirus, those in MSA2 were from rotaviruses with diverse hosts. Individual mutations identified from the MSAs were filtered by AF2 prediction and Rosetta energy evaluation (see Methods). A filtered set of 6 and 15 mutations from each MSA became the MSA1 and MSA2 mutations, respectively (Fig 1-2A). The mutations in these sets reduced solvent-exposed hydrophobics, improved surface electrostatic interactions, and altered packing of hydrophobic residues in the core of VP5*. Upon introduction of these mutations into VP5*-base, VP5*-MSA1 and VP5*-MSA2 both displayed substantially higher reactivity to our panel of VP5*-specific mAbs during ELISA of IMAC eluates.

We scaled up expression and purification of VP5*-MSA1 and VP5*-MSA2 so that we could compare them to VP5*-base in more detail. Size exclusion chromatography (SEC)

revealed that the two designs purified in higher yield than VP5*-base and adopted the intended dimeric state, whereas VP5*-base predominantly exists as a monomer (Fig 1-2B). VP5*-MSA1 and VP5*-MSA2 possessed 4.5 and 10.5 °C increases in melting temperature, respectively, as assessed by nanoDSF with SYPRO dye (Fig 1-2D). NanoDSF analysis indicated similar increases in the temperature of aggregation. Concentration normalized ELISAs revealed that VP5*-base displayed poor reactivity with all tested nAbs, suggesting that recombinantly purified VP5*-base does not preserve a native-like conformation. VP5*-MSA1 and VP5*-MSA2 possessed higher reactivity by ELISA with tested nAbs, indicating they likely preserve a native-like conformation. Of these designed antigens, VP5*-MSA1 displays higher reactivity, suggesting stabilizing mutations in the VP5*-MSA2 design may diminish antigenicity (Fig 1-3B). An MSA-based approach to stabilizing VP5* facilitates the purification of native-like VP5* dimers that possess improved yield, thermal stability, and ELISA reactivity.

We also designed VP4-like antigens that include the VP8* domain by altering the topology of VP4 to remove the trypsin-cleavable linker, instead connecting the C terminus of VP5*-base, VP5*-MSA1, or VP5*-MSA2 to residue 42 of VP8* by a tetraglycine linker (Fig 1-1A). As previous work has demonstrated that native VP8* displays sufficient soluble yield and thermal stability to be manufactured and evaluated clinically, we kept the native sequence of VP8*. SEC revealed that, unlike for VP5* antigens, the two designs purified in similar yield to VP4-like-base (Fig 1-2C). Analysis by nanoDSF with SYPRO dye revealed that VP4-like-MSA1 and VP4-like-MSA2

possessed 4.4 and 9.6 °C increases in melting temperature respectively (Fig 1-2E). The fact that the VP4-like antigens displayed the same gains in thermal stability and temperature of aggregation onset as the designed VP5* antigens suggests that the VP5* portion of the antigen was the less stable component and was not further stabilized by the presence of VP8*. While MSA-based stabilizing mutations were critical for preserving ELISA reactivity for VP5* designs, they had differing effects for VP4-like designs, with the nine additional mutations in VP4-like-MSA2 improving thermal stability but diminishing antigenicity. VP4-like-MSA1 had slightly lower antigenicity than VP4-like-base (Fig 1-3C).

We determined crystal structures of both VP4-like dimers at 1.78 and 2.70 Å resolution for VP4-like-MSA1 and-MSA2, respectively. The structures revealed close agreement across the VP5* dimer from PDB ID 7UMS, with backbone RMSDs of 1.10 and 1.14 Å, respectively (Fig 1-3A). Deviations did occur in the VP8* region of the complex relative to the cryoEM structure of 7UMS. As no mutations were introduced into VP8* and the structure of 7UMS was determined in complex with an antibody that contacts both VP5* and VP8*, it is unclear whether the observed deviations were due to absence of the antibody or to mutations introduced into VP5*. Together, these data demonstrate that MSA-based redesign of VP5* alone or in a VP4-like context resulted in upright dimer antigens that produced in higher yield and were more monodisperse and stable than native VP5* and VP4.

As we were now able to recombinantly produce both VP5* and VP4-like antigens, we wanted to examine their ability to elicit protective antibody responses. Groups of 10 mice were immunized with VP5* antigens, VP4-like antigens, or an irrelevant protein control, for a total of 7 groups (Fig 1-4A). Mice were immunized with 5 μg of antigen in Addavax at weeks 0 and 4, with terminal sera collected at week 6. Sera were evaluated for neutralizing activity against a vaccine-matched strain of attenuated rotavirus in cultured MA104 cells. The experiment allowed us to compare VP5* and VP4-like antigens, which possess trade-offs for immunogenicity. VP5* antigens only contain the portion of VP4 targeted by broadly neutralizing antibodies, while VP4-like antigens contain the full antibody accessible portion of the rotavirus spike and displayed higher ELISA reactivity with nAbs. Among mice immunized with VP5* designed antigens, only those immunized with VP5*-MSA2 possessed significantly higher titers post-immunization. Mice immunized with all of the VP4-like antigens possessed significantly higher neutralizing antibody titers, revealing the protection elicited by these antigens and their superior performance compared to VP5* antigens (Fig 1-4B).

References

1. Hallowell, B. D., Chavers, T., Parashar, U. & Tate, J. E. Global Estimates of Rotavirus Hospitalizations Among Children Below 5 Years in 2019 and Current and Projected Impacts of Rotavirus Vaccination. *J Pediatric Infect Dis Soc* **11**, 149–158 (2022).
2. Asensio-Cob, D., Rodríguez, J. M. & Luque, D. Rotavirus Particle Disassembly and Assembly In Vivo and In Vitro. *Viruses* **15**, (2023).
3. Ward, R. L. & Bernstein, D. I. Rotarix: a rotavirus vaccine for the world. *Clin. Infect. Dis.* **48**, 222–228 (2009).
4. Jonesteller, C. L., Burnett, E., Yen, C., Tate, J. E. & Parashar, U. D. Effectiveness of Rotavirus Vaccination: A Systematic Review of the First Decade of Global Postlicensure Data, 2006-2016. *Clin. Infect. Dis.* **65**, 840–850 (2017).
5. Desselberger, U. Differences of Rotavirus Vaccine Effectiveness by Country: Likely Causes and Contributing Factors. *Pathogens* **6**, (2017).
6. Varghese, T., Kang, G. & Steele, A. D. Understanding Rotavirus vaccine efficacy and effectiveness in countries with high child mortality. *Vaccines (Basel)* **10**, 346 (2022).
7. Parker, E. P. *et al.* Causes of impaired oral vaccine efficacy in developing countries. *Future Microbiol.* **13**, 97–118 (2018).
8. Crawford, S. E. *et al.* Characterization of virus-like particles produced by the expression of rotavirus capsid proteins in insect cells. *J. Virol.* **68**, 5945–5952 (1994).
9. Prasad, B. V., Burns, J. W., Marietta, E., Estes, M. K. & Chiu, W. Localization of VP4 neutralization sites in rotavirus by three-dimensional cryo-electron microscopy. *Nature* **343**, 476–479 (1990).
10. Gilbert, J. M. & Greenberg, H. B. Cleavage of rhesus rotavirus VP4 after arginine 247 is essential for rotavirus-like particle-induced fusion from without. *J. Virol.* **72**, 5323–5327 (1998).
11. Yoder, J. D. & Dormitzer, P. R. Alternative intermolecular contacts underlie the rotavirus VP5* two- to three-fold rearrangement. *EMBO J.* **25**, 1559–1568 (2006).
12. Denisova, E. *et al.* Rotavirus capsid protein VP5* permeabilizes membranes. *J. Virol.* **73**, 3147–3153 (1999).
13. Golantsova, N. E., Gorbunova, E. E. & Mackow, E. R. Discrete domains within the rotavirus VP5* direct peripheral membrane association and membrane permeability. *J. Virol.* **78**, 2037–2044 (2004).
14. de Sautu, M., Herrmann, T., Scanavachi, G., Jenni, S. & Harrison, S. C. The rotavirus VP5*/VP8* conformational transition permeabilizes membranes to Ca2. *PLoS Pathog.* **20**, e1011750 (2024).
15. Dormitzer, P. R., Greenberg, H. B. & Harrison, S. C. Purified recombinant rotavirus

VP7 forms soluble, calcium-dependent trimers. *Virology* **277**, 420–428 (2000).
16. Aoki, S. T. *et al.* Cross-linking of rotavirus outer capsid protein VP7 by antibodies or disulfides inhibits viral entry. *J. Virol.* **85**, 10509–10517 (2011).
17. Shaw, R. D., Vo, P. T., Offit, P. A., Coulson, B. S. & Greenberg, H. B. Antigenic mapping of the surface proteins of rhesus rotavirus. *Virology* **155**, 434–451 (1986).
18. Aoki, S. T. *et al.* Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. *Science* **324**, 1444–1447 (2009).
19. Kirkwood, C. D., Bishop, R. F. & Coulson, B. S. Human rotavirus VP4 contains strain-specific, serotype-specific and cross-reactive neutralization sites. *Arch. Virol.* **141**, 587–600 (1996).
20. Coulson, B. S. & Kirkwood, C. Relation of VP7 amino acid sequence to monoclonal antibody neutralization of rotavirus and rotavirus monotype. *J. Virol.* **65**, 5968–5974 (1991).
21. Gorrell, R. J. & Bishop, R. F. Homotypic and heterotypic serum neutralizing antibody response to rotavirus proteins following natural primary infection and reinfection in children. *J. Med. Virol.* **57**, 204–211 (1999).
22. Westerman, L. E., McClure, H. M., Jiang, B., Almond, J. W. & Glass, R. I. Serum IgG mediates mucosal immunity against rotavirus infection. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7268–7273 (2005).
23. Nair, N. *et al.* VP4- and VP7-specific antibodies mediate heterotypic immunity to rotavirus in humans. *Sci. Transl. Med.* **9**, (2017).
24. Wen, X. *et al.* Construction and characterization of human rotavirus recombinant VP8* subunit parenteral vaccine candidates. *Vaccine* **30**, 6121–6126 (2012).
25. Wen, X. *et al.* Inclusion of a universal tetanus toxoid CD4(+) T cell epitope P2 significantly enhanced the immunogenicity of recombinant rotavirus ΔVP8* subunit parenteral vaccines. *Vaccine* **32**, 4420–4427 (2014).
26. Herrmann, T. *et al.* Functional refolding of the penetration protein on a non-enveloped virus. *Nature* **590**, 666–670 (2021).
27. Jenni, S. *et al.* Rotavirus VP4 Epitope of a Broadly Neutralizing Human Antibody Defined by Its Structure Bound with an Attenuated-Strain Virion. *J. Virol.* **96**, e0062722 (2022).
28. Zeller, M. *et al.* Genetic analyses reveal differences in the VP7 and VP4 antigenic epitopes between human rotaviruses circulating in Belgium and rotaviruses in Rotarix and RotaTeq. *J. Clin. Microbiol.* **50**, 966–976 (2012).
29. Matthijnssens, J. *et al.* Molecular and biological characterization of the 5 human-bovine rotavirus (WC3)-based reassortant strains of the pentavalent rotavirus vaccine, RotaTeq. *Virology* **403**, 111–127 (2010).
30. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
31. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using

ProteinMPNN. *Science* **378**, 49–56 (2022).
32. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
33. Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
34. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res.* **53**, D20–D29 (2025).
35. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
36. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
37. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
38. The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC.

Figures

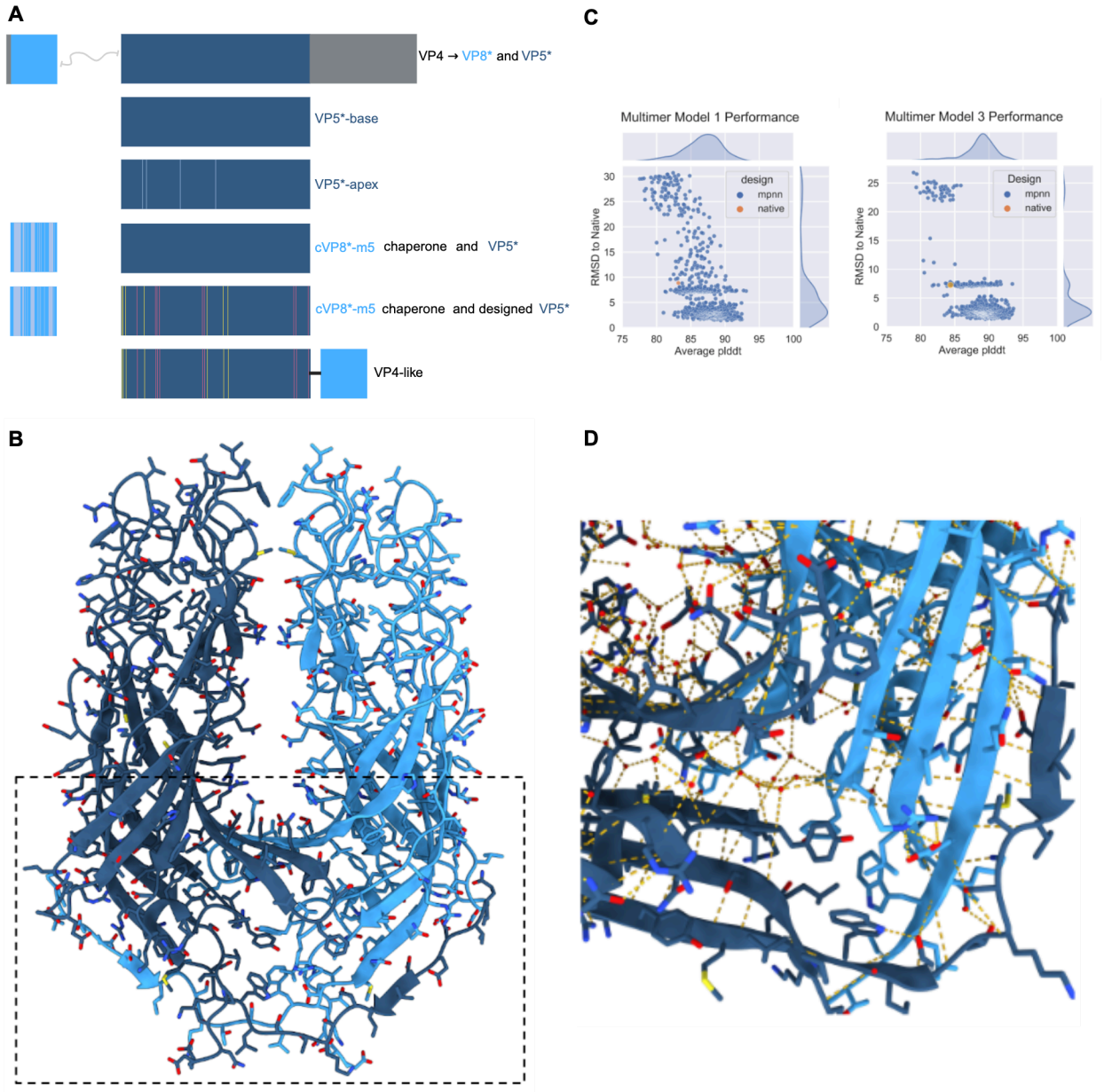


Figure 1-1. VP4 design scheme and ProteinMPNN redesign. A) Schematic of native VP4 proteolytically processed into VP5* and VP8*, the domain boundaries of VP5* antigens, introduction of mutations into VP5*, co-expression of a redesigned VP8* as a chaperone during purification, and the domain boundaries and topology of VP4-like antigens. B) Structure of the VP5* dimer that was the focus of design efforts. In the dashed box is the region redesigned with ProteinMPNN. C) Plots of AF2 prediction RMSD to native structure vs pLDDT for both models 1 and 3. Blue dots are individual ProteinMPNN-redesigned sequences, and orange dot is VP5*-base. D) Structure of simian rotavirus at 1.6 Å resolution (PDB ID 2B4H) reveals extensive interactions between the protein backbone and solvent molecules.

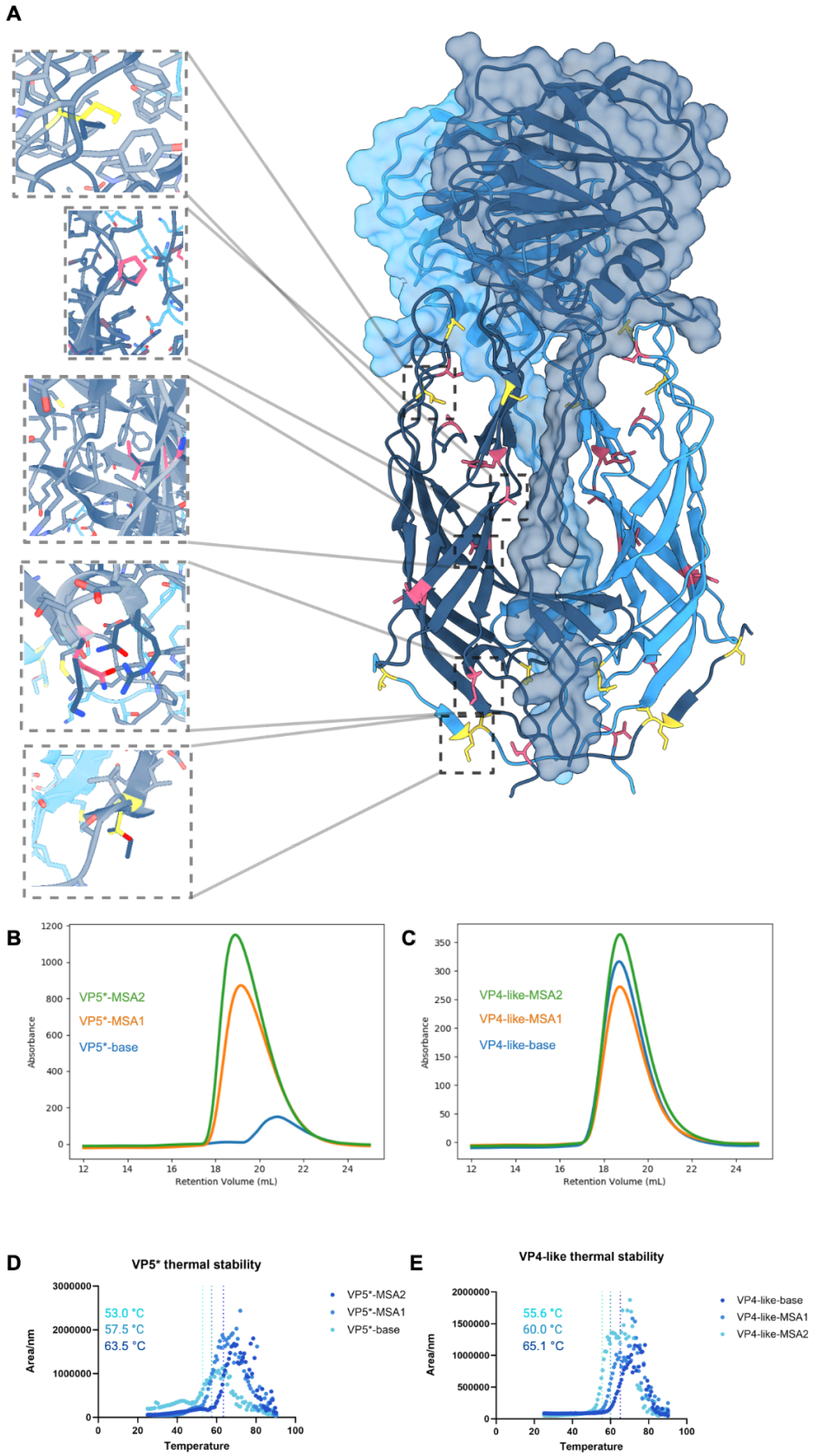

A
B
VP5*-MSA2
VP5*-MSA1
VP5*-base
Absorbance
Retention Volume (mL)
C
VP4-like-MSA2
VP4-like-MSA1
VP4-like-base
Absorbance
Retention Volume (mL)
D
VP5* thermal stability
53.0 °C
57.5 °C
63.5 °C
VP5*-MSA2
VP5*-MSA1
VP5*-base
Area/nm
Temperature
E
VP4-like thermal stability
55.6 °C
60.0 °C
65.1 °C
VP4-like-base
VP4-like-MSA1
VP4-like-MSA2
Area/nm
Temperature

Figure 1-2. MSA based mutations improve yield, dimeric state, and thermal stability of VP5* and VP4-like antigens. A) Structure of VP4-like antigen with VP5* shown in cartoon representation and the VP8* portion shown with cartoon and surface representation. The MSA1 set of mutations are shown in yellow, while the MSA2 mutations are shown in pink. Magnified insets of select mutations show the original residue in blue and the introduced residue in yellow or pink. From top, mutations are I397M, A298P, V353I, K360Q, and I255T. B) Size exclusion chromatography (SEC) of VP5* antigens on Superdex 200 Increase 10/300 GL column. C) Size exclusion chromatography (SEC) of VP4-like antigens on Superdex 200 Increase 10/300 GL column. D) NanoDSF with SYPRO dye of VP5* antigens, with melting temperature listed and shown by a vertical dashed line. E) NanoDSF with SYPRO dye of VP4-like antigens, with melting temperature listed and shown by a vertical dashed line.

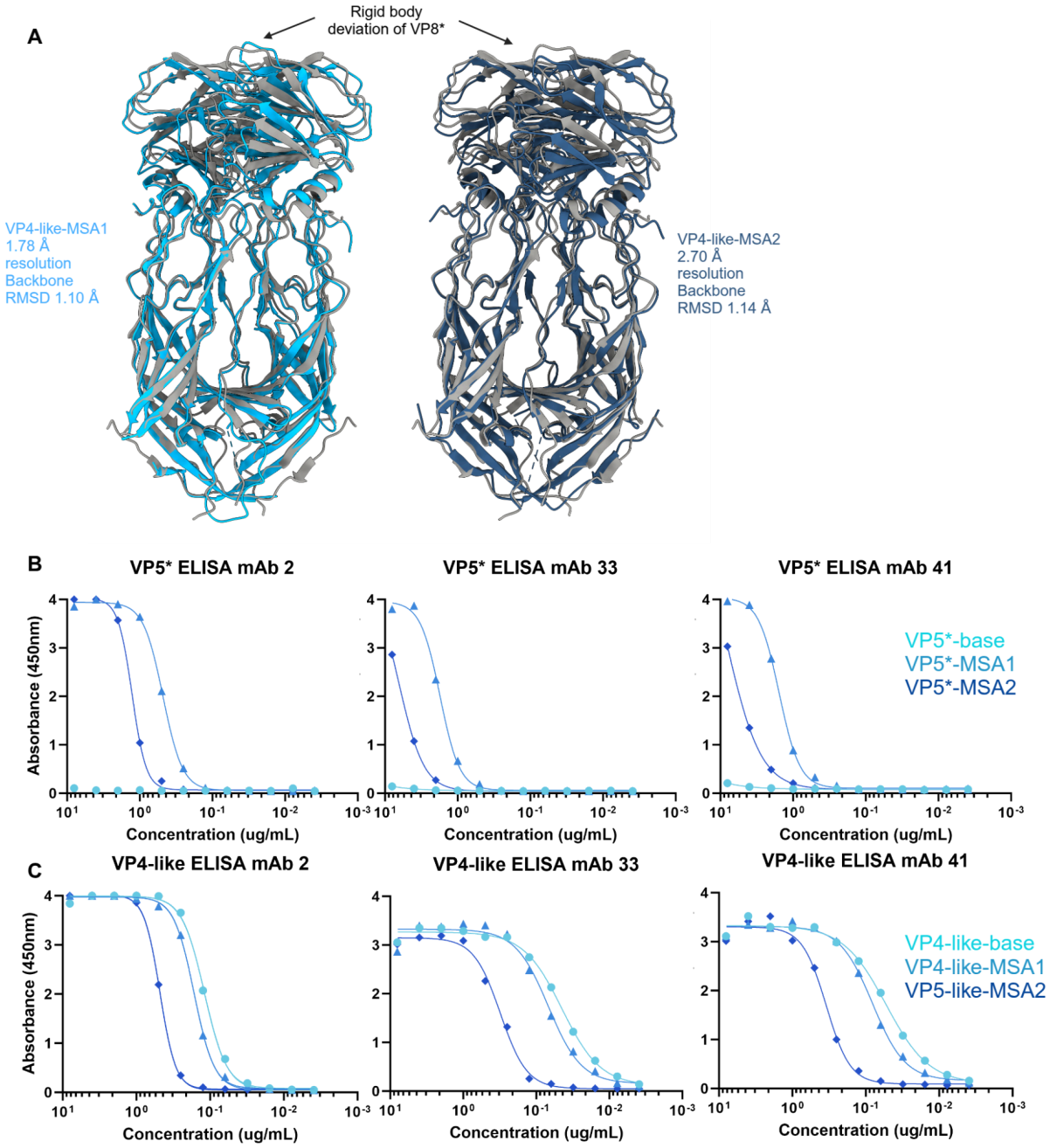


Figure 1-3. Structure and antigenicity of designed antigens. A) Crystal structures of VP4-like-MSA1 and VP4-like-MSA2 in light and dark blue, respectively, superimposed on VP4 of PDB ID 7UMS in gray. Backbone RMSD was across entire VP5* region. Rigid body deviations in VP8* region were observed. B) ELISA dilution curves for VP5* antigens. C) ELISA dilution curves for VP4-like antigens.

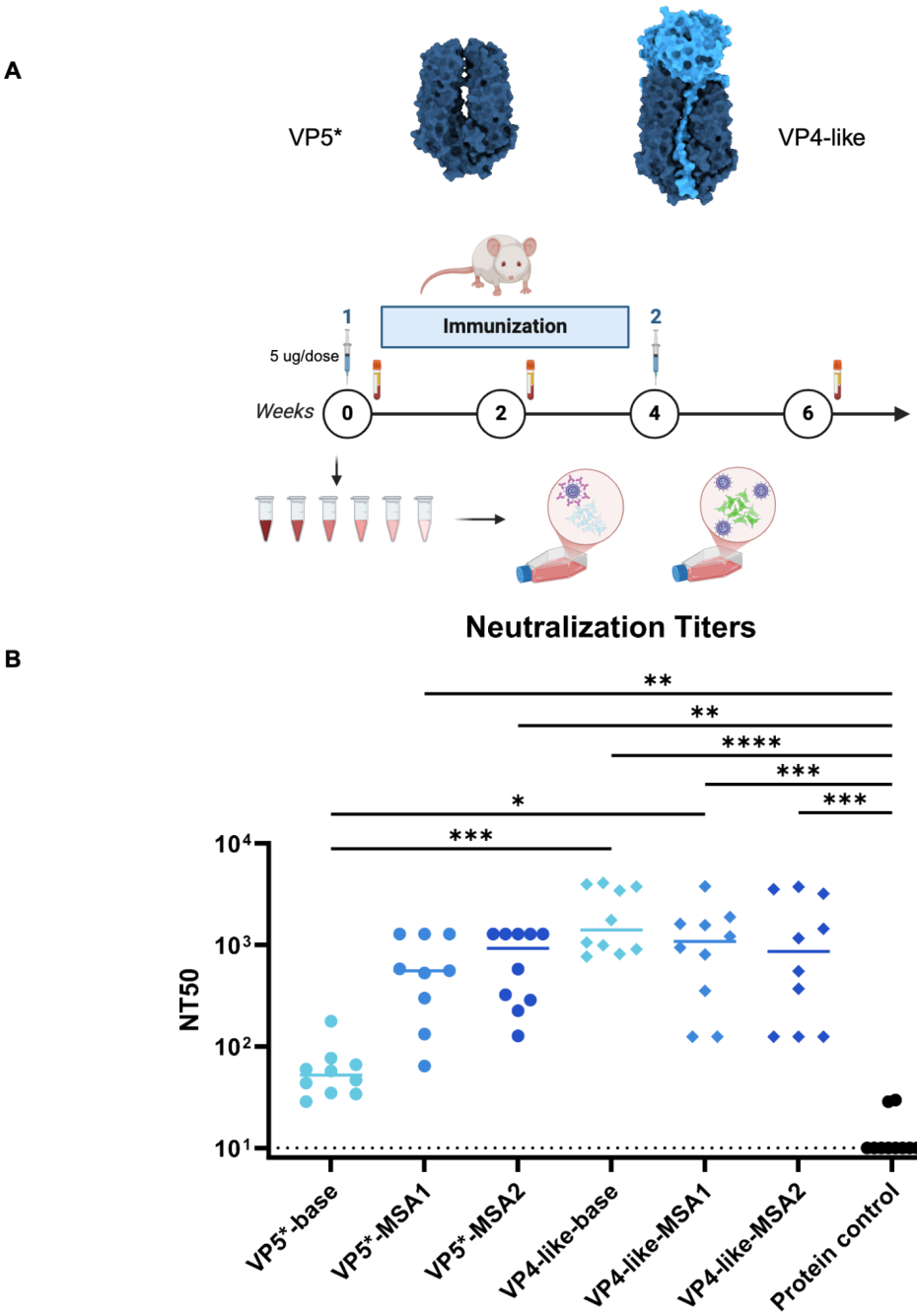
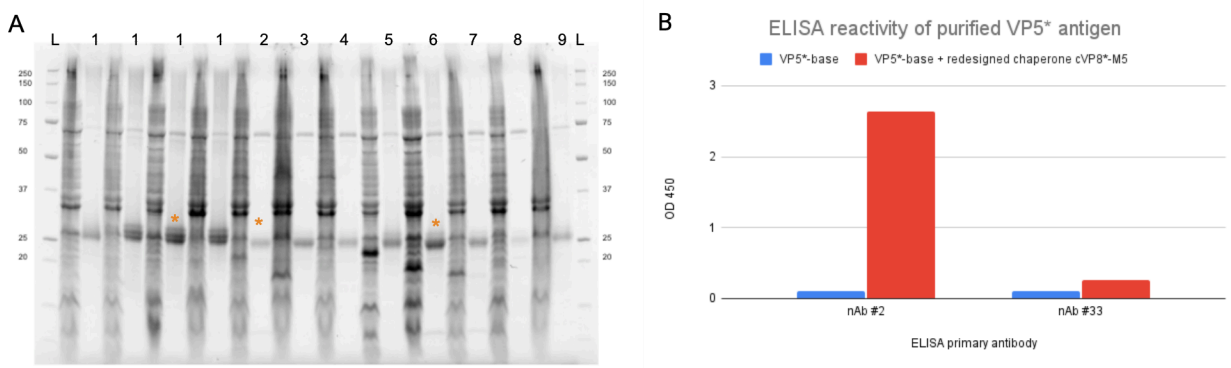
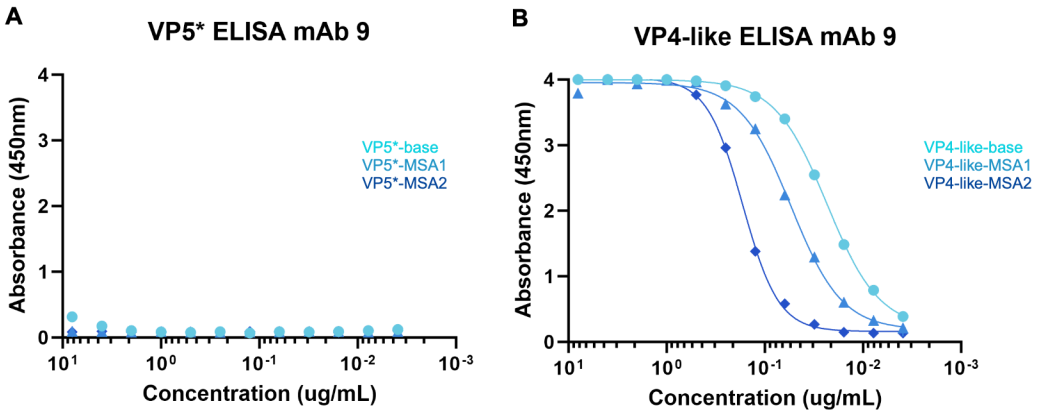


Figure 1-4. Immunogenicity of designed antigens. A) Scheme of immunization study comparing VP5* (top left) and VP4-like (top right) antigens. Mice were immunized twice and neutralization titers were measured using pre-immunization sera from week 0 or post-immunization sera from week 6. B) Neutralization titers were measured in post-immunization sera from individual mice using MA104 cells with a Rotarix strain of rotavirus. Groups are labeled by the immunogen used, with RABV-G used as a protein control. Statistical significance assessed using Kruskal-Wallis test.



Supplementary Figure 1-4. VP8* chaperone improves yield of VP5* while preserving antigenicity. A) SDS-PAGE gel following purification of VP5*-apex (1) under multiple conditions, VP5*-base (2), and co-expression of native VP8* (3) or ProteinMPNN-redesigned VP8* (4-9) with VP5*-base. For each construct, clarified lysate was loaded on the left, and eluate on the right. VP5*-apex under optimal conditions, VP5*-base, and VP5* with cVP8*-M5 co-expression are marked by an *. B) Preliminary screening ELISA of eluate (not concentration normalized) from VP5*-base or VP5* with cVP8*-M5 co-expression using the neutralizing antibodies #2 and #33.



Supplementary Figure 1-4. No trace of VP8* after purification of VP5* designed antigens. A) ELISA with VP8* specific mAb #9 shows no reactivity with VP5* designs that were co-expressed with redesigned VP8* chaperone. B) ELISA with VP8* specific mAb #9 against VP4-like-designs.

Materials and Methods

AlphaFold Prediction

The ColabFold implementation of AlphaFold was downloaded from the github repository into an apptainer environment with the necessary dependencies[32]. All predictions were made on either A4000 or A6000 GPUs on a custom high performance computing cluster. Sequences of candidate designed antigens were input into the ColabFold implementation to create MSAs using mmseqs2 and to generate structural predictions of the dimer using AlphaFold2_multimer_v3. Predictions were made with structural templates and with an amber relaxation step. Predictions were initially made using one structural template (PDB ID 7UMS) for filtering ProteinMPNN designs, but for all subsequent work transitioned to using all available PDB templates found by ColabFold during structure prediction by including the --templates flag. For each sequence predicted, structural predictions were obtained with all 5 AlphaFold models.

Following structure prediction, analysis with python scripts in a jupyter notebook was performed to determine prediction confidence by examining average pLDDT and PAE across the dimer. All 5 predictions for a given sequence were used to determine the mean and standard deviation for pLDDT and PAE across the different models. Python scripts were used to calculate root mean square deviation (RMSD) between the experimental structure of human rotavirus VP4 (PDB ID 7UMS) and the structures predicted for designed sequences.

Rosetta energy evaluation

Relaxed AlphaFold predictions were input into a PyRosetta script that calculated a biophysical Rosetta energy score using the ref2015 score function[33]. No Rosetta minimization or relaxation steps were undertaken. Rosetta energy scores for predicted structures from all 5 ColabFold models were used to calculate the mean energy and standard deviation for each sequence.

Multiple sequence alignment (MSA) based design

Standard MSAs were generated using NCBI protein BLAST with the standard databases and a maximum number of 500 sequences[34–37]. Clustered MSAs were generated using NCBI protein BLAST with the experimental databases option and a maximum number of 250 sequences. MSAs were filtered to greater than 70% identity to the query sequence with coverage greater than 50%, and downloaded as aligned FASTAs. Analysis with python scripts in a jupyter notebook was performed to determine the residue positions where more than 25% of the sequences contained the same amino acid which differed from the corresponding amino acid in the query sequence. The standard and clustered MSAs were used to generate two distinct sets of suggested mutations. Experimental structures or high-confidence predicted structures of the corresponding antigen were inspected in PyMol with the mutagenesis wizard tool to determine if the mutation would be tolerated[38]. If the mutation did not result in large steric clashes or the loss of two or more hydrogen bonds, it was retained in a list of MSA suggested mutations, resulting in a standard and clustered list of potential mutations.

Both lists of potential mutations were used to generate point mutations to the sequence of the native antigen. These sequences differing by one amino acid from the native antigen were input into AlphaFold, and the output structural predictions were evaluated for their RMSD, pLDDT, PAE, and Rosetta energies. Mutations that resulted in worse structural predictions relative to the prediction for the native sequence (RMSD > 1 Å higher, pLDDT 0.5 lower, PAE 0.5 higher, or 3 REU higher) were discarded. Remaining mutations were incorporated into MSA based designed antigens.

ProteinMPNN

ProteinMPNN was run using an implementation downloaded from the github repository[31]. All sequence design was conducted on CPUs or an A4000 GPU on a custom high performance computing cluster. Experimental structures of antigens were used when available, and all 5 high-confidence ColabFold predictions for the native sequence were used for antigens lacking experimental structures. Epitope residues were identified from structures of antigen in complex with monoclonal antibody #41 or from the literature. Epitope residues and other surface exposed residues near epitopes were disallowed for design. For VP5*, these residues were input as the following list: [281, 282, 283, 284, 285, 287, 289, 297, 306, 329, 330, 331, 332, 333, 334, 335, 337, 384, 385, 386, 388, 390, 391, 392, 393, 394, 396, 398, 399, 429, 434, 435, 436, 438, 440, 441, 442, 443, 444, 445, 446, 447, 448, 454, 455, 459]. ProteinMPNN was run with a sampling temperature of 0.2, and an amino acid bias of -0.8 for glycine and alanine.

ProteinMPNN was employed in a 2-step process. In the first step, structures served as inputs to ProteinMPNN for sequence design with all residues allowed to mutate that

were not disallowed for design. ProteinMPNN was run for 200 batches with five sequences each, to generate 1000 designed sequences for each input structure. A python script was used to analyze the frequency of mutations from the native amino acid at each position. Where ProteinMPNN chose a mutation at a given position in more than 25% of the sequences, the mutation was manually inspected with the PyMol mutagenesis wizard in the context of the input structure. If any suggested mutation caused large steric clashes or the loss of at least two hydrogen bonds, the residue was no longer considered for design. In this way, the first round of sequence design was used to identify designable positions where ProteinMPNN suggested mutations that did not appear to be deleterious.

In the second round, only these designable residues were allowed to mutate, and ProteinMPNN was run for 40 batches with five sequences each to generate 200 potential sequences. These sequences were then evaluated by AlphaFold structure prediction. Predicted structures with the lowest RMSD, highest pLDDT, and lowest PAE were selected as candidate designs, and their sequences were ordered for experimental evaluation.

Protein Expression

For initial screening of designs in *E. coli*, codon optimized DNA sequences were ordered as gene fragments (IDT or Twist), cloned into a pET-29b+ based plasmid using golden gate cloning with BsaI-HFv2 (NEB), and transformed into chemical competent BL21(DE3) cells (NEB). Cells were grown for 16 hours at 37°C in 1 mL of LB broth with

antibiotic before inoculating 500 uL of culture into 50 mL of Terrific Broth II containing antibiotic. Cells were grown at 37°C to an OD of 0.6-0.8 before adding IPTG to a final concentration of 0.25 mM and cells were grown at 18°C for 18 hours. Cells were centrifuged at 4000 g for 15 minutes to harvest, and pellets were stored at -80°C.

For constructs used for more detailed comparisons with the native antigen, plasmids encoding codon optimized sequences were ordered from GenScript. Protein expression was conducted in the same manner as above.

Purification

For purification of proteins from *E. coli*, cell pellets were resuspended in 8 mL of lysis buffer (50 mm Tris, 250 mM NaCl, 20 mM imidazole, 100 mM Arginine, 250 mM mannitol, 0.5 mg/mL lysozyme, 1 U / mL benzonase, 0.02 tablets/mL of protease inhibitor cocktail) and lysed by ultrasonication (Qsonica 24-tip sonicator, 6 minutes total, 5s on, 5s off, 80% amplitude). Lysate was clarified by centrifugation at 14000 g for 30 min. Clarified lysate was applied to 0.5 mL of Ni-NTA resin (Qiagen) equilibrated with wash buffer 1 (50 mm Tris, 250 mM NaCl, 20 mM imidazole, 100 mM Arginine, 250 mM mannitol). Resin was then washed with 20 column volumes of wash buffer 1 for general protein purification or wash buffer 2 (50 mm Tris, 250 mM NaCl, 20 mM imidazole, 0.75% CHAPS) for low endotoxin protein purification. Protein was eluted with 2.4 column volumes of elution buffer (50 mm Tris, 250 mM NaCl, 500 mM imidazole) and analyzed by SDS-PAGE gel. Eluted proteins were further filtered by size exclusion

chromatography using a Superdex 200 Increase 10/300 GL column (Cytiva) on an ÄKTA pure (Cytiva).

Melting and aggregation temperature determination:

Determination of melting and aggregation temperatures were done by differential scanning fluorimetry with the UNcle instrument and software (UNchained Labs). SEC purified proteins at a concentration between 0.25-1 mg/mL were added to 8.8 uL quartz capillary wells in a 16-well cassette (UNchained Labs). With the temperature of the cassette changing from 25 to 95 °C at a thermal ramp rate of 1°C/min, intrinsic changes in protein fluorescence emission were used to evaluate melting and aggregation temperatures. As an orthogonal measurement of the melting temperature, protein was mixed with the hydrophobic dye Sypro Orange (Thermo Fisher Scientific) to evaluate change in fluorescence emission during the thermal ramp.

Enzyme-linked immunosorbent assay (ELISA):

For initial screening of constructs, proteins were expressed in *E. coli* and then purified by IMAC. Eluate was then diluted 50-100 fold in 50 mM Tris, 300 mM NaCl. All incubation steps occurred while shaking at 600 rpm, and all wash steps involved 3 washes with ELISA wash buffer (25 mM Tris, 150 mM NaCl, 0.2% Tween 20, pH 8.0) by an automated plate washer (Biotek). 100 uL of the solution was incubated onto a His binding plate (Qiagen) for 1.5 hours. 200 uL of ELISA blocking buffer (25 mM Tris, 150 mM NaCl, 0.2% Tween 20, 5% nonfat milk, pH 8.0) was added to each well and incubated for 1 hour. Plates were washed and 100 uL of ELISA blocking buffer

containing primary antibody at 4 ug/mL was added to each well and incubated for 1 hour. Plates were washed and 100 uL of ELISA blocking buffer containing secondary antibody (goat anti-human IgG-HRP, Southern Biotech at a 1:5000 dilution). Plates were washed, 100 uL of TMB (3,30,50,5-tetramethylbenzidine, SeraCare) was added for 3 minutes, the reaction was quenched with the addition of 100 uL of 1N HCl, and the absorbance at 450 nm of each well was measured with an Epoch plate reader (Biotek).

Antigenicity comparisons for VP5* and VP4-like antigens were done using a similar ELISA protocol, but beginning with SEC purified protein. Antigen was added to the first well of each row at 8 ug/mL and subsequently diluted 2 fold in each subsequent well of the row. The following steps were conducted as described above.

Sequences for all antibodies used can be found in Nair et al. 2017. Antibodies were ordered from GenScript using the TurboCHO protein expression service.

Crystal structure preparation and determination

For rotavirus antigens, protein was SEC purified in 25 mM Tris, 150 mM NaCl, 5% glycerol, pH 8.0 and concentrated to 500 uL at 6-10 mg/mL in 3K concentrators (Amicon). All crystallization experiments were conducted using the sitting drop vapor diffusion method. Crystallization trials were set up in 200 nL drops using the 96-well plate format at 20 ˚C. Crystallization plates were set up using a Mosquito (SPT Labtech) then imaged using UVEX microscopes and UVEX PS-256 (JAN Scientific). Diffraction quality crystals formed in 0.1 M Sodium HEPES pH 7.5 and 15 %(w/v) PEG 20000 for

VP4-like-MSA1. Diffraction quality crystals formed in 0.2 M Magnesium chloride, 0.1 M Tris pH 8 and 20% (w/v) PEG 6000 for VP4-like-MSA2.

Diffraction data was collected either at National Light Synchrotron Light Source II AMX/FMX or the Advanced Light Source beamline on 8.2.1. X-ray intensities and data reduction were evaluated and integrated using XDS and merged/scaled using Pointless/Aimless in the CCP4 program suite(Kabsch 2010; Winn et al. 2011). Structure determination and refinement starting phases were obtained by molecular replacement using Phaser using the design model for the structures(McCoy et al. 2007). Following molecular replacement, the models were improved using phenix.autobuild(Adams et al. 2010); efforts were made to reduce model bias by setting rebuild-in-place to false, and using simulated annealing and prime-and-switch phasing. Structures were refined in Phenix(Adams et al. 2010). Model building was performed using COOT(Emsley and Cowtan 2004). The final model was evaluated using MolProbity(Williams et al. 2018). Data collection and refinement statistics are recorded in Table S1. Data deposition, atomic coordinates, and structure factors will be deposited in the Protein Data Bank (PDB).

Immunogenicity studies

For immunogenicity studies, 7 week old female Balb/cAnNHsd mice were purchased from Envigo (order code 047). Mice were housed in a specific-pathogen free facility and all animal studies were in accordance with the University of Washington's Institutional Animal Care and Use Committee. For immunizations, 5 ug/dose of each protein was

diluted in dPBS buffer and mixed 1:1 with Addavax adjuvant (InvivoGen). At 8 weeks of age, animals were immunized with 100 uL of the mixture (week 0). 4 weeks later, mice were immunized again. At both weeks 0 and 4, blood was collected by the submental route. At week 6, a terminal bleed was conducted via cardiac puncture.

Neutralization titers

Sera from the immunization study was sent to the lab of Sasirekha Ramani to determine neutralization titers. Rotarix G1P[8] rotavirus strain at titers of approximately 10^6 focus forming units per mL (FFU/mL) were used to infect MA104 cultured cells. Infection without serum control was used as a negative control, while infection with dilutions of the monoclonal antibody mAb #41 were used as a positive control to ensure that each plate remained within reference values. Serum samples were diluted 2-fold starting at a 1:20 dilution, and each dilution was incubated with virus and cultured cells. Cultured cells were then stained to determine the percentage of infected cells at each dilution, and neutralization titers were calculated as an NT50, the dilution of serum at which each serum sample prevented 50% of infection. Values were reported as NT50 values and plotted. A two-way ANOVA was used to determine statistical significance.

Table S1. Data collection and refinement statistics

	VP4-like-MSA1	**VP4-like-MSA2**
Data collection		
Space group	$P2_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	51.93, 84.16, 99.46	75.82, 89.16, 138.36
a, b, g (°)	90, 104.38, 90	90, 90, 90

Resolution (Å)	32.19 - 1.78 (1.87 - 1.78)	34.59 - 2.70 (2.85 - 2.70)
R_{merge}	0.170 (1.323)	0.544 (2.828)
I / sI	6.6 (1.2)	5.1 (1.2)
Completeness (%)	97.5 (99.9)	100 (100)
Redundancy	5.3 (5.2)	13.5 (13.4)
Refinement		
Resolution (Å)	32.19 - 1.78 (1.82 - 1.78)	34.59 - 2.70 (2.77 - 2.70)
No. reflections	77382 (5609)	26414 (1835)
R_{work} / R_{free}	0.1944 (0.2836)/ 0.2442 (0.3322)	0.2303 (0.3287)/ 0.2740 (0.3798)
No. atoms		
Protein	6597	6599
Ligand/ion	16	n/a
Water	657	29
B-factors		
Protein	25	41
Ligand/ion	29	n/a
Water	29	31
R.m.s. deviations		
Bond lengths (Å)	0.006	0.002
Bond angles (°)	0.790	0.470

*Values in parentheses are for highest-resolution shell.

Acknowledgements

Thank you to all of the scientists whose work on rotavirus, its lifecycle, and immune responses makes structure-based vaccine design possible.

Funding for all rotavirus experimental work came from the Gates Foundation.

Maintenance of the computational infrastructure and technical help establishing the Colabfold environment were generously provided by Luki Goldschmidt and Patrick Vecchiato.

Crystallization and structure determination were performed by Asim Bera, Alex Kang, Emily Joyce, and Evans Brackenbrough.
Crystallographic data collected at the National Synchrotron Light Source II beamline. The Center for Bio-Molecular Structure (CBMS) is primarily supported by the NIH-NIGMS through a Center Core P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1607011). NSLS2 is a U.S.DOE Office of Science User Facility operated under Contract No. DE-SC0012704. This publication resulted from the data collected using the beamtime obtained through NECAT BAG proposal # 311950. We also want to thank the Advanced Light Source (ALS) beamline 8.2.1 at Lawrence Berkeley National Laboratory for X-ray crystallography data collection. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health (NIH), National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences and US Department of Energy (DOE) (DE-AC02-05CH11231).

All neutralization titers were performed by Pallavi Sharma with input from Nicole Hayes and Sasirekha Ramani.

Sequences

VP4-like-antigens

>VP4-like-base
MSHHHHHHSGGSGSGSAQANEDIIVSKTSLWKEMQYNRDIIIRFKFGNSIVKMGGLGYKWSEISYKAANY
QYNYLRDGEQVTAHTTCSVNGVNNFSYNGGSLPTDFGISRYEVIKENSYVYVDYWDDSKAFRNMVYVKS
LAANLNSVKCTGGSYDFSLPVGAWPVINGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILR
TRTVNLYGLPAANPNNGNEYYEISGRFSLISLVPTNGGGGQTRYAPVNWGHGEINDSTTVEPILDGPYQP
TTFTPPNDYWILINSNTNGVVYESTNNSDFWTAVVAIEPHVTPVDRQYMIFGESKQFNVSNDSNKWKFLE
MFRSSSQNEFYNRRTLTSDTRLVGILKYGGRVWTFHGETPRATTDSSSTANLNNISITIHSEFYIIPRSQES
KCNEYINNGLPPIQNTR
>VP4-like-MSA1
MSHHHHHHSGGSGSGSAQVNEDITISKTSLWKEMQYNRDIIIRFKFGNSIIKMGGLGYKWSEISYKAANYQ
YNYLRDGEQVTAHTTCSVNGVNNFSYNGGSLPTDFGISRYEVIKENSYVYVDYWDDSKAFRNMVYVRSL
AANLNSVKCTGGSYDFSIPVGAWPVMNGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILRT
RTVNLYGLPAANPNNGNEYYEISGRFSLISLVPTNGGGGQTRYAPVNWGHGEINDSTTVEPILDGPYQPT
TFTPPNDYWILINSNTNGVVYESTNNSDFWTAVVAIEPHVTPVDRQYMIFGESKQFNVSNDSNKWKFLEM
FRSSSQNEFYNRRTLTSDTRLVGILKYGGRVWTFHGETPRATTDSSSTANLNNISITIHSEFYIIPRSQESKC
NEYINNGLPPIQNTR
>VP4-like-MSA2
MSHHHHHHSGGSGSGSAQVNEDITISKTSLWKEMQYNRDITIRFKFGNSIIKMGGLGYKWSEISFKPANY
QYTYLRDGEQVTAHTTCSVNGVNNFSYNGGSLPTDFGISRYEVIKENSYVYIDYWDDSQAFRNMVYVRS
LAANLNSVKCTGGSYDFSIPVGAWPVMNGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILR
TRVVNLYGLPAVNPNNGNEYYEISGRFSLISLVPSNGGGGQTRYAPVNWGHGEINDSTTVEPILDGPYQP
TTFTPPNDYWILINSNTNGVVYESTNNSDFWTAVVAIEPHVTPVDRQYMIFGESKQFNVSNDSNKWKFLE
MFRSSSQNEFYNRRTLTSDTRLVGILKYGGRVWTFHGETPRATTDSSSTANLNNISITIHSEFYIIPRSQES
KCNEYINNGLPPIQNTR

VP5* antigens

>VP5*-base
MSHHHHHHSGGSGSGSAQANEDIIVSKTSLWKEMQYNRDIIIRFKFGNSIVKMGGLGYKWSEISYKAANY
QYNYLRDGEQVTAHTTCSVNGVNNFSYNGGSLPTDFGISRYEVIKENSYVYVDYWDDSKAFRNMVYVKS
LAANLNSVKCTGGSYDFSLPVGAWPVINGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILR
TRTVNLYGLPAANPNNGNEYYEISGRFSLISLVPTNDD
>VP5*-apex
MSHHHHHHSGGSGSGSAQANEDIIVSKTSLWKEMQYNRDIIIRFKFGNSIVKSGGSGYKWSEISYKAANY
QYNYLRDGEQVTAHTTCSVNGVNNFSYNGGSSPTDFGISRYEVIKENSYVYVDYWDDSKAFRNMVYVKS
LAANLNSVKCTGGSYDFSLPSGAWPVINGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILR
TRTVNLYGLPAANPNNGNEYYEISGRFSLISLVPTNDD
>VP5*-MSA1
MSHHHHHHSGGSGSGSAQVNEDITISKTSLWKEMQYNRDIIIRFKFGNSIIKMGGLGYKWSEISYKAANYQ
YNYLRDGEQVTAHTTCSVNGVNNFSYNGGSLPTDFGISRYEVIKENSYVYVDYWDDSKAFRNMVYVRSL
AANLNSVKCTGGSYDFSIPVGAWPVMNGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILRT
RTVNLYGLPAANPNNGNEYYEISGRFSLISLVPTNDD
>VP5*-MSA2
MSHHHHHHSGGSGSGSAQVNEDITISKTSLWKEMQYNRDITIRFKFGNSIIKMGGLGYKWSEISFKPANY
QYTYLRDGEQVTAHTTCSVNGVNNFSYNGGSLPTDFGISRYEVIKENSYVYIDYWDDSQAFRNMVYVRS

LAANLNSVKCTGGSYDFSIPVGAWPVMNGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILR
TRVVNLYGLPAVNPNNGNEYYEISGRFSLISLVPSNDD

VP5* antigens redesigned with 2-step ProteinMPNN process

>VP5_mpnn_00
MSGHHHHHHSGAQVDEDIIVSKTSLWKEMQYNRDIIIRFKFGNSIVKMGGLGYKWSEISYKAANYQYNYL
RDGEQVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYWDDSKAFRNIVYVRSLAANLN
SVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGNEYYEISGRFSLISLVPTNDD
>VP5_mpnn_459
MSGHHHHHHSGAQVDEDEVVSEASAWKERLYTTDLVIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
NRDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVEVDFSDDSKAFKNLKFVRSAEAN
FDTVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTDFKDGKPNNSLRVRFSLTVDEPPFSILRTRTV
NLYGLPAANPNNGKEYYEISGTFKLISLVPTNDD
>VP5_mpnn_536
MSGHHHHHHSGAQMDEDEDVSEPSFWKVKKYTVDITIRFAFGNSIVKMGGLGYKWSEISYKAASMSYSY
KRDGKEVKAHTTVSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSADDKAFKALKYVRSLSANF
TDVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGLTLTTKFSNGTSFNSLRATFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEITGTLTLESLVEEPDD
>VP5_mpnn_178
MSGHHHHHHSGAQVDEDEDVSEASAWKERKFTVDLVIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKELKLVRSASANF
DTVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLETRFKDGKADNSLRVIFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEISGTFTLISLVPTNDD
>VP5_mpnn_844
MSGHHHHHHSGAQVNEDIVVSKTSLWKEMQYNRDITIRFKFGNSIVKMGGLGYKWSEISYKAANYQYSY
TRDGEEVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYYDDSQAFRNMVYVRSLTAEF
NEVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLLTQFTDFVALNTLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGREYYEISGRFSLISLVPSNDD
>VP5_mpnn_693
MSGHHHHHHSGAQMDEDEDVSPPSLWKTKRYTVDITIRFSFGNSIVKMGGLGYKWSEISYKAASMSYSY
ERDGEKVNAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNIDFSEDSKAFQALKYVRSLSANFS
TVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLETRFKDGESENSLRLTFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEITGTLTLESLTEEPDD
>VP5_mpnn_262
MSGHHHHHHSGAQVDEDKIVSEASAWKEVEYTVDLIIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNYN
RDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFGDDTKAFKNLKFVRSITANFD
TVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTNFKDGKINNSLRFKFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGTFKLISLVPTNDD
>VP5_mpnn_923
MSGHHHHHHSGAQVNEDIVISKTSLWKEMQYNRDITIRFKFGNSIVKMGGLGYKWSEISYKAASYQYSYT
RDGEKVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYYDDSQAFRNMVYVRSLRAEF
NSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLITQFTDFVSLNTLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEISGRFSLISLVPSNDD
>VP5_mpnn_240
MSGHHHHHHSGAQVDEDEDVSGASAWKEKKFTTDLVIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDVSDDTKAFKELKFVRSASANF
DSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITIVTDFKDGQISNSLRIRFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGKFTLISLVPTNDD

>VP5_mpnn_193
MSGHHHHHHSGAQEDEDRDVSEPSFWKTKLYDVDITIRFKFGNSIVKMGGLGYKWSEISYKAASMSYSY
ERDGKKVTAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSADDKAFKALKYVRSASANFT
DVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLETTFSNGESKNSLRLTFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEITGRFRLESLVEDPDD
>VP5_mpnn_477
MSGHHHHHHSGAQVDEDVDVSDASAWKERRFTVDLIIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
LRDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKELKLVRSASANF
DSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTDFKDGKIDNSLRIRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEISGTFDLISLVPTNDD
>VP5_mpnn_594
MSGHHHHHHSGAQMDEDEDIEEPSFWKTKLYKKDITIRFKFGNSIVKMGGLGYKWSEISYKAASMSYSY
ERDGKPVKAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSADDKAFQALKYVRSLSANF
TDVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGLTLETKFSNGQSQNSLRATFSLTVDEPPFSILRTRTVN
LYGLPAANPNNGKEYYEIKGTLNLESLTEDPDD
>VP5_mpnn_994
MSGHHHHHHSGAQVNEDIVISKTSLWKEMQYNRDITIRFKFGNSIVKMGGLGYKWSEISYKAASYQYSYT
RDGEEVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYYDDSQAFRNMVYVRSLTADFN
DVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLITQFTDFVALNTLRFRFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGRFSLISLVPSNDD
>VP5_mpnn_765
MSGHHHHHHSGAQMDEDEEVSEPSKWKTVRYNKDIVIRFSFGNSIVKMGGLGYKWSEISYKAASYSYSY
ERDGKKINAHTTVSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSADDKAFQALKYVRSLEANF
TPVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLETKFKNGESINSLRLTFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEIKGTLVLESLVEEPDD
>VP5_mpnn_103
MSGHHHHHHSGAQVDEDVEVSGSSSWKERRYTVDLIIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGEQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVEVDFSDDTKAFKELRFVRSASANF
DSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTDFSDGKARNSLRFRFSLTVDEPPFSILRTRTVN
LYGLPAANPNNGKEYYEISGTFDLISLVPTNDD
>VP5_mpnn_509
MSGHHHHHHSGAQMDEDEDVSEPSLWKTVLYKTDIIIRFRFGNSIVKMGGLGYKWSEISYKAASYSYSYE
RDGKKVNAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSADDKAFQSLKYVRSLSANFT
DVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGLTLETRFENGKSINSLRATFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEITGSLSLESLVEDPDD
>VP5_mpnn_259
MSGHHHHHHSGAQVDEDEVVGGASAWKEVKYTVDVKIRFKFGNSIVKMGGLGYKWSEISYKASSVSYN
YERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVEVDFSDDTKAFKALKFVRSITANF
DPVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLITKFKDGKSDNSLRLRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEISGTFNLISLVPTNDD
>VP5_mpnn_934
MSGHHHHHHSGAQMDEDEDVEEPSKWKAKLYKKDIIIRFSFGNSIVKMGGLGYKWSEISYKAASMSYSY
ERDGKKVNAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSSDDKAFKALKYVRSLSANFT
DVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGLTLETEFKDGKSINSLRATFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGTFTLTSLVEDEDD
>VP5_mpnn_598
MSGHHHHHHSGAQVDEDEVVSEASAWKEVRYTVDLVIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
DRDGEQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKALKFVRSASANF

DPVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITIVTEFKDGKIKNSLRIRFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGTFTLISLVPTNDD
>VP5_mpnn_647
MSGHHHHHHSGAQVDEDEVVSGPSLWKEVRHTVDLIIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKSLKFVRSASANF
DEVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLETLFEDGKIKNSLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEISGTFDLISLEPTNDD
>VP5_mpnn_843
MSGHHHHHHSGAQVDEDRDVSGASAWKEREYIVDLKIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGRQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKALKFVRSASANF
DSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLITRFKDGKIDNSLRIEFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGTFKLISLVPTNDD
>VP5_mpnn_603
MSGHHHHHHSGAQVDEDKEVSGASAWKEVEYTVDLIIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDSKAFKNLRFVRSASAN
FDSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTEFKDGEVRNSLRIRFSLTVDEPPFSILRTRTVN
LYGLPAANPNNGKEYYEISGTFKLISLVPTNDD
>VP5_mpnn_683
MSGHHHHHHSGAQMDEDEELSEPSYWKTKLYKTDIKIRFAFGNSIVKMGGLGYKWSEISYKAASMSYSY
ERDGKKVNAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSEDSKAFKALKYVRSLSANFT
TVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGLTLETSFSNGSSTNSLRATFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEITGTLTLESLVEEPDD
>VP5_mpnn_128
MSGHHHHHHSGAQVDEDKIVSGASKWKEVEYTVDVIIRFKFGNSIVKMGGLGYKWSEISYKASSTSYNY
ERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVEVDFSDDTKAFKNLKFVRSASANF
DPVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTDFKDGKASNSLRVRFSLTVDEPPFSILRTRTVN
LYGLPAANPNNGKEYYEISGTFKLISLVPTNDD
>VP5_mpnn_97
MSGHHHHHHSGAQVDEDVEVSGASAWKEVRYTVDLVIRFKFGNSIVKMGGLGYKWSEISYKASSVSYN
YERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVEVDFSDDTKAFKNLKFVRSASAN
FDPVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLITEFEDGKIKNSLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEISGTFDLISLVPTNDD
>VP5_mpnn_0
MSGHHHHHHSGAQADEDIIVSKTSLWKEMQYNRDIIIRFKFGNSIVKMGGLGYKWSEISYKAANYQYNYL
RDGEQVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYWDDSKAFRNIVYVRSLAANLN
SVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLSTQFTDFVSLNSLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGNEYYEISGRFSLISLVPTNDD
>VP5_mpnn_746
MSGHHHHHHSGAQVDEDVVVSEASAWKEVEYTVDLIIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVEVDFSDDTKAFQNLKFVRSASAN
FDTVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTEFKDGEAHNSLRFRFSLTVDEPPFSILRTRTV
NLYGLPAANPNNGKEYYEISGTFRLISLVPTNDD
>VP5_mpnn_297
MSGHHHHHHSGAQVNEDIVISKTSLWKEMQYNRDIIIRFKFGNSIVKMGGLGYKWSEISYKAASYQYTYE
RDGEKVTAHTTVSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYYDDSQAFRNMVYVRSLTAEFN
EVRCTGGSYHFSLPVGAWPVINGGAVSLHYAGVTLITQFTDFVALNTLRFRFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGRFSLISLVPSNDD
>VP5_mpnn_918

MSGHHHHHHSGAQVDEDEDVSGASAWKERKYVVDLKIRFKFGNSIVKMGGLGYKWSEISYKASSVSYN
YERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKDLKFVRSAKAN
FDEVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLETSFSDGQVNNSLRIRFSLTVDEPPFSILRTRTV
NLYGLPAANPNNGKEYYEISGKFELISLVPTNDD
>VP5_mpnn_619
MSGHHHHHHSGAQVDEDSVVSGASKWKEVEYKVDLTIRFKFGNSIVKMGGLGYKWSEISYKASSVSYN
YERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKSLKLVRSASAN
FDSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTNFKDGVAENSLRVKFSLTVDEPPFSILRTRTV
NLYGLPAANPNNGKEYYEISGTFRLISLEPTNDD
>VP5_mpnn_174
MSGHHHHHHSGAQVNEDIVISKTSLWKEMQYNRDITIRFKFGNSIVKMGGLGYKWSEISYKAASYQYSYT
RDGEEVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYYDDSQAFRNMVYVRSLTADFN
DVRCTGGSYHFSLPVGAWPVINGGAVSLHYAGVTLITQFTDFVSLNTLRFRFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGREYYEISGRFSLISLVPSNDD
>VP5_mpnn_523
MSGHHHHHHSGAQVNEDIVISKTSLWKEMQYNRDITIRFKFGNSIVKMGGLGYKWSEISYKAASYQYSYT
RDGEEVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYWDDSQAFRNMVYVRSLTADF
NTVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLITQFTDFVALNTLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGREYYEISGRFSLISLVPSNDD
>VP5_mpnn_439
MSGHHHHHHSGAQMDEDEDIGEPSLWKTVRYKKDITIRFAFGNSIVKMGGLGYKWSEISYKAASMSYSY
ERDGKKVNAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSADDKAFKALKYVRSLNANF
STVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLETKFSNGESQNSLRLTFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEITGTFELESLVEDPDD
>VP5_mpnn_63
MSGHHHHHHSGAQMDEDRDISEPSIWKTKLYKKDITIRFAFGNSIVKMGGLGYKWSEISYKAASMSYSYT
RDGKKINAHTTVSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSSDDKAFQSLKYVRSLKANFT
DVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTDFNEGQSTNSLRLTFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEITGTLELESLEEDPDD
>VP5_mpnn_618
MSGHHHHHHSGAQMDEDKDISEPSLWKTKLYKKDITIRFAFGNSIVKMGGLGYKWSEISYKAASMSYSYE
RDGKKVNAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNIDFSSDDKAFKALKYVRSLSANFSS
VRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLTTNFSNGTSNNSLRLTFSLTVDEPPFSILRTRTVNLYG
LPAANPNNGKEYYEITGTLTLESLVEEPDD
>VP5_mpnn_655
MSGHHHHHHSGAQMDEDEDIEDASKWKTRKYTIDITIRFAFGNSIVKMGGLGYKWSEISYKAASMSYSYE
RDGKTVNAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVLVDFSADDKAFQALKYVRSLSANFTP
VRCTGGSYHFSLPVGAWPVINGGAVSLHFAGLTLETKFDNGKSSNSLRATFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEITGTLTLESLVEEADD
>VP5_mpnn_50
MSGHHHHHHSGAQKDEDEDISEPSYWKTKRYKVDITIRFKFGNSIVKMGGLGYKWSEISYKAASHSYSY
ERDGKKINAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVLVDFSADDKAFKELKYVRSLSANFT
DVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGLTLTTKFNNGESENSLRLTFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEITGTLTLESLVEDPDD
>VP5_mpnn_862
MSGHHHHHHSGAQMDEDEDVEEPSYWKTKRYTIDITIRFSFGNSIVKMGGLGYKWSEISYKAASMSYSY
ERDGKTVNAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVNVDFSADDKAFKALKYVRSLTANFT
PVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGLTLETKFSNGKSINSLRATFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEITGTLTLESLTEEEDD

>VP5_mpnn_819
MSGHHHHHHSGAQVDEDEEVSGSSSWKEKKYTVDLRIRFKFGNSIVKMGGLGYKWSEISYKASSVSYN
YERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKALKFVRSASAN
FDTVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTDFKDGESKNSLRIRFSLTVDEPPFSILRTRTVN
LYGLPAANPNNGKEYYEISGTFTLISLVPTNDD
>VP5_mpnn_845
MSGHHHHHHSGAQVDEDREVSGSSSWKEREYTVDLRIRFKFGNSIVKMGGLGYKWSEISYKASSVSYN
YERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKNLKFVRSASAN
FDSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTEFKDGEIKNSLRIRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEISGTFRLISLVPTNDD
>VP5_mpnn_915
MSGHHHHHHSGAQVDEDEDVSGASAWKERKFTVDLKIRFKFGNSIVKMGGLGYKWSEISYKASSVSYN
YDRDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKNLNLVRSASAN
FDSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITIITKFKDGKIENSLRIEFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGTFTLISLVPTNDD
>VP5_mpnn_800
MSGHHHHHHSGAQVDEDKIVSGSSYWKEVEYTVDLKIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVIDFSDDTKAFKNLKFVRSASANF
DPVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTNFKDGKSDNSLRFVFSLTVDEPPFSILRTRTVN
LYGLPAANPNNGKEYYEISGTFKLISLVPTNDD
>VP5_mpnn_989
MSGHHHHHHSGAQVDEDEVVSGASYWKEKKFKVDLIIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
ERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKNLKFVRSAEANF
DSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTTFKDGKSNNSLRFRFSLTVDEPPFSILRTRTVN
LYGLPAANPNNGKEYYEISGTFTLISLVPTNDD
>VP5_mpnn_85
MSGHHHHHHSGAQVNEDIVISKTSLWKEMQYNRDITIRFKFGNSIVKMGGLGYKWSEISYKAASYQYSYT
RDGEEVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYYDDSQAFRNMVYVRSLTAEFN
DVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLITQFTDFVSLNTLRFRFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGRFSLISLVPSNDD
>VP5_mpnn_649
MSGHHHHHHSGAQVDEDREVSGASAWKEVEYTVDLVIRFKFGNSIVKMGGLGYKWSEISYKASSVSYN
YERDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVVVDFSDDTKAFKNLKFVRSASAN
FDSVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLVTDFKDGEINNSLRFRFSLTVDEPPFSILRTRTVN
LYGLPAANPNNGKEYYEISGTFRLISLVPTNDD
>VP5_mpnn_869
MSGHHHHHHSGAQVNEDIVVSKTSLWKEMQYNRDITIRFKFGNSIVKMGGLGYKWSEISYKAASYQYSY
TRDGEEVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYYDDSQAFRNMVYVRSLTAEF
NEVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLITQFTDFVSLNTLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGREYYEISGRFSLISLVPSNDD
>VP5_mpnn_198
MSGHHHHHHSGAQMDEDRDVSEPSLWKTKLYKKDITIRFAFGNSIVKMGGLGYKWSEISYKAASMSYSY
ERDGKKINAHTTISVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYIDFSSDDKAFKALKYVRSLEANFSD
VRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLETRFSNGVSTNSLRLTFSLTVDEPPFSILRTRTVNLYG
LPAANPNNGKEYYEITGTFTLESLVEDPDD
>VP5_mpnn_974
MSGHHHHHHSGAQVNEDIVISKTSLWKEMQYNRDITIRFKFGNSIVKMGGLGYKWSEISYKAASYQYSYI
RDGEEVTAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVYVDYWDDSQAFRNMVYVRSLTAEF

NDVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGVTLITQFTDFVALNTLRFRFSLTVDEPPFSILRTRTVNL
YGLPAANPNNGKEYYEISGRFSLISLVPSNDD
>VP5_mpnn_772
MSGHHHHHHSGAQVDEDEVVSGPSLWKEIKYTVDLKIRFKFGNSIVKMGGLGYKWSEISYKASSVSYNY
LRDGKQVNAHTTCSVNGVNNFSYNGGFLPTDFGISRYEVIKENSYVEVDFSDDTKAFKNLKFVRSASANF
DTVRCTGGSYHFSLPVGAWPVINGGAVSLHFAGITLITNFKDGKIDNSLRIRFSLTVDEPPFSILRTRTVNLY
GLPAANPNNGKEYYEISGTFKLISLVPTNDD

VP8* chaperone
>cVP8*-base
MLDGPYQPTTFTPPNDYWILINSNTNGVVYESTNNSDFWTAVVAIEPHVTPVDRQYMIFGESKQFNVSND
SNKWKFLEMFRSSSQNEFYNRRTLTSDTRLVGILKYGGRVWTFHGETPRATTDSSSTANLNNISITIHSEF
YIIPRSQESKCNEYINNGLPPIQN
>cVP8*-M5
MESGWYPPTTWSPPENYYVIVRSTKEGVILSISNNKNLKIFAVVIKPKTPKKEVELPIDGKKEKIYVENTSDK
YKIIFVYKTSENDEWKIIESYETTNNLFAITLKDNKLIYSTGTTPNSTFSSISSSDLENYTVTTYGEYTVRPIEN
EEEILKLLQSGDPPIVP
>cVP8*-M1
MEPIEPNNVVWTSGDPLTTTVEQPPLSGWYPPTTFSPPENYWIIVKSTKEGVLFEISNNKNFYFSVVKIPP
NTPKETKTFKINGKEEKFEVSNTSSKYKYIFLEKTSENDEWKIKSSFETSTNLFAFTIENNKLYYSTGTTPNV
TFTSISIPNLENLTVKLYGRYTVLPKENREEVLELLKNGLPPIVP
>cVP8*-M3
MEPIPPNNVVWVPGDPLTTNVEQPPFSGWYPPTTFSPPENYWIIIKSEKDGLVLEVSNDKNLYLAVVKIPP
NTKEEEKEFPLRGKKETFPVSNTSDKYKYIFLLKTSKNDEWKIIDSFETTNNLFAISLENNKLYYSYGTDPN
NTFKSIESENLENYKITTYGRYTVVPSSQRDQVLEWLRNGEPPVVP
>cVP8*-M13
MLSGPYQPTTFSPPNDYWILINSNTEGVVYESTNNEDFWTAVVAIPPNTPPQDVQLPIFGKSETFNVSNDS
PKWKFLEMVRYSKNSPWINRRSLESDTRLVGILKKGGRLYTFHGTTPNASIDSISIPNLDNISITIYSDFYIIP
KSQESKILELIKNGLPPIVN
>cVP8*-M17
MSGEPIPPNNVVWTSGDPLTSTTEQPPLSGWYPPTKFSPPENYYIIIKSEKDGIVLRIWNKKNLYIAVVIIKP
NTPKEKKTFDLFGKKEEFYVENTSDKYKYIFLYKTSENDEWKIIDYFETKNRLFAITLKNNKLYYSTGVDPN
NTFKSIESEDLENYEVEIYGRYTVLPSSQEDTVLKLLQNGEPPVVP
>cVP8*-M19
MSGEPIPPNNVVWQPGDPLTTTVEQPPLSGWYPPTTFSPPENYYLIVRVNKDGLVLEISNDKNFRLAVVYI
PPNTPEKEVTLPLFGKEETLPVSNTSDKYKYIFILKTSENDEWKIVDSFETKNRLFAITLKDNKLYYSTGTDP
NSTFTSISLPDIENYTVTVYNDYTVLPIDREEEVLELLRNGRPPVVP

Group A Streptococcus

Group A *Streptococcus* (GAS) causes a wide range of clinical diseases, from pharyngitis and skin infections to sepsis, due to the numerous virulence factors employed by the bacterium[1–3]. Even when acute infection resolves itself, complications from GAS infection can include autoimmune responses which lead to rheumatic heart disease (RHD)[4]. GAS is estimated to cause over 500,000 deaths every year, with a majority of these due to RHD, and mortality is projected to increase or remain constant over the next 15 years, highlighting the urgent need for a vaccine[2,4,5]. An incomplete understanding of the autoimmune response to GAS motivates the development of subunit vaccines, which offer the opportunity to exclude antigens that might contribute to autoimmunity[6].

Most vaccine design efforts have focused on the M protein, an important component of the GAS cell wall, which offers potent yet strain-specific protection[6,7]. Unfortunately, M protein is hypervariable, with more than 200 strains identified worldwide[7]. Other vaccine design approaches have sought to identify conserved, highly-expressed, and immunogenic proteins on the surface of GAS[8,9]. These approaches have identified four proteins that fulfil these criteria and upon immunization protect mice from GAS challenge[8]. Only a single example of structure-based vaccine design, on the antigen SpyCEP, has been reported[10], leaving open the possibility that the manufacturability, stability, and immunogenicity of these antigens could be improved. Because of the distinct protein design challenges they pose, we chose to focus on two of these, Streptolysin O (SPy0167) and SpyAD (SPy0269), which are present in multiple leading vaccine candidates for GAS[3,11–13]. Epitopes targeted by antibodies have not been well

mapped for either antigen, with only a single report of an SLO epitope bound by a neutralizing antibody[14].

Streptolysin O (SLO), a member of the family of bacterial cholesterol-dependent cytolysins, is a potent toxin and major virulence factor of GAS[15,16]. SLO monomers bind cholesterol in host cell membranes, form an oligomeric prepore complex, and then undergo a conformational change that allows for insertion into the membrane and host cell lysis[17,18]. As a result, the development of subunit vaccines containing SLO is complicated by the cytotoxicity of the protein and its tendency to oligomerize and aggregate[19]. We reasoned that locking SLO in its monomeric state might improve manufacturability and the potency of vaccine-elicited antibody responses.

We set out to stabilize the monomeric form of SLO, which could prevent the conformational change that facilitates prepore formation (Fig 2-1A). We began with an experimentally determined structure of SLO at 2.1 Å resolution (PDB ID 4HSC)[17], focusing on residues 103-571 of SLO and using the same ProteinMPNN-based two-step process as for rotavirus VP5* to redesign the core of the protein along residues known to be important for cytotoxicity and oligomerization[20]. We filtered sequences by AlphaFold 2 (AF2) prediction, selecting for experimental screening those with the highest model confidence (pLDDT > 92.5) and lowest backbone root mean square deviation (RMSD) to the experimental structure (< 1.8 Å). 48 top-ranking designed antigens were ordered that contained an average of 25 mutations. The mutations predominantly altered hydrophobic packing within the core of SLO, changed

surface exposed residues to more soluble polar residues, or introduced steric changes to residues that play a role in cytotoxicity and oligomerization. These designs were purified by IMAC alongside the sequence from PDB ID 4HSC (denoted SLO-base) and screened by SDS-PAGE to assess soluble yield. Numerous designs displayed higher soluble yield than the native sequence, suggesting that ProteinMPNN sequence redesign and AF2 filtering was able to generate antigens with improved properties (Fig 2-1B).

We then identified a set of preferred mutations that commonly appeared within the twelve designed antigens with the highest yield. In addition, we constructed a clustered MSA and used the MSA, AF2 predictions, and Rosetta energies in the same manner as described for rotavirus VP5* to identify six mutations that might stabilize SLO (Fig 2-2A). In a second round of designs, these MSA-based mutations were combined with varying numbers of the preferred ProteinMPNN-derived mutations. In addition, disulfides identified by DisulfidebyDesign 2.0 and predicted to form in all five AF2 models were introduced into three of the designs[21]. Seven designed variants of SLO were purified alongside SLO-base and the previously reported variant dmSLO, which contains two mutations to attenuate cytotoxicity[19]. Size exclusion chromatography indicated that SLO-base largely formed aggregate, with a smaller dimer peak, while dmSLO largely formed aggregate. By contrast, all seven designed SLO antigens appeared by SEC to be monodisperse monomers (Fig 2-2B). We were unable to purify enough SLO-base or dmSLO for downstream analysis, so we compared our designed antigens to commercially obtained recombinant SLO. Our designed antigens displayed higher

melting and aggregation onset temperatures to recombinant SLO (46-56 and 47-55 °C compared to 41.8 and 42.2 °C, respectively). We chose to focus on SLO-d2 as all of our designed antigens displayed similar properties and SLO-d2 contained the fewest mutations, comprising nine mutations from ProteinMPNN, two found by both ProteinMPNN and MSA-based methods, and four from the MSA alone. DSF analysis revealed that SLO2-d2 possessed higher T_m and T_{agg}, with 10 and 9 °C improvements, respectively, compared to recombinant SLO (Fig 2-2C). These mutations improve packing interactions within the core of the protein, change residues known to mediate cytotoxicity, and alter packing interactions and reduce β-strand propensity within conformationally dynamic helices.

No sequences of conformation-specific mAbs against SLO have been published, complicating analysis of whether our designs retained native-like antigenicity. Nevertheless, we reasoned that we could conduct ELISAs akin to the anti-streptolysin O (ASO) test used during diagnosis of GAS infection[22]. ELISAs with serial dilutions of human convalescent sera indicated that SLO-d2 had comparable or slightly improved reactivity relative to recombinant SLO, indicating that it is recognized by polyclonal antibodies elicited by natural infection (Fig 2-4B). Together, these data show that combining ProteinMPNN- and MSA-derived stabilizing mutations improved the yield, monodispersity, and stability of SLO, a potentially important vaccine antigen for GAS, without deleteriously affecting its antigenicity.

The role of SpyAD in GAS pathogenesis is less well understood, though reports indicate that it may play roles in preventing cell aggregation, facilitating cell division, and enabling adhesion to mammalian epithelial cells[23]. No experimentally determined structure of SpyAD is available, but sequence analysis and structures of homologous PrgA from *Enterococci* indicate that the extracellular portion of SpyAD contains regions at the N and C terminus that interact to form a helical coiled-coil, separated by an intervening globular domain[23,24]. SpyAD can be recombinantly produced in *E. coli* and is conjugated to GAS carbohydrates in a leading vaccine candidate, but its low thermal stability (reported T_m = 37 ºC) might diminish its immunogenicity[12].

We set out to improve the thermal stability of SpyAD, beginning by constructing an atomic model using AF2. We focused on residues 57-743 of the protein, as these form the extracellular portion containing secondary structure, and used all five models of AF2 with homologous *Enterococcus* PrgA (PDB ID 6Z9L) serving as a structural template[23,24]. The resultant predictions displayed reasonable agreement to PrgA in each coiled coil domain, with more divergence across the globular domain (backbone RSMDs of 2.0-3.8 and 5.3 Å for coiled coil domains and the globular domain, respectively), consistent with the 22.6% identity between *Enterococcus* PrgA and SpyAD, and high model confidence throughout the protein (pLDDTs of 86-88.5) (Fig 2-3A). The only areas of deviation and low confidence were in the short, flexible loops that connect coiled-coil dimers. These results suggest that the AF2-predicted structures might be sufficiently accurate to enable structure-guided design.

We used all five predicted structures of SpyAD as inputs to ProteinMPNN, designing both the coiled-coil and globular domains. 200 output sequences for each predicted structure were compared to look for mutations that occurred with frequency of more than 25%, increasing our confidence that the mutation might be beneficial and not an artifact of one particular model's prediction. Positions where mutations did not appear to be deleterious by visual inspection in PyMOL for all five structures were designed with ProteinMPNN in a subsequent step, while all other residues were fixed. In that step, ProteinMPNN was used to obtain 200 output sequences where only one of the coiled coil or globular domains was varied. We predicted these sequences and filtered them by AF2 model confidence (pLDDT > 90) and backbone RMSD (< 1.5 Å for coiled-coil dimers, < 2.5 Å for globular domain) within the coiled-coil and globular domains. RMSD across the entire protein was not informative, as small changes in the flexible loops lead to large changes in the position of downstream domains. 48 top-ranking designs were ordered alongside a construct comprising residues 57-743 of SpyAD (SpyAD-base). The introduced mutations predominantly altered hydrophobic packing and reduced the number of solvent-exposed hydrophobic residues. Multiple designs displayed comparable expression to SpyAD-base, similar monodispersity, and improved thermal stability in a preliminary screen. As these designs only contained mutations in one of the coiled-coil dimers or the globular domain, these data allowed us to pinpoint which portions of SpyAD contributed the most to its thermal instability. Designs with mutations in the globular domain had the clearest improvements in thermal stability (data not shown).

Unlike SLO, the clustered MSA for SpyAD only contained sequences with greater than 98% identity or less than 70% identity, so it was not useful for identifying potential mutations. The ProteinMPNN-derived mutations common to the SpyAD designs with the highest thermal stability were identified in different regions of the protein. Designs that combined increasing numbers of these mutations with disulfides suggested by DisulfidebyDesign 2.0 were ordered and characterized alongside SpyAD-base. Of these, SpyAD-d3 had comparable thermal stability to the other designed antigens while containing fewer mutations and was thus characterized in more detail. SpyAD-d3 possessed comparable yield and significantly higher T_m and T_{agg} than SpyAD-base (18 and 22 °C improvements, respectively) (Fig 2-3C and 2-3D). SpyAD-d3 contains twelve mutations, all of which are in the globular domain, that alter the hydrophobicity and packing of solvent-exposed residues, introduce three disulfide bonds, and may reduce conformational flexibility (Fig 2-3B).

Monoclonal antibodies against SpyAD have not been reported, so we used ELISAs with serial dilutions of human convalescent sera to assess antigenicity. We found that SpyAD-d3 had comparable but slightly reduced reactivity to SpyAD-base, indicating that it is similarly recognized by polyclonal antibodies elicited by natural infection (Fig 2-4C). Together, these data show that AF2-predicted structures can enable structure-based design, providing a model accurate enough for the introduction of ProteinMPNN-derived stabilizing mutations and disulfides that improved the thermal stability of the vaccine candidate antigen SpyAD while preserving antigenicity.

References

1. Walker, M. J. *et al.* Disease manifestations and pathogenic mechanisms of Group A Streptococcus. *Clin. Microbiol. Rev.* **27**, 264–301 (2014).
2. Walkinshaw, D. R. *et al.* A Strep A vaccine global demand and return on investment forecast to inform industry research and development prioritization. *npj Vaccines* **8**, 1–13 (2023).
3. Bi, S. *et al.* A Multicomponent Vaccine Provides Immunity against Local and Systemic Infections by Group A Streptococcus across Serotypes. *MBio* **10**, (2019).
4. Watkins, D. A. *et al.* Global, regional, and national burden of rheumatic heart disease, 1990-2015. *N. Engl. J. Med.* **377**, 713–722 (2017).
5. Vekemans, J. *et al.* The Path to Group A Streptococcus Vaccines: World Health Organization Research and Development Technology Roadmap and Preferred Product Characteristics. *Clin. Infect. Dis.* **69**, 877–883 (2019).
6. Walkinshaw, D. R. *et al.* The Streptococcus pyogenes vaccine landscape. *NPJ Vaccines* **8**, 16 (2023).
7. Rivera-Hernandez, T. *et al.* Differing efficacies of lead group A streptococcal vaccine candidates and full-length M protein in cutaneous and invasive disease models. *MBio* **7**, (2016).
8. Bensi, G. *et al.* Multi high-throughput approach for highly selective identification of vaccine candidates: the Group A Streptococcus case. *Mol. Cell. Proteomics* **11**, M111.015693 (2012).
9. Davies, M. R. *et al.* Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. *Nat. Genet.* **51**, 1035–1043 (2019).
10. McKenna, S. *et al.* Structure, dynamics and immunogenicity of a catalytically inactive CXC chemokine-degrading protease SpyCEP from Streptococcus pyogenes. *Comput. Struct.*

Biotechnol. J. **18**, 650–660 (2020).

11. Rivera-Hernandez, T. *et al.* Vaccine-induced Th1-type response protects against invasive group A Streptococcus infection in the absence of opsonizing antibodies. *MBio* **11**, (2020).
12. Di Benedetto, R. *et al.* Rational Design of a Glycoconjugate Vaccine against Group A Streptococcus. *Int. J. Mol. Sci.* **21**, (2020).
13. McCabe, S. *et al.* The group A streptococcal vaccine candidate VAX-A1 protects against group B Streptococcus infection via cross-reactive IgG targeting virulence factor C5a peptidase. *Vaccines (Basel)* **11**, 1811 (2023).
14. Tang, D. *et al.* Multimodal mass spectrometry identifies a conserved protective Epitope in S. pyogenes Streptolysin O. *Anal. Chem.* **96**, 9060–9068 (2024).
15. Kehoe, M. A., Miller, L., Walker, J. A. & Boulnois, G. J. Nucleotide sequence of the streptolysin O (SLO) gene: structural homologies between SLO and other membrane-damaging, thiol-activated toxins. *Infect. Immun.* **55**, 3228–3232 (1987).
16. Tweten, R. K. Cholesterol-dependent cytolysins, a family of versatile pore-forming toxins. *Infect. Immun.* **73**, 6199–6209 (2005).
17. Feil, S. C., Ascher, D. B., Kuiper, M. J., Tweten, R. K. & Parker, M. W. Structural studies of Streptococcus pyogenes streptolysin O provide insights into the early steps of membrane penetration. *J. Mol. Biol.* **426**, 785–792 (2014).
18. van Pee, K. *et al.* CryoEM structures of membrane pore and prepore complex reveal cytolytic mechanism of Pneumolysin. *Elife* **6**, (2017).
19. Chiarot, E. *et al.* Targeted amino acid substitutions impair streptolysin O toxicity and group A Streptococcus virulence. *MBio* **4**, e00387–12 (2013).
20. Magassa, N. 'goundo, Chandrasekaran, S. & Caparon, M. G. Streptococcus pyogenes cytolysin-mediated translocation does not require pore formation by streptolysin O. *EMBO Rep.* **11**, 400–405 (2010).
21. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for disulfide

engineering in proteins. *BMC Bioinformatics* **14**, 346 (2013).

22. Sen, E. S. & Ramanan, A. V. How to use antistreptolysin O titre. *Arch. Dis. Child. Educ. Pract. Ed.* **99**, 231–238 (2014).
23. Gallotta, M. *et al.* SpyAD, a moonlighting protein of group A Streptococcus contributing to bacterial division and host cell adhesion. *Infect. Immun.* **82**, 2890–2901 (2014).
24. Schmitt, A. *et al.* Enterococcal PrgA extends far outside the cell and provides surface exclusion to protect against unwanted conjugation. *J. Mol. Biol.* **432**, 5681–5695 (2020).
25. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
26. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
27. Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
28. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res.* **53**, D20–D29 (2025).
29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
30. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
31. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
32. The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC.
33. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).

Figures

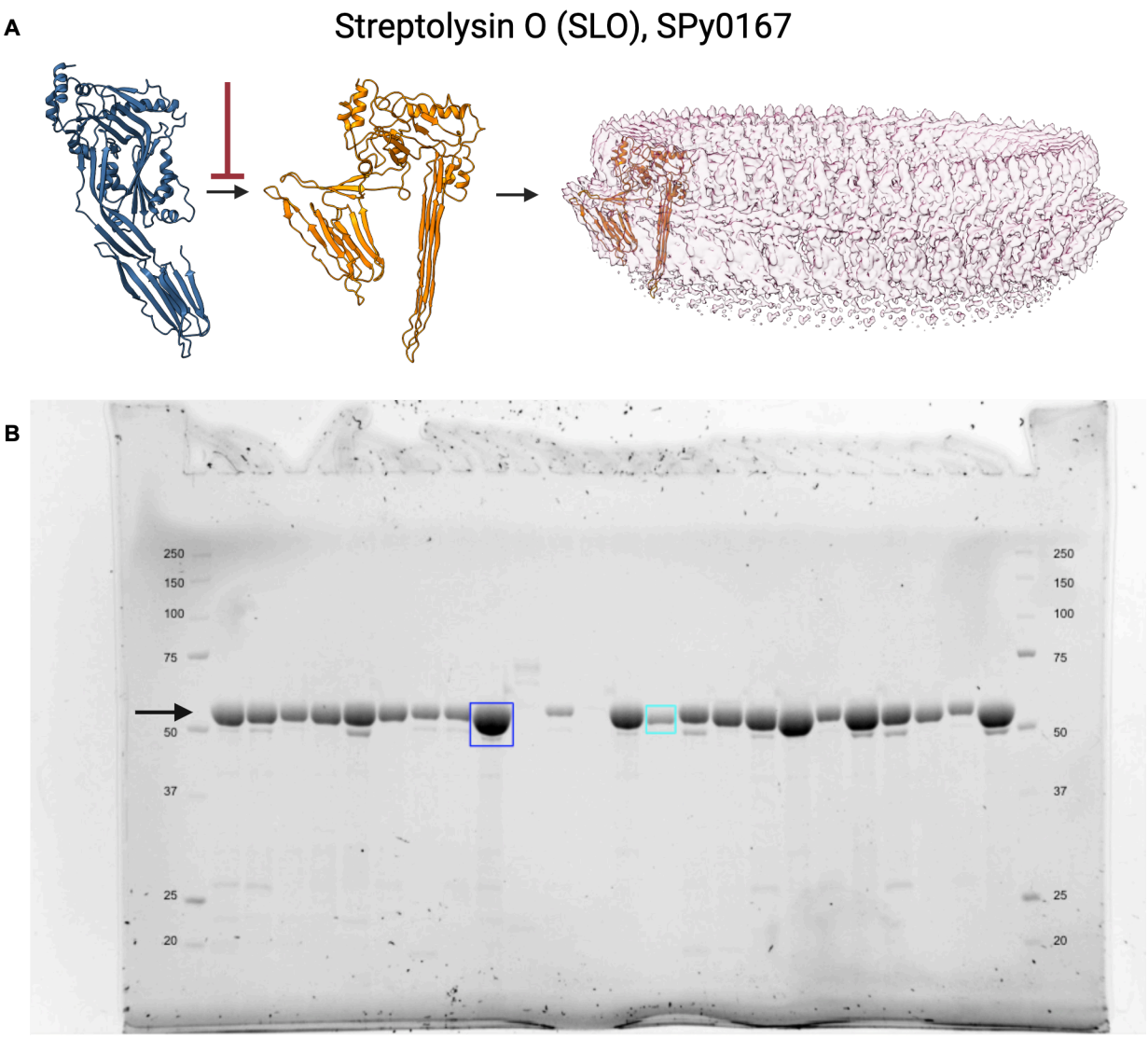


Figure 2-1. Scheme of stabilization goal and improved yield with redesign. A) SLO transitions from a monomeric conformation in blue (left) to a prepore conformation in orange (center) that assembles into a large, oligomeric, and cytotoxic pore. The primary design goal was to stabilize the monomer to inhibit changes in conformation and oligomerization. B) SDS-PAGE gel with each lane loaded with IMAC eluate from a separate SLO design. 23 ProteinMPNN-redesigned SLO antigens were loaded alongside SLO-base. SLO-base is highlighted with a light blue box, while a highly expressing design is highlighted with a dark purple box. The arrow on the left indicates the expected molecular weight of SLO.

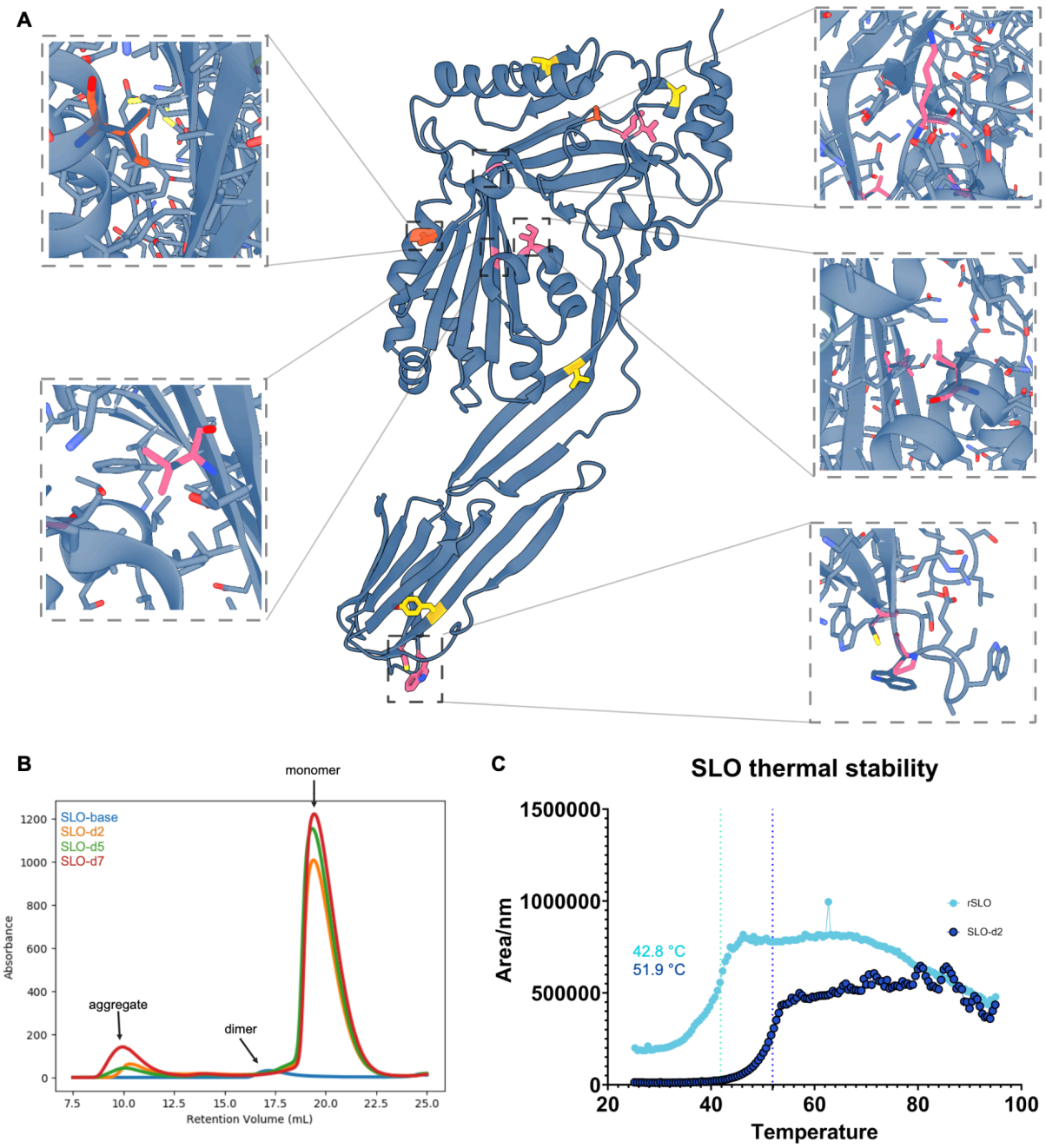


Figure 2-2. ProteinMPNN- and MSA- derived mutations improve yield, monodispersity, and thermal stability of SLO. A) Structure of SLO antigen (PDB ID 4HSC) shown in cartoon representation with positions in SLO-d2 with ProteinMPNN-derived mutations shown as pink sticks and positions with MSA-derived mutations shown as yellow sticks. Mutations identified by both approaches are shown as orange sticks. Magnified insets of select mutations show the original residue in blue and the introduced residue in yellow or pink. Clockwise from top right, mutations are G396K, I273L, W537P and C530S, A296V, and V415I. B) Size exclusion chromatography (SEC) of SLO antigens on Superdex 200 Increase 10/300 GL column. C) NanoDSF with SYPRO dye of recombinant SLO (rSLO) from Sigma Aldrich and SLO-d2, with melting temperature listed and shown by a vertical dashed line.

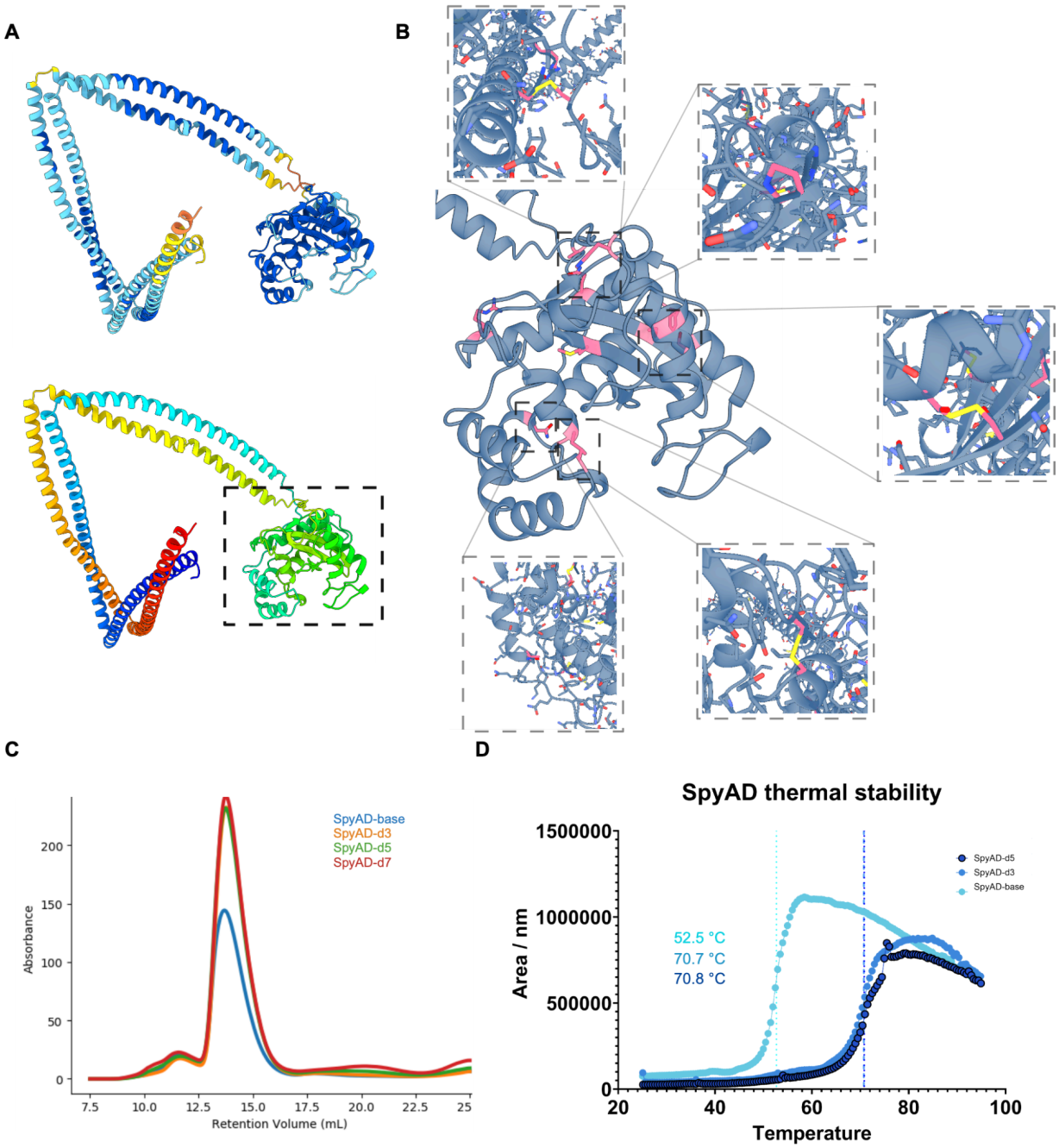


Figure 2-3. AlphaFold 2 prediction enables structure-based design of SpyAD, and ProteinMPNN-derived mutations improve thermal stability. A) Structure of SpyAD antigen from AlphaFold 2 (AF2) prediction is shown in cartoon representation. On top, the structure is colored by AF2 pLDDT, with portions ranging from high confidence (dark blue and blue) to intermediate (yellow) and low confidence (orange). On the bottom, the structure is colored by sequence position, with the N terminus in blue and the C terminus in red. The globular domain, the domain with the lowest thermal stability, is noted with a dashed box. B) The predicted structure of the globular domain is shown with cartoon representation in blue, with positions mutated in SpyAD-d3 shown as pink sticks. Magnified insets of select mutations show the original residue in blue and the introduced residue in pink. Clockwise from top left, mutations are Q349C and A525C, H515P, S381C and S493C, I274C and G471C, and N318L. C)

Size exclusion chromatography (SEC) of SpyAD antigens on Superdex 200 Increase 10/300 GL column. D) NanoDSF with SYPRO dye of SpyAD antigens, with melting temperature listed and shown by a vertical dashed line. Due to similarity of melting temperature for SpyAD-d3 and SpyAD-d5, it is difficult to distinguish their corresponding lines.

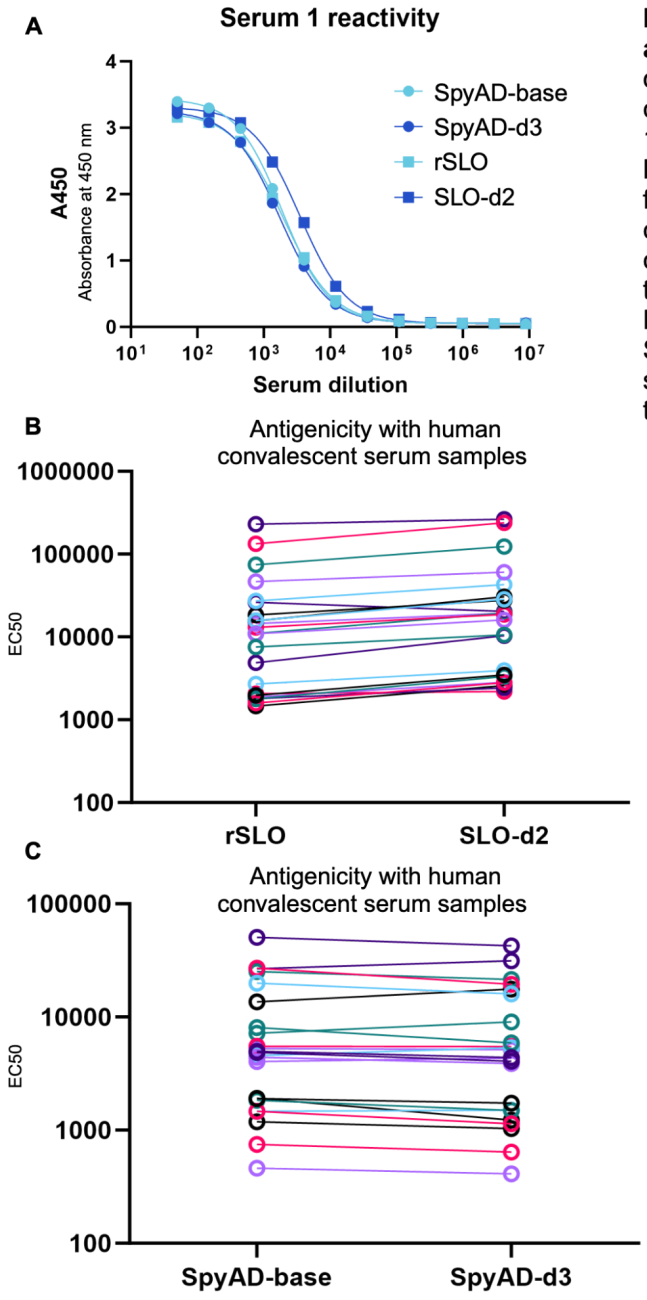


Figure 2-4. Designed antigens retain antigenicity. A) Representative ELISA against designed SLO and SpyAD antigens with human convalescent sera. Initial serum dilution was 1:50, with eleven subsequent three-fold dilutions. B) EC50 values of 23 human convalescent sera for rSLO and SLO-d2. EC50 values were calculated in GraphPad Prism using ELISA curves. For each serum sample, a line connects the EC50 value for rSLO to that for SLO-d2. C) EC50 values of 23 human convalescent sera for SpyAD-base and SpyAD-d3. For each serum sample, a line connects the EC50 value for rSLO to that for SLO-d2.

Materials and Methods

AlphaFold Prediction

The ColabFold implementation of AlphaFold was downloaded from the github repository into an apptainer environment with the necessary dependencies[25,26]. All predictions were made on either A4000 or A6000 GPUs on a custom high performance computing cluster. Sequences of candidate designed antigens were input into the ColabFold implementation to create MSAs using mmseqs2 and to generate structural predictions using AlphaFold2_ptm. Predictions were made with structural templates and with an amber relaxation step. For both SLO and SpyAD antigens, predictions were made with all available PDB templates found by ColabFold during structure prediction by including the --templates flag. For each sequence predicted, structural predictions were obtained with all 5 AlphaFold models.

Following structure prediction, analysis with python scripts in a jupyter notebook was performed to determine prediction confidence by examining average pLDDT and PAE across the monomer or oligomers. All 5 predictions for a given sequence were used to determine the mean and standard deviation for pLDDT and PAE across the different models. Python scripts were used to calculate root mean squared deviation (RMSD) between the desired structure, either an experimental structure or high confidence prediction of the native sequence, and the structures predicted for designed sequences.

Rosetta energy evaluation

Relaxed AlphaFold predictions were input into a PyRosetta script that calculated a biophysical Rosetta energy score using the ref2015 score function[27]. Rosetta energy scores for predicted structures from all 5 ColabFold models were used to calculate the mean energy and standard deviation for each sequence.

Multiple sequence alignment (MSA) based design

Standard MSAs were generated using NCBI protein BLAST with the standard databases and a maximum number of 500 sequences[28–31]. Clustered MSAs were generated using NCBI protein BLAST with the experimental databases option and a maximum number of 250 sequences. MSAs were filtered to greater than 70% identity to the query sequence with coverage greater than 50%, and downloaded as aligned FASTAs. Analysis with python scripts in a jupyter notebook was performed to determine the residue positions where more than 25% of the sequences contained the same amino acid which differed from the corresponding amino acid in the query sequence. The standard and clustered MSAs were used to generate two distinct sets of suggested mutations. Experimental structures or high-confidence predicted structures of the corresponding antigen were inspected in PyMol with the mutagenesis wizard tool to determine if the mutation would be tolerated[32]. If the mutation did not result in large steric clashes or the loss of two or more hydrogen bonds, it was retained in a list of MSA suggested mutations, resulting in a standard and clustered list of potential mutations.

Both lists of potential mutations were used to generate point mutations to the sequence of the native antigen. These sequences differing by one amino acid from the native antigen were input into AlphaFold, and the output structural predictions were evaluated for their RMSD, pLDDT, PAE, and Rosetta energies. Mutations that resulted in worse structural predictions relative to the prediction for the native sequence (RMSD > 1 Å higher, pLDDT 0.5 lower, PAE 0.5 higher, or 3 REU higher) were discarded. Remaining mutations were incorporated into an MSA based designed antigen.

ProteinMPNN

ProteinMPNN was run using an implementation downloaded from the github repository[33]. All sequence design was conducted on CPUs or an A4000 GPU on a custom high performance computing cluster. An experimental structure, and all 5 high-confidence ColabFold predictions for the native sequence were used for antigens lacking experimental structures. ProteinMPNN was run with a sampling temperature of 0.2, and an amino acid bias of -0.8 for glycine and alanine. Cysteines were disallowed during design.

ProteinMPNN was employed in a 2-step process for antigens. The process for SpyAD is described in the results, and the process for SLO is described here. In the first step, structures served as inputs to ProteinMPNN for sequence design with all residues allowed to mutate that were not disallowed for design. ProteinMPNN was run for 200 batches with 5 sequences each, to generate 1000 designed sequences for each input structure. A python script was used to analyze the frequency of mutations from the

native amino acid at each position. Where ProteinMPNN chose a mutation at a given position in more than 25% of the sequences, the mutation was manually inspected with the PyMol mutagenesis wizard in the context of the input structure. If any suggested mutation caused large steric clashes or the loss of at least two hydrogen bonds, the residue was no longer considered for design. In this way, the first round of sequence design was used to identify designable positions where ProteinMPNN suggested mutations that did not appear to be deleterious.

In the second round, only these designable residues were allowed to mutate, and ProteinMPNN was run for 40 batches with 5 sequences each to generate 200 potential sequences. Residues G396 and C530 were explicitly designed to reduce oligomerization and cytotoxicity. These sequences were then evaluated by AlphaFold structure prediction. Predicted structures with the lowest RMSD and highest pLDDT were selected as candidate designs, and their sequences were ordered for experimental evaluation.

Protein Expression

For initial screening of designs in *E. coli*, codon optimized DNA sequences were ordered as gene fragments (IDT or Twist), cloned into a pET-29b+ based plasmid using golden gate cloning with BsaI-HFv2 (NEB), and transformed into chemical competent BL21(DE3) cells (NEB). Cells were grown for 16 hours at 37°C in 1 mL of LB broth with antibiotic before inoculating 500 uL of culture into 50 mL of Terrific Broth II containing antibiotic. Cells were grown at 37°C to an OD of 0.6-0.8 before adding IPTG to a final

concentration of 0.25 mM and cells were grown at 37°C for 4 hours. Cells were centrifuged at 4000 g for 15 minutes to harvest, and pellets were stored at -80°C.

For constructs used for more detailed comparisons with the native antigen, plasmids encoding codon optimized sequences were ordered from GenScript. Protein expression was conducted in the same manner as above.

Purification

For purification of proteins from *E. coli*, cells were resuspended in 8 mL of lysis buffer (50 mm Tris, 250 mM NaCl, 20 mM imidazole, 0.5 mg/mL lysozyme, 1 U / mL benzonase, 0.02 tablets/mL of protease inhibitor cocktail) and lysed by ultrasonication (Qsonica 24-tip sonicator, 6 minutes total, 5s on, 5s off, 80% amplitude). Lysate was clarified by centrifugation at 14000 g for 30 min. Clarified lysate was applied to 0.5 mL of Ni-NTA resin (Qiagen) equilibrated with a wash buffer (50 mm Tris, 250 mM NaCl, 20 mM imidazole). Resin was then washed with 20 column volumes of wash buffer. Protein was eluted with 2.4 column volumes of elution buffer (50 mm Tris, 250 mM NaCl, 500 mM imidazole) and analyzed by SDS-PAGE gel. Eluted proteins were further filtered by size exclusion chromatography using a Superdex 200 Increase 10/300 GL column (Cytiva) on an ÄKTA pure (Cytiva).

Melting and aggregation temperature determination:

Determination of melting and aggregation temperatures were done by differential scanning fluorimetry with the UNcle instrument and software (UNchained Labs). SEC

purified proteins at a concentration between 0.25-1 mg/mL were added to 8.8 uL quartz capillary wells in a 16-well cassette (UNchained Labs). With the temperature of the cassette changing from 25 to 95 C at a thermal ramp rate of 1C/min, intrinsic changes in protein fluorescence emission were used to evaluate melting and aggregation temperatures. As an orthogonal measurement of the melting temperature, protein was mixed with the hydrophobic dye Sypro Orange to evaluate change in fluorescence emission during the thermal ramp.

Enzyme-linked immunosorbent assay (ELISA):

Antigenicity evaluations for GAS SLO and PrgA were conducted in a biological safety cabinet. All incubation steps occurred while shaking at 600 rpm, and all wash steps involved 3 washes with ELISA wash buffer (25 mM Tris, 150 mM NaCl, 0.2% Tween 20, pH 8.0) by an automated plate washer (Biotek). 50 uL of SEC purified antigen at 2 ug/mL was incubated onto a half area well high binding plate (Corning) for 1.5 hours. 70 uL of ELISA blocking buffer (25 mM Tris, 150 mM NaCl, 0.2% Tween 20, 5% nonfat milk, pH 8.0) was added to each well and incubated for 1 hour. At the primary antibody step of the ELISA, 3 fold serial dilutions of human convalescent sera were made starting at 1:50 in ELISA blocking buffer, and 50 uL was added to each well. Plates were incubated for 1 hour. Plates were then washed and 50 uL of ELISA blocking buffer containing secondary antibody (goat anti-human IgG-HRP, Southern Biotech at a 1:5000 dilution) was added for a 30 minute incubation. Plates were washed, 50 uL of TMB (3,30,50,5-tetramethylbenzidine, SeraCare) was added for 2.5 minutes, the

reaction was quenched with the addition of 50 uL of 1N HCl, and the absorbance at 450 nm of each well was measured with an Epoch plate reader (Biotek).

Acknowledgements
Thank you to all of the scientists whose work on Group A Streptococcus, its lifecycle, and immune responses makes structure-based vaccine design possible.

De-identified human convalescent sera were generously given by Alex Greninger, and we are thankful to the Greninger lab and each of the patients who provided sera.

Maintenance of the computational infrastructure and technical help establishing the Colabfold environment were conducted by Luki Goldschmidt and Patrick Vecchiato.

Suna Cheng kindly conducted some of the replicates of ELISAs using human convalescent sera.

Sequences

Designed SLO antigens

>SLO-base
MSGEDHTEEINDKIYSLNYNELEVLAKNGETIENFVPKEGVKKADKFIVIERKKKNINTTPVDISIIDSVTDRT
YPAALQLANKGFTENKPDAVVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWHDNYSGG
NTLPARTQYTESMVYSKSQIEAALNVNSKILDGTLGIDFKSISKGEKKVMIAAYKQIFYTVSANLPNNPADVF
DKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGGDAAEHNKVVTKDFDVIRNVIKDNATFSRKNPAYPISYTSVFLKNNKIAGVNNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARE
CTGLAWEWWRKVIDERDVKLSKEINVNISGSTLSPYGSITYKGSGSGSGSHHHHHH
>dmSLO
MSGEDHTEEINDKIYSLNYNELEVLAKNGETIENFVPKEGVKKADKFIVIERKKKNINTTPVDISIIDSVTDRT
YPAALQLANKGFTENKPDAVVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWHDNYSGG
NTLPARTQYTESMVYSKSQIEAALNVNSKILDGTLGIDFKSISKGEKKVMIAAYKQIFYTVSANLPNNPADVF
DKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGGDAAEHNKVVTKDFDVIRNVIKDNATFSRKNLAYPISYTSVFLKNNKIAGVNNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARE
CTGLAFEWWRKVIDERDVKLSKEINVNISGSTLSPYGSITYKGSGSGSGSHHHHHH
>SLO-d2
MSGEDHTEEINDKIYSLNYNELEVLAKNGETIENFVPKEGVKKADKFIVIERKKKNINTTPVDISIIDSVTDRT
YPGAIQLANKGFTENKPDALVTKRNPQKIHIDLPGMGDKATIEVNDPTYANVSTAIDNLVNQWHDNYSGGN
TLPARTQYTESMVYSKSQIEAALNVNSKLLDGTLGIDFKSISKGEKKVMIAVYKQIFYTVSANLPNNPSDVF
DKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATFSRKNPAYPISYTSVFLKNNKIAGVNNRTEYIETTSTE
YTSGKINLSHQGAYVAQFEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARES
TGLAWEPWRKVIDERDVKLSKEINVNISGSTLSPYGSITYKGSGSGSGSHHHHHH
>SLO-d3
MSGEDHTEEINDKIYSLNYNELEVLAKNGETIENFVPKEGVKKADKFIVIERKKKNINTTPVDISIIDSVTDRT
YPGAIQLANKGFTENKPDALVTKRNPQKIHIDLPGMGDKATIEVNDPTYANVSTAIDNLVNQWHDNYSGGN
TLPARTQYTESMVYSKSQIEAALNVNSKLLDGTLGIDFKSISKGEKKVMIAVYKQIFYTVSANLPNNPSDVF
DKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATFSRKNPAYPISYTSVFLKNNKIAGINNRTEYIETTSTEY
TSGKINLSHQGAYVAQFEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMAREST
GLAWEPWRKVIDERDVPLSKEINVNISGSTLSPYGSITYKGSGSGSGSHHHHHH
>SLO-d4
MSGEDHTEEINDKIYSLNYNELEVLAKNGETIENFVPKEGVKKADKFIVVEREKKSINTTPVDISIIDSVTDRT
YPGAIQLANKGFTENKPDALVTKRNPQKIHIDLPGMGDKATIEVNDPTYANVSTAIDNLVNQWHDNYSGGN
TLPARTQYTESMVYSKSQIEAALNVNSKLLDGTLGIDFKSISKGEKKVMIAVYKQIFYTVSANLPNNPSDVF
DKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATFSRKNPAYPISYTSVFLKNNKIAGVNNRTEYIETTSTE
YTSGKINLSHQGAYVAQFEISWDEVNYDDKGKEVITKKSWDNNWHSKTSPFSTVIPLGANSRNIRIMARES
TGLAWEPWRKVIDERDVKLSKEINVNISGSTLSPYSSITYKGSGSGSGSHHHHHH
>SLO-d5
MSGEDHTEEINDKIYSLNYNELEVLAKNCETIENFVPKEGVKKADKFIVIERKKKNINTTPVDISIIDSVTDRT
YPGAIQLANKGFTENKPDALVTKRNPQKIHIDLPGMGDKATIEVNDPTYANVSTAIDNLVNQWHDNYSGGN
TLPARTQYTESMVYSKCQIEAALNVNSKLLDGTLGIDFKSISKGEKKVMIAVYKQIFYTVSANLPNNPSDVF
DKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN

SSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATFSRKNPAYPISYTSVFLKNNKIAGVNNRTEYIETTSTE
YTSGKINLSHQGAYVAQFEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARES
TGLAWEPWRKVIDERDVKLSKEINVNISGSTLSPYGSITYKGSGSGSGSHHHHHH
>SLO-d6
MSGEDHTEEINDKIYSLNYNELEVLAKNGETIENFVPKEGVKKADKCIVIERKKKNINTTPVDISIIDSVTDRT
YPGAIQLANKGFTENKPDALVTKRNPQKIHIDLPGMGDKATIEVNDPTYANVSTAIDNLVNQWHDNYSGGN
TLPARTQYTESMVYSKSQIEAALNVNSKLLDGTLGIDFKSISKGEKKVMIAVYKQIFYTVSANLPNNPSDVF
DKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATFSRKNPAYPISYTSVFLKNNKIAGVNNRTEYIETTSTE
YTSGKINLSHQGAYVAQFEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGCNSRNIRIMARES
TGLAWEPWRKVIDERDVKLSKEINVNISGSTLSPYGSITYKGSGSGSGSHHHHHH
>SLO-d7
MSGEDHTEEINDKIYSLNYNELEVLAKNGETIENFVPKEGVKKADKFIVIERKKKNINTTPVDISIIDSVTDRT
YPGAIQLANKGFTENKPDALVTKRNPQKIHIDLPGMGDKATIEVNDPTYANVSTAIDNLVNQWHDNYSGGN
TLPARTQYTESMVYSKSQIEAALNVNSKLLDGTLGIDFKSISKGEKKVMIAVYKQIFYTVSANLPNNPSDVF
DKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSCSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATFSRKNPAYPISYTSVFLKNNKIAGVNNRTEYIETTSTC
YTSGKINLSHQGAYVAQFEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARES
TGLAWEPWRKVIDERDVKLSKEINVNISGSTLSPYGSITYKGSGSGSGSHHHHHH
>SLO-d8
MSGEDHTEEINDKIYSLNYNELEVLAKNGETIENFVPKEGVKKADKFIVIERKKKNINTTPVDISIIDSVSDRT
YPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWHDNYSGG
NTLPARTQYTESMVYSKSQIEAALNVNSKLLDGTLGIDFKSISKGEKKVMIAVYKQIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATFSRKNPAYPISYTSVFLKNNKIAGINNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEVNYDDKGKEVITKRSWDNNWHSKTSPFSTVIPLGANSRNIRIMARE
STGLAWEPWRKVIDERDVKLSKEINVNISGSTLSPSGSITYKGSGSGSGSHHHHHH

Designed SpyAD antigens

>SpyAD-base
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPDN
LTPEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYG
VGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRV
DKHNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKV
SSLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEA
TRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTT
SSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSHHHHHH
>SpyAD-d2
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNTMKAPQGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPD
NLTPEVQNELACFAAHMINSVRRQLGLPPVTVTAGSQEFARLLCTSYKKTHGNTRPSFVYGQPGVSGHY
GVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYCHAINFLR
VDKHNPNAPVYLGFCTSNVGSLNEHFVMFPESNIANHQRFNKTPIKCVGSTKDYAQRVGTVSDTIAAIKG

KVSSLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQL
EATRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAK
TTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSHHHHHH
>SpyAD-d3
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNTMKAPQGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGLQLNQYQDIPADRNRKVDPD
NLTPEVQNELACFAAHMINSVRKQLGLPPVTVTAGSQEFARLLCTSYKKTHGNTRPSFVYGQPGVSGHY
GVGPHDKTIIEDSAEASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYCHAINFLR
VDKHNPNAPVYLGFCTSNVGSLNEHFVIFPESNIANPQRFNKTPIKCVGSTKDYAQRVGTVSDTIAAIKGK
VSSLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLE
ATRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTT
SSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSHHHHHH
>SpyAD-d4
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQCDNEQKALTSAQEIYTNTLASSEET
LLAQGAECQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTCNDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNTMKAPQGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPD
NLTPEVQNELACFAAHMINSVRRQLGLPPVTVTAGSQEFARLLCTSYKKTHGNTRPSFVYGQPGVSGHY
GVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYCHAINFLR
VDKHNPNAPVYLGFCTSNVGSLNEHFVMFPESNIANHQRFNKTPIKCVGSTKDYAQRVGTVSDTIAAIKG
KVSSLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQCDSLNLQVRQLNDTKGSLRTELLAAKAKQAQL
EATRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLCDFKLNPNRLQVIRERIDNTKQDLAK
CTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSHHHHHH
>SpyAD-d5
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQCDNEQKALTSAQEIYTNTLASSEET
LLAQGAECQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTCNDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNTMKAPQGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGLQLNQYQDIPADRNRKVDPD
NLTPEVQNELACFAAHMINSVRKQLGLPPVTVTAGSQEFARLLCTSYKKTHGNTRPSFVYGQPGVSGHY
GVGPHDKTIIEDSAEASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYCHAINFLR
VDKHNPNAPVYLGFCTSNVGSLNEHFVIFPESNIANPQRFNKTPIKCVGSTKDYAQRVGTVSDTIAAIKGK
VSSLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQCDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLE
ATRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLCDFKLNPNRLQVIRERIDNTKQDLAKC
TSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSHHHHHH
>SpyAD-d6
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQCDNEQKALTSAQEIYTNTLASSEET
LLAQGAECQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTCNDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNTMKAPQGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPD
NLTPEVQNELACFAAHMINSVRRQLGLPPVTVTAGSQEFARLLCTSYKKTHGNTRPSFVYGQPGVSGHY
GVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYCHAINFLR
VDKHNPNAPVYLGFCTSNVGSLNEHFVMFPESNIANHQRFNKTPIKCVGSTKDYAQRVGTVSDTIAAIKG
KVSSLENRLSAIEQEADIMAAQAKVSQLQGKLASTLKQCDSLNEQVRQLNDTKGSLRTELKAAKAKQAQL
EATLDQSLAKLASLKAALHQTEADAEQAAARVTALVAKKAHLQYLCDFKLNPNRLQVIRERIDNLKQDLAK
CTSSLKNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEKEHKDEDEATGSGSGSHHHHHH
>SpyAD-d7

MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQCDNEQKALTSAQEIYTNTLASSEET
LLAQGAECQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTCNDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNTMKAPQGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGLQLNQYQDIPADRNRKVDPD
NLTPEVQNELACFAAHMINSVRKQLGLPPVTVTAGSQEFARLLCTSYKKTHGNTRPSFVYGQPGVSGHY
GVGPHDKTIIEDSAEASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYCHAINFLR
VDKHNPNAPVYLGFCTSNVGSLNEHFVIFPESNIANPQRFNKTPIKCVGSTKDYAQRVGTVSDTIAAIKGK
VSSLENRLSAIEQEADIMAAQAKVSQLQGKLASTLKQCDSLNEQVRQLNDTKGSLRTELKAAKAKQAQLE
ATLDQSLAKLASLKAALHQTEADAEQAAARVTALVAKKAHLQYLCDFKLNPNRLQVIRERIDNLKQDLAKC
TSSLKNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEKEHKDEDEATGSGSGSHHHHHH

2-step ProteinMPNN-derived SLO antigens

>SLO_mpnn_13
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQVEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNVFYTVSANLPNNPA
DVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDIL
ENSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARV
CTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYKGSGSGSGSHHHHHH
>SLO_mpnn_21
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFLVIERKKKNINTTPVDISIDDSVSD
RTYPGSLQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWYDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGADAAEHNKVVTKDFDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGPNSRNIRIMARV
CTGLAWEWWRKVIDERDVKLSKEINVNISGSTLSPSGKITYKGSGSGSGSHHHHHH
>SLO_mpnn_32
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEY
TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_63
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEYT
SGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCTG
LAWEWWRKVIDERDVKLSKEINVNISGSTLNPTGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_68
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIEIERKKKNINTTPVDISIDDSVSDR
TYPGALQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWYDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD

VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYKGSGSGSGSHHHHHH
>SLO_mpnn_89
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVSD
RTYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWYDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPARPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_93
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFYVIERKKKNINTTPVDISIDDSVSD
RTYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYSGSGSGSGSHHHHHH
>SLO_mpnn_103
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGALQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYEGSGSGSGSHHHHHH
>SLO_mpnn_153
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFFVIERKKKNINTTPVDISIDDSVAD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNVFYTVSANLPNNPA
DVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDIL
ENSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARV
CTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYEGSGSGSGSHHHHHH
>SLO_mpnn_155
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVADR
TYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGADAAEHNKVVTKDIDVVRNIIKDNATYSRKNPAIPISYTSVFLKNNKIAGVNNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_177
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFKVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE

NSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYEGSGSGSGSHHHHHH
>SLO_mpnn_188
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWYDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKHIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEY
TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYKGSGSGSGSHHHHHH
>SLO_mpnn_209
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATFSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEYT
SGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCTG
LAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_218
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATFSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEYT
SGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCTG
LAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGNITYKGSGSGSGSHHHHHH
>SLO_mpnn_220
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYNGSGSGSGSHHHHHH
>SLO_mpnn_229
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWYDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGADAAEHNKVVTKDIDVIRNIIKDNATFSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEYT
SGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCTG
LAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYKGSGSGSGSHHHHHH
>SLO_mpnn_246
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSLQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATYSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE

YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_266
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFLVIERKKKNINTTPVDISIDDSVSD
RTYPGALQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYS
GGNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNP
ADVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSD
ILENSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTS
TEYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMAR
VCTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_328
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFYVIERKKKNINTTPVDISIDDSVSD
RTYPGSLQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYS
GGNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNP
ADVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSD
ILENSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTS
TEYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMAR
VCTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYSGSGSGSGSHHHHHH
>SLO_mpnn_332
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEYT
SGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCTG
LAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_352
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVSD
RTYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_364
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIEIERKKKNINTTPVDISIDDSVSDR
TYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEY
TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLLPSGKITYKGSGSGSGSHHHHHH
>SLO_mpnn_382
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATFSRKNPAIPISYQSVFLKNNKIAGINNRTEYVETTSTEY

TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGPNSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLSPSGSITYSGSGSGSGSHHHHHH
>SLO_mpnn_428
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKHIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEYT
SGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCTG
LAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYSGSGSGSGSHHHHHH
>SLO_mpnn_448
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVTDR
TYPGALQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWYDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNVFYTVSANLPNNPA
DVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDIL
ENSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATGSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARV
CTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGKITYKGSGSGSGSHHHHHH
>SLO_mpnn_469
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFLVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNVFYTVSANLPNNPA
DVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDI
LENSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATFSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGPNSRNIRIMARV
CTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_501
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEYT
SGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCTG
LAWEWWRKVIDERDVKLSKEINVNISGSTLDPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_534
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVIIERKKKNINTTPVDISIDDSVSDR
TYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNVFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_540
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFFVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWYDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATGSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE

YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLKPSGSITYEGSGSGSGSHHHHHH
>SLO_mpnn_612
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVTERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWYDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYQSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGPNSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_667
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFFVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_694
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGPNSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGKITYKGSGSGSGSHHHHHH
>SLO_mpnn_722
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVIIERKKKNINTTPVDISIDDSVSDR
TYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWYDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEY
TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYNGSGSGSGSHHHHHH
>SLO_mpnn_732
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFFVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATFSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_735
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEY

TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYNGSGSGSGSHHHHHH
>SLO_mpnn_821
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWYDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGVNNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYDGSGSGSGSHHHHHH
>SLO_mpnn_823
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKILDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEYT
SGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCTG
LAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_825
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFFVIERKKKNINTTPVDISIDDSVSD
RTYPGALQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYS
GGNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNP
ADVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSD
ILENSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTS
TEYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMAR
VCTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_842
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYNGSGSGSGSHHHHHH
>SLO_mpnn_852
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFFVIERKKKNINTTPVDISIDDSVSD
RTYPGAIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNVFYTVSANLPNNPA
DVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVFVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDI
LENSSFTAVVLGADAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARV
CTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYSGSGSGSGSHHHHHH
>SLO_mpnn_881
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWWDNYS
GGNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNP
ADVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSD
ILENSSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATASRKNPARPISYTSVFLKNNKIAGINNRTEYVETT

STEYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMA
RVCTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPTGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_892
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNVFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATFSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTE
YTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVC
TGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGKITYKGSGSGSGSHHHHHH
>SLO_mpnn_898
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVKERKKKNINTTPVDISIDDSVSD
RTYPGSVQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWEDNYS
GGGNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNP
ADVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSD
ILENSSFTAVVLGKDAAEHNKVVTKDFDVIRNIIKDNATYSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETT
STEYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMA
RVCTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGQITYKGSGSGSGSHHHHHH
>SLO_mpnn_948
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVAD
RTYPGALQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYS
GGGNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNP
ADVFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSD
ILENSSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTS
TEYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMAR
VCTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGKITYDGSGSGSGSHHHHHH
>SLO_mpnn_956
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEY
TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYKGSGSGSGSHHHHHH
>SLO_mpnn_965
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFVVIERKKKNINTTPVDISIDDSVSD
RTYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSG
GNTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPAD
VFDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILE
NSSFTAVVLGKDAAEHNKVVTKDFDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTST
EYTSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARV
CTGLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGSITYKGSGSGSGSHHHHHH
>SLO_mpnn_968
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSLQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLINQWEDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVIRNIIKDNATYSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEY

TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYEGSGSGSGSHHHHHH
>SLO_mpnn_980
MSGEDHTEEINDKIYSLNFNELEVLAKNGETVENFVPKEGVKKADKFIVIERKKKNINTTPVDISIDDSVSDR
TYPGSIQLANKGFTENKPDAIVTKRNPQKIHIDLPGMGDKATVEVNDPTYANVSTAIDNLVNQWFDNYSGG
NTLPARTQYTESMVYSKSQTEAALNVNSKLLDGTLGIDYKSISKGEKKVMIAVYKNIFYTVSANLPNNPADV
FDKSVTFKELQRKGVSNEAPPLFVSNVAYGRTVLVKLETSSKSNDVEAAFSAALKGTDVKTNGKYSDILEN
SSFTAVVLGKDAAEHNKVVTKDIDVVRNIIKDNATLSRKNPAIPISYTSVFLKNNKIAGINNRTEYVETTSTEY
TSGKINLSHQGAYVAQYEILWDEINYDDKGKEVITKRRWDNNWYSKTSPFSTVIPLGANSRNIRIMARVCT
GLAWEWWRKVIDERDVKLSKEINVNISGSTLNPSGEITYKGSGSGSGSHHHHHH

ProteinMPNN-derived SpyAD antigens with mutations in single segment

Note: cc1 is coiled coil domain 1, cc2 is coiled domain 2, etc. and glob is globular domain
Some designs only contain the globular domain and one or two of the coiled coil domains

>SpyAD_cc3-base_cc2-base_cc1-d15_glob-d29
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIF
GNNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHKDQIWAKSSPGNQLNQYQDIPAYRNVEVDPDN
LTPEVLATLANFARHLINSVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVG
PHDKTIIEDSAGASGLIRNPDNDYLVIGATNDDHTVDGIMRNIFDFIKYALFDSHEHGNTFEHAIRLLMTDKH
NPNAPVYLGLAFSNVGSLNVHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSLE
NRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEATRD
QSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLL
NAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSCETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTCNDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVDK
KNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQCDSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALCEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_glob-base
MSGPSTQDSIVGNNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPA
DRNRFVDPDNLTPEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYG
QPGVSGHYGVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNT
YGHAINFLRVDKHNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKGSGSGSGSH
HHHHH
>SpyAD_cc3-base_cc2-d14_cc1-d4_glob-d2
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
RLAQAAETQRELTAAETELHNAQADLHSKETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNP
DAITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIV
GNNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNL

SPEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGV
GPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVD
KKNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSS
LENRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEAT
RDQALAKLASLKAALHQQEALAEQAAARVTALEAKIAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTS
SLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSCETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTCNDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYCHAINFLRVDK
KNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQCDSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALCEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSCETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYCHAINFLRVDK
KNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALCEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-base_cc1-base_glob-d26
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSL
FGNNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEVDPD
NLTPEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGV
GPHDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLLMVD
KHNPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQRVGTVSDTIAAIKGKVS
SLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEAT
RDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTS
SLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-d14_cc1-d4_glob-d2
MSGSEETRLAQAAETQRELTAAETELHNAQADLHSKETALSERKASISAETTRVQDLEEQVKTAEQNIAKL
NALLSNPDAITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAP
STQDSIVGNNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNRE
VDPDNLSPEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGV
NGHYGVGPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHA
INFLRVDKKNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAAL
KGKLSSLENRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELEAAKAKL
AQLEATRDQALAKLASLKAALHQQEALAEQAAARVTALEAKIAHLQYLRDFKLNGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-d16_cc1-base_glob-d3

MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETR
LAQAAETQRELTATETELHNAQADYHSKETALSEQKASISAETTRAQDLEEQVKTLEQNIAKLNALLSNPD
AITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSIV
GNNSMKAPEGYPLEELKKLAASGYIGSASYNNYYKEHADQIKAKASPGLSLNQYQDIPADRNREVDPDNL
SPEVQNELAQFAAHLINSVRRQLGLPPVTVTAGSQEFARTLSTSYKKTHGNTRPSFVYGQPGVNGHYGV
GPHDKTIIEDSAGASGLIRNDDFMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGNAINFLRVD
KKNPNAPIYLGFSTSNVGSLNYHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQRVGTVSDTIAAIKGKVSS
LENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELEAAKAKLAQAEATLD
QTLAKLASLKAALHQKEALAEQAAARVTALEAKIAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_glob-d26
MSGPSTQDSLFGNNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIP
AYRNVEVDPDNLTPEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYG
QPGVSGHYGVGPHDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNN
FSHAIRLLMVDKHNPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKGSGSGSGSH
HHHHH
>SpyAD_cc3-base_cc2-base_cc1-d15_glob-d26
MSGPDAITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPST
QDSLFGNNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEV
DPDNLTPEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGH
YGVGPHDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLL
MVDKHNPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKG
KLSSLENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNGSGSGSGSHHHHHH
>SpyAD_glob-d1
MSGPSTQDSIVGNNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIKAKASPGNQLNQYQDIPA
DRNREVDPDNLSPEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVY
GQPGVNGHYGVGPHDKTIIEDSAAASGLIKNDDNMYENIGATNDDHTVNGIKRGIYDTIKYMLFTDHEHGN
TYGHAINFLRVDKHNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAKGSTKGSGSGSGS
HHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSKETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVDK
KNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALAEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-base_cc1-d15_glob-d26
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSLF
GNNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEVDPDNL
TPEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVG
PHDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLLMVDK
HNPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSS
LENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEAT

RDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTS
SLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQCDNLQKALTSAQEEYTNLLASSEET
RLAQAAETQRELTAAETELHNAQADLHSKETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNP
DAITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIV
GNNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNL
SPEVQNELAQFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGV
GPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVD
KKNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSS
LENRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEAT
RDQALAKLASLKAALHQQEALAEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKCTSS
LKNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-d14_cc1-d15_glob-d3
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSKETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLAASGYIGSASYNNYYKEHADQIKAKASPGLSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHLINSVRRQLGLPPVTVTAGSQEFARTLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAGASGLIRNDDFMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGNAINFLRVDK
KNPNAPIYLGFSTSNVGSLNYHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSL
ENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALAEQAAARVTALEAKIAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSCETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTCNDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVDK
KNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQCDSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALCEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-base_cc1-d15_glob-d2
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQCDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSCETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTCNDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIV
GNNSMKAPEGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDN
LSPEVQNELACFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYG
VGPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYCHAINFLRV
DKKNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKCKGSTKDYAQREGTLSDTIAALKGKL
SSLENRLSALHQHADVMAAQAKVSQLQGKLASQLKQCDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLE
ATRDQSLAKLASLKAALHQTEALCEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKC
TSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-base_cc1-d15_glob-d1
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETLL
AQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDA

ITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVGN
NTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIKAKASPGNQLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIKNDDNMYENIGATNDDHTVNGIKRGIYDTIKYMLFTDHEHGNTYGHAINFLRVDK
HNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSS
LENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEAT
RDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSS
LKNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-base_cc1-d15_glob-base
MSGSEETLLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKL
NAMISNPDAITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAP
STQDSIVGNNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNR
EVDPDNLTPEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGV
SGHYGVGPHDKTIIEDSAGASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHA
INFLRVDKHNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIA
ALKGKLSSLENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKA
KQAQLEATRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNGSGSGSGSHHHHH
H
>SpyAD_cc3-d8_cc2-base_cc1-d15_glob-d3
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETL
LAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPD
AITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLAASGYIGSASYNNYYKEHADQIKAKASPGLSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHLINSVRRQLGLPPVTVTAGSQEFARTLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAGASGLIRNDDFMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGNAINFLRVDK
KNPNAPIYLGFSTSNVGSLNYHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSL
ENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEATR
DQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-base_cc1-d15_glob-d2
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIV
GNNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNL
SPEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGV
GPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVD
KKNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLS
SLENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEA
TRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTT
SSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-base_cc1-base_glob-base
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
LLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNP
DAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPDN
LTPEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYG
VGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRV
DKHNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKV
SSLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEA

TRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTT
SSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d16_cc1-base_glob-d3
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTATETELHNAQADYHSKETALSEQKASISAETTRAQDLEEQVKTLEQNIAKLNALLSNPD
AITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSIV
GNNSMKAPEGYPLEELKKLAASGYIGSASYNNYYKEHADQIKAKASPGLSLNQYQDIPADRNREVDPDNL
SPEVQNELAQFAAHLINSVRRQLGLPPVTVTAGSQEFARTLSTSYKKTHGNTRPSFVYGQPGVNGHYGV
GPHDKTIIEDSAGASGLIRNDDFMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGNAINFLRVD
KKNPNAPIYLGFSTSNVGSLNYHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQRVGTVSDTIAAIKGKVSS
LENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELEAAKAKLAQAEATLD
QTLAKLASLKAALHQKEALAEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-base_cc1-d15_glob-base
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETL
LAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPD
AITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNREVDPDNLT
PEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVG
PHDKTIIEDSAGASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVDK
HNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSS
LENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEAT
RDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSS
LKNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-base_cc1-d15_glob-d2
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETLL
AQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDA
ITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVGN
NSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLSP
EVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVGP
HDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVDKK
NPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSLE
NRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEATRD
QSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-base_cc1-base_glob-d26
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETL
LAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPD
AITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSLF
GNNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEVDPDNL
TPEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVG
PHDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLLMVDK
HNPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQRVGTVSDTIAAIKGKVSSL
ENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRD
QSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-d14_cc1-base_glob-d3
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEET
RLAQAAETQRELTAAETELHNAQADLHSKETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNP

DAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSI
VGNNSMKAPEGYPLEELKKLAASGYIGSASYNNYYKEHADQIKAKASPGLSLNQYQDIPADRNREVDPDN
LSPEVQNELAQFAAHLINSVRRQLGLPPVTVTAGSQEFARTLSTSYKKTHGNTRPSFVYGQPGVNGHYG
VGPHDKTIIEDSAGASGLIRNDDFMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGNAINFLRV
DKKNPNAPIYLGFSTSNVGSLNYHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQRVGTVSDTIAAIKGKVS
SLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALAEQAAARVTALEAKIAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSL
LNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-base_cc1-d15_glob-base
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQCDNLQKALTSAQERYTNLLASSEETL
LAQGAEHQRELTATETELHNAQADQHSCETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPD
AITKAAQTCNDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNTMKAPQGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNREVDPDNLT
PEVQNELACFAAHMINSVRRQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVG
PHDKTIIEDSAGASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYCHAINFLRVDK
HNPNAPVYLCFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKCKGSTKDYAQREGTLSDTIAALKGKLSS
LENRLSALHQHADVMAAQAKVSQLQGKLASQLKQCDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEAT
RDQSLAKLASLKAALHQTEALCEQAAARVTALVAKKAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKCTS
SLKNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-base_cc1-d15_glob-d1
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQCDNLQKALTSAQEEYTNLLASSEETL
LAQGAEHQRELTATETELHNAQADQHSCETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPD
AITKAAQTCNDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNTMKAPQGYPLEELKKLEASGYCGSASYNNYYKEHADQIKAKASPGNQLNQYQDIPADRNREVDPDNL
SPEVQNELACFAAHMINSVRRQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGV
GPHDKTIIEDSAAASGLIKNDDNMYENIGATNDDHTVNGIKRGIYDTIKYMLFTDHEHGNTYCHAINFLRVD
KHNPNAPVYLCFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKCKGSTKDYAQREGTLSDTIAALKGKLS
SLENRLSALHQHADVMAAQAKVSQLQGKLASQLKQCDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEA
TRDQSLAKLASLKAALHQTEALCEQAAARVTALVAKKAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKCT
SSLKNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-d16_cc1-d15_glob-d26
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETR
LAQAAETQRELTATETELHNAQADYHSKETALSEQKASISAETTRAQDLEEQVKTLEQNIAKLNALLSNPD
AITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSLFG
NNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEVDPDNLT
PEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGP
HDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLLMVDKH
NPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSLE
NRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELEAAKAKLAQAEATLD
QTLAKLASLKAALHQKEALAEQAAARVTALEAKIAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_glob-d2
MSGPSTQDSIVGNNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPA
DRNREVDPDNLSPEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVY
GQPGVNGHYGVGPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHG
NTYGHAINFLRVDKKNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKGSGSGSGS
HHHHHH
>SpyAD_cc3-d1_cc2-d14_cc1-d4_glob-d3

MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSKETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLAASGYIGSASYNNYYKEHADQIKAKASPGLSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHLINSVRRQLGLPPVTVTAGSQEFARTLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAGASGLIRNDDFMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGNAINFLRVDK
KNPNAPIYLGFSTSNVGSLNYHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSLE
NRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATRD
QALAKLASLKAALHQQEALAEQAAARVTALEAKIAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc1-base_glob-base
MSGPDAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPST
QDSIVGNNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFV
DPDNLTPEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVS
GHYGVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAI
NFLRVDKHNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIA
AIKGKVSSLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-base_cc1-base_glob-d26
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETLL
AQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDA
ITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSLFG
NNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEVDPDNLT
PEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGP
HDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLLMVDKH
NPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQRVGTVSDTIAAIKGKVSSLE
NRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQ
SLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSLKN
AQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d16_cc1-d15_glob-d26
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTATETELHNAQADYHSKETALSEQKASISAETTRAQDLEEQVKTLEQNIAKLNALLSNPD
AITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSLFG
NNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEVDPDNLT
PEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGP
HDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLLMVDKH
NPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSLE
NRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELEAAKAKLAQAEATLD
QTLAKLASLKAALHQKEALAEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-base_cc1-d15_glob-d3
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETLL
AQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDA
ITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVGN
NSMKAPEGYPLEELKKLAASGYIGSASYNNYYKEHADQIKAKASPGLSLNQYQDIPADRNREVDPDNLSP
EVQNELAQFAAHLINSVRRQLGLPPVTVTAGSQEFARTLSTSYKKTHGNTRPSFVYGQPGVNGHYGVGP
HDKTIIEDSAGASGLIRNDDFMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGNAINFLRVDKK
NPNAPIYLGFSTSNVGSLNYHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSLE
NRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEATRD

QSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-base_cc1-d15_glob-base
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETLL
AQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDA
ITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVGN
NTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNREVDPDNLTP
EVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGP
HDKTIIEDSAGASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVDKH
NPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSL
ENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEATR
DQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSCETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTCNDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYCHAINFLRVDK
KNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQCDSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALCEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-base_cc1-d4_glob-d26
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETL
LAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPD
AITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSLF
GNNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEVDPDNL
TPEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVG
PHDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLLMVDK
HNPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATR
DQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-base_cc1-d15_glob-d2
MSGPDAITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPST
QDSIVGNNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREV
DPDNLSPEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVN
GHYGVGPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAI
NFLRVDKKNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAAL
KGKLSSLENRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-d14_cc1-d4_glob-d2
MSGPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQCDNEQKALTSAQEIYTNTLASSEET
RLAQAAETQRELTAAETELHNAQADLHSCETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNP
DAITKAAQTCNDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIV
GNNSMKAPEGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDN
LSPEVQNELACFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYG
VGPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYCHAINFLRV

DKKNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKCKGSTKDYAQREGTISDTIAALKGKLS
SLENRLSALHQHADVMAQQAKVSQLQGKLASQLKQCDSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEA
TRDQALAKLASLKAALHQQEALCEQAAARVTALEAKIAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKCT
SSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSCETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAINFLRVDK
KNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALCEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d1_cc2-d16_cc1-d4_glob-d3
MSGMETLQELKAEIDAVEKTISQLKAELTELATAITKITAEINHLKEQIDNLQKALTSAQERYTNLLASSEETR
LAQAAETQRELTATETELHNAQADYHSKETALSEQKASISAETTRAQDLEEQVKTLEQNIAKLNALLSNPD
AITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLAASGYIGSASYNNYYKEHADQIKAKASPGLSLNQYQDIPADRNREVDPDNLS
PEVQNELAQFAAHLINSVRRQLGLPPVTVTAGSQEFARTLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAGASGLIRNDDFMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGNAINFLRVDK
KNPNAPIYLGFSTSNVGSLNYHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSLE
NRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNDTKGSLRTELEAAKAKLAQAEATLD
QTLAKLASLKAALHQKEALAEQAAARVTALEAKIAHLQYLRDFKLNPNELQKIREEIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-base_cc1-d15_glob-d26
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETL
LAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPD
AITKAAQTANDNVKALSSELEKAKADLENIKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSLFG
NNTMKAPQGYPLEELKKLDASGYIGSASYNNYYKEHEDQIVAKSSPGNQLNQYQDIPAYRNVEVDPDNLT
PEVLASLANFAQHLINDVRRQLGLPPLTVTAGSLEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGP
HDKTIIEDAAGASGLIRNPDNDYLVIGATNDDHTVDGLQRNIFDTIKYHLFDSHEHGNNFSHAIRLLMVDKH
NPNAPVYLGFSVDNVGSVNNHLVIFPESAIANHQRFNKTPIKAKGSTKDYAQREGTLSDTIAALKGKLSSLE
NRLSALHQHADVMAAQAKVSQLQGKLASQLKQRDSLNLKVRQLNDTKGSLRTELLAAKAKQAQLEATRD
QSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSLK
NAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-base_cc2-d14_cc1-d4_glob-d2
MSGPDAITKAAQTANDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPST
QDSIVGNNSMKAPEGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREV
DPDNLSPEVQNELAQFAAHMINSVRKQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVN
GHYGVGPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYGHAI
NFLRVDKKNPNAPVYLGFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAAL
KGKLSSLENRLSALHQHADVMAQQAKVSQLQGKLASQLKQADSLNLQVRQLNGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQIDNLQKALTSAQEEYTNLLASSEETR
LAQAAETQRELTAAETELHNAQADLHSKETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNPD
AITKAAQTCNDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIVG
NNSMKAPEGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDNLS

PEVQNELAQFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYGVG
PHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYCHAINFLRVDK
KNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKAKGSTKDYAQREGTISDTIAALKGKLSSL
ENRLSALHQHADVMAQQAKVSQLQGKLASQLKQCDSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEATR
DQALAKLASLKAALHQQEALAEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKATSSL
KNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH
>SpyAD_cc3-d8_cc2-d14_cc1-d4_glob-d2
MSGMETLQELKAEIDAVEKTISQVKAELTELATAITKITAEINHLKEQCDNLQKALTSAQEEYTNLLASSEET
RLAQAAETQRELTAAETELHNAQADLHSCETALSERKASISAETTRVQDLEEQVKTAEQNIAKLNALLSNP
DAITKAAQTCNDNVKALSSELEKAKADLENTKAKVKKQLTEELAAVKAALAEKEAELSRLKSSAPSTQDSIV
GNNSMKAPEGYPLEELKKLEASGYCGSASYNNYYKEHADQIIAKASPGNSLNQYQDIPADRNREVDPDN
LSPEVQNELACFAAHMINSVRKQLGLPPVTVTAGCQEFARLLSTSYKKTHGNTRPSFVYGQPGVNGHYG
VGPHDKTIIEDSAAASGLIRNDDNMYENIGATNDDHTVNGIKRGIYDSIKYMLFTDHEHGNTYCHAINFLRV
DKKNPNAPVYLCFSTSNVGSLNEHFVIFPESNIANHQRFNKTPIKCKGSTKDYAQREGTISDTIAALKGKLS
SLENRLSALHQHADVMAQQAKVSQLQGKLASQLKQCDSLNLQVRQLNDTKGSLRTELEAAKAKLAQLEA
TRDQALAKLASLKAALHQQEALCEQAAARVTALEAKIAHLQYLRDFKLNPNKLQEIREKIDNLKQDLAKCTS
SLKNAQEALAALQAEKSSLEATLATLEHQLTLLKTLANEKEYRLLDEDKATGSGSGSGSHHHHHH

Tuberculosis

Mycobacterium tuberculosis (Mtb) is the most deadly infectious disease globally[1]. It is estimated that in 2023, tuberculosis caused 10.8 million illnesses and the deaths of 1.25 million people[1,2]. Furthermore, Mtb's morbidity and mortality are projected to grow, as the WHO recently reported an 8.3% reduction in TB incidence in 2023 relative to a goal of 50% reduction by 2025[1]. More than 100 years after its introduction, Bacille Calmette-Guérin (BCG) remains the only vaccine for Mtb[3]. The live-attenuated vaccine is administered to infants and is estimated to possess 70-80% efficacy in preventing meningitis and systemic TB in children[4,5]. However, BCG only has 30% efficacy in preventing the progression of asymptomatic, latent infection into active pulmonary disease in adults[4,6]. As Mtb spreads via aerosolized droplets and approximately 25% of the world's population is estimated to be latently infected, reactivation of latent infection into active pulmonary disease serves as the main route of transmission[7,8]. Novel and effective vaccines are urgently needed to prevent the continued, devastating burden of Mtb[1].

Current Mtb vaccine design efforts employ a diverse array of vaccine modalities and antigens[2], but in the absence of well-defined correlates of protection, it is difficult to judge which antigens and type of immune response to target[9–11]. While the role of $CD4^+$ T cells in controlling Mtb infection is well established, the role of $CD8^+$ T cells and B cells is less clear, with a variety of recent reports suggesting potential roles for these components of the adaptive immune response[12–22]. Given the lack of well-defined correlates of protection, several vaccine candidates are currently in Phase 3 clinical trials to evaluate efficacy[2,3]. Of these, M72/AS01E, an adjuvanted protein subunit

vaccine, has demonstrated 54% efficacy in a Phase 2b trial for preventing progression from latent infection to active pulmonary disease[23,24]. M72 is a fusion of two proteins, Mtb32A and PPE18 (also known as Mtb39A). To facilitate recombinant production, the protein is arranged as residues 224-355 of Mtb32A, followed by PPE18, followed by residues 33-227 of Mtb32A[23]. The arrangement of each antigen in the fusion protein might result in a non-native conformation, and recombinant production of the protein is challenging.

We set out to separate the two antigens that make up M72 and stabilize each in a native-like conformation. Mtb32A, a secreted serine protease with an unknown role in the Mtb lifecycle, was originally selected because it strongly stimulated naive $CD4^+$ T cells from uninfected humans[25]. Recombinant production of Mtb32A in *E. coli* results in cleaved fragments of the antigen and cell toxicity, likely due to its serine protease activity[25]. We predicted the structure of Mtb32A with AlphaFold2 (AF2) using all five models and structural templates from the PDB (Fig 3-1). Although homologs in the PDB display 37% identity or less, all five predicted structures showed close agreement and possessed high model confidence, with root mean squared deviation (RMSD) < 0.9 Å between models and pLDDT > 90. Additionally, examination of the predicted structures clearly identified the catalytic triad, in agreement with bioinformatic analysis of sequences alone[25] (Fig 3-1). Examination also revealed a well-packed core and a surface displaying numerous polar residues, suggesting that Mtb32A may be a well-behaved protein apart from its protease activity. We introduced an S208A mutation at the catalytic serine to create an inactive version, Mtb32A_S208A, which was readily

purified with high yield after expression in *E. coli*: we typically obtained 40-80 mg of purified protein from one liter of culture using a default, unoptimized process. These data, including the accurate hydrogen bond arrangements of the catalytic triad in predicted structures, further demonstrated the ability to successfully perform structure-based antigen design from AF2 predictions.

PPE18 is a member of the large PE:PPE family of heterodimeric antigens that comprise approximately 10% of the Mtb genome, having undergone extensive duplication relative to less virulent *Mycobacterium* species[26–29]. PPE18 has been reported to enhance Mtb survival within macrophages, and knock-out of PPE18 reduces infection and survival of Mtb *in vivo*[30]. PPE18 was selected for inclusion in M72 because it strongly stimulated naive $CD4^+$ T cells from uninfected humans and a DNA vaccine encoding PPE18 reduced bacterial load in a mouse challenge model of Mtb infection[31]. To stabilize PPE18 and enable recombinant expression, we began by predicting the structure with AF2 using all five models and structural templates from the PDB (Fig 3-2A). AF2 predictions enabled us to develop three approaches to improve the soluble yield of PPE18 (Fig 3-2B). First, all five models predicted the PPE domain, which comprises the first 180 residues of PPE18, with high confidence and close agreement. However, the predictions differed considerably across residues 181-391 of PPE18, and displayed much lower model confidence (pLDDT < 70). The predictions bolster previous reports that the first 180 residues of PPE18 display high conservation with other PPE proteins, while the C-terminal portion displays little conservation and is likely disordered[32]. Therefore, we truncated PPE18 at residue 180. Second, predicted structures of the first

180 residues of PPE18 revealed an extensive hydrophobic surface on one portion of the antigen. We reasoned that the hydrophobic surface might constitute the interface with PE13, the known binding partner of PPE18. We used AF2 to generate five predictions of the PE13:PPE18 heterodimer, all five of which showed close agreement to experimentally determined PE:PPE structures (RMSD < 1.5 Å to PDB ID 5XFS) and indicated that the hydrophobic surface of PPE18 forms an extensive interface with PE13[33]. We hypothesized that a PE13:PPE18 heterodimeric antigen might better preserve the native conformation and facilitate recombinant production by occluding the hydrophobic surface of PPE18. Third, PPE18, like other PPE proteins, contains hydrophobic, surface-exposed residues that recruit the EspG5 chaperone in Mtb[34,35]. As the PPE protein family has duplicated and diverged extensively in Mtb, MSAs of PPE18 either returned sequences with >99% identity or other, diverged PPE proteins, precluding the use of MSA-based methods to design PPE18. To facilitate improved soluble yield of designed antigens, we used ProteinMPNN to mutate several surface-exposed hydrophobic residues in PPE18, obtaining 200 output sequences for each of the five predicted structures. We incorporated the most frequent mutation at each position if it arose in more than 25% of output sequences for each input structure and did not appear to cause steric clashes by manual inspection in PyMOL. The process identified thirteen mutations predicted to improve solubility (Fig 3-2D).

To determine how each potential design strategy affected the soluble yield and monodispersity of PPE18, we compared M72 to PPE18 constructs with or without truncation at residue 180, co-expression of PE13, and solubilizing mutations at surface

hydrophobic residues. We expressed these antigens in *E. coli* and purified them by IMAC. Size exclusion chromatography (SEC) revealed that M72 possessed moderate soluble yield due to the inclusion of Mtb32A but formed aggregates. PPE18 displayed very low soluble yield, with or without PE13. Truncated PPE18, with or without PE13, behaved similarly to M72, with higher yield and the formation of aggregates. However, PPE18 with solubilizing mutations and coexpressed PE13 demonstrated higher soluble yield and a clear peak at the expected size for the heterodimer. Truncation of this PPE18 variant at residue 180 further improved the soluble yield (PPE18:PE13-base). However, the majority of protein was still aggregated, with a smaller fraction existing as a heterodimer (Fig 3-2C). These results indicated that a combination of all three design strategies resulted in the best yield for PPE18, but the resulting construct was still aggregation-prone, hindering recombinant production.

We set out to stabilize the PPE18:PE13-base heterodimer by designing a linker domain to genetically fuse PE13 and PPE18. We explored multiple linker design strategies (Fig 3-3A). As the C terminus of PPE18-180 was close to the N terminus of PE13 in our AF2 predictions, we linked the two by a flexible $(GS)_4$ linker (PPE18-PE13-d1) or used RFdiffusion and ProteinMPNN to design a *de novo* linker that made packing interactions with the antigens (PPE18-PE13-d2). We also explored an alternative arrangement in which the C terminus of PE13 was linked to the N terminus of PPE18 by a longer linker. The linker was either taken from the Mtb protein EspB (PPE18-PE13-d3), which has both PPE-like and PE-like domains within the same protein[34], or designed with RFdiffusion and ProteinMPNN to have high secondary structure and form packing

interactions with the antigens (PPE18-PE13-d4 and -d5). The designed antigens containing *de novo* linkers were filtered by AF2 prediction for high model confidence (pLDDT > 84) and low RMSD (backbone RMSD < 2.0 Å) to the prediction of the native heterodimer. Following expression in *E. coli* and purification by IMAC, SEC indicated that PPE18-d2 possessed improved yield and less aggregates relative to the other designs, suggesting that the short *de novo* linker outperforms a short GS linker or longer linker domains. PPE18-PE13-d2 also displayed higher yield and substantially less aggregate than the PPE18:PE13-base heterodimer (Fig 3-3B). However, the fusion protein displayed two peaks on SEC that likely represent the designed monomer and an unintended dimer of fusion proteins. To better understand the cause of this off-target species, we crystallized PPE18-PE13-d2 and determined its structure at a resolution of 2.61 Å. Although the majority of the structure aligned closely to our AF2 design model (backbone RMSD of 1.38 Å over 274 residues), it also illuminated a deviation between the design model and our experimental structure at a tryptophan in a conserved loop among PPE proteins. That small deviation in the antigen created a clash with the *de novo* linker, causing a larger deviation as the α-helix within the *de novo* linker flipped away from the antigens (Fig 3-3C). We speculate that deviation may lead to domain swapping between the PE13 domain in one molecule and the PPE18 domain in another, causing the unintended dimer to form. The difference in the tryptophan position might be due to inaccuracy in the predicted structure, flexibility of the residue, or deviations caused by our *de novo* linker.

We examined the position of the conserved tryptophan among four PPE:PE heterodimers whose structures have been experimentally determined (PDB IDs 2G38, 4W4K, 5XFS, and 6UUJ)[33,36–38]. Both the tryptophan position in our crystal structure and that in the AF2 design model are seen in experimental structures. Both the tryptophan position in our crystal structure and that in the AF2 design model were seen among the experimental structures. To further examine this question, we crystallized a construct in which PE13 and PPE18 truncated at residue 180 were flexibly linked to the components of the designed heterodimer synzipE2[39]. We reasoned that flexible linkage to a heterodimeric scaffold might better preserve the native interaction of PE13 and PPE18. We also introduced a disulfide bond into synzipE2 that was suggested by DisulfidebyDesign 2.0 and predicted to form by all 5 models of AF2. SEC indicated that coexpression of the fusion proteins PE13-synzipE2a and PPE18-180-synzipE2b resulted in a monodisperse heterodimer, and we were able to obtain crystals that diffracted to 2.8 Å resolution. A resultant structure showed no interaction between synzipE2 and the PE13:PPE18 complex, both of which were resolved, and the flexible glycine-serine linkers between the antigen and synzipE2 domains were not resolved (not shown). In this structure the conserved tryptophan and three adjacent residues were not resolved, indicating that the loop was flexible. In addition, the structure revealed that the introduced disulfide in synzipE2 did not form, such that the construct did not contain a covalent linkage between PE13 and PPE18.

Hypothesizing that the loop containing the tryptophan is flexible in native PPE18, and wanting to include a covalent linkage between PE13 and PPE18 in our designed

antigens, we set out to redesign PPE18-PE13-d2. To reduce the number of introduced mutations, we used the crystal structure of PPE18-PE13-d2 to identify seven of the introduced solubilizing mutations, and two mutations introduced at the N terminus of PE13 during design of the linker domain, that removed favorable packing interactions of the native solvent-exposed hydrophobic residue. We reasoned that given the favorable yield of PPE18-PE13-d2.2, reverting some of the solubilizing mutations would better preserve antigenicity while still allowing recombinant production. We tested two designs containing reversions at these nine positions to the native residues. One retained the original linker (PPE18-PE13-d2.1), while the other substituted 11 residues in the *de novo* linker with $(GS)_6$ to provide flexibility compatible with the deviation at the beginning of the linker domain (PPE18-PE13-d2.2). Following expression and purification by IMAC, SEC revealed that both designed antigens were predominantly in the intended monomeric state, with the highest yield and monodispersity for PPE18-PE13-d2.2 (Fig 3-3D). Differential scanning fluorimetry (DSF) analysis of PPE18-PE13-d2.2 and PPE18:PE13-base demonstrated comparable melting temperatures (T_m of 44.4 and 42.5 °C, respectively), indicating that truncation, solubilizing mutations, and an introduced covalent linkage did not reduce the T_m (not shown). The designed antigen enables recombinant production of a monodisperse PPE18 antigen. The inclusion of PE13 provides additional B and T cell epitopes, and CD4+ T cell responses against PE13 are enriched among those who control latent infection relative to those who progress to active disease[40].

Several preclinical vaccine design efforts for Mtb have focused on mRNA delivery to elicit protective $CD4^+$ T cell responses[14]. A recent report investigating the presentation of Mtb proteins to $CD4^+$ T cells by antigen presenting cells (APCs) revealed substantial enrichment of substrates of the Type 7 secretion system (T7SS), including PE:PPE proteins[41]. In addition, delivery to APCs of mRNA encoding EsxA:EsxB, a heterodimeric substrate of the T7SS analogous to PE13:PPE18, led to MHC II display of EsxB peptides, while mRNA encoding EsxB alone did not[41]. We reasoned that proper folding and improved solubility of the EsxA:EsxB heterodimer might explain the difference. We set out to examine how design of Mtb39 might alter its expression and cell-surface display, which may correlate with presentation on MHC II and subsequent $CD4^+$ T cell responses during delivery as an mRNA vaccine. Transfection of DNA plasmids encoding M72, M72 with two mutations designed to remove cryptic transmembrane domains[42], or PPE18-PE13-d2.2 with two mutations designed to remove cryptic transmembrane domains revealed that five-fold more cells displayed the designed PPE18-PE12-d2.2 antigen (Fig 3-4). The higher levels of cell-surface display for the designed antigen relative to M72 are promising, but further work is necessary to determine if increased expression leads to improved vaccine-elicited $CD4^+$ T cell and B cell responses.

Due to the lack of validated monoclonal antibodies for Mtb32A and PPE18, we were unable to assess the antigenicity of these designed antigens. Nevertheless, our data showed that AF2 structure predictions and design of a *de novo* linker between PPE18 and its binding partner PE13 enabled the development of antigens that are

monodisperse and can be purified in high yield in both cases. Furthermore, our structural data demonstrated that our designed PPE18 antigen recapitulates the intended structure.

References

1. World Health Organization. *Global Tuberculosis Report 2024*. (2024).
2. Zhuang, L., Ye, Z., Li, L., Yang, L. & Gong, W. Next-Generation TB Vaccines: Progress, Challenges, and Prospects. *Vaccines (Basel)* **11**, (2023).
3. Lai, R., Ogunsola, A. F., Rakib, T. & Behar, S. M. Key advances in vaccine development for tuberculosis—success and challenges. *npj Vaccines* **8**, 1–10 (2023).
4. Mangtani, P. *et al.* Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin. Infect. Dis.* **58**, 470–480 (2014).
5. Martinez, L. *et al.* Infant BCG vaccination and risk of pulmonary and extrapulmonary tuberculosis throughout the life course: a systematic review and individual participant data meta-analysis. *The Lancet Global Health* **10**, e1307–e1316 (2022).
6. Katelaris, A. L. *et al.* Effectiveness of BCG Vaccination Against Mycobacterium tuberculosis Infection in Adults: A Cross-sectional Analysis of a UK-Based Cohort. *J. Infect. Dis.* **221**, 146–155 (2020).
7. Pai, M. *et al.* Tuberculosis. *Nat Rev Dis Primers* **2**, 16076 (2016).
8. Houben, R. M. G. J. & Dodd, P. J. The global burden of latent tuberculosis infection: A re-estimation using mathematical modelling. *PLoS Med.* **13**, e1002152 (2016).
9. Nemes, E. *et al.* The quest for vaccine-induced immune correlates of protection against tuberculosis. *Vaccine Insights* **1**, 165–181 (2022).
10. Wang, J., Fan, X.-Y. & Hu, Z. Immune correlates of protection as a game changer in tuberculosis vaccine development. *NPJ Vaccines* **9**, 208 (2024).
11. Kaufmann, S. H. E. Fact and fiction in tuberculosis vaccine research: 10 years later. *Lancet Infect. Dis.* **11**, 633–640 (2011).
12. Darrah, P. A. *et al.* Prevention of tuberculosis in macaques after intravenous BCG immunization. *Nature* **577**, 95–102 (2020).

13. Simonson, A. W. *et al.* CD4 T cells and CD8α+ lymphocytes are necessary for intravenous BCG-induced protection against tuberculosis in macaques. *bioRxiv* (2024) doi:10.1101/2024.05.14.594183.
14. Larsen, S. E. *et al.* An RNA-Based Vaccine Platform for Use against Mycobacterium tuberculosis. *Vaccines (Basel)* **11**, (2023).
15. Behar, S. M., Woodworth, J. S. M. & Wu, Y. Next generation: tuberculosis vaccines that elicit protective CD8+ T cells. *Expert Rev. Vaccines* **6**, 441–456 (2007).
16. Irvine, E. B. *et al.* Robust IgM responses following intravenous vaccination with Bacille Calmette–Guérin associate with prevention of Mycobacterium tuberculosis infection in macaques. *Nat. Immunol.* **22**, 1515–1523 (2021).
17. Wang, S. *et al.* Markov field network model of multi-modal data predicts effects of immune system perturbations on intravenous BCG vaccination in macaques. *Cell Syst.* (2024) doi:10.1016/j.cels.2024.10.001.
18. Li, H. *et al.* Latently and uninfected healthcare workers exposed to TB make protective antibodies against Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5023–5028 (2017).
19. Lu, L. L. *et al.* A functional role for antibodies in tuberculosis. *Cell* **167**, 433–443.e14 (2016).
20. Li, H. & Javid, B. Antibodies and tuberculosis: finally coming of age? *Nat. Rev. Immunol.* **18**, 591–596 (2018).
21. Achkar, J. M., Chan, J. & Casadevall, A. B cells and antibodies in the defense against Mycobacterium tuberculosis infection. *Immunol. Rev.* **264**, 167–181 (2015).
22. Watson, A. *et al.* Human antibodies targeting a Mycobacterium transporter protein mediate protection against tuberculosis. *Nat. Commun.* **12**, 602 (2021).
23. Skeiky, Y. A. W. *et al.* Differential immune responses and protective efficacy induced by components of a tuberculosis polyprotein vaccine, Mtb72F, delivered as naked DNA or recombinant protein. *J. Immunol.* **172**, 7618–7628 (2004).

24. Van Der Meeren, O. *et al.* Phase 2b Controlled Trial of M72/AS01E Vaccine to Prevent Tuberculosis. *N. Engl. J. Med.* **379**, 1621–1634 (2018).
25. Skeiky, Y. A. *et al.* Cloning, expression, and immunological evaluation of two putative secreted serine protease antigens of Mycobacterium tuberculosis. *Infect. Immun.* **67**, 3998–4007 (1999).
26. Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537–544 (1998).
27. Fishbein, S., van Wyk, N., Warren, R. M. & Sampson, S. L. Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity: Evolution of PE/PPE-associated virulence. *Mol. Microbiol.* **96**, 901–916 (2015).
28. Ates, L. S. New insights into the mycobacterial PE and PPE proteins provide a framework for future research. *Mol. Microbiol.* **113**, 4–21 (2020).
29. D'Souza, C., Kishore, U. & Tsolaki, A. G. The PE-PPE Family of Mycobacterium tuberculosis: Proteins in Disguise. *Immunobiology* **228**, 152321 (2023).
30. Bhat, K. H., Ahmed, A., Kumar, S., Sharma, P. & Mukhopadhyay, S. Role of PPE18 protein in intracellular survival and pathogenicity of Mycobacterium tuberculosis in mice. *PLoS One* **7**, e52601 (2012).
31. Dillon, D. C. *et al.* Molecular characterization and human T-cell responses to a member of a novel Mycobacterium tuberculosis mtb39 gene family. *Infect. Immun.* **67**, 2941–2950 (1999).
32. Hakim, J. M. C. & Yang, Z. Predicted Structural Variability of Mycobacterium tuberculosis PPE18 Protein With Immunological Implications Among Clinical Strains. *Front. Microbiol.* **11**, 595312 (2020).
33. Chen, X. *et al.* Structural basis of the PE-PPE protein interaction in Mycobacterium tuberculosis. *J. Biol. Chem.* **292**, 16880–16890 (2017).
34. Korotkova, N. *et al.* Structure of the Mycobacterium tuberculosis type VII secretion system

chaperone EspG5 in complex with PE25-PPE41 dimer. *Mol. Microbiol.* **94**, 367–382 (2014).

35. Phan, T. H., Ummels, R., Bitter, W. & Houben, E. N. G. Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. *Sci. Rep.* **7**, 42704 (2017).
36. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8060–8065 (2006).
37. Ekiert, D. C. & Cox, J. S. Structure of a PE-PPE-EspG complex from Mycobacterium tuberculosis reveals molecular specificity of ESX protein secretion. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14758–14763 (2014).
38. Williamson, Z. A., Chaton, C. T., Ciocca, W. A., Korotkova, N. & Korotkov, K. V. PE5-PPE4-EspG3 heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems. *J. Biol. Chem.* **295**, 12706–12715 (2020).
39. Thompson, K. E., Bashor, C. J., Lim, W. A. & Keating, A. E. SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth. Biol.* **1**, 118–129 (2012).
40. Musvosvi, M. *et al.* T cell receptor repertoires associated with control and disease progression following Mycobacterium tuberculosis infection. *Nat. Med.* **29**, 258–269 (2023).
41. Leddy, O. *et al.* Immunopeptidomics informs discovery and delivery of*Mycobacterium tuberculosis*MHC-II antigens for vaccine design. *bioRxiv* 2024.10.02.616386 (2024) doi:10.1101/2024.10.02.616386.
42. Wang, J. Y. J. *et al.* Improving the secretion of designed protein assemblies through negative design of cryptic transmembrane domains. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2214556120 (2023).
43. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,

583–589 (2021).

44. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
45. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
46. The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC.
47. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
48. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
49. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
50. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
51. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
52. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
53. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

Figures

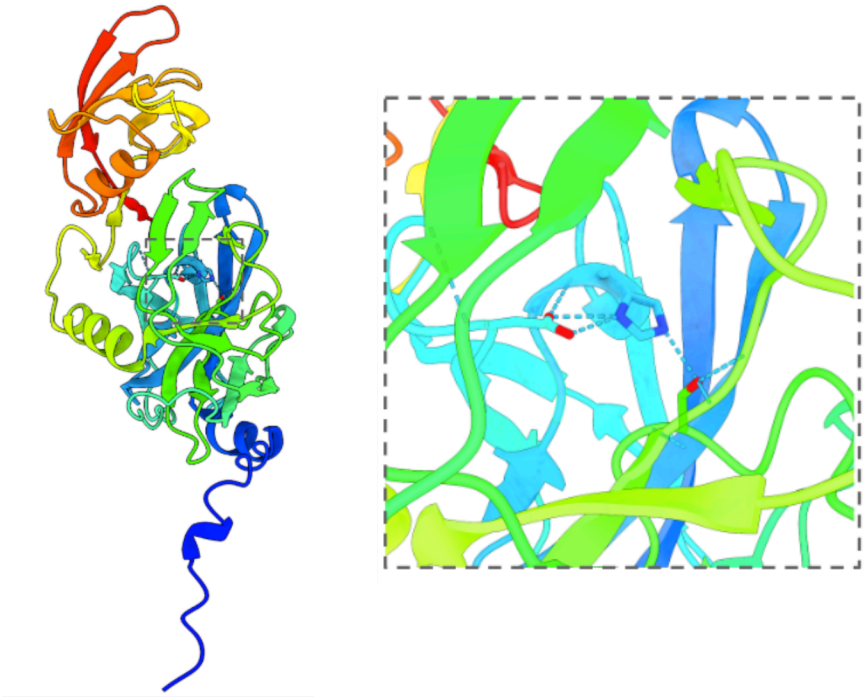
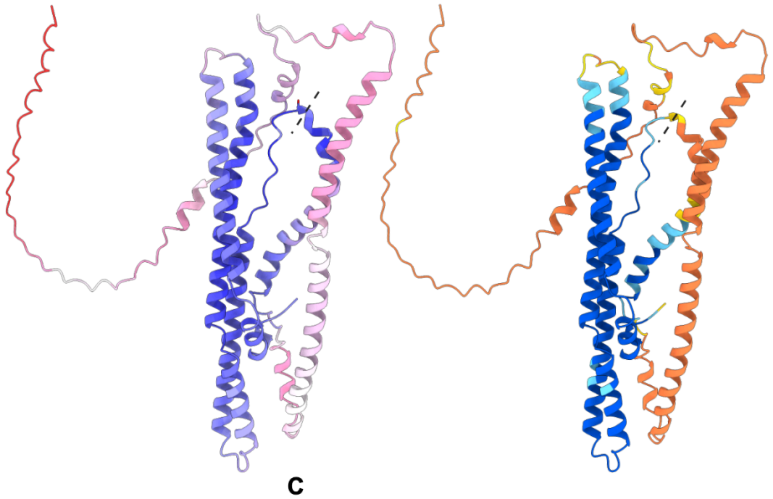
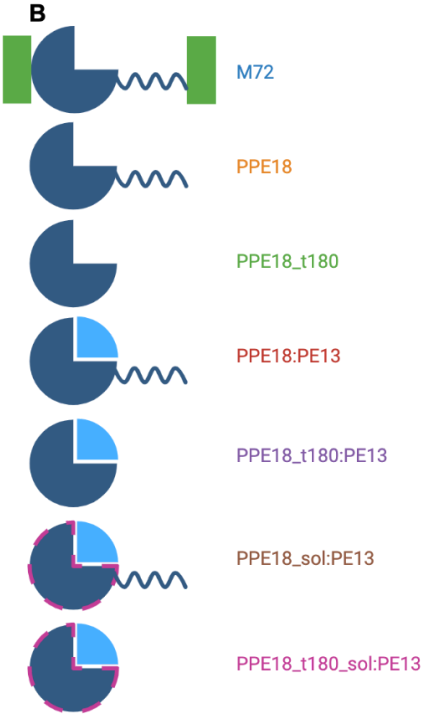


Figure 3-1. AlphaFold 2 (AF2) prediction of Mtb32A. AF2 prediction of Mtb32 is shown colored by sequence position (N terminus in blue, C terminus in red). The predicted structure revealed an unstructured N terminus and two domains within Mtb32A. The catalytic triad is magnified, illustrating accurate prediction of the hydrogen bonds and rotamer positions.

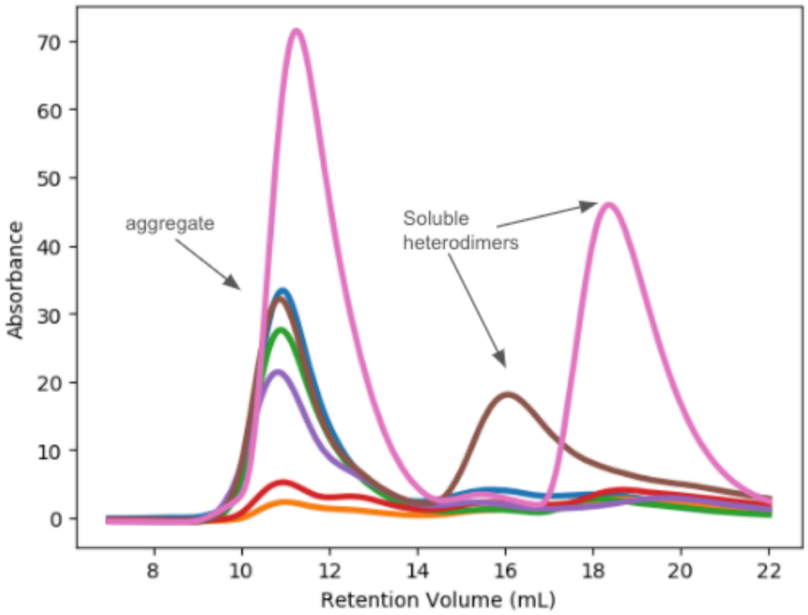
A



B


M72
PPE18
PPE18_t180
PPE18:PE13
PPE18_t180:PE13
PPE18_sol:PE13
PPE18_t180_sol:PE13

C


Absorbance
Retention Volume (mL)
aggregate
Soluble
heterodimers
0
10
20
30
40
50
60
70
8
10
12
14
16
18
20
22

D

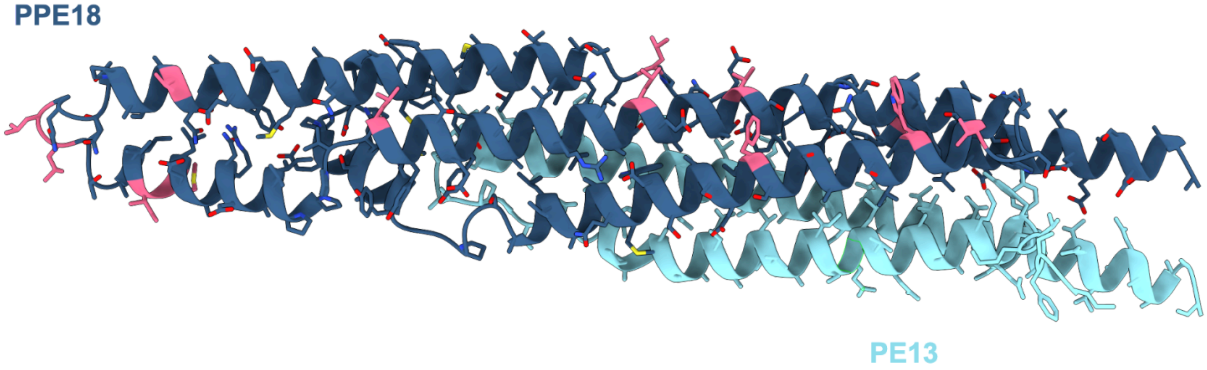

PPE18
PE13

Figure 3-2. The impact of three design strategies on yield of PPE18. A) AF2 prediction of the full-length PPE18 antigen. The structure on the left was colored by RMSD between five AF2 predictions (low RMSD in dark purple, intermediate RMSD in white, high RMSD in pink). The structure on the right was colored by average AF2 pLDDT as a measure of model confidence (high confidence in blue, intermediate in yellow, low in orange). Dashed lines denote residue 180 in PPE18. B) Scheme of initial designs. M72 has split Mtb32A (green boxes), the ordered PPE domain of PPE18 from residues 1-180 (dark blue portion of circle), and disordered residues 181-391 of PPE18 (dark blue curves). Various designed constructs co-expressed binding partner PE13 (light blue portion of circle), truncation at residue 180, and ProteinMPNN-derived solubilizing mutations at solvent-exposed hydrophobic residues (dashed pink border to circle). C) Size exclusion chromatography traces of M72 and designed PPE18 antigens on Superdex 200 Increase 10/300 GL column. Colors of traces match with the scheme in panel B. D) AF2 predicted structure of best construct, PPE18_t180_sol:PE13 (denoted in text as PPE18:PE13-base). Residues 1-180 of PPE18 in dark blue, PE13 in light blue, with hydrophobic residues mutated to more soluble polar amino acids highlighted in pink.

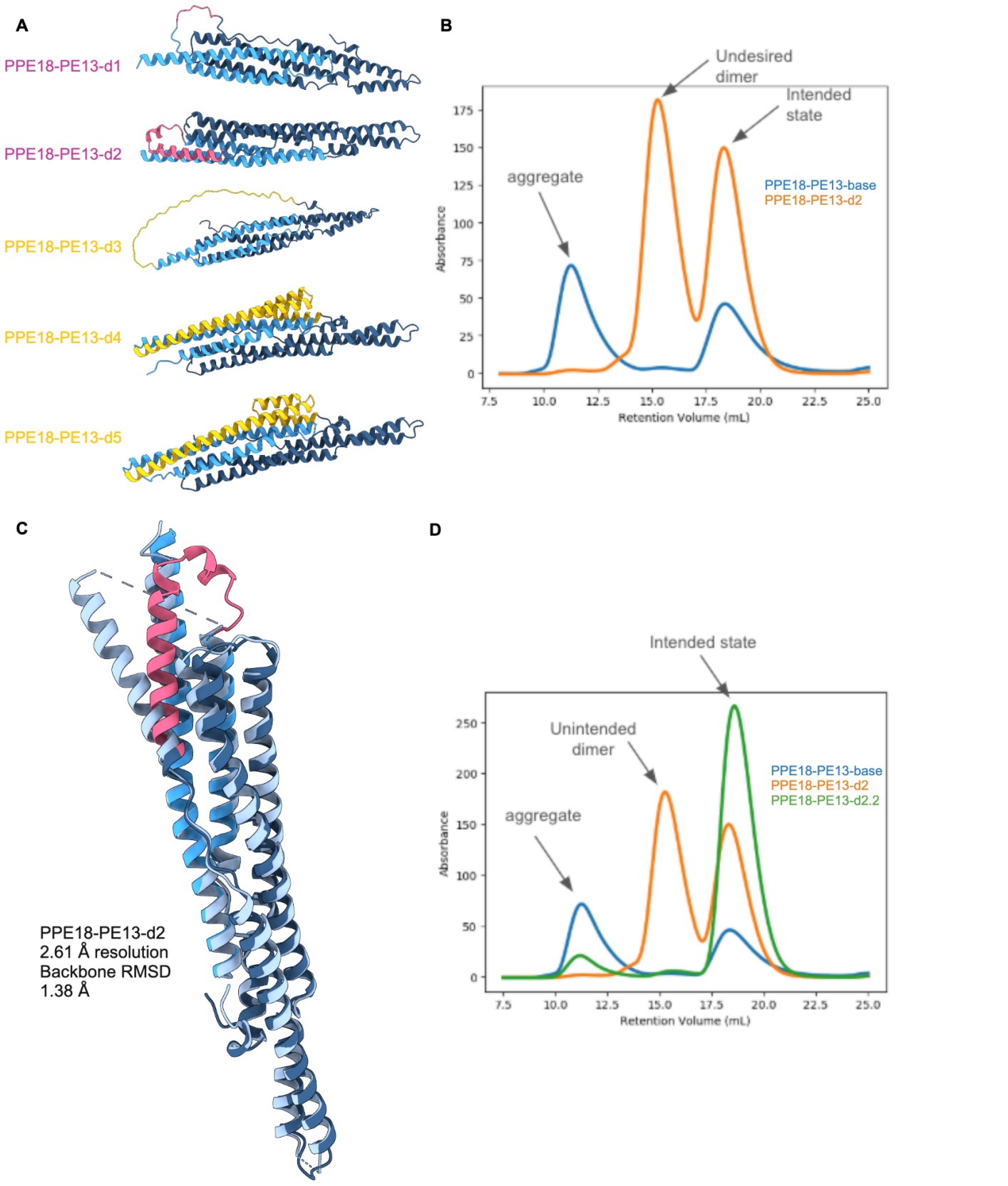

A
PPE18-PE13-d1
PPE18-PE13-d2
PPE18-PE13-d3
PPE18-PE13-d4
PPE18-PE13-d5
B
Undesired dimer
Intended state
aggregate
PPE18-PE13-base
PPE18-PE13-d2
Absorbance
Retention Volume (mL)
C
PPE18-PE13-d2
2.61 Å resolution
Backbone RMSD
1.38 Å
D
Intended state
Unintended dimer
aggregate
PPE18-PE13-base
PPE18-PE13-d2
PPE18-PE13-d2.2
Absorbance
Retention Volume (mL)

Figure 3-3. Fusion proteins display improved yield and monodispersity. A) Scheme of five designed fusion proteins between PPE18 (dark blue) and PE13 (light blue). Top two designs, which display construct names and linker regions in pink, involved a short linker between the C terminus of PPE18 and the N terminus of PE13. Bottom three designs, which display construct names and linker regions in yellow, involved a long linker between the C terminus of PE13 and the N terminus of PPE18. B) Size exclusion chromatography traces of PPE18:PE13-base in blue and PPE18-PE13-d2 in orange on Superdex 200 Increase 10/300 GL column. C) Crystal structure of PPE18-PE13-d2 in silver superimposed on the AF2 predicted structure of the antigen, which has the PPE18 domain colored in dark blue, the PE13 domain colored in a bright blue, and the *de novo* linker colored in pink. Dashed line indicates unresolved, flexible residues in linker. Backbone RMSD shown was across the PPE18 and PE13 regions. B) Size exclusion chromatography traces of PPE18:PE13-base in blue, PPE18-PE13-d2 in orange, and PPE18-PE13-d2.2 in green on Superdex 200 Increase 10/300 GL column.

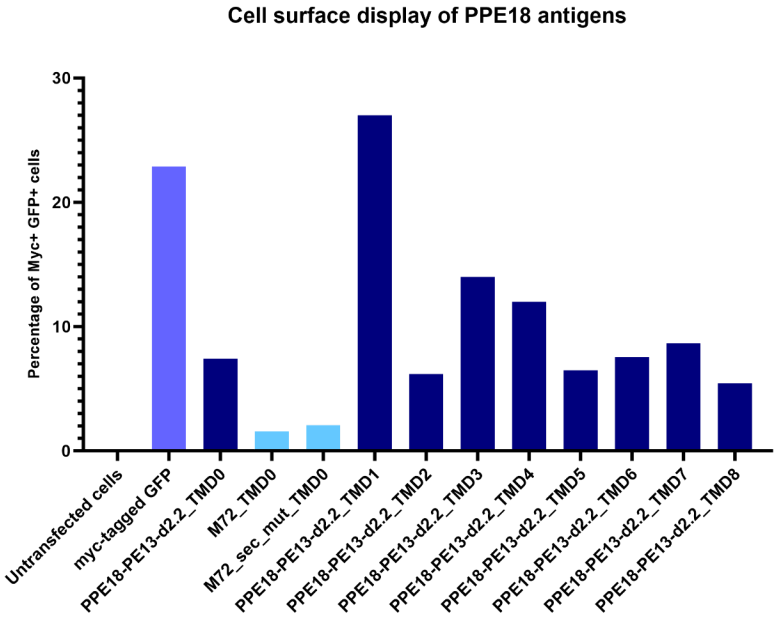


Figure 3-4. Cell surface display of designed PPE18 antigen. Comparison between M72, M72 with two mutations to improve secretion, and PPE18-PE13-d2.2 with the same two mutations to improve secretion. All three were compared with a human PDGFR transmembrane domain. Constructs on the right comprised PPE18-PE13-d2.2 with a different transmembrane domain (TMD). All antigens were myc-tagged. The degree of cell surface display was examined via flow cytometry by examining the percentage of $Myc^{+}GFP^{+}$ cells. GFP fluorescence was used as a transfection control. TMD1, from mouse CD80, shared similarity to a human protein, but TMD2 - TMD8 displayed no similarity to the human genome, making them viable for an mRNA vaccine candidate.

Materials and Methods

AlphaFold Prediction

The ColabFold implementation of AlphaFold was downloaded from the github repository into an apptainer environment with the necessary dependencies[43,44]. All predictions were made on either A4000 or A6000 GPUs on a custom high performance computing cluster. Sequences of candidate designed antigens were input into the ColabFold implementation to create MSAs using mmseqs2 and to generate structural predictions using AlphaFold2_ptm for monomeric proteins or the AlphaFold2_multimer_v3 for oligomers. Predictions were made with structural templates and with an amber relaxation step. For *Mycobacterium tuberculosis* antigens, predictions were made with all available PDB templates found by ColabFold during structure prediction by including the --templates flag. For each sequence predicted, structural predictions were obtained with all 5 AlphaFold models.

Following structure prediction, analysis with python scripts in a jupyter notebook was performed to determine prediction confidence by examining average pLDDT and PAE across the monomer or oligomers. All 5 predictions for a given sequence were used to determine the mean and standard deviation for pLDDT and PAE across the different models. Python scripts were used to calculate root mean squared deviation (RMSD) between the desired structure, either an experimental structure or high confidence prediction of the native sequence, and the structures predicted for designed sequences.

De novo linkers were evaluated using the same ColabFold protocol in the context of native or designed antigens. In these cases, pLDDT and PAE were evaluated across the de novo region of the protein, not across the entire protein.

ProteinMPNN

ProteinMPNN was run using an implementation downloaded from the github repository[45]. All sequence design was conducted on CPUs or an A4000 GPU on a custom high performance computing cluster.

For redesign of solvent-exposed hydrophobic residues, all 5 high-confidence ColabFold predictions for the native PPE18 sequence were used and ProteinMPNN was run with a sampling temperature of 0.2, and an amino acid bias of -0.8 for glycine and alanine. Cysteines were disallowed during design. ProteinMPNN was run for 40 batches with 5 sequences each, to generate 200 designed sequences for each input structure. A python script was used to analyze the frequency of mutations from the native amino acid at each position. Where ProteinMPNN chose a mutation at a given position in more than 25% of the sequences, the mutation was manually inspected with the PyMol mutagenesis wizard in the context of the input structure[46]. If any suggested mutation caused large steric clashes or the loss of at least two hydrogen bonds, the residue was no longer considered for design. Thirteen solubilizing mutations were identified that reduced hydrophobicity and did not appear deleterious in PyMol.

For design of *de novo* scaffolds or linkers generated in RFdiffusion, output structures from RFdiffusion served as inputs to ProteinMPNN. The *de novo* portion of the protein generated by RFdiffusion was designed, while all residues in the antigen were kept native. ProteinMPNN was run with a sampling temperature of 0.1 with 8 batches of 1 sequence each to generate 8 output sequences per design. These sequences were evaluated by ColabFold structure prediction, and the sequences with the highest pLDDT and lowest PAE across the *de novo* regions were ordered for experimental evaluation.

RFdiffusion

RFdiffusion was run using an implementation downloaded from the github repository and run on A4000 GPUs on a custom high performance computing cluster[47].

To generate linkers between the PPE18:PE13 heterodimeric antigen, the highest ranked AlphaFold model of the heterodimer was used as an input structure. Varying numbers of intervening residues were specified between the two proteins. Both PPE18 and PE13 were specified as the first protein in the fusion in separate runs. RFdiffusion was run with the base configuration. The input motif was provided to RFdiffusion with either 50, 100, or 200 timesteps of denoising with a varying noise scale (0,0.25,0.5, or 1). Output structures were filtered using Pymol scripts for a high proportion of secondary structure and proximity of the linker residues to the antigens.

Protein Expression

For initial screening of designs in *E. coli*, codon optimized DNA sequences were ordered as gene fragments (IDT or Twist), cloned into a pET-29b+ based plasmid using golden gate cloning with BsaI-HFv2 (NEB), and transformed into chemical competent BL21(DE3) cells (NEB). Cells were grown for 16 hours at 37°C in 1 mL of LB broth with antibiotic before inoculating 500 uL of culture into 50 mL of Terrific Broth II containing antibiotic. Cells were grown at 37 C to an OD of 0.6-0.8 before adding IPTG to a final concentration of 0.25 mM and cells were grown at 18°C for 18 hours. Cells were centrifuged at 4000 g for 15 minutes to harvest, and pellets were stored at -80°C.

For constructs used for more detailed comparisons with the native antigen, plasmids encoding codon optimized sequences were ordered from GenScript. Protein expression was conducted in the same manner as above.

Protein purification

For purification from *E. coli*, cells were resuspended in 8 mL of lysis buffer 1 for Mtb32A antigens (50 mm Tris, 250 mM NaCl, 20 mM imidazole, 0.5 mg/mL lysozyme, 1 U / mL benzonase, 0.02 tablets/mL of protease inhibitor cocktail) or lysis buffer 2 for Mtb39 antigens (50 mm Tris, 250 mM NaCl, 20 mM imidazole, 0.375% CHAPS, 0.5 mg/mL lysozyme, 1 U / mL benzonase, 0.02 tablets/mL of protease inhibitor cocktail) and lysed by ultrasonication (Qsonica 24-tip sonicator, 6 minutes total, 5s on, 5s off, 80% amplitude). Lysate was clarified by centrifugation at 14000 g for 30 min. Clarified lysate was applied to 0.5 mL of Ni-NTA resin (Qiagen) equilibrated with wash buffer (50 mm Tris, 250 mM NaCl, 20 mM imidazole). Resin was then washed with 20 column volumes

of wash buffer. Protein was eluted with 2.4 column volumes of elution buffer (50 mm Tris, 250 mM NaCl, 500 mM imidazole) and analyzed by SDS-PAGE gel. Eluted proteins were further filtered by size exclusion chromatography using a Superdex 200 Increase 10/300 GL column (Cytiva) on an ÄKTA pure (Cytiva).

Melting and aggregation temperature determination:

Determination of melting and aggregation temperatures were done by differential scanning fluorimetry with the UNcle instrument and software (UNchained Labs). SEC purified proteins at a concentration between 0.25-1 mg/mL were added to 8.8 uL quartz capillary wells in a 16-well cassette (UNchained Labs). With the temperature of the cassette changing from 25 to 95 °C at a thermal ramp rate of 1°C/min, intrinsic changes in protein fluorescence emission were used to evaluate melting and aggregation temperatures.

Mammalian cell surface display

CMV based plasmids encoding the antigens, an internal ribosomal entry site, and a downstream mNeonGreen that served as a transfection control, were ordered from GenScript. Constructs encoded the designed antigens, a glycine-serine linker containing a c-myc tag, and the transmembrane and cytoplasmic domains from human PDGFR or other membrane-anchored proteins. For mammalian expression, Expi293 suspension cells were transfected at 3 million cells/mL with the following per mL of culture: 1 ug of plasmid DNA, 3 ug PEI-max, and 100 uL Opti-MEM (Thermo Scientific). Cells were subsequently cultured for 3 days at 37C and pelleted at 300 g for 5 minutes.

Cells were washed (1X PBS, 0.5% BSA, 5 mM EDTA, pH 7.4), labeled with the Zombie live/dead dye (BioLegend) according to manufacturer's instructions, and labeled with AlexaFluor 594 tagged anti-myc antibody at 10 ug/mL (Abcam). Cells were analyzed using an Attune NxT flow cytometer (Thermo Fisher Scientific). Approximately 300,000 events were collected for each group.

Crystallization

The proteins PPE18-PE13-d2 and PE13-synzipE2A:PPE18-synzipE2B were purified by SEC and concentrated to 8 mg/mL in 25 mM Tris, 150 mM NaCl, 10% glycerol, pH 8.0 using 3K concentrators (Amicon). All crystallization experiments were conducted using the sitting drop vapor diffusion method. Crystallization trials were set up in 200 nL drops using the 96-well plate format at 20 ˚C. Crystallization plates were set up using a Mosquito (SPT Labtech), then imaged using UVEX microscopes and UVEX PS-256 (JAN Scientific). Diffraction quality crystals formed in 0.06 M of 0.3M Magnesium chloride hexahydrate and 0.3M Calcium chloride dihydrate; 0.12 M of 0.2M 1,6-Hexanediol, 0.2M 1-Butanol, 0.2M 1,2-Propanediol, 0.2M 2-Propanol, 0.2M 1,4-Butanediol and 0.2M 1,3-Propanediol; 0.1 M of 0.4M Imidazole; 0.6M MES monohydrate (acid) pH 6.5; 30 % of 40% (v/v) PEG 500 MME; 20 % (w/v) PEG 20000 for PPE18-PE13-d2. Diffraction quality crystals formed in 11 % w/v PEG 8000, 0.1 M Sodium cacodylate pH 6.2 and 0.4 M Sodium chloride for the PE13-synzipE2a:PPE18-180-synzipE2b heterodimer.

Diffraction data was collected either at National Light Synchrotron Light Source II AMX/FMX or the Advanced Light Source beamline on 8.2.1. X-ray intensities and data reduction were evaluated and integrated using XDS and merged/scaled using Pointless/Aimless in the CCP4 program suite[48,49]. Structure determination and refinement starting phases were obtained by molecular replacement using Phaser using the designed model for the structures[50]. Following molecular replacement, the models were improved using phenix.autobuild[51]; efforts were made to reduce model bias by setting rebuild-in-place to false, and using simulated annealing and prime-and-switch phasing. Structures were refined in Phenix[51]. Model building was performed using COOT[52]. The final model was evaluated using MolProbity[53]. Data collection and refinement statistics are recorded in Table S1. Data deposition, atomic coordinates, and structure factors will be deposited in the Protein Data Bank (PDB).

Table S1. Data collection and refinement statistics

	PPE18-PE13-d2	**PE13-synzipE2a: PPE18-180-synzipE2b**
Data collection		
Space group	C2	$P2_1 2_1 2_1$
Cell dimensions		
a, b, c (Å)	88.65, 29.86, 105.40	34.56, 87.83, 187.64
a, b, g (°)	90, 103.24, 90	90, 90, 90
Resolution (Å)	51.30 - 2.61 (2.75 - 2.61)	44.91 - 2.81 (2.96 - 2.81)
R_{merge}	0.318 (2.910)	0.491 (1.722)
I / sI	4.3 (0.8)	6.4 (1.7)
Completeness (%)	99.8 (100)	99.8 (100)
Redundancy	5.7 (6.0)	10.6 (11.2)

Refinement		
Resolution (Å)	51.30 - 2.61 (2.77 - 2.61)	46.91 - 2.81 (2.91 - 2.81)
No. reflections	8441 (1386)	13518 (1241)
R_{work} / R_{free}	0.2675 (0.3690)/ 0.3178 (0.3975)	0.2288 (0.3265)/ 0.2887 (0.4095)
No. atoms		
Protein	2050	2603
Ligand/ion	n/a	n/a
Water	22	21
B-factors		
Protein	60	51
Ligand/ion	n/a	n/a
Water	46	32
R.m.s. deviations		
Bond lengths (Å)	0.003	0.003
Bond angles (°)	0.510	0.480

*Values in parentheses are for highest-resolution shell.

Acknowledgements

Thank you to all of the scientists whose work on tuberculosis, its lifecycle, and immune responses makes structure-based vaccine design possible.

Maintenance of the computational infrastructure and technical help establishing the Colabfold environment were generously provided by Luki Goldschmidt and Patrick Vecchiato.

Crystallization and structure determination were performed by Asim Bera, Alex Kang, Emily Joyce, and Evans Brackenbrough.

Crystallographic data collected at the National Synchrotron Light Source II beamline. The Center for Bio-Molecular Structure (CBMS) is primarily supported by the NIH-NIGMS through a Center Core P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1607011). NSLS2 is a U.S.DOE Office of Science User Facility operated under Contract No. DE-SC0012704. This publication resulted from the data collected using the beamtime obtained through NECAT BAG

proposal # 311950. We also want to thank the Advanced Light Source (ALS) beamline 8.2.1 at Lawrence Berkeley National Laboratory for X-ray crystallography data collection. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health (NIH), National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences and US Department of Energy (DOE) (DE-AC02-05CH11231).

Sebastian Ols graciously helped conduct cell surface display experiments.

Sequences

M72

>m72

MSTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLGLGVVDNNGNGARVQ
RVVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQTKSGGTRTGNVTL
AEGPPAEFMVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTV
GSWIGSSAGLMVAAASPYVAWMSVTAGQAELTAAQVRAAAAYETAYGLTVPPPVIAENRAELMI
LIATNLLGQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQA
AAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISNMVSMANN
HMSMTNSGVSMTNTLSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSSGLGGGVAANL
GRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVGQMGARAGGG
LSGVLRVPPRPYVMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQVGPQVVNINTKLGYN
NAVGAGTGIVIDPNGVVLTNNHVIAGATDINAFSVGSGQTYGVDVVGYDRTQDVAVLQLRGAGG
LPSAAIGGGVAVGEPVVAMGNSGGQGGTPRAVPQRVVALGQTVQASDSLTGAEETLNGLIQFD
AAIQPGDSGGPVVNGLGQVVGMNTAASGSGSHHHHHH

Mtb32A antigens

>mtb32A

MSAPPALSQDRFADFPALPLDPSAMVAQVGPQVVNINTKLGYNNAVGAGTGIVIDPNGVVLTNN
HVIAGATDINAFSVGSGQTYGVDVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMG
NSGGQGGTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQV
VGMNTAASDNFQLSQGGQGFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLGLGVVDNNGNGA
RVQRVVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQTKSGGTRTGN
VTLAEGPPAGGGSGSHHHHHH

>mtb32A_S208A

MSAPPALSQDRFADFPALPLDPSAMVAQVGPQVVNINTKLGYNNAVGAGTGIVIDPNGVVLTNN
HVIAGATDINAFSVGSGQTYGVDVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMG
NSGGQGGTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDAGGPVVNGLGQV
VGMNTAASDNFQLSQGGQGFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLGLGVVDNNGNGA
RVQRVVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQTKSGGTRTGN
VTLAEGPPAGGGSGSHHHHHH

PPE18 antigens **(**Note: bicistronic constructs separated by a //—//)

>PPE18

MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIGSS
AGLMVAAASPYVAWMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMILIATNLL
GQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEA
SDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISNMVSMANNHMSMT
NSGVSMTNTLSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSSGLGGGVAANLGRAAS
VGSLSVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVGQMGARAGGGLSGVLR
VPPRPYVMPHSPAAGGSGSHHHHHH

>PPE18_180

MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIGSS
AGLMVAAASPYVAWMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMILIATNLL
GQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEMTSAGGGSGSHHHHHH
>PE13:PPE18
MSFVMAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAMYQS
VSARAAAIHDQFVATLASSASSYAATEVANAAAAS//—//MVDFGALPPEINSARMYAGPGSASLV
AAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIGSSAGLMVAAASPYVAWMSVTAGQAELTA
AQVRVAAAAYETAYGLTVPPPVIAENRAELMILIATNLLGQNTPAIAVNEAEYGEMWAQDAAAMF
GYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQG
TTPSSKLGGLWKTVSPHRSPISNMVSMANNHMSMTNSGVSMTNTLSSMLKGFAPAAAAQAVQ
TAAQNGVRAMSSLGSSLGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLT
SLTSAAERGPGQMLGGLPVGQMGARAGGGLSGVLRVPPRPYVMPHSPAAGGSGSHHHHHH
>PE13:PPE18_solubilizing_mutations
MSFVMAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAMYQS
VSARAAAIHDQFVATLASSASSYAATEVANAAAAS//—//MVDFGALPPEINSARMYAGPGSASLV
AAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSAGLMVAAASPYVAWMSATAGQAEETA
AQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNENGQNTPAIATNEAEYGEMWAQDAAAM
FGYAAATATATATLTPFEEAPEMTNAGGLLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQ
GTTPSSKLGGLWKTVSPHRSPISNMVSMANNHMSMTNSGVSMTNTLSSMLKGFAPAAAAQAV
QTAAQNGVRAMSSLGSSLGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPL
TSLTSAAERGPGQMLGGLPVGQMGARAGGGLSGVLRVPPRPYVMPHSPAAGSGSHHHHHH
>PE13:PPE18_180
MSFVMAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAMYQS
VSARAAAIHDQFVATLASSASSYAATEVANAAAAS//—//MVDFGALPPEINSARMYAGPGSASLV
AAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIGSSAGLMVAAASPYVAWMSVTAGQAELTA
AQVRVAAAAYETAYGLTVPPPVIAENRAELMILIATNLLGQNTPAIAVNEAEYGEMWAQDAAAMF
GYAAATATATATLLPFEEAPEMTSAGGGSGSHHHHHH
>PE13:PPE18_180_solubilizing_mutations
MSFVMAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAMYQS
VSARAAAIHDQFVATLASSASSYAATEVANAAAAS//—//MVDFGALPPEINSARMYAGPGSASLV
AAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSAGLMVAAASPYVAWMSATAGQAEETA
AQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNENGQNTPAIATNEAEYGEMWAQDAAAM
FGYAAATATATATLTPFEEAPEMTNAGGGSGSHHHHHH
>PPE18-PE13-d1
MVDFGALPPEINSARMYAGPGSASLVAAAQAWDSLASDLRSAASAFQSVVDGLTTGSWIGSSA
ALFVAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNENG
QNTPAIAKNEAEYGEMWAQDAAAMFGYAAATATATATLTPFEEAPEMTNAGGSGSGSGMSFVM
AYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAQFQSVSARA
AAIHDQFVATLASSASSYAATEKANAAAASGSGSHHHHHH
>PPE18-PE13-d2
MVDFGALPPEINSARMYAGPGSASLVAAAQAWDSLASDLRSAASAFQSVVDGLTTGSWIGSSA
ALFVAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNENG
QNTPAIAKNEAEYGEMWAQDAAAMFGYAAATATATATLTPFEEAPEEGVEKMQEAHQQLLQDL

EKLSNDLDALRKTHLKPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFA
AHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGSGSHHHHHH
>PPE18-PE13-d3
MSFVMAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAQFQS
VSARAAAIHDQFVATLASSASSYAATEKANAAAAKGKPGEAATANDNDGEGTAQAESAGASGG
DSSAELTDTPRSVDFGALPPEINSARMYAGPGSASLVAAAQAWDSLASDLRSAASAFQSVVDG
LTTGSWIGSSAALFVAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRA
ELETLIATNENGQNTPAIAKNEAEYGEMWAQDAAAMFGYAAATATATATLTPFEEAPEMTNAGG
SGGSHHHHHH
>PPE18-PE13-d4
MSFVMAYPEMLAAAADTLASIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAQFQS
VSARAAAIHVQFVATLASSASSYAATEKANAAAASQKELEKEAQKILNRINERKKRLEQLGQENL
QTAQETLEQSKELVKKVLEVVREVLKDEEVSKKLRELVERIIEGDKEAWEELVKLFQETLKSVDF
GALPPEINSARMYAGPGSASLVAAAQAWDSLASDLRSAASAFQSVVDGLTTGSWIGSSAALFVA
AASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNENGQNTP
AIAKNEAEYGEMWAQDAAAMFGYAAATATATATLTPFEEAPEMTNAGGSGGSHHHHHH
>PPE18-PE13-d5
MSFVMAYPEMLAAAADTLASIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAQFQS
VSARAAAIHVQFVATLASSASSYAATEKANAAAASNQELEERAREILKKIEERREEIEKRAEENLK
LSEETLRLSKELTLEVLRRTKELVPEELQPEVEEIIKLVEEGKFKEAFEKMKELVEKMLETVDFGA
LPPEINSARMYAGPGSASLVAAAQAWDSLASDLRSAASAFQSVVDGLTTGSWIGSSAALFVAAA
SPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNENGQNTPAIA
KNEAEYGEMWAQDAAAMFGYAAATATATATLTPFEEAPEMTNAGGSGGSHHHHHH
>PPE18-PE13-d2.1
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMVAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAAATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGSGSHHHHHH
>PPE18-PE13-d2.2
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMVAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAAATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLGGSGGSGGSGGSAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAA
HFAAHAAQFQSVSARAAAIHDQFVATLASSASSYAALEKANAAAASGSGSHHHHHH
>PE13-synzipE2a:PPE18-180-synzipE2b
MSFVMAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHFAAHAAQFQS
VSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSNLVAQLENEVASCENENETLKKK
NLHKKDLIAYLEKEIANLRKKIEE//—//MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVAS
DLRSAASAFQSVVDGLTTGSWIGSSAGLMVAAASPYVAWMSATAGQAEETAAQVRVAAAAYET
AYGQTVPPPVIAENRAELETLIATNENGQNTPAIATNEAEYGEMWAQDAAAMFGYAAATATATAT
LTPFEEAPEMTNAGGSGSGSGSGSGSGSGSGGSGGSARNAYLRKKIARCKKDNLQLERDEQNLE
KIIANLRDEIARLENEVASHEQGSGSHHHHHH

Membrane anchored PPE18 antigens

>M72_hPDGFR_TMD
MTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLGLGVVDNNGNGARVQR
VVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQTKSGGTRTGNVTLA
EGPPAEFMVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTV
GSWIGSSAGLMVAAASPYVAWMSVTAGQAELTAAQVRAAAAYETAYGLTVPPPVIAENRAELMI
LIATNLLGQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQA
AAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISNMVSMANN
HMSMTNSGVSMTNTLSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSSGLGGGVAANL
GRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVGQMGARAGGG
LSGVLRVPPRPYVMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQVGPQVVNINTKLGYN
NAVGAGTGIVIDPNGVVLTNNHVIAGATDINAFSVGSGQTYGVDVVGYDRTQDVAVLQLRGAGG
LPSAAIGGGVAVGEPVVAMGNSGGQGGTPRAVPQRVVALGQTVQASDSLTGAEETLNGLIQFD
AAIQPGDSGGPVVNGLGQVVGMNTAASGGSGGSGGSEQKLISEEDLGGSGGSVVVISAILALV
VLTIISLIILIMLWQKK
>M72_V67D_A157E_hPDGFR_TMD
MTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLGLGVVDNNGNGARVQR
VVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQTKSGGTRTGNVTLA
EGPPAEFMVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTV
GSWIGSSAGLMDAAASPYVAWMSVTAGQAELTAAQVRAAAAYETAYGLTVPPPVIAENRAELMI
LIATNLLGQNTPAIAVNEAEYGEMWAQDAAAMFGYAEATATATATLLPFEEAPEMTSAGGLLEQA
AAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISNMVSMANN
HMSMTNSGVSMTNTLSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSSGLGGGVAANL
GRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVGQMGARAGGG
LSGVLRVPPRPYVMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQVGPQVVNINTKLGYN
NAVGAGTGIVIDPNGVVLTNNHVIAGATDINAFSVGSGQTYGVDVVGYDRTQDVAVLQLRGAGG
LPSAAIGGGVAVGEPVVAMGNSGGQGGTPRAVPQRVVALGQTVQASDSLTGAEETLNGLIQFD
AAIQPGDSGGPVVNGLGQVVGMNTAASGGSGGSGGSEQKLISEEDLGGSGGSVVVISAILALV
VLTIISLIILIMLWQKK
>PPE18-PE13-d2.2_hPDGFR_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL
GGSGGSVVVISAILALVVLTIISLIILIMLWQKK
>PPE18-PE13-d2.2_mouseCD80_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL
GGSGGSPPDSKNTLVLFGAGFGAVITVVVIVVIIKCFCKHRSCFRRNEASRETNNSLTFGPEEAL
AEQTVFL

>PPE18-PE13-d2.2_VSVG_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL
GGSGGSSSAKSSIASFFFIIGLIIGLFLVLRVGIHLCIKLKHTKKRQIYTDIEMNRLGK
>PPE18-PE13-d2.2_Dmel_PS4_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL
GGSGGGSHAEIAIWKIIVSVIVGILVFSAATYALYKRGFFKRAIKDDLKQLIRDSFEDGIIRTEMEENA
QSQGDADLDEKLDAYADTTGKCTHV
>PPE18-PE13-d2.2_Dmel_PS5_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL
GGSGGSSETPLPIWYIILSLIAGHLLLGAMTYILYKLRFFKRGKKEELKRLLEEHRSETKEPATDC
EGNQEEINVEMHSDLEN
>PPE18-PE13-d2.2_ZYE4_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL
GGSGGSDMVIPIIIISTVLLLVCIMCISITIGCVYQMKRIKRNQRHHAISEKTMRNQTFDLVNQSLL
>PPE18-PE13-d2.2_EPS9_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL
GGSGGSVNKITIVVASAVVFYIVITVFVCILKHKECYQSKRKRSMKKETNDSRETDHFITFKK
>PPE18-PE13-d2.2_L7I1_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL

GGSGGSYFSRELLAIVMLAVPALIFLVALLIQILKSRPDRSCIFSSERKRFQKEESNSENDINERQ
QSTHCYATTDLRYD
>PPE18-PE13-d2.2_NC05_TMD
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLRSAASAFQSVVDGLTTGSWIGSSA
GLMDAAASPYVAWMSATAGQAEETAAQVRVAAAAYETAYGQTVPPPVIAENRAELETLIATNEN
GQNTPAIATNEAEYGEMWAQDAAAMFGYAEATATATATLTPFEEAPEEGVEKMQEAHQQLLQD
LEKLSNDLDALRKTHAYPEMLAAAADTLQSIGATTVASNAAAAAPTTGVVPPAADEVSALTAAHF
AAHAAQFQSVSARAAAIHDQFVATLASSASSYAATEKANAAAASGGSGGSGGSEQKLISEEDL
GGSGGSSGGGKAGIVIFVLIVVAIGIGIGIGFFLWRRKEEEKKKEQEKDPEKEPLSPLQIEFEDLA
KKINGYTRGPFLGWR

Rabies

Rabies virus (RABV) causes progressive encephalitis that is almost always fatal if untreated[1,2]. The predominant route of transmission to humans is through the saliva of an infected animal[1,3]. Once in skin or muscle, the negative-strand RNA virus infects peripheral motor neurons and the central nervous system during an asymptomatic period that typically lasts 1-3 months and then causes the rapid onset of neurological symptoms, coma, and death[4]. Following a suspected exposure, administration of a four-dose vaccine series and purified rabies immunoglobulin (RIG) effectively prevents infection and disease[5–7]. The success of RIG demonstrates that antibodies are sufficient to mediate protection, and the WHO considers vaccinated individuals fully protected if they possess serum virus neutralizing antibody (VNA) titers above 0.5 international units (IU) per mL[1,5,8]. Despite a well-defined correlate of protection and effective treatment, it is estimated that 59,000 people die of rabies virus every year, though the true figure is likely much higher due to widespread underreporting[9]. Most of these people live in rural communities that lack sufficient public health infrastructure to afford the high cost of current, whole inactivated vaccines and to administer treatment in the narrow time frame between exposure and symptom onset[5,8,10,11]. An affordable vaccine with more durable vaccine-elicited antibody responses would enable routine immunizations to protect people before exposure[12,13].

Ongoing vaccine design efforts use multiple delivery modalities to elicit protective antibodies against rabies glycoprotein (RABV-G), the sole target of antibodies on the surface of enveloped virions[12,14–19]. A class III viral fusion protein that mediates cell attachment and entry[20–24], RABV-G forms trimers that display a pH-dependent

equilibrium between prefusion and postfusion conformations[25–30]. RABV-G contains central, pleckstrin homology, and fusion domains (CD, PHD, and FD, respectively)[29]. These domains do not change conformation during fusion, but the arrangement between them changes dramatically due to changes in the conformation of short regions that link the domains[29]. Insertion of the hydrophobic fusion loops into the viral membrane stabilizes both the trimer and the prefusion conformation[31]. Removal of both the transmembrane domain and the hydrophobic fusion loops enables recombinant production of soluble RABV-G. However, the protein is monomeric and conformationally dynamic[29]. Though virions display both prefusion and postfusion RABV-G, the most potently neutralizing antibodies target epitopes on the CD and PHD that are only accessible on the prefusion timer[30–35]. Due to its conformational heterogeneity, only a small proportion of recombinant RABV-G protein will present these epitopes, likely eliciting suboptimal antibody responses upon vaccination[31,32].

To develop a subunit vaccine that elicits potent antibody responses, we set out to stabilize the ectodomain of RABV-G in its trimeric, prefusion conformation, the structure of which was recently determined[31,32]. We utilized the Pasteur vaccins (PV) strain, truncated at residue 402 (the last residue resolved in structures of the prefusion trimer), and adopted the previously reported fusion loop substitution strategy. To obtain monodisperse trimers, we flexibly linked the well-validated trimerization domains GCN4 and foldon to the C terminus of RABV-G[36,37] (Fig 4-1A). Constructs were screened in Expi293F cells, with the amount of secreted, antigenically intact RABV-G assessed by cell supernatant ELISA with the neutralizing antibodies RVC20, RVC58, and RVA122,

along with the non-conformational antibody RVC68[6]. Designs that appeared promising in preliminary screening were then further characterized. While fusion to foldon did not improve ELISA reactivity, GCN4 fusion did, likely because its narrow diameter is compatible with the narrow cavity at the base of RABV-G between the FDs. Addition of a GCN4 trimerization domain resulted in improvement in ELISA reactivity against RVA122, which recognizes a quaternary epitope across the trimer[31] (Fig 4-1B). As linkage of GCN4 to the C terminus was successful, subsequent work involved comparison to a base construct (G-d1) containing the C-terminal GCN4 domain to stabilize trimeric RABV-G. Comparisons to *de novo* trimerization domains generated through deep learning methods are in the supplementary note. Though the addition of a GCN4 domain stabilized the desired trimeric state, ELISA reactivity remained quite low to all tested antibodies, indicating that further stabilizing mutations would be necessary to address the conformational heterogeneity of RABV-G.

Recent work introduced mutations within RABV-G at histidines or conformationally dynamic linker regions and used cell surface display to identify two mutations, H261L and H270P, that stabilize the prefusion conformation[32]. Of these, H270P has the largest prefusion-stabilizing effect and is being developed for both mRNA and viral-vectored vaccines[19,38]. RABV-G containing H270P provides a benchmark for our efforts and highlights the classical approach: identify individual mutations at positions with a clear role in RABV-G's pH-dependent conformational heterogeneity. To investigate the use of AlphaFold 2 (AF2) to identify prefusion-stabilizing mutations, we first experimented with AF2 inputs while predicting the structure of G-d1 and sequences containing the H261L

and H270P mutations (G-d1-H261L and G-d1-H270P). With default settings, all five models predict each sequence in the postfusion conformation. We adjusted both the MSA and structural templates to improve discrimination. We were unsuccessful in using MSA subsampling to predict conformational heterogeneity, as the predicted structure depended on the particular MSA but was unaffected by known stabilizing mutations, in line with a recent publication on the limits of subsampling MSAs[39]. Choice of structural template altered AF2 predictions, and we settled on use of two templates, prefusion VSV-G (PDB ID 6TIT) and a dimeric form of CHPV-G with both prefusion and postfusion conformations (PDB ID 5MDM)[40,41]. Predictions made with these templates displayed conformational heterogeneity between models and differences in model confidence that reflected the prefusion-stabilizing effect of H270P, with one of five predictions in prefusion conformation for G-d1 (pLDDT 66) and two for G-d1-H270P (pLDDT 72-74).

Having established an AF2 workflow to assess conformational heterogeneity, we focused on two strategies to stabilize the prefusion conformation of RABV-G. We used ProteinMPNN to either identify isolated stabilizing mutations within conformationally dynamic regions or mutations distributed throughout the protein (Fig 4-2A).

For the first strategy, which followed the classical approach to prefusion stabilization, we began with an experimental structure of trimeric, prefusion RABV-G in the membrane at 3.4 Å resolution (PDB ID 7U9G), along with a high-confidence AF2 model of postfusion RABV-G that displays close agreement to the crystallized extended intermediate (PDB ID 6LGW)[29,31]. Both structures served as inputs into ProteinMPNN to identify mutations

within conformationally dynamic regions (residues 48-54, 184-188, and 259-276)[29] (Fig 4-2A). We obtained 1000 output sequences for each structure to identify mutations that were frequently suggested by ProteinMPNN for the prefusion conformation but were rare or unobserved for the postfusion conformation. Encouragingly, our strategy identified H270P and multiple mutations at H261. In addition, DisulfidebyDesign 2.0 suggested a paired mutation that would introduce a potentially stabilizing disulfide[42]. We created sequences containing a random combination of one to three of the identified mutations and evaluated these sequences using the AF2 workflow we developed. We filtered designs based on the number of AF2 predictions in the prefusion conformation (keeping those with at least 2 predictions in prefusion conformation) and the model confidence of those prefusion predictions (pLDDT > 68) . The top 9 sequences, all of which contained mutations in residues 260-276, were chosen for screening (G-d2 through G-d10).

For the second strategy, we used the same ProteinMPNN-based two-step process as for rotavirus VP5* and GAS SLO to design the prefusion RABV-G structure (see Methods) (Fig 4-2A). We allowed design at all residues except those involved in known epitopes, disulfide bonds, or glycans in the first step. In the second round of ProteinMPNN, we output 200 sequences, all of which were evaluated using the AF2 workflow developed. We ordered three sequences (G-d11 through G-d13) predicted to be better than G-d1-H270P based on the number of AF2 predictions in prefusion conformation (at least 3) and model confidence (pLDDT > 75). Furthermore, to assess the interpretability of the mutations chosen, we manually inspected each design to

identify common mutations and try to rationalize them in PyMOL. We created two additional designs, one with only the mutations we thought had a clear role in stabilization by visual inspection in PyMOL (G-d14), and one with those mutations and others that improved solubility (G-d15).

Screening the nine sequences from the first strategy along with five sequences from the second approach revealed that two of the three designed antigens with distributed mutations, G-d12 and G-d13, displayed improved ELISA reactivity relative to G-d1 and sequences containing isolated stabilizing mutations (Fig 4-2B). G-d12 and G-d13 contain 28 and 45 mutations relative to G-d1, respectively, suggesting that a large number of distributed mutations across RABV-G identified by deep learning methods can outperform classical stabilizing mutations(Fig 4-2C). G-d14 and G-d15, which contain smaller, manually curated sets of mutations, did not outperform G-d1, indicating that the relative contribution of these mutations is not always easily interpretable by manual inspection. Among sequences from the first strategy, one containing the single mutation V273P performed comparably to G-d1-H270P, suggesting that ProteinMPNN and AF2 can identify stabilizing mutations in conformationally dynamic regions in the same manner as previous experimental approaches.

To determine if we could further improve G-d12 and G-d13, we recombined them, focusing on twenty positions where only one of the two designs differed from the G-d1 residue. At each position, we decided to revert to the native residue or introduce the ProteinMPNN-derived mutation based on visual inspection in PyMOL. The process led

to constructs (G-d12R and G-d13R) with 31 and 35 mutations relative to G-d1. G-d12R and G-d13R displayed improved ELISA reactivity relative to G-d12 and G-d13, suggesting that distinct sets of ProteinMPNN-derived stabilizing mutations could be combined for further stabilization (Fig 4-2B). We then assessed whether isolated stabilizing mutations in residues 260-276 could additively stabilize our designs with distributed mutations throughout RABV-G. These constructs, including designs containing H270P, displayed comparable ELISA reactivity to G-d12R and G-d13R, indicating that isolated stabilizing mutations likely had no additive effects for these designed antigens. However, we carried forward G-d12R and G-d13R, with or without the H270P mutation, to examine how addition of isolated stabilizing mutations affected immunogenicity. Size exclusion chromatography (SEC) revealed that all constructs had considerably higher yield than G-d1 and G-d1-H270P (Fig 4-3A).

To assess the immunogenicity of the designed antigens, we purified soluble RABV-G trimers via IMAC and SEC, and then immunized mice at weeks 0 and 4 with 3 μg of antigen, collecting sera at weeks 0, 2, and 6 (Fig 4-3B). The study contained seven groups of 10 mice, with six RABV-G antigens and one irrelevant protein control. We sent pooled pre-immunization sera and week 6 sera from individual mice to the Kansas State University Rabies Laboratory, which performed RFFIT neutralization assays. As expected, no VNA was detected among pre-immunized mice or those immunized with an irrelevant protein control. Most mice did not have detectable VNA after immunization with G-d1 (mean VNA 0.057), as has been previously reported, and none had VNA above the WHO level of 0.5 IU/mL (Fig 4-3C). Similar results were seen for mice

immunized with G-d1-H270P (mean VNA 0.078), though one mouse had VNA above the WHO threshold for protection. These results are in line with an mRNA immunogenicity study in which RABV-G containing H270P led to 2.4-fold improvements in VNA relative to native RABV-G. However, mice immunized with G-d12R and G-d13R possessed significantly higher VNA (mean VNA 12.06 and 19.51, respectively), with all mice in each group possessing VNA above 0.5 IU/mL. Immunization with G-d12R-H270P and G-d13R-H270P resulted in comparable mean levels of VNA (mean VNA 29.9 and 19.31, respectively) but higher variability, suggesting that its inclusion did not further improve immunogenicity (Fig 4-3C). Together, these results indicate that stabilization of RABV-G in its prefusion conformation with numerous, distributed mutations improved vaccine-elicited protection.

We sought to further stabilize our constructs using mutations suggested by MSAs and a recent deep mutational scan (DMS) of RABV-G[43]. Since RABV-G has evolved for conformational heterogeneity, mutations in the MSA are unlikely to be prefusion-stabilizing. However, we reasoned that mutations to consensus residues within the MSA might improve the breadth of vaccine-elicited protection and solubility. Six mutations from a standard MSA of RABV-G were identified, and all were included in a single construct as they are surface-exposed and did not appear deleterious by visual inspection in PyMOL. A recent DMS provided a mean effect on functional entry of every mutation at each position in RABV-G[43]. Since functional entry requires RABV-G to adopt a postfusion conformation, we looked for mutations that had strongly negative effects on cell entry but did not appear to be deleterious to the prefusion conformation by visual

inspection in PyMOL and AF2 prediction. As our AF2 workflow already predicted that the designed constructs G-d12R and G-d13R stabilized the prefusion conformation, it failed to identify any additional stabilizing effects of DMS mutations. To evaluate these mutations, 14 sequences which each contained a single DMS mutation in addition to the mutations in G-d12R were empirically screened. In addition, we screened nine reversions of ProteinMPNN mutations, as they appeared to be functionally neutral by DMS. The MSA mutations, five of the 14 DMS mutations, and one reversion of a ProteinMPNN mutation appeared to further improve cell supernatant ELISA reactivity relative to the control, G-d12R (Supp Fig 4-5). A sequence containing all of these mutations (G-d12R.8) resulted in further improvements in ELISA reactivity. Furthermore, differential scanning fluorimetry (DSF) analysis revealed that G-d12R.8 possessed a T_m of 70ºC (Fig 4-4D). While we were unable to obtain G-d1 at a sufficient concentration for DSF, full-length RABV-G extracted from virions into detergent micelles was reported to have a T_m of 44-50ºC (depending on the detergent)[44], indicating a substantial improvement in thermal stability in G-d12R.8. These results highlight how deep learning approaches, experimental DMS, and MSA consensus mutations can be complementary in antigen stabilization.

We measured the affinity between G-d12R.8 and the fragment antigen-binding (Fab) region of the neutralizing antibodies RVA122, RVC20, and RVA58 using surface plasmon resonance (SPR). SPR analysis revealed K_D values of 3.89 nM, 1.56 nM, and 15.7 nM, respectively (Fig 4-4 A-C). These values are consistent with published characterization of these antibodies against native RABV-G, but quantitative

comparison of affinity is difficult because published reports used IgG, which introduces avidity effects[6]. To further inspect the interaction between G-d12R.8 and RVA122, we determined the cryo-EM structure of a complex between RVA122 Fab and G-d12R.8 at 3.3 Å resolution (Fig 4-4E). The FDs and GCN4 trimerization domains were unresolved, indicating that they are flexible. We speculate that the flexibility in the base of G-d12R.8 likely arises from the absence of FD insertion into the viral membrane, though it is also possible that the entire protein is flexible but RVA122 binding introduces rigidity in the apical region. The CD and PHD domains were both resolved and revealed that the designed antigen preserves the intended conformation and interaction with RVA122[31]. Together, these data show that distributed ProteinMPNN-derived stabilizing mutations improved the yield, stability, and vaccine-elicited antibody response of RABV-G.

References

1. Liu, C. & Cahill, J. D. Epidemiology of Rabies and Current US Vaccine Guidelines. *R. I. Med. J.* **103**, 51–53 (2020).
2. Fooks, A. R. *et al.* Current status of rabies and prospects for elimination. *Lancet* **384**, 1389–1399 (2014).
3. Fooks AR, Cliquet F, Finke S, Freuling C, Hemachudha T, Mani RS, Müller T, Nadin-Davis S, Picard-Meyer E, Wilde H, Banyard AC. Rabies. *Nat. Rev. Dis. Primers* **3**, (2017).
4. Hemachudha, T. *et al.* Human rabies: neuropathogenesis, diagnosis, and management. *Lancet Neurol.* **12**, 498–513 (2013).
5. World Health Organization. *WHO Expert Consultation on Rabies: Third Report*. (World Health Organization, Genève, Switzerland, 2018).
6. De Benedictis, P. *et al.* Development of broad-spectrum human monoclonal antibodies for rabies post-exposure prophylaxis. *EMBO Mol. Med.* **8**, 407–421 (2016).
7. de Melo, G. D., Hellert, J., Gupta, R., Corti, D. & Bourhy, H. Monoclonal antibodies against rabies: current uses in prophylaxis and in therapy. *Curr. Opin. Virol.* **53**, 101204 (2022).
8. Johnson, N., Cunningham, A. F. & Fooks, A. R. The immune response to rabies virus infection and vaccination. *Vaccine* **28**, 3896–3901 (2010).
9. Hampson, K. *et al.* Estimating the global burden of endemic canine rabies. *PLoS Negl. Trop. Dis.* **9**, e0003709 (2015).
10. Changalucha, J. *et al.* The need to improve access to rabies post-exposure vaccines: Lessons from Tanzania. *Vaccine* **37 Suppl 1**, A45–A53 (2019).
11. Lodha, L., Manoor Ananda, A. & Mani, R. S. Rabies control in high-burden countries: role of universal pre-exposure immunization. *Lancet Reg. Health Southeast Asia* **19**, 100258

(2023).

12. Fooks, A. R., Banyard, A. C. & Ertl, H. C. J. New human rabies vaccines in the pipeline. *Vaccine* **37 Suppl 1**, A140–A145 (2019).
13. Zhang, X., Zhu, Z. & Wang, C. Persistence of rabies antibody 5 years after postexposure prophylaxis with vero cell antirabies vaccine and antibody response to a single booster dose. *Clin. Vaccine Immunol.* **18**, 1477–1479 (2011).
14. Maki, J. *et al.* Oral vaccination of wildlife using a vaccinia-rabies-glycoprotein recombinant virus vaccine (RABORAL V-RG®): a global review. *Vet. Res.* **48**, 57 (2017).
15. Li, Q. *et al.* Immunogenicity and antigenicity of the ectodomain of rabies virus glycoprotein stably expressed in HEK293T cells. *Int. J. Med. Sci.* **20**, 1282–1292 (2023).
16. Li, Q. *et al.* A single dose of recombinant adenoviral vector rabies vaccine expressing two copies of glycoprotein protects mice from lethal virus challenge. *J. Infect. Dev. Ctries.* **18**, 1281–1290 (2024).
17. Aldrich, C. *et al.* Proof-of-concept of a low-dose unmodified mRNA-based rabies vaccine formulated with lipid nanoparticles in human volunteers: A phase 1 trial. *Vaccine* **39**, 1310–1318 (2021).
18. Hellgren, F. *et al.* Unmodified rabies mRNA vaccine elicits high cross-neutralizing antibody titers and diverse B cell memory responses. *Nat. Commun.* **14**, 3713 (2023).
19. Cao, H. *et al.* A rabies mRNA vaccine with H270P mutation in its glycoprotein induces strong cellular and humoral immunity. *Vaccine* **42**, 1116–1121 (2024).
20. Harrison, S. C. Viral membrane fusion. *Nat. Struct. Mol. Biol.* **15**, 690–698 (2008).
21. Backovic, M. & Jardetzky, T. S. Class III viral membrane fusion proteins. *Curr. Opin. Struct. Biol.* **19**, 189–196 (2009).
22. Baquero, E., Albertini, A. A. V. & Gaudin, Y. Recent mechanistic and structural insights on class III viral fusion glycoproteins. *Curr. Opin. Struct. Biol.* **33**, 52–60 (2015).
23. Lentz, T. L., Burrage, T. G., Smith, A. L. & Tignor, G. H. The acetylcholine receptor as a

cellular receptor for rabies virus. *Yale J. Biol. Med.* **56**, 315–322 (1983).

24. Thoulouze, M. I. *et al.* The neural cell adhesion molecule is a receptor for rabies virus. *J. Virol.* **72**, 7181–7190 (1998).
25. Whitt, M. A., Buonocore, L., Prehaud, C. & Rose, J. K. Membrane fusion activity, oligomerization, and assembly of the rabies virus glycoprotein. *Virology* **185**, 681–688 (1991).
26. Gaudin, Y., Ruigrok, R. W., Tuffereau, C., Knossow, M. & Flamand, A. Rabies virus glycoprotein is a trimer. *Virology* **187**, 627–632 (1992).
27. Gaudin, Y., Tuffereau, C., Segretain, D., Knossow, M. & Flamand, A. Reversible conformational changes and fusion activity of rabies virus glycoprotein. *J. Virol.* **65**, 4853–4859 (1991).
28. Gaudin, Y., Ruigrok, R. W., Knossow, M. & Flamand, A. Low-pH conformational changes of rabies virus glycoprotein and their role in membrane fusion. *J. Virol.* **67**, 1365–1372 (1993).
29. Yang, F. *et al.* Structural Analysis of Rabies Virus Glycoprotein Reveals pH-Dependent Conformational Changes and Interactions with a Neutralizing Antibody. *Cell Host Microbe* **27**, 441–453.e7 (2020).
30. Cai, X. *et al.* Structural heterogeneity of the rabies virus virion. *Viruses* **16**, 1447 (2024).
31. Callaway, H. M. *et al.* Structure of the rabies virus glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Sci Adv* **8**, eabp9151 (2022).
32. Ng, W. M. *et al.* Structure of trimeric pre-fusion rabies virus glycoprotein in complex with two protective antibodies. *Cell Host Microbe* **30**, 1219–1230.e7 (2022).
33. Zorzan, M. *et al.* Antiviral mechanisms of two broad-spectrum monoclonal antibodies for rabies prophylaxis and therapy. *Front. Immunol.* **14**, 1186063 (2023).
34. Hellert, J. *et al.* Structure of the prefusion-locking broadly neutralizing antibody RVC20 bound to the rabies virus glycoprotein. *Nat. Commun.* **11**, 596 (2020).
35. Kedari, A. *et al.* Structural insight into rabies virus neutralization revealed by an engineered

antibody scaffold. *Structure* (2024) doi:10.1016/j.str.2024.10.002.

36. Harbury, P. B., Kim, P. S. & Alber, T. Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80–83 (1994).
37. Güthe, S. *et al.* Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *J. Mol. Biol.* **337**, 905–915 (2004).
38. Jenkin, D. *et al.* Safety and immunogenicity of a simian-adenovirus-vectored rabies vaccine: an open-label, non-randomised, dose-escalation, first-in-human, single-centre, phase 1 clinical trial. *Lancet Microbe* **3**, e663–e671 (2022).
39. Chakravarty, D. *et al.* AlphaFold predictions of fold-switched conformations are driven by structure memorization. *Nat. Commun.* **15**, 7296 (2024).
40. Beilstein, F. *et al.* Identification of a pH-sensitive switch in VSV-G and a crystal structure of the G pre-fusion state highlight the VSV-G structural transition pathway. *Cell Rep.* **32**, 108042 (2020).
41. Baquero, E. *et al.* Structural intermediates in the fusion-associated transition of vesiculovirus glycoprotein. *EMBO J.* **36**, 679–692 (2017).
42. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **14**, 346 (2013).
43. Aditham, A. K. *et al.* Deep mutational scanning of rabies glycoprotein defines mutational constraint and antibody-escape mutations. *bioRxivorg* (2024) doi:10.1101/2024.12.17.628970.
44. Clénet, D. *et al.* Full-length G glycoprotein directly extracted from rabies virus with detergent and then stabilized by amphipols in liquid and freeze-dried forms. *Biotechnol. Bioeng.* **118**, 4317–4330 (2021).
45. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
46. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**,

679–682 (2022).

47. Wayment-Steele, H. K. *et al.* Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 832–839 (2024).
48. Monteiro da Silva, G., Cui, J. Y., Dalgarno, D. C., Lisi, G. P. & Rubenstein, B. M. High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nat. Commun.* **15**, 2464 (2024).
49. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res.* **53**, D20–D29 (2025).
50. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
51. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
52. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
53. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
54. Fallahi, F., Wandeler, A. I. & Nadin-Davis, S. A. Characterization of epitopes on the rabies virus glycoprotein by selection and analysis of escape mutants. *Virus Res.* **220**, 161–171 (2016).
55. Shi, C. *et al.* Research progress on neutralizing epitopes and antibodies for the Rabies virus. *Infect Med (Beijing)* **1**, 262–271 (2022).
56. Wang, W. *et al.* Antigenic variations of recent street rabies virus. *Emerg. Microbes Infect.* **8**, 1584–1592 (2019).
57. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
58. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of

specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

59. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

Figures

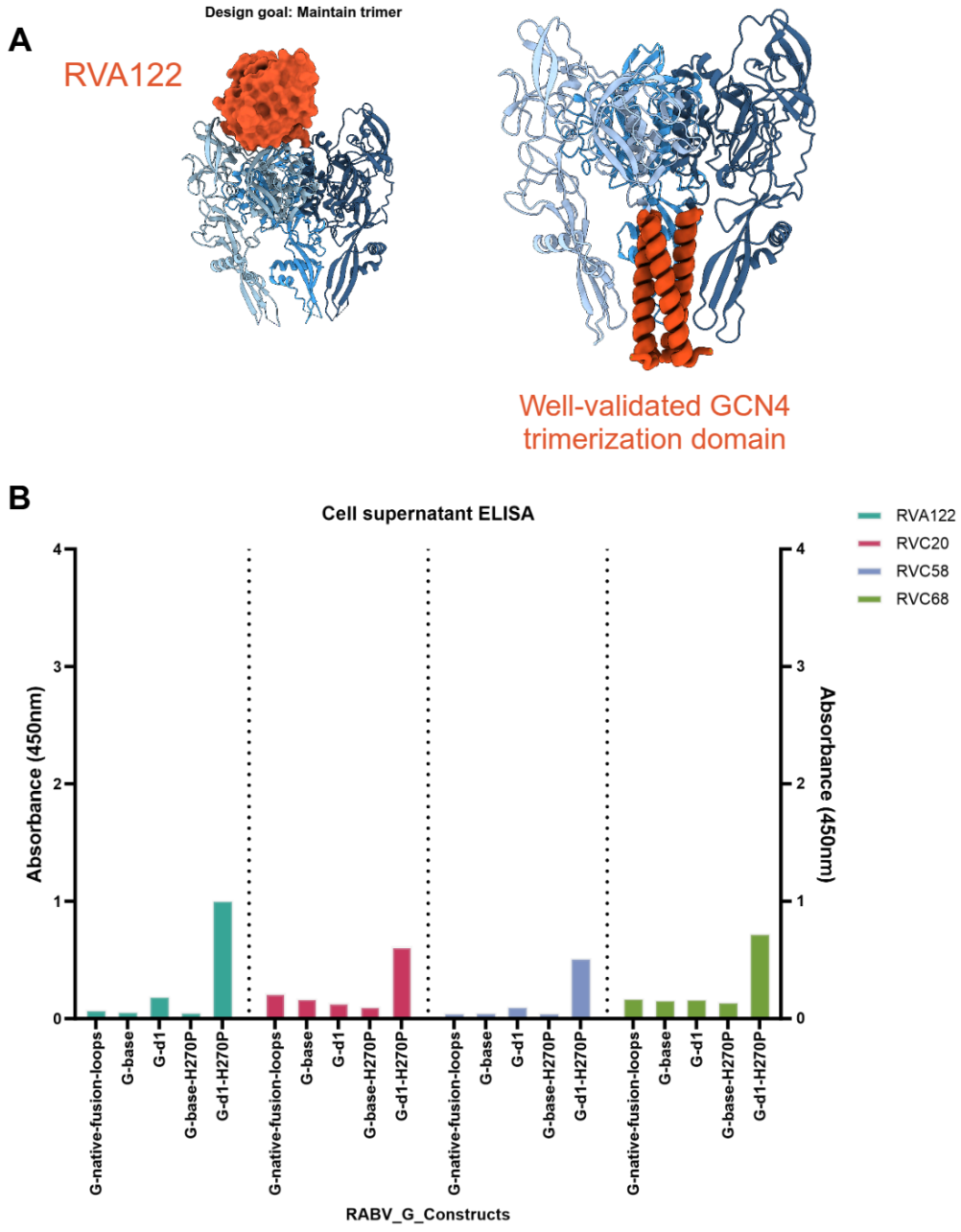


Figure 4-1. Fusion of GCN4 domain stabilizes trimeric form of RABV-G. A) Scheme of design goal to elicit potent neutralizing antibodies like RVA122, which recognizes a quaternary epitope within trimeric RABV-G, via fusion of well-validated GCN4 trimerization domain. RVA122 shown as an orange surface, the three chains of RABV-G in cartoon representation with varying shades of blue, and GCN4 domains with modified ribbon representation in orange. B) Cell supernatant ELISA reactivity against three

neutralizing antibodies (RVA122, RVC20, and RVC58) along with RVC68, which recognizes a linear epitope. G-d1 constructs (with or without known prefusion-stabilizing mutation H270P) contain the GCN4 trimerization domain.

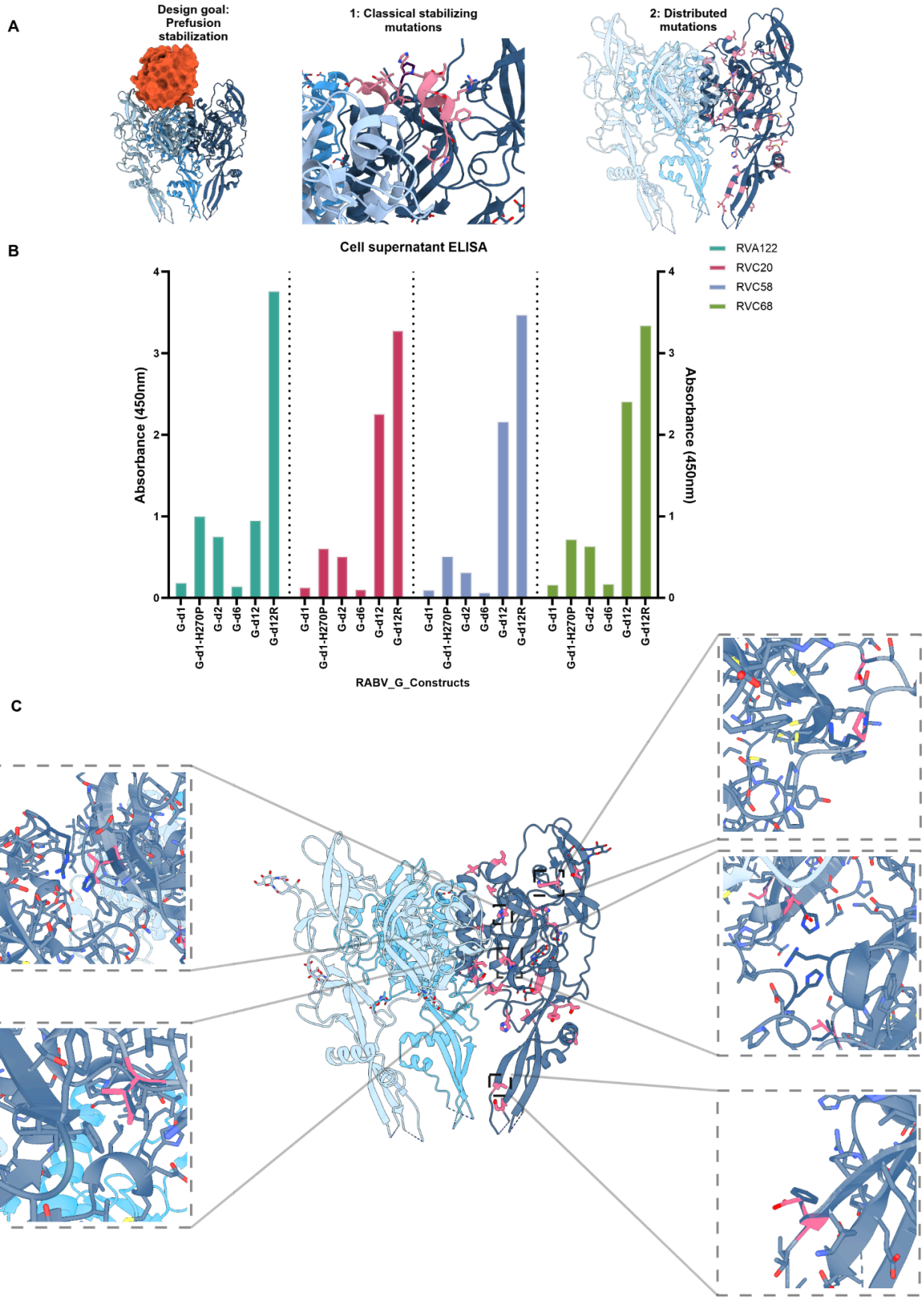

A
Design goal: Prefusion stabilization
1: Classical stabilizing mutations
2: Distributed mutations
B
Cell supernatant ELISA
RVA122
RVC20
RVC58
RVC68
Absorbance (450nm)
Absorbance (450nm)
0
1
2
3
4
G-d1
G-d1-H270P
G-d2
G-d6
G-d12
G-d12R
RABV_G_Constructs
C

Figure 4-2. Distributed mutations outperform the classical approach for prefusion-stabilization. A) Scheme of design goal to elicit potent neutralizing antibodies like RVA122, and two approaches for prefusion stabilization shown on the structure of trimeric, prefusion RABV-G (PDB ID 7UMS). The first (central) was a classical approach of isolated mutations with large effect within the conformationally dynamic region of residues 259-273 (highlighted in pink), with the known prefusion mutation H270P highlighted in dark purple. The second (right) was distributed mutations shown as pink sticks introduced throughout RABV-G by ProteinMPNN. B) Cell supernatant ELISA reactivity against three neutralizing antibodies (RVA122, RVC20, and RVC58) along with RVC68, which recognizes a linear epitope. C) ProteinMPNN-derived mutations within the construct G-d12R are shown as pink sticks. Selected mutations are magnified within insets, with the native residue in dark blue and the introduced residue in pink. Clockwise from top right, K250P, H21D, F82E, V358I, and H328I.

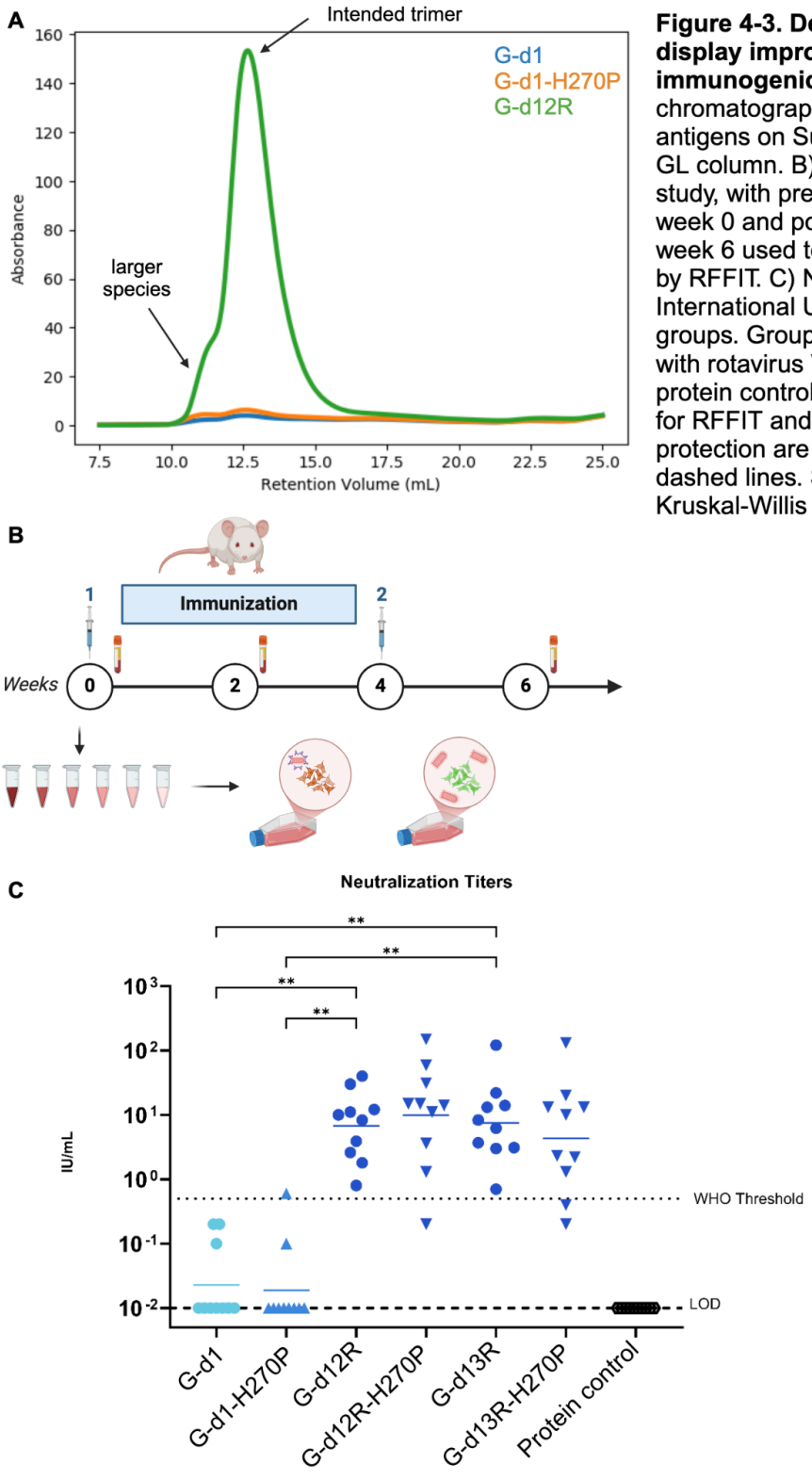
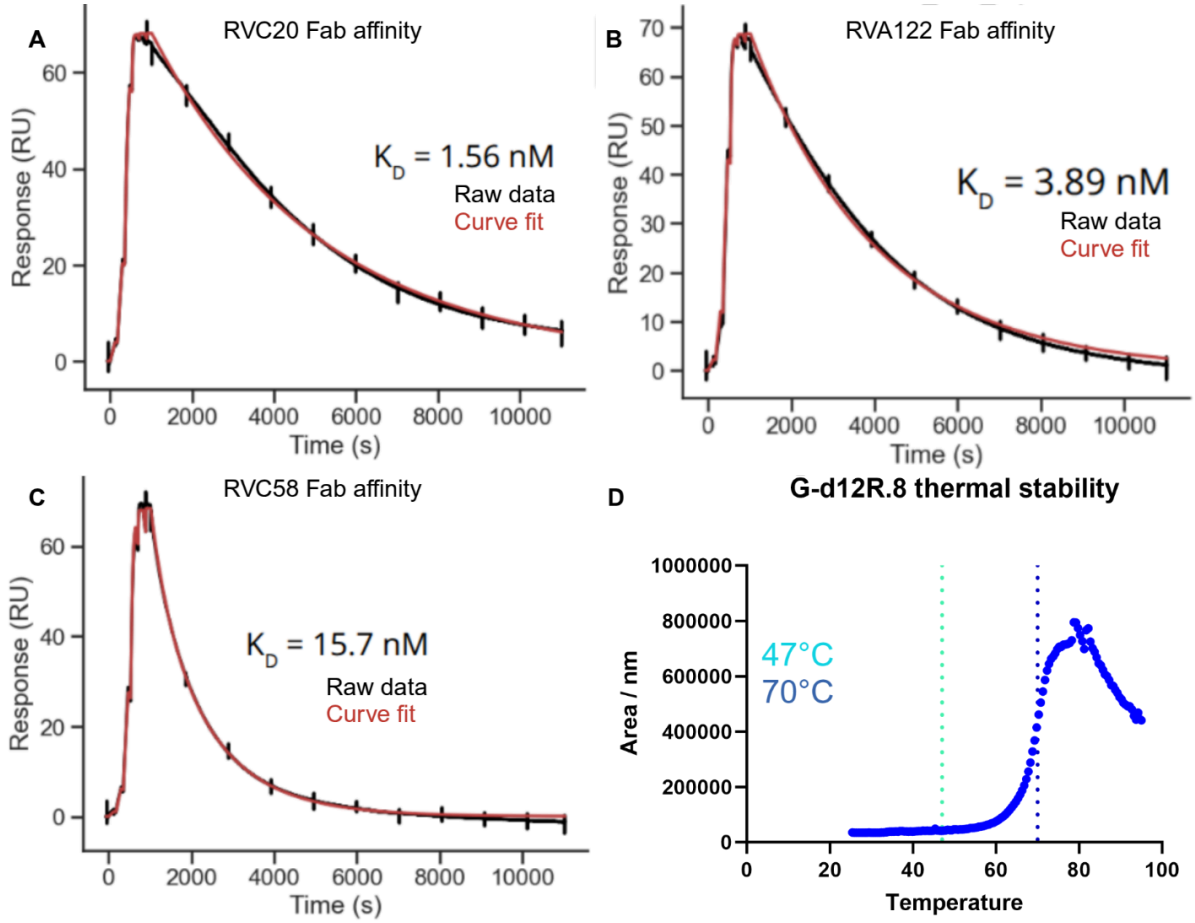


Figure 4-3. Designed RABV-G antigens display improved manufacturability and immunogenicity. A) Size exclusion chromatography traces of three RABV-G antigens on Superdex 200 Increase 10/300 GL column. B) Scheme of immunization study, with pre-immunization sera from week 0 and post-immunization sera from week 6 used to measure neutralization titers by RFFIT. C) Neutralization titers plotted in International Units per mL (IU/mL) for groups. Groups are labeled by immunogen, with rotavirus VP4-like antigen used as protein control. Both limit of detection (LOD) for RFFIT and the WHO threshold for protection are indicated by horizontal dashed lines. Statistical analysis using Kruskal-Willis test.


A
RVC20 Fab affinity
Response (RU)
Time (s)
K_D = 1.56 nM
Raw data
Curve fit
B
RVA122 Fab affinity
Response (RU)
Time (s)
K_D = 3.89 nM
Raw data
Curve fit
C
RVC58 Fab affinity
Response (RU)
Time (s)
K_D = 15.7 nM
Raw data
Curve fit
D
G-d12R.8 thermal stability
Area / nm
Temperature
47°C
70°C

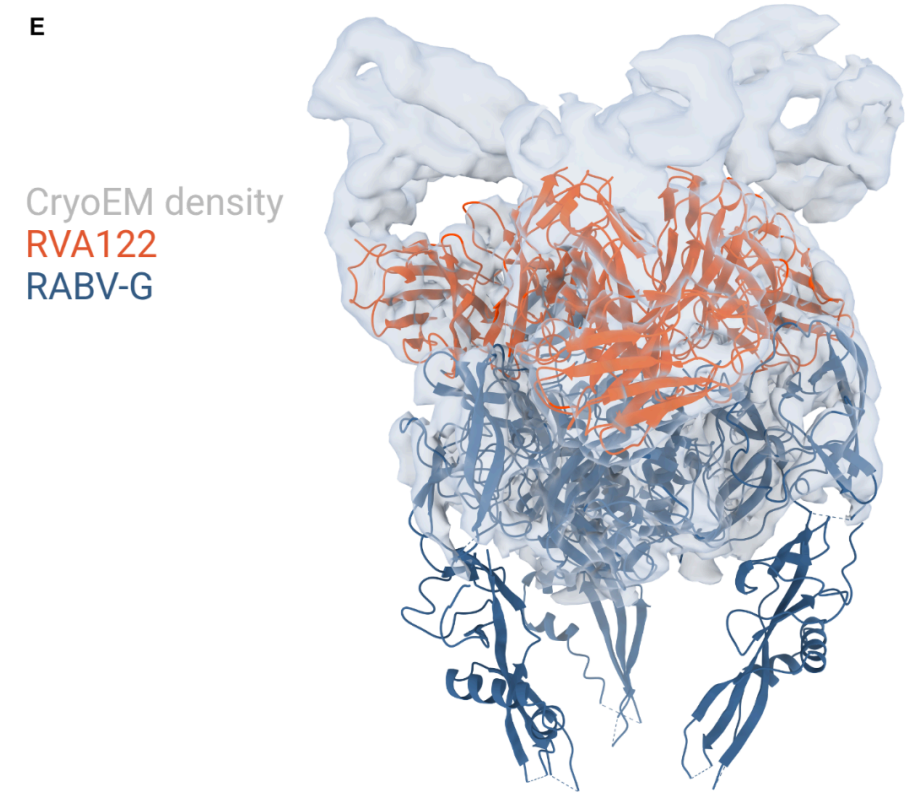
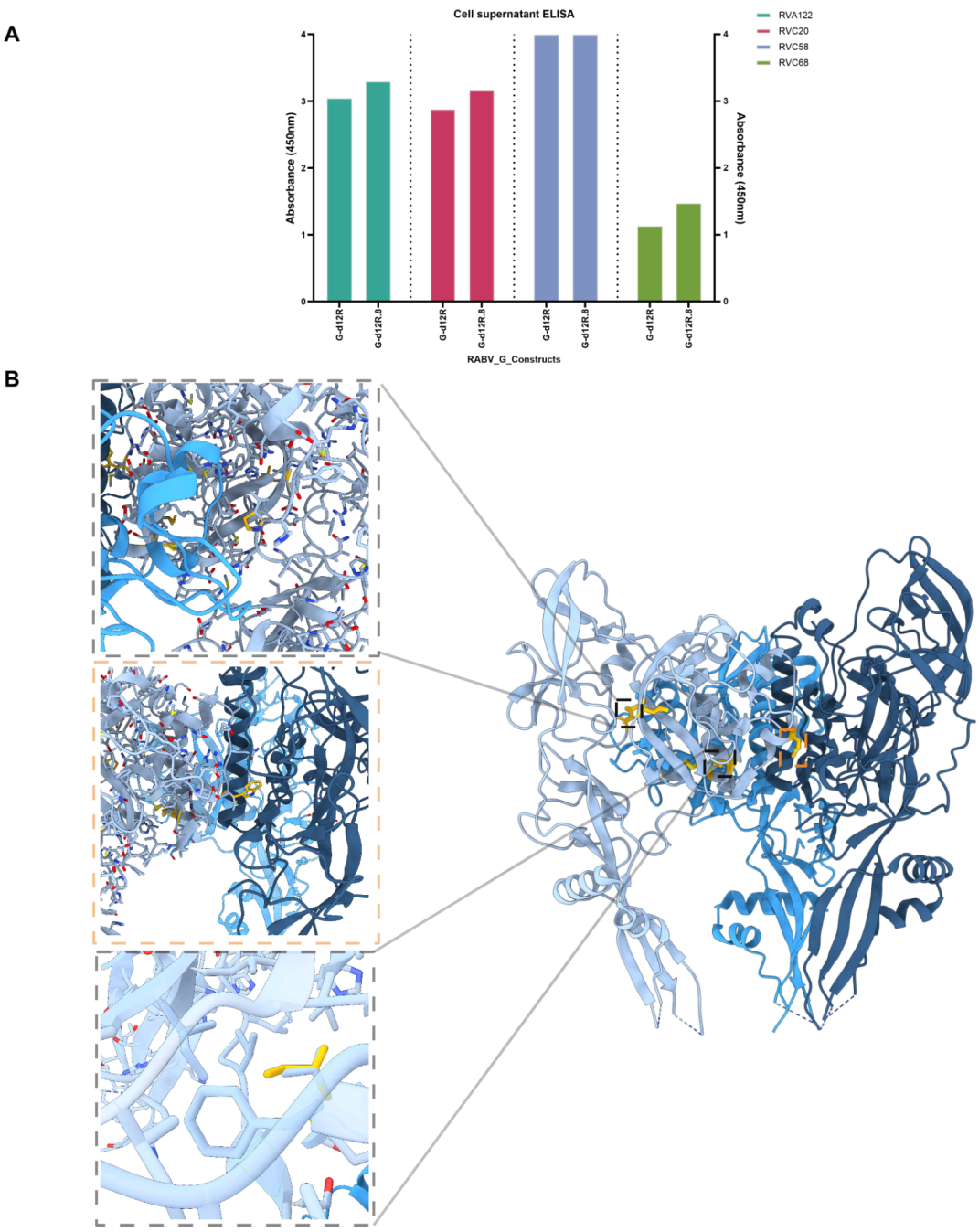

E
CryoEM density
RVA122
RABV-G

Figure 4-4. Designed antigen G-d12R.8 preserves antigenicity and intended structure. A-C) SPR binding data for captured G-d12R.8 with RVC20 (A), RVA122 (B) or RVC58 (C) flowed over. K_D calculated from curve fit. D) NanoDSF with SYPRO dye, with melting temperatures noted in text and by vertical dashed line. Native antigen melting temperature taken from literature for full-length RABV-G solubilized from virions in detergent micelles.



Supplementary Figure 4-5. Introduction of DMS mutations improves ELISA reactivity. A) Cell supernatant ELISA reactivity against three neutralizing antibodies (RVA122, RVC20, and RVC58) along with RVC68, which recognizes a linear epitope. On the left is the designed antigen G-d12R containing ProteinMPNN-derived mutations, while on the right is G-d12R.8 containing ProteinMPNN- and DMS-derived mutations B) Structure of RABV-G trimer (PDB ID 7UMS) in cartoon representation, with each monomer shown in a different shade of blue. The five DMS mutations are shown as yellow sticks in the overall view. Three mutations are shown in magnified insets with the native residue in pale blue and the introduced residue in yellow. From top, L260P, M385W, and V296I.

Supplementary note

To assess how deep learning approaches might enable stabilization of RABV-G, we explored two strategies for the *de novo* design of a trimerization domain. In the first approach, we designed trimerization domains at the C terminus of RABV-G. Using the structure of the prefusion trimer as our input, we generated backbones with RFdiffusion that were compatible with the termini and fusion domains (FDs) of RABV-G. Following sequence design with ProteinMPNN, we predicted the full length proteins with AF2 and selected designs to test that displayed the highest pLDDT and PAE. Additional designs were selected that displayed lower pLDDT and PAE but were predicted to form packing or electrostatic interactions with the FDs, which might stabilize the prefusion conformation in addition to forming trimers. In the second approach, we explored insertion of trimerization domains into the central helices of RABV-G. RABV-G lacks interactions between its central helices, which are too far apart to mediate intra-trimer interactions. However, the homologous VSV-G trimer is stabilized by interactions between residues in its central helices. We reasoned that design of a larger trimerization domain that extends from the base of the central helices could improve trimerization and soluble yield. Our input structure to RFdiffusion contained truncations in the FD, as we wanted to accommodate a larger trimerization domain and few neutralizing antibodies have been identified that target the FD. To select candidate sequences, we utilized RFdiffusion, ProteinMPNN, and AF2 in the same manner as for our first approach. All designs tested were unsuccessful, showing no improvement over G-d1.

To explore potential reasons for the failure of these *de novo* designs, we predicted these sequences in isolation with AF2, using single-sequence mode and providing no structural templates. AF2 confidently predicts the GCN4 trimerization domain with high pLDDT and low PAE. However, tested sequences from our first approach, which appeared promising in the context of RABV-G, displayed low model confidence or deviation from the intended backbone when predicted in isolation. Their poor AF2 prediction confidence corresponds to their inability to stabilize RABV-G trimers. To test the hypothesis that model confidence metrics from single-sequence prediction by AF2 can identify designs that might work, we again generated backbones with RFdiffusion. However, in this attempt trimerization domains were generated unconditionally, with no input structure of antigen affecting RFDiffusion outputs. Backbones compatible with the termini and FDs were designed by ProteinMPNN. Sequences were then predicted in isolation and filtered purely by pLDDT and PAE. Of four designs screened, three displayed higher ELISA reactivity than ectodomain lacking a trimerization domain, indicating successful stabilization of trimeric RABV-G.

Structure predictions of trimers that extend from the central helix indicate higher model confidence, with some designs having pLDDT and PAE values comparable to GCN4. However, these designs contain truncations to the FD which may disrupt folding of RABV-G. To test the hypothesis that truncations to the FD disrupted folding of designs that extended the central helix, we screened constructs that contained the GCN4 trimerization domain and truncations of the fusion loop. Each truncation displayed considerably lower ELISA reactivity (data not shown).

Therefore, both hypothesized failure modes were validated. For this particular antigen, one of two well-validated trimerization domains effectively stabilized the desired state. However, many antigens are incompatible with well-validated oligomerization domains due to their termini or desired oligomerization state. For these, unconditional generation of *de novo* trimerization domains, selection of those compatible with full-length antigen, and filtering by single-sequence AF2 prediction may represent a reasonable strategy to identify new oligomerization domains.

Materials and Methods

AlphaFold Prediction

The ColabFold implementation of AlphaFold was downloaded from the github repository into an apptainer environment with the necessary dependencies[45,46]. All predictions were made on either A4000 or A6000 GPUs on a custom high performance computing cluster. Sequences of candidate designed antigens were input into the ColabFold implementation to create MSAs using mmseqs2 and to generate structural predictions for trimers with AlphaFold2_multimer_v3. Predictions were made with structural templates and with an amber relaxation step. For rabies G, custom templates from the PDB were used during prediction with the --custom-templates-path flag. For each sequence predicted, structural predictions were obtained with all 5 AlphaFold models.

Following structure prediction, analysis with python scripts in a jupyter notebook was performed to determine prediction confidence by examining average pLDDT and PAE across the monomer or oligomers. All 5 predictions for a given sequence were used to determine the mean and standard deviation for pLDDT and PAE across the different models. Python scripts were used to calculate root mean squared deviation (RMSD) between the desired structure, either an experimental structure or high confidence prediction of the native sequence, and the structures predicted for designed sequences.

MSA depth was tuned using the --msa-seq and --msa-extra-seq arguments while running ColabFold. MSA subsampling to examine conformational sampling was done

using the Colab notebook for AF-Cluster, and all output MSAs were used as inputs for structural prediction by ColabFold [47,48].

De novo scaffolds were initially evaluated using the same ColabFold protocol in the context of native or designed antigens. In these cases, pLDDT and PAE were evaluated across the *de novo* region of the protein, not across the entire protein. In the subsequent unconstrained generation of *de novo* scaffolds, prediction occurred in single-sequence mode, without any MSA inputs.

Multiple sequence alignment (MSA) based design

Standard MSAs were generated using NCBI protein BLAST with the standard databases and a maximum number of 500 sequences[49–52]. All sequences in the MSA displayed greater than 90% identity to the query sequence with coverage greater than 50%, and were downloaded as aligned FASTAs. Analysis with python scripts in a jupyter notebook was performed to determine the residue positions where more than 25% of the sequences contained the same amino acid which differed from the corresponding amino acid in the query sequence. All six were examined by visual inspection in PyMOL and did not appear to cause steric clashes or lead to the loss of hydrogen bonds, leading to six mutations in the MSA set.

ProteinMPNN

ProteinMPNN was run using an implementation downloaded from the github repository[53]. All sequence design was conducted on CPUs or an A4000 GPU on a

custom high performance computing cluster. Experimental structures of antigens were used when available, and all 5 high-confidence ColabFold predictions for the native sequence were used for antigens lacking experimental structures.

For RABV-G, epitope residues were identified from structures of antigen in complex with antibody or from the literature[31,32,34,35,54–56]. Epitope residues, disulfide bonds, native glycan sites, and other surface exposed residues near epitopes were disallowed for design. ProteinMPNN was run with a sampling temperature of 0.2, and an amino acid bias of -0.8 for glycine and alanine. Cysteines were disallowed during design. ProteinMPNN was employed in a 2-step process for RABV-G. In the first step, structures served as inputs to ProteinMPNN for sequence design with all residues allowed to mutate that were not disallowed for design. ProteinMPNN was run for 200 batches with 5 sequences each, to generate 1000 designed sequences for each input structure. A python script was used to analyze the frequency of mutations from the native amino acid at each position. Where ProteinMPNN chose a mutation at a given position in more than 25% of the sequences, the mutation was manually inspected with the PyMOL mutagenesis wizard in the context of the input structure. If any suggested mutation caused large steric clashes or the loss of at least two hydrogen bonds, the residue was no longer considered for design. In this way, the first round of sequence design was used to identify designable positions where ProteinMPNN suggested mutations that did not appear to be deleterious.

In the second round, only these designable residues were allowed to mutate, and ProteinMPNN was run for 40 batches with 5 sequences each to generate 200 potential

sequences. These sequences were then evaluated by AlphaFold structure prediction using our specific templates. Predicted structures with the lowest RMSD, highest pLDDT, and lowest PAE were selected as candidate designs, and their sequences were ordered for experimental evaluation.

For design of *de novo* scaffolds or linkers generated in RFDiffusion, output structures from RFDiffusion served as inputs to ProteinMPNN. The *de novo* portion of the protein generated by RFDiffusion was designed, while all residues in the antigen were kept native. ProteinMPNN was run with a sampling temperature of 0.1 with 8 batches of 1 sequence each to generate 8 output sequences per design. These sequences were evaluated by ColabFold structure prediction, and the sequences with the highest pLDDT and lowest PAE across the *de novo* regions were ordered for experimental evaluation.

RFdiffusion

RFdiffusion was run using an implementation downloaded from the github repository and run on A4000 GPUs on a custom high performance computing cluster[57].

To stabilize the desired trimeric state of RABV-G, RFdiffusion was used to generate novel backbones that could function as scaffold domains. Novel backbones were generated in both constrained and unconstrained cases with the symmetry setting in RFdiffusion with the desired oligomeric state specified. In both cases, the number of denoising timesteps (50,100, or 200) and noise scale (0,0.25,0.5, or 1) were varied. In the constrained case, either the entire symmetric antigen or a portion of it were centered

on the origin with the axis of rotation aligned to the Z-axis and provided to RFdiffusion as the input motif. Output structures were filtered using PyMOL scripts for a high proportion of secondary structure and a sufficient fraction of residues in proximity to (alpha carbon within 8 Å of alpha carbons from) diffused regions of other chains in the protein. In the unconstrained case, RFDiffusion was not provided with an input structure, and both timesteps and noise scales were varied identically to the constrained case. Output structures were filtered using PyMOL scripts for a high proportion of secondary structure, a sufficient fraction of residues in proximity to other chains, and termini compatible with fusion to the antigen.

Protein Expression

To initially screen designs in mammalian cells, codon optimized DNA sequences were ordered as gene fragments (Twist), cloned into a pCMV based plasmid using golden gate cloning with BsaI-HFv2 (NEB), and transformed into 5-alpha cells (NEB). Cells were grown for 16 hours in 4 mL of LB broth containing antibiotic. Plasmid was purified via miniprep (Qiagen). Expi293 cells were transfected at 3 million cells/mL with the following per mL of culture: 1 ug of plasmid DNA, 3 ug PEI-max, and 100 uL Opti-MEM (Thermo Scientific). Cells were subsequently cultured for 3 days at 37C, supernatant harvested following centrifugation at 4000 g for 10 minutes, and supernatant screened by ELISA.

For constructs used for more detailed comparisons with the native antigen, plasmids encoding codon optimized sequences were ordered from GenScript. Protein expression was conducted in the same manner as above.

Purification

For purification of proteins from Expi293 cells, cell supernatant was adjusted to 50 mM Tris, 500 mM NaCl, pH 8.0 and batch bound with Ni excel resin (Cytiva) by shaking for 1.5 hours at 20C. Resin was collected in a chromatography column and washed with 20 column volumes of wash buffer (50 mm Tris, 500 mM NaCl, 30 mM imidazole). Protein was eluted with 2.4 column volumes of elution buffer (50 mm Tris, 500 mM NaCl, 300 mM imidazole) and analyzed by SDS-PAGE gel. Eluted proteins were further filtered by size exclusion chromatography using a Superdex 200 Increase 10/300 GL column (Cytiva) on an ÄKTA pure (Cytiva).

Melting and aggregation temperature determination:

Determination of melting and aggregation temperatures were done by differential scanning fluorimetry with the UNcle instrument and software (UNchained Labs). SEC purified proteins at a concentration between 0.25-1 mg/mL were added to 8.8 uL quartz capillary wells in a 16-well cassette (UNchained Labs). With the temperature of the cassette changing from 25 to 95 ºC at a thermal ramp rate of 1ºC/min, intrinsic changes in protein fluorescence emission were used to evaluate melting and aggregation temperatures. As an orthogonal measurement of the melting temperature, protein was

mixed with the hydrophobic dye Sypro Orange (Thermo Fisher Scientific) to evaluate change in fluorescence emission during the thermal ramp.

Enzyme-linked immunosorbent assay (ELISA):

For initial screening of constructs, Expi293 cell supernatant was diluted 64 or 512 fold in 50 mM Tris, 300 mM NaCl. All incubation steps occurred while shaking at 600 rpm, and all wash steps involved 3 washes with ELISA wash buffer (25 mM Tris, 150 mM NaCl, 0.2% Tween 20, pH 8.0) by an automated plate washer (Biotek). 100 uL of the solution was incubated onto a His binding plate (Qiagen) for 1.5 hours. 200 uL of ELISA blocking buffer was added to each well and incubated for 1 hour. Plates were washed and 100 uL of ELISA blocking buffer (25 mM Tris, 150 mM NaCl, 0.2% Tween 20, 5% nonfat milk, pH 8.0) containing primary antibody at 4 ug/mL was added to each well and incubated for 1 hour. Plates were washed and 100 uL of ELISA blocking buffer containing secondary antibody (goat anti-human IgG-HRP, Southern Biotech at a 1:5000 dilution). Plates were washed, 100 uL of TMB (3,30,50,5-tetramethylbenzidine, SeraCare) was added for 3 minutes, the reaction was quenched with the addition of 100 uL of 1N HCl, and the absorbance at 450 nm of each well was measured with an Epoch plate reader (Biotek).

Full-length IgG and Fabs were ordered from Genscript. Full-length antibodies were used for ELISAs, while Fabs were used for SPR and cryo-EM.

Immunogenicity studies

For immunogenicity studies, 7 week old female Balb/cAnNHsd mice were purchased from Envigo (order code 047). Mice were housed in a specific-pathogen free facility and

all animal studies were in accordance with the University of Washington's Institutional Animal Care and Use Committee. For immunizations, 3 µg/dose of each protein were diluted in dPBS buffer and mixed 1:1 with Addavax adjuvant (InvivoGen). At 8 weeks of age, animals were immunized with 100 uL of the mixture (week 0). 4 weeks later, mice were immunized again. At both weeks 0 and 4, blood was collected by the submental route. At week 6, a terminal bleed was conducted via cardiac puncture.

Neutralization titers

Sera from the immunization study were shipped to Kansas State University's commercial Rabies laboratory, where rapid fluorescent foci inhibition tests (RFFITs). Briefly, sera were serially diluted 5-fold starting at a 1:5 dilution, and each dilution was incubated with live rabies virus. Following incubation, cultured cells were added to the mixture and stained for rabies infection. The percentage of infected cells at each dilution was used to calculate neutralization titers in International Units per milliliter (IL/mL). RFFITs were performed for terminal sera (week 6) from each mouse in every study group. RFFITs were also performed for pooled pre-immunization sera for each study group. Values were reported as IU/mL for each serum sample, and plotted via GraphPad Prism. Statistical analysis was performed in GraphPad Prism using a Kruskal-Wallis test.

Surface Plasmon Resonance

Binding kinetics were analyzed via Surface Plasmon Resonance (SPR) on a Biacore 8K (Cytiva). His-tagged G-d12R.8 at 0.125 µg/mL in HBS-EP+ (0.01 M HEPES pH 7.4,

0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20) was captured with His binding plates to obtain a capture level of approximately 250 response units. Fabs diluted in HBS-EP+ were injected to monitor association. HBS-EP+ was used as a running buffer during dissociation. Dissociation was allowed to occur over three hours to obtain full curves . Binding kinetics were determined by global fitting of curves to k_{on} and k_{off} using the Cytiva evaluation software.

Cryo-EM sample preparation, data collection and processing

Cryo-EM samples were prepared on glow-discharged CFLAT 1.2/1.3 holey-carbon grids (Electron Microscopy Sciences). Protein sample at 1 mg/ml was applied to each grid, incubated for 30 seconds, and then blotted using Whatman filter paper; this process was performed twice, followed by a third application of sample and final blotting and plunge freezing in liquid-ethane using a Vitrobot (FEI). Data were acquired on a Titan Krios operating at 300 kV equipped with a K3 Summit Direct Detector and a Quantum GIF energy filter (Gatan) operating in zero-loss mode with a 20-eV slit width. Movies were collected in counting mode, collecting 50 frames with a total dose of 52 e−/Å−2, with a defocus range of −0.5 to −2 µm. Automated data collection was carried out using SerialEM at a nominal magnification of 105,000× with a pixel size of 0.843 Å[58]. All data processing was carried out using cryoSPARC[59]. Movie frame alignment and dose-weighting summation were performed using Patch Motion, and CTF estimation was done using patch CTF. Particles were picked using Blob picker.

An initial dataset was collected on the G-d12R.8 structure alone, however this resulted in poor ice quality and incoherent 2D class averages. Another dataset was collected on

G-d12R.8 in complex with RVC20 and RVA122 Fabs[6], but this resulted in class averages showing a preferred orientation of only top views, likely due to the perpendicular angle of approach of the RVC20 Fab. Finally, a dataset with G-d12R.8 in complex with only the RVA122 Fab resulted in good ice quality and full distribution of particle views. For this dataset, three rounds of 2D classification and selection were done to isolate well-aligned, single particles. Initial models were generated by ab-initio reconstruction with C1 symmetry. Non-uniform 3D refinement was then carried out with C3 symmetry. This resulted in a structure where the RBV core was resolved at 3.3 A, while the membrane fusion loops at the protein's base remained unresolved due to their flexibility.

Acknowledgements
Thank you to all of the scientists whose work on rabies virus, its lifecycle, and immune responses makes structure-based vaccine design possible.

Work on rabies antigens was done with Bianka Haro, who initially worked with me during her rotation and then continued to work with me on the project when she joined the lab. The effort would not have progressed without her work on it.

Maintenance of the computational infrastructure and technical help establishing the Colabfold environment were generously provided by Luki Goldschmidt and Patrick Vecchiato.

Annie Dosey kindly agreed to perform cryo-EM structure determination.

All neutralization titers were performed by Jason Defisher at KSVDL.

Sequences

RABV G constructs **(all tested with VSV-G signal peptide)**

>G-original-fusion-loops
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETYT
NFVGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPYPDYHWLRTVKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEHLVVEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSGSGSHHHHHH
>G-base
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETYT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPYPDGSGSGTVKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEHLVVEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSGSGSHHHHHH
>G-d1
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEHLVVEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-base-H270P
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETYT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPYPDGSGSGTVKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEPLVVEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSGSGSHHHHHH
>G-d1-H270P
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEPLVVEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d2_V273P
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEHLVPEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH

>G-d3_H261A_D262E
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLAEFRSDEIEHLVPEEYVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d4_V273P_G12C_V48C
KFPIYTIPDKLCPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKCGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEHLVPEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d5_H261E_D262E_V273P
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLEEFRSDEIEHLVPEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d6_V273P_L276Y
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEHLVPEEYVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d7_H261A_V273P_L276Y
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLADFRSDEIEHLVPEEYVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d8_H270P_V273P
KFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETKT
NGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVIISPSVADL
DPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEPLVPEELVKK
REECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGR
CHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d9_H270P_G12C_V48C

MKCLLYLAFLFICVNCKFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKCGYISAIKMN
GFTCTGVVTEAETKTNGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTT
KTTKESLVIISPSVADLDPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSR
GKRASKGSETCGFVDERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFR
SDEIEPLVVEELVKKREECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWN
EIIPSKGCLRVGGRCHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSG
SRMKQLEDKIEEILSKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d10
MKCLLYLAFLFICVNCKFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKCGYISAIKMN
GFTCTGVVTEAETKTNGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTT
KTTKESLVIISPSVADLDPYDRSLHSRVFPGGNCSGVAVSSTYCSTNHDYTIWMPENPRLGMSCDIFTNSR
GKRASKGSETCGFVDERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFR
SDEIEHLVVEELVKKREECLDALESIMTTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWN
EIIPSKGCLRVGGRCHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPGSGSG
SRMKQLEDKIEEILSKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d11
KFPPYTIPDKLGPWSPIDLSDLECPNNLVIEDEGCTNLSGFSYMELKVGYISAIKVNGFTCTGVVHEAETKT
NGSGYVTNEEKEKEFRPTPDECRKAYNLKKSGDPRYEESLHNPDPDGSGSGTTKTTKYSLVILSPVQADL
DPYDRSLHSRVFPGGNCSGVDESSSTCSTNNDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGIYKSLKGACKLKLCGVLGLRLMDGTWVAMQVSNETPWCPPGQLTNDADVRSDEEEHLDVEELIKK
REECLDALESIMTTKSVSFRRLSKLRKLVPGFGKAYTIFNNTLMEADAIYKSVRTWNEIIPSKGCLRVGGKC
HPHVNGIFFNGIILGPDGNVLIPEMQSSLLQQHMELLVEIVIPDMHPLADPGSGSGSRMKQLEDKIEEILSKI
YHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d12
KFPPYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKMNGFTCTGVVTEAETK
TNGSGYVTTTFKRKHFRPTPDACRAAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKYSLVILSPVSA
DLDPYDKSLHSRVFPGGNCSGVDESSPTCSTNNDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCG
FVDERGNYKSLKGACKLKLCGVLGLRLMDGTWVAMQVSNETPWCPPGQLRNDADVRSDEEEHLDVEE
LVKKREECLDALESIMTTKSVSFRRLSKLRKLVPGFGKAYTIFNNTLMEADAIYKSVRTWNEIIPSKGCLRV
GGKCHPHVNGIFFNGIILGPDGNVLIPEMQSSLLQQHMELLVEIVIPPMHPLADPGSGSGSRMKQLEDKIE
EILSKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d13
KFPPYTIPDKLGPWSPIDLSDLECPNNLVIEDEGCTNLSGFSYMELKVGYISAIKVNGFTCTGVVHEAETKT
NGSGYVTNEEKEKEFRPTPDECRKAYNLKKSGDPRYEESLHNPDPDGSGSGTTKTTKRSLVILSPVTADL
DPYDKSLHSRVFPGGNCSGVDESSPTCSTNNDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFV
DERGIYKSLKGACKLKLCGVLGLRLMDGTWVAMQVSNETPWCPPGQLTNDADVRSDEEEHLDVEELVK
KREECLDALESIMTTKSVSFRRLSKLRKLVPGFGKAYTIFNNTLMEADAIYKSVRTWNEIIPSKGCLRVGGK
CHPHVNGIFFNGIILGPDGNVLIPEMQSSLLQQHMELLVEIVIPDMHPLADPGSGSGSRMKQLEDKIEEILS
KIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d12R
KFPPYTIPDKLGPWSPIDIHDLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKVNGFTCTGVVTEASTKT
NGSGYVTTTEKRKHFRPTPDACRKAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVILSPVSAD
LDPYDKSLHSRVFPGGNCSGVDESSPTCSTNNDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGF
VDERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQVSNETPWCPPGQLRNDADVRSDEEEHLDVEELV
KKREECLDALESIMTTKSVSFRRLSKLRKLVPGFGKAYTIFNNTLMEADAIYKSVRTWNEIIPSKGCLRVGG
KCHPHVNGIFFNGIILGPDGNVLIPEMQSSLLQQHMELLVEIVIPPMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d13R

KFPPYTIPDKLGPWSPIDIHDLSCPNNLVIEDEGCTNLSGFSYMELKVGYISAIKVNGFTCTGVVTEASTKT
NGSGYVTTTEKEKEFRPTPDACRKAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVILSPVTAD
LDPYDKSLHSRVFPGGNCSGVDESSPTCSTNNDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGF
VDERGIYKSLKGACKLKLCGVLGLRLMDGTWVAMQVSNETPWCPPGQLTNDADVRSDEEEHLDVEELV
KKREECLDALESIMTTKSVSFRRLSKLRKLVPGFGKAYTIFNNTLMEADAIYKSVRTWNEIIPSKGCLRVGG
KCHPHVNGIFFNGIILGPDGNVLIPEMQSSLLQQHMELLVEIVIPDMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d12R-H270P
KFPPYTIPDKLGPWSPIDIHDLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKVNGFTCTGVVTEASTKT
NGSGYVTTTEKRKHFRPTPDACRKAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVILSPVSAD
LDPYDKSLHSRVFPGGNCSGVDESSPTCSTNNDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGF
VDERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQVSNETPWCPPGQLRNDADVRSDEEEPLDVEELV
KKREECLDALESIMTTKSVSFRRLSKLRKLVPGFGKAYTIFNNTLMEADAIYKSVRTWNEIIPSKGCLRVGG
KCHPHVNGIFFNGIILGPDGNVLIPEMQSSLLQQHMELLVEIVIPPMHPLADPGSGSGSRMKQLEDKIEEIL
SKIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d13R-H270P
KFPPYTIPDKLGPWSPIDIHDLSCPNNLVIEDEGCTNLSGFSYMELKVGYISAIKVNGFTCTGVVTEASTKT
NGSGYVTTTEKEKEFRPTPDACRKAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVILSPVTAD
LDPYDKSLHSRVFPGGNCSGVDESSPTCSTNNDYTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGF
VDERGIYKSLKGACKLKLCGVLGLRLMDGTWVAMQVSNETPWCPPGQLTNDADVRSDEEEPLDVEELVK
KREECLDALESIMTTKSVSFRRLSKLRKLVPGFGKAYTIFNNTLMEADAIYKSVRTWNEIIPSKGCLRVGGK
CHPHVNGIFFNGIILGPDGNVLIPEMQSSLLQQHMELLVEIVIPDMHPLADPGSGSGSRMKQLEDKIEEILS
KIYHIENEIARIKKLVGERGSGSGSGSGSHHHHHH
>G-d12R.8
KFPPYTIPDKIGPWSPIDIHDLSCPNNLVVEDEGCTNLSGFSYMELKVGYISAIKVNGFTCTGVVTEASTKT
DGSGYVTTTEKRKHFRPTPDACRKAYNWKMAGDPRYEESLHNPDPDGSGSGTTKTTKESLVILSPVSAD
LDPYDKSLHSRVFPGGNCSGITESSPTCSTNNDYTIWMPENPRLGMSCDIFTNSRGKRASKGSKTCGFV
DERGLYKSLKGACKLKLCGVLGLRLMDGTWVAMQVSNETPWCPPGQLVNPADFRSDEEEHLDVEELVK
KREECLDALESIITTKSISFRRLSKLRKLVPGFGKAYTIFNNTLMEADAIYKSVRTWNEIIPSKGCLRVGGKC
HPHVNGIFFNGIILGPDGHVLIPEMQSSLLQQHWELLESIVIPPMHPLADPGSGSGSRMKQLEDKIEEILSKI
DHIEGEIARIKKLVGERGSGSGSGSGSHHHHHH

Discussion

We set out to stabilize antigens from four pathogens that pose distinct protein design challenges. Our results suggest that the designed antigens possess improved properties and could become potential vaccine candidates with further development. During this work, we learned principles that may generalize for antigen stabilization with deep learning methods.

ProteinMPNN-derived mutations resulted in misfolding of rotavirus VP5*, while MSA-based approaches enabled stabilization of the antigen. Current deep learning methods for protein design are insufficient for antigens which form extensive contacts with solvent molecules. Methods for sequence redesign and structure prediction, including ProteinMPNN and AF2, do not model solvent[1–6], likely explaining their inability to identify redesigned sequences for VP5*. While there have been a few recent examples of deep learning methods to explicitly model solvent molecules[7,8], and one approach using Rosetta software[9], these have not been used for protein design. Development of methods that combine tools that explicitly consider solvent with methods for protein design could enable the stabilization of antigens whose conformation depends on extensive interactions with solvent. Currently, MSA-based approaches might be the most tractable for this type of antigen.

The diminished ELISA reactivity of VP4-like-MSA2 despite the low RMSD of an experimentally determined structure demonstrates that MSA-based stabilization can still introduce mutations that subtly alter conformation and reduce antigenicity. Both MSA-based approaches and deep learning methods identify mutations which must be

iteratively refined to arrive at stabilized antigens that elicit more potent antibody responses. As with sequential rounds of stabilization for the RSV F protein[10–12], multiple rounds of mutation, characterization, and reversion of introduced mutations led to designed antigens with improved properties for SLO, SpyAD, PPE18, and RABV-G. While significant advancements in predicting the effect of point mutations and modeling protein dynamics could enable a better understanding of the contribution of each potential mutation to an antigen, deep learning methods for protein design currently possess poor performance on these tasks[13–15]. As a result, these methods extend our capabilities past historical approaches, but they do not yet represent a paradigm shift that enables the one-step design of stabilized antigens. However, current deep learning methods considerably advance our capability to stabilize challenging antigens for subunit vaccines.

AF2 predictions provided atomically-accurate starting points for structural hypotheses and sequence redesign of SpyAD, Mtb32A, and PPE18, all of which lack experimentally determined structures. A crystal structure of a designed PPE18 antigen attests to the accuracy of the AF2 predictions outside of a short, flexible loop. AF2 predictions, along with Rosetta energies, identified which MSA-based mutations would be compatible with the structure of VP5* and SLO. Traditional approaches to antigen stabilization are reliant on experimental structures which can be slow or challenging to obtain, and design based on sequence alone could enable a rapid response to emerging pathogens which possess low homology to well-studied pathogens.

Deep learning methods can outperform standard approaches for traditional antigen stabilization tasks. For RABV-G, ProteinMPNN redesign and AF2 filtering resulted in the identification of three distinct mutation sets. Two of these displayed better prefusion stabilization than introduction of prolines or a potential disulfide bond into the conformationally dynamic region of RABV-G. These mutation sets contained 28-45 mutations in a single design without causing misfolding of the protein, allowing the identification of combinations of mutations that may display epistasis. Many of these mutations may not have been selected if tested in isolation, as has historically been the process. The ability to identify and accurately model combinations of these mutations *in silico* is an advantage of deep learning approaches. For fusion of PE13 and PPE18 into a single protein, generation of a *de novo* linker tailored to the antigen outperformed a flexible glycine-serine linker. In addition, an MSA-based approach is insufficient to address particular antigen stabilization challenges.

An MSA is constrained by the evolutionary history of each antigen. For both SpyAD and PPE18, few identified homologs with moderate (70-90%) identity meant insufficient diversity was present in the MSA to suggest stabilizing mutations. That would likely be an issue for other antigens with similarly shallow MSAs. For RABV-G, the protein has evolved to dynamically switch between a prefusion and postfusion conformation to mediate cell entry. This evolutionary pressure makes any prefusion stabilizing mutations, which are necessary to improve vaccine-elicited antibody responses, highly deleterious, preventing them from being common in an MSA. Similarly, residues that mediate cytotoxicity and conformational transitions in SLO are highly conserved. More

broadly, any antigen stabilization effort that aims to reduce a function the protein has evolved for pathogen fitness will need to identify mutations that are rare or unobserved in an MSA. In each of the cases tested, ProteinMPNN-derived mutations effectively addressed the protein design challenges these proteins posed.

Specifically, designed antigens displayed improved soluble yield (SLO and PPE18), thermal stability (SpyAD, SLO, and RABV-G), and prefusion stabilization (RABV-G), and possessed comparable antigenicity (SLO, SpyAD, and RABV-G). Furthermore, these ProteinMPNN-derived mutations can act synergistically with mutations from other approaches, highlighting how deep learning methods can complement historical approaches. MSA-based mutations for SLO, disulfides introduced by DisulfidebyDesign 2.0, and DMS-identified mutations for RABV-G all further improved designed antigens.

Future work is necessary to improve the designed antigens to the point that they could become potential vaccine candidates. For rotavirus, neutralization titers suggest VP4-like antigens should be prioritized over VP5* antigens for eliciting protective antibody responses. Reversion of specific mutations in the VP4-like-MSA2 designs might lead to antigens with the thermal stability of VP4-like-MSA2 but comparable antigenicity to VP4-like-base. For designed SLO, SpyAD, Mtb32A, and PPE18 antigens, immunization studies using a well-studied challenge model are necessary to determine the functional protection they elicit, as bacterial pathogens lack a clear neutralization assay to assess antibody responses. At that point, protection could inform further antigen design. For RABV-G, future work to address the flexibility of the fusion domains

and GCN4 trimerization domains and to investigate the durability of vaccine-elicited antibody responses will be valuable. For each antigen, we see genetic fusion to protein nanoparticles as an important next step. Past work has shown that protein nanoparticles elicit more potent antibody responses than soluble antigens through multivalent display of antigens and altered trafficking in vivo[16–19]. Our hope is that displaying stabilized antigens on protein nanoparticles could enable the development of safe and effective subunit vaccines against these deadly pathogens.

References

1. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
2. Akpinaroglu, D. *et al.* Structure-conditioned masked language models for protein sequence design generalize beyond the native sequence space. *bioRxiv* 2023.12.15.571823 (2023) doi:10.1101/2023.12.15.571823.
3. Hsu, C. *et al.* Learning inverse folding from millions of predicted structures. *bioRxiv* (2022) doi:10.1101/2022.04.10.487779.
4. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
5. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
6. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
7. Park, S. & Seok, C. GalaxyWater-CNN: Prediction of water positions on the protein structure by a 3D-convolutional neural network. *J. Chem. Inf. Model.* **62**, 3157–3168 (2022).
8. Zamanos, A., Ioannakis, G. & Emiris, I. Z. HydraProt: A new deep learning tool for fast and accurate prediction of water molecule positions for protein structures. *J. Chem. Inf. Model.* **64**, 2594–2611 (2024).
9. Lai, J. K., Ambia, J., Wang, Y. & Barth, P. Enhancing structure prediction and design of soluble and membrane proteins with explicit solvent-protein interactions. *Structure* **25**, 1758–1770.e8 (2017).

10. McLellan, J. S. *et al.* Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* **342**, 592–598 (2013).
11. Krarup, A. *et al.* A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism. *Nat. Commun.* **6**, 8143 (2015).
12. Joyce, M. G. *et al.* Iterative structure-based improvement of a fusion-glycoprotein vaccine against RSV. *Nat. Struct. Mol. Biol.* **23**, 811–820 (2016).
13. Mansoor, S., Baek, M., Juergens, D., Watson, J. L. & Baker, D. Zero-shot mutation effect prediction on protein stability and function using RoseTTAFold. *Protein Sci.* **32**, e4780 (2023).
14. Pak, M. A. *et al.* Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLoS One* **18**, e0282689 (2023).
15. Chakravarty, D. *et al.* AlphaFold predictions of fold-switched conformations are driven by structure memorization. *Nat. Commun.* **15**, 7296 (2024).
16. Marcandalli, J. *et al.* Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus. *Cell* **176**, 1420–1431.e17 (2019).
17. Walls, A. C. *et al.* Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382.e17 (2020).
18. Song, J. Y. *et al.* Immunogenicity and safety of SARS-CoV-2 recombinant protein nanoparticle vaccine GBP510 adjuvanted with AS03: interim results of a randomised, active-controlled, observer-blinded, phase 3 trial. *EClinicalMedicine* **64**, 102140 (2023).

19. Ols, S. *et al.* Multivalent antigen display on nanoparticle immunogens increases B cell clonotype diversity and neutralization breadth to pneumoviruses. *Immunity* **56**, 2425–2441.e14 (2023).