Expanding beyond the natural protein repertoire to engineer targeted vaccines and diagnostics

Thèse N° 7515

Présentée le 7 novembre 2019

à la Faculté des sciences et techniques de l'ingénieur
Laboratoire de conception de protéines et d'immuno-ingénierie
Programme doctoral en biotechnologie et génie biologique

pour l'obtention du grade de Docteur ès Sciences

par

Fabian SESTERHENN

Acceptée sur proposition du jury

Prof. M. Dal Peraro, président du jury
Prof. B. E. Ferreira De Sousa Correia, directeur de thèse
Prof. B. Höcker, rapporteuse
Dr D. Angeletti, rapporteur
Prof. C. Heinis, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Acknowledgements

The work presented in this thesis would have not been possible without the support of many fantastic people.

First and foremost, I would like to thank my thesis advisor, Bruno Correia, for entrusting me with this exciting scientific project, and his unconditional and incredibly enthusiastic support throughout the last four years. Taking on the challenge of being the first PhD student of the newly founded lab has been an invaluable experience that I have never regretted. You have been an absolutely outstanding advisor and mentor, who helped me grow both as a scientist, as well as a human being. Your intellectual brilliance, paired with an always positive mindset, unlimited energy and a strong will to solve problems are absolutely unique, and I cannot imagine a better PhD advisor and mentor. Thank you for the enormous trust you placed in me from the first day, for finding the right balance between guidance and freedom to work in my own way, and for being there to suffer and to celebrate together. I am incredibly grateful for this experience and for the many things I have learned from you – both personally and professionally.

Next, I would like to thank Che Yang, who started his PhD in the lab at the same time I did, and worked closely with me during the last four years. You are an incredibly talented scientist, and it was an honor to work, learn and grow together with you. The perfect team-dynamics we developed made working extremely productive and fun. I have always appreciated your ideas, feedback, and your pragmatic and creative approach to solve problems. On a personal level, you became a really good friend, and I very much enjoyed the time with you inside and outside of the lab. I cannot imagine working towards this PhD without you, and it was a special privilege to share with you every exciting – as well as the less exciting – moment of the last four years.

Special thanks also go to Jaume Bonet, a postdoc in the lab. Professionally, I owe you my deepest gratitude for the countless hours during which you helped me with the computational aspects of this thesis. More importantly, you motivated me early on to learn programming, and you were always incredibly supportive. Also, I very much appreciate the rich scientific discussions we had throughout the last four years, and thank you for your empathy during more difficult times, for your continuous support, and for teaching me all these little tricks! Thank you for becoming a great friend - I will miss our lunch and coffee breaks, filled with endless jokes, laughs, and hilarious theories about life!

Similarly, I would like to thank Pablo Gainza, another postdoc in the lab. You are an exceptionally talented scientist, and I very much value the numerous scientific discussions we had. Thank you for always believing in my work, and for all these brilliant questions you asked. Thank you for being a great friend, for the fun lunch and coffee breaks, and after-work-beers!

I am also very grateful to Sabrina Vollers, a former postdoc of the lab. You have provided essential support for much of the work presented, and thank you so much for all the time you spent proofreading the manuscripts and my thesis! Not to forget of course - thank you for all the fun times we had together in the lab, and during after-work-beers!

Furthermore, a lot of the presented work would have not been possible without the very talented and hardworking technicians of the lab – Stéphane Rosset, Patricia Corthésy, Sandrine Georgeon and Mélanie Villard. Thank you very much for the work you have done – all this would have not been possible without you!

I would like to take this opportunity to also thank every other member of the lab for the great atmosphere, the collaborative spirit, and lots of fun during the past four years.

I am also grateful to my friends from around the world who have supported me throughout the years. Thank you all for the many precious and fulfilling moments we shared together in Lausanne and elsewhere, and for lots of fun!

Finally, I would like to thank my family. Danke an meine Eltern, Birgit und Ewald, für eure bedingungslose Unterstützung in all den Jahren. Ihr habt immer an mich geglaubt, und mir alle Möglichkeiten geboten, um mich selbst zu verwirklichen. Danke für all die mutmachenden Worte, euer Verständnis, und jeden einzelnen Besuch egal wo auf der Welt.

Last, but foremost to my heart, I would like to thank the most important person in my life – Julie. Thank you so much for your unconditional, everlasting support and love throughout the past almost 9 years. Thank you for always listening so attentively, and for all your moral support – I could have not done this PhD without you. Thank you for being there when it was time to celebrate successes, and above all, for always encouraging and standing behind me in difficult times. I know that this was not always an easy journey, especially while you were also struggling with your PhD. Not only you have been an incredible motivation coach, but I also owe you my sincere gratitude for your scientific support – your exceptional skill to ask valuable and critical questions guided me through this project, and the many brilliant ideas and suggestions you had were an invaluable contribution to this work. Words cannot express how thankful and lucky I am to be with you, and I am very much looking forward to the many exciting adventures ahead of us!

Abstract

Proteins have evolved over millions of years to carry out the vast majority of biological functions that are fundamental to life. Their three-dimensional structures and functions have been in the focus of biomedical research for many decades, and substantial progress has been made in understanding protein folding and structure-function relationships.

The inverse problem of protein folding is called *protein design* – the design of novel protein sequences that fold into predefined three-dimensional structures. Towards this aim, computational tools have emerged as a powerful tool for the *de novo* design of proteins with structural and biophysical properties that are not found in nature. To date, however, the vast majority of *de novo* designed proteins has been deprived of biological functions. Recently, the *de novo* design of proteins with customized molecular and biological functions has gained momentum, aiming to exploit them to tackle outstanding biotechnological and biomedical challenges of the 21st century.

One of these grand biomedical challenges that could be transformed by *de novo* protein design is the design of novel and more effective vaccines, especially for pathogens where traditional approaches for vaccine development have failed. Among these pathogens is the respiratory syncytial virus (RSV), which causes severe lower respiratory tract infections in young children and the elderly. Recently, numerous broadly protective, RSV neutralizing antibodies (nAbs) have been isolated from humans, and their structural characterization in complex with their target epitopes has greatly improved our molecular understanding of an effective nAb response. A remaining challenge is the design of immunogens that effectively spotlight these antigenic sites, and elicit targeted nAb responses *in vivo*.

My thesis work leverages *de novo* protein design for the design of epitope-focused immunogens that induce nAbs *in vivo*. Strikingly, we show how a cocktail of three *de novo* designed immunogens presenting selected neutralization epitopes elicit RSV nAbs in non-human primates. Furthermore, the designed immunogens bear unique potential as boosting immunogens in non-naïve subjects, allowing the focusing of nAbs onto defined antigenic sites. Together, these represent a substantial step forward for the use of immunogens based on computationally designed proteins, and offers a roadmap to employ computational protein design in the engineering of precision immunogens for other pathogens.

From a protein design perspective, my work introduces a ‘bottom-up’ approach towards the *de novo* design of functional proteins. The bottom-up approach presents a general computational protocol to build *de novo* proteins with embedded binding motifs, including those that are structurally irregular or discontinuous, i.e. consist of multiple segments. Beyond immunogens, we exploit the designed proteins as biosensors to detect and quantify epitope-specific antibody responses, providing a practical diagnostic tool to enable high-resolution immune monitoring.

Altogether, my thesis showcases a versatile, function-centric *de novo* protein design approach, applicable to address challenges including, but not limited to, the design of immunogens and antibody biosensors.

Keywords: protein design, Rosetta, reverse vaccinology, vaccine, immunofocusing, respiratory syncytial virus, antibody biosensing.

Zusammenfassung

Proteine haben sich über viele Millionen Jahre der Evolution entwickelt, um die überwiegende Mehrheit der biologischen Funktionen zu erfüllen, die für das Leben essentiell sind. Ihre dreidimensionalen Strukturen standen in den letzten Jahrzehnten im Mittelpunkt der biomedizinischen Forschung, was zu Fortschritten beim Verständnis der Proteinfaltung und ihrer Struktur-Funktions-Beziehungen geführt hat.

Das umgekehrte Problem der Proteinfaltung wird als Proteindesign bezeichnet - das Design neuartiger Proteinsequenzen, die in eine vorgegebene dreidimensionale Struktur falten. Computeralgorithmen dienen als Werkzeug für das *de novo* Design von Proteinen mit strukturellen und biophysikalischen Eigenschaften, die in der Natur nicht zu finden sind. Bis heute waren jedoch die meisten *de novo* designten Proteine ohne biologische Funktion. In jüngster Zeit hat das *de novo* Design von Proteinen mit maßgeschneiderten molekularen und biologischen Funktionen an Dynamik gewonnen, um sie für die Bewältigung von biomedizinischen Herausforderungen des 21. Jahrhunderts zu nutzen.

Eine der großen biomedizinischen Herausforderungen, zu denen *de novo* Proteindesign einen Beitrag leisten könnte, ist die Entwicklung neuartiger, wirksamer Impfstoffe - insbesondere für Krankheitserreger, bei denen traditionelle Ansätze für die Impfstoffentwicklung gescheitert sind. Zu diesen Erregern gehört das Respiratorische Synzytial Virus (RSV), das bei kleinen Kindern und älteren Menschen schwere Infektionen der unteren Atemwege verursacht. In letzter Zeit wurden zahlreiche Antikörper aus Menschen isoliert, die RSV wirksam neutralisieren, und deren strukturelle Charakterisierung hat unser molekulares Verständnis einer wirksamen, neutralisierenden Antikörperreaktion erheblich verbessert. Eine Herausforderung bleibt jedoch das Design von Immunogenen, die gezielt neutralisierende Antikörper gegen diese Epitope *in vivo* hervorrufen.

Meine Dissertation zielt darauf ab, mit Hilfe von computer-basiertem Proteindesign Antigene zu entwickeln, welche neutralisierende Antikörper gegen gezielt ausgesuchte Epitope hervorrufen. Wir zeigen, wie ein Cocktail bestehend aus drei computer-designten Antigenen RSV-neutralisierende Antikörper in nichtmenschlichen Primaten hervorruft. Weiterhin zeigen wir deren einzigartiges Potenzial, eine bereits bestehende Antikörperantwort auf ausgewählte Epitope zu fokussieren. Diese Ergebnisse sind von Bedeutung für eine Impfstrategie basierend auf Computer-designten Antigen, und zeigt Wege auf, um ähnliche Antigene für andere Infektionserreger zu entwickeln.

Für das Gebiet des Proteindesigns wird in dieser Arbeit ein „Bottom-Up“ Ansatz für das *de novo* Design von funktionellen Proteinen vorgestellt. Diese vielseitig einsetzbare Strategie erlaubt es, Proteine mit eingebauten funktionellen Bindungsmotiven zu designen (einschließend strukturell unregelmäßige oder diskontinuierliche Motive). Neben dem Design von Antigenen zeigen wir eine Anwendung dieser Designstrategie für die Entwicklung von protein-basierten, diagnostischen Tests, mit denen spezifische Antikörperantworten detektiert und quantifiziert werden können.

Insgesamt zeigt diese Dissertation einen vielseitigen einsetzbaren, funktionsorientierten Ansatz für das *de novo* Design von Proteinen, anwendbar für die Entwicklung von Antigenen, Biosensoren und darüber hinaus.

Schlüsselworte: Proteindesign, Rosetta, Reverse Vakzinologie, Impfstoff, Immunfokussierung, Respiratorisches Synzytial Virus, Antikörper-Biosensor

Table of contents

ACKNOWLEDGEMENTS .... **III**

ABSTRACT .... **V**

ZUSAMMENFASSUNG .... **VI**

TABLE OF CONTENTS .... **VII**

LIST OF FIGURES .... **IX**

LIST OF SUPPLEMENTARY FIGURES .... **X**

LIST OF TABLES .... **XI**

LIST OF ABBREVIATIONS .... **XI**

CHAPTER 1 INTRODUCTION **1**

1.1 COMPUTATIONAL PROTEIN DESIGN 3
1.2 *DE NOVO* PROTEIN DESIGN – STATE OF THE ART 5
1.3 TOWARDS FUNCTIONAL PROTEIN DESIGN 6
1.4 STRUCTURE-BASED IMMUNOGEN DESIGN STRATEGIES 12
1.5 RESPIRATORY SYNCYTIAL VIRUS 21

OBJECTIVES **24**

CHAPTER 2 BOOSTING SUBDOMINANT NEUTRALIZING ANTIBODY RESPONSES WITH A COMPUTATIONALLY DESIGNED EPITOPE-FOCUSED IMMUNOGEN **26**

2.1 ABSTRACT 26
2.2 INTRODUCTION 27
2.3 RESULTS 29
2.4 DISCUSSION 39
2.5 METHODS 41
2.6 SUPPLEMENTARY INFORMATION 48

CHAPTER 3 TRIVALENT COCKTAIL OF *DE NOVO* DESIGNED IMMUNOGENS ENABLES THE ROBUST INDUCTION AND FOCUSING OF FUNCTIONAL ANTIBODIES *IN VIVO* **55**

3.1 ABSTRACT 56
3.2 INTRODUCTION 56
3.3 RESULTS 58
3.4 DISCUSSION AND CONCLUSIONS 67
3.5 METHODS 69
3.6 SUPPLEMENTARY INFORMATION 82

CHAPTER 4 A BOTTOM-UP DESIGN APPROACH FOR THE *DE NOVO* DESIGN OF FUNCTIONAL PROTEINS **99**

4.1 INTRODUCTION 100
4.2 RESULTS 102
4.3 DISCUSSION 109
4.4 METHODS 111
4.5 SUPPLEMENTARY INFORMATION 115

CHAPTER 5 CONCLUSIONS & PERSPECTIVES **119**
5.1 STRUCTURE-BASED VACCINE DESIGN 119
5.2 PROTEIN DESIGN FIELD 123
REFERENCES **127**

CURRICULUM VITAE

List of Figures

FIGURE 1.1: ORDERS OF PROTEIN STRUCTURE. 1
FIGURE 1.2: SAMPLING AND SCORING STRATEGIES FOR PROTEIN DESIGN. 3
FIGURE 1.3: STRUCTURAL DETERMINANTS OF PROTEIN FUNCTION USING THE EXAMPLE OF THE EPO – EPOR RECOGNITION. 6
FIGURE 1.4: STRATEGIES FOR FUNCTIONAL SITE TRANSPLANTATION. 9
FIGURE 1.5: AVAILABLE TEMPLATE STRUCTURES FOR BINDING MOTIFS OF DIFFERENT STRUCTURAL COMPLEXITY. 10
FIGURE 1.6: STRUCTURE-BASED APPROACHES FOR IMMUNOGEN DESIGN. 14
FIGURE 1.7: NEUTRALIZATION EPITOPES ON PRE- AND POSTFUSION RSVF. 22
FIGURE 2.1: DESIGN OF AN RSV-BASED NANOPARTICLE DISPLAYING A SITE II EPITOPE-FOCUSED IMMUNOGEN. 30
FIGURE 2.2: IMMUNOGENICITY AND QUANTIFICATION OF SITE II-SPECIFIC ANTIBODY RESPONSES. 31
FIGURE 2.3: RSVF CROSS-REACTIVITY AND SERUM NEUTRALIZATION. 33
FIGURE 2.4: HETEROLOGOUS PRIME BOOST RESHAPES ANTIBODY RESPONSES ENHANCING LEVELS OF SITE II-SPECIFIC ANTIBODIES. 36
FIGURE 2.5: BOOSTED SITE II-SPECIFIC ANTIBODIES ARE FUNCTIONAL AND MEDIATE INCREASED NEUTRALIZATION ACTIVITY. 38
FIGURE 3.1: COMPUTATIONAL DESIGN OF IMMUNOGENS TO ELICIT RSV NABS FOCUSED ON THREE DISTAL EPITOPES. 58
FIGURE 3.2: TEMPLATED COMPUTATIONAL DESIGN AND BIOPHYSICAL CHARACTERIZATION OF SYNTHETIC IMMUNOGENS. 59
FIGURE 3.3: MOTIF-CENTRIC *DE NOVO* DESIGN OF EPITOPE-FOCUSED IMMUNOGENS. 61
FIGURE 3.4: STRUCTURAL CHARACTERIZATION OF *DE NOVO* DESIGNED IMMUNOGENS. 63
FIGURE 3.5: SYNTHETIC IMMUNOGENS ELICIT NEUTRALIZING SERUM RESPONSES IN MICE AND NHPS, AND FOCUS PRE-EXISTING IMMUNITY ON SITES 0 AND II. 66
FIGURE 4.1: A BOTTOM-UP *DE NOVO* PROTEIN DESIGN STRATEGY FOR THE DESIGN OF FUNCTIONAL PROTEINS. 103
FIGURE 4.2: BIOPHYSICAL CHARACTERIZATION OF LEAD VARIANTS FROM EACH TOPOLOGY. 104
FIGURE 4.3: STRUCTURAL CHARACTERIZATION OF 4E1H_95. 106
FIGURE 4.4: BIOSENSOR FOR THE DETECTION OF ANTIBODY RESPONSES BASED ON *DE NOVO* DESIGNED PROTEINS. 108
FIGURE 5.1: SYNTHETIC IMMUNOGENS AS BOOSTING IMMUNOGENS UNDER CONDITIONS OF PRE-EXISTING IMMUNITY. 120
FIGURE 5.2: SINGLE- VERSUS MULTI-SITE IMMUNOFOCUSING USING SYNTHETIC IMMUNOGENS. 122
FIGURE 5.3: BOTTOM-UP *DE NOVO* DESIGN OF FUNCTIONAL PROTEINS. 125
FIGURE 5.4: POTENTIAL FUTURE APPLICATION OF *DE NOVO* PROTEINS FOR SYNTHETIC BIOLOGY APPLICATIONS. 126

List of Supplementary Figures

FIGURE S 2.1: ADJUVANT SCREEN FOR FFL_001 IMMUNOGEN. 48
FIGURE S 2.2: HOMOLOGY-GUIDED RESURFACING OF FFL_001. 49
FIGURE S 2.3: SPR SENSORGRAMS FOR SITE II NABS. 50
FIGURE S 2.4: SCHEMATIC REPRESENTATION OF THE SPR SERUM COMPETITION ASSAY. 51
FIGURE S 2.5: FAR-ULTRAVIOLET CIRCULAR DICHROISM SPECTRUM OF ANTIGENIC SITE II PEPTIDE. 51
FIGURE S 2.6: SERUM REACTIVITY WITH RSVF AND VIRAL LYSATE. 52
FIGURE S 2.7: OVERLAPPING ANTIBODY CLONOTYPES OF MICE IMMUNIZED WITH RSVF AND NRM. 52
FIGURE S 2.8: CORRELATION OF SITE II PEPTIDE SPECIFIC SERUM TITER WITH RSV NEUTRALIZATION IC_{50}. 53
FIGURE S 2.9: ENRICHMENT OF SITE II-SPECIFIC ANTIBODIES FROM MOUSE SERA. 53
FIGURE S 2.10: COMPETITION ELISA WITH MOTAVIZUMAB ANTIBODY CONTROL. 54
FIGURE S 3.1: COMPUTATIONAL DESIGN AND EXPERIMENTAL OPTIMIZATION OF S4_1 DESIGN SERIES. 82
FIGURE S 3.2: EXPERIMENTAL CHARACTERIZATION OF S4_1 DESIGN SERIES. 83
FIGURE S 3.3: COMPUTATIONAL DESIGN AND EXPERIMENTAL OPTIMIZATION OF S0_1 DESIGN SERIES. 84
FIGURE S 3.4: BIOPHYSICAL CHARACTERIZATION OF THE S0_1 DESIGN SERIES. 85
FIGURE S 3.5: SHAPE MIMICRY OF COMPUTATIONALLY DESIGNED IMMUNOGENS COMPARED TO PREFUSION RSVF. 86
FIGURE S 3.6: TOPOBUILDER DESIGN STRATEGY. 87
FIGURE S 3.7: *DE NOVO* BACKBONE ASSEMBLY FOR SITE IV IMMUNOGEN. 87
FIGURE S 3.8: BIOPHYSICAL CHARACTERIZATION OF S4_2 DESIGN SERIES. 88
FIGURE S 3.9: BIOPHYSICAL CHARACTERIZATION OF S4_2.45 AND S0_2.126. 89
FIGURE S 3.10: *DE NOVO* TOPOLOGY ASSEMBLY TO STABILIZE SITE 0 USING TOPOBUILDER. 90
FIGURE S 3.11: BIOPHYSICAL CHARACTERIZATION OF S0_2 DESIGN SERIES. 91
FIGURE S 3.12: BINDING AFFINITY TOWARDS PANELS OF HUMAN NEUTRALIZING ANTIBODIES AND HUMAN SERA. 92
FIGURE S 3.13: COMPARISON OF S0_2.126 ROSETTA SCORES AGAINST NATURAL PROTEINS OF SIMILAR SIZE. 93
FIGURE S 3.14: ELECTRON MICROSCOPY ANALYSIS OF SITE-SPECIFIC ANTIBODIES IN COMPLEX WITH PREFUSION RSVF. 94
FIGURE S 3.15: COMPOSITION AND EM ANALYSIS OF TRIVAX1 RSVN NANOPARTICLES. 95
FIGURE S 3.16: EM ANALYSIS OF TRIVAX2 FERRITIN NANOPARTICLES. 96
FIGURE S 3.17: NHP NEUTRALIZATION TITER MEASURED BY AN INDEPENDENT LABORATORY. 97
FIGURE S 3.18: MOUSE IMMUNIZATION STUDIES WITH TRIVAX1. 97
FIGURE S 3.19: NHP SERUM REACTIVITY WITH DESIGNED IMMUNOGENS. 98
FIGURE S 4.1: TOPOBUILDER WORKFLOW. 111
FIGURE S 4.2: DESIGN AND HIGH-THROUGHPUT SCREENING FOR 3H1L_02 TOPOLOGY. 115
FIGURE S 4.3: BIOPHYSICAL CHARACTERIZATION OF BEST 3H1L_01 VARIANTS. 116
FIGURE S 4.4: EXTENDED BIOPHYSICAL CHARACTERIZATION OF 3H1L_02.395. 117
FIGURE S 4.5: TOPOLOGICAL TUNING OF THE 4H TOPOLOGY FOR OPTIMAL PRESENTATION OF BOTH BINDING MOTIFS. 118

List of Tables

TABLE 1.1: GENERAL DESIGN STRATEGIES FOR ANTIGENIC CONFORMATIONAL STABILIZATION OF VIRAL FUSION PROTEINS. 16
TABLE 1.2: OVERVIEW OF IMMUNOGEN ENGINEERING APPROACHES AND THEIR IMMUNOLOGICAL OUTCOMES. 20

TABLE S 3.1: REFINEMENT STATISTICS OF THE S0_2.126 NMR STRUCTURE. 77
TABLE S 3.2: X-RAY DATA COLLECTION AND REFINEMENT STATISTICS OF S0_2.126 CRYSTAL STRUCTURE. 79
TABLE S 3.3: X-RAY DATA COLLECTION AND REFINEMENT STATISTICS OF S4_2.45 CRYSTAL STRUCTURE. 81
TABLE S 4.1: FOLD DEFINITION AND SPATIAL POSITIONING OF EACH SSE IN THE DESIGNED TOPOLOGIES. 113

List of abbreviations

Å	Angstrom
bnAb	Broadly neutralizing antibody
BRET	Bioluminescence resonance energy transfer
EM	Electron microscopy
FunFolDes	Functional folding and design algorithm
HA	Influenza hemagglutinin
nAb	Neutralizing antibody
NHP	Non-human primate
RMSD	Root mean square deviation
RSV	Respiratory syncytial virus
RSVF	Respiratory syncytial virus fusion protein
RSVN	Respiratory syncytial virus nucleoprotein
SPR	Surface plasmon resonance
SSE	Secondary structure element

Chapter 1 Introduction

Proteins are the most versatile macromolecules in living systems and are fundamental to enable life. The word ‘protein’ derives from the Greek work ‘*proteios’*, meaning ‘of first importance’. Proteins are extremely abundant in nature, making up half the dry weight of a cell. They have evolved throughout millions of years, and represent the ‘molecular machines’ that exert virtually any function within a cell, tissue and the entire organism. Key functions include, for example, mechanical roles, catalysis of chemical reactions, information transfer, molecular recognition of other proteins, and many others (*6*).

Proteins are composed of a sequence of amino acids, which are covalently linked through peptide bonds (Fig 1.1). The sequence of amino acids is determined by the genetic information, which contains – in most organisms – information for 20 chemically different amino acids (*6, 7*). All 20 amino acids share the same backbone atoms, but carry side chains with different chemical properties (charged, polar, hydrophobic). Upon translation of the genetic information into its corresponding amino acid sequence (‘primary structure’), proteins collapse spontaneously into a defined three-dimensional structure that corresponds to their lowest free energy state (*8*). Thus, the information for the three-dimensional structure of a protein is solely contained in its amino acid sequence, and protein structure is described at three basic levels (Fig 1.1). The secondary structure refers to the formation of a local sub-structure, such alpha-helices, beta strands and loops. Alpha helices stabilized predominantly by local hydrogen bonds formed by the main-chain peptide groups, whereas beta-strands are stabilized both by local and non-local hydrogen bonds. Loops are less ordered secondary structures, and connect strands to strands, strands to helices, and helices to helices. The tertiary structure describes the three-dimensional organization between all secondary structures and loops, which is largely determined and stabilized by non-local interactions. The main driving forces of this thermodynamically

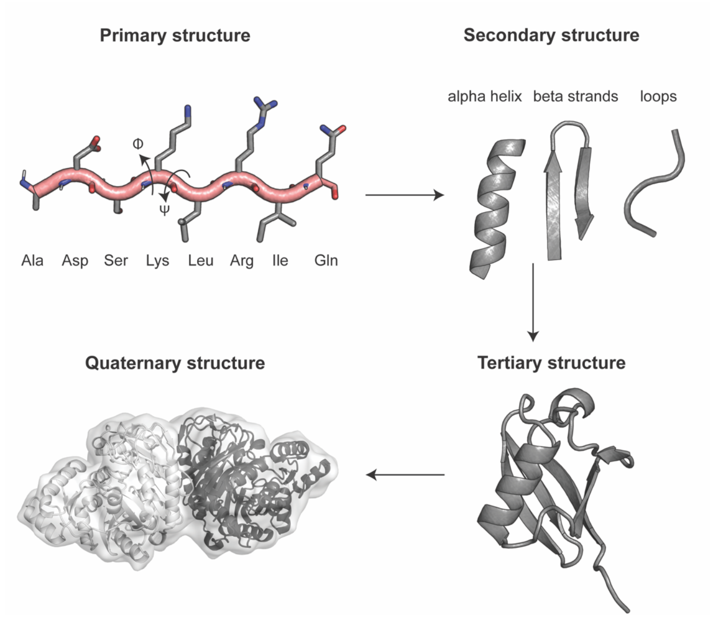


Figure 1.1: Orders of protein structure.
The primary structure corresponds to the amino acid sequence as encoded by the genetic information. The main chain (‘backbone’) is shown in light red, side chains in grey, and backbone torsion angles (Φ and ψ) are indicated. There are three types of secondary structures: alpha helices, beta strands and loops. The tertiary structure refers to the spatial organization between all secondary structures, and is primarily stabilized by non-local interactions. The quaternary structure is formed by a complex of multiple protein molecules.

favorable process are the burial of hydrophobic side chains, the formation of intramolecular hydrogen bonds, salt bridges, and van-der-Waals forces (*9*). Lastly, the quaternary structure is formed by the assembly of multiple proteins, forming multi-subunit complexes held together by hydrogen bonds, salt bridges, hydrophobic interactions, and sometimes disulfide bridges (*9*).

The molecular functions of proteins are encoded in their three-dimensional structures. Thus, a detailed understanding of protein structure is key for our understanding of biology. To date, the scientific community has determined more than 150,000 protein structures using experimental techniques (X-ray crystallography, nuclear magnetic resonance (NMR), and high-resolution cryo-electron microscopy), which are deposited in the Protein Data Bank (PDB), a constantly growing database (*10*).

When considering the vast amount of amino acid sequences found in nature, one would expect a similarly large number of different protein tertiary structures. However, upon classification of the experimentally determined protein structures, it was found that nature re-uses a restricted set of protein structures. These have been catalogued in databases aiming to provide a comprehensive description of the structural and evolutionary relationship between all proteins. Among the most noteworthy protein structural classification databases are the CATH and SCOP databases (*11, 12*). For example, the CATH database classifies proteins hierarchically according to four attributes: class, architecture, topology/fold, and homologous superfamily. The class refers to the secondary structure composition (e.g. mostly alpha, mixed alpha/beta, mostly beta), whereas the architecture describes their overall shape and the orientation between secondary structures in the three-dimensional space, but ignoring connectivity. The topology or fold definition categorizes structures by overall shape and connectivity between secondary structures, and the homologous superfamily groups together proteins which share a common ancestor.

Depending on the exact definition, nature uses between 1,000 and 10,000 discrete protein folds, indicating that the same fold can be used for multiple functional purposes (*13*). An illustrative example for this statement is the TIM-barrel fold, a mixed alpha-beta fold that is found in at least 15 different enzyme families with diverse enzymatic functions (*14*). Another example is the immunoglobulin fold, found in hundreds of proteins with different functions, including antibodies, various cell surface receptors, enzymes and many others (*15*).

Beyond nature's relatively limited set of protein folds, a long-standing goal for bioengineers was to create novel protein sequences that fold into a desired structure and perform a given biological function, known as *protein design* (*16, 17*). Generally, protein design endeavors can be classified into two main categories – (I) protein re-design and (II) *de novo* protein design. Protein re-design refers to improving or altering a desired property of an existing protein (e.g. solubility or binding affinity to another molecule) by introducing mutations in relevant regions. These efforts can either be guided by a known 3D structure, or alternatively, can be attempted by introducing a large number of (random) mutations, which are subsequently experimentally screened using directed evolution approaches (*18*). Mostly, protein re-design tends to have little influence on the overall protein structure, but is rather an optimization of a particular biophysical or functional property.

In contrast, *de novo* design refers to the design of novel proteins 'from scratch', based on the physical principles of protein structure and using a sequence unrelated to those observed in nature (*19*). When considering the vast combinatorial amount of possible amino acid sequences (a protein of 100 residues has 20^{100} possible sequences as each residue can be one of the 20 amino acids), it seems unlikely that nature has sampled the sequence space exhaustively (*19*). Instead, only an infinitesimal fraction of the theoretically possible sequence and protein fold space has been explored by evolution (*11, 20*). In other words, the evolution of natural proteins has been of incremental nature, likely driven by the recombination of small protein fragments, rather than a systematic exploration of the sequence-structure space (*21, 22*).

Expanding towards this unknown territory of sequences and structures that are not found in nature is what defines the arena for *de novo* protein design (*19*). However, due to the vast number of possible sequences and structures, exploring this sequence space in a random fashion (e.g. using directed evolution) would be a quite inefficient and non-systematic approach. Therefore, computational methods that are based on the physical understanding of protein folding have been developed, as detailed in the following section.

1.1 Computational protein design

In general, computational protein design algorithms perform two basic operations: (I) sampling and (II) scoring. Sampling refers to, depending on the allowed degress of freedom, changing the sequence as well as backbone and/or side chain conformations. For each sampling state, a scoring function computes (or approximates) the energy, together resulting in an energy landscape that maps each conformational state to an energy level (Fig 1.2). According to Anfinsen's dogma, a given protein sequence will fold into its lowest energy state, and thus, the native structure is found in the global minimum of the energy landscape (*8*). However, the exhaustive sampling and scoring of all possible sequences and their conformational states is computationally expensive, thus requiring strategies to reduce this vast search space, as described below.

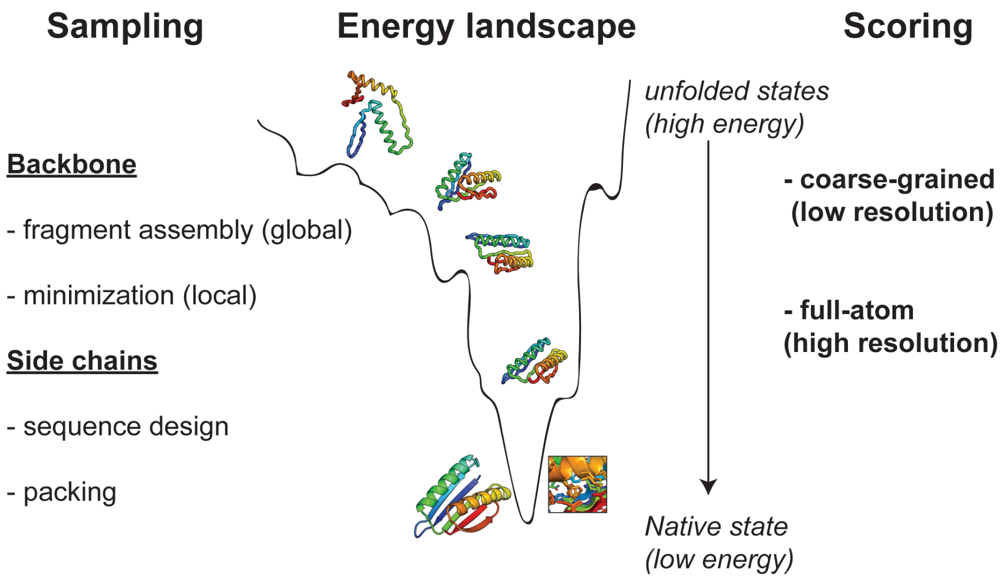


Figure 1.2: Sampling and scoring strategies for protein design.
Backbone and side chain sampling strategies to explore the sequence-structure space (left) and to calculate the energy of the sampled states using coarse-grained and full atom energy functions (right). Unfolded states are high in energy, whereas the native state corresponds to its lowest energy state. The shown structure is Top7, from PDB 1QYS.

The Rosetta macromolecular modeling suite is currently the most widely used software package for *de novo* protein design, and is used throughout the different chapters of my thesis. Rosetta started as a sequence-based, *de novo* structure-prediction tool, and has since proven to be one of the most successful methods for structure prediction in the community-wide critical assessment of structure prediction (CASP) competition

(*23*). Given its success in sequence-based structure prediction, much effort has been undertaken to develop protocols for protein design. To date, Rosetta is being developed by a large scientific community to perform a wide variety of macromolecular modeling tasks, including *de novo* protein design, ligand docking, loop modeling, and many others (*24*). Since a detailed review of protocols implemented in Rosetta would exceed the scope of this thesis, the description is limited to *de novo* protein design protocols used in this work.

Rosetta, like all structure prediction and protein design softwares, performs two tasks: first, searching the conformational space by coarsely sampling backbone degrees of freedom; and, in case of design, sampling the sequence space. For each structural model, the energy must be computed, thus allowing to rank the models and find the global energy minimum. However, the sequence-structure space is vast because the degrees of freedom in a polypetide chain scale exponentially with the length of the amino acid sequence. Thus, Rosetta reduces this vast search space through stochastical Monte Carlo sampling, and by using knowledge-based sampling and scoring strategies, as described below.

Depending on the design objective, backbone degrees of freedom can either be sampled on a global or on a more local scale. One of the most important strategies for the large-scale sampling of backbone conformations is referred to as fragment assembly, which was originally developed for *de novo* structure prediction. A protein fragment is a particular conformation of a short, continuous backbone stretch. These fragments, typically in the length of 3 and 9 amino acids, are collected for each position from a database of known, high quality protein structures. From this database, fragments are picked based on similar local sequences, hypothesizing that these will adopt a similar local structure (*25*). Subsequently, the fragments are assembled, and a low-resolution scoring function assesses whether the assembly resulted in a conformation allowing for favorable non-local interactions. This coarse-grained scoring function uses a reduced description of protein structure, treating side chains as centroids. It includes statistical potentials for solvation energy, electrostatics, hydrogen bonding as well as a repulsive term for steric clashes, and is mostly knowledge-based. Importantly, this low-resolution scoring function enables the exploration of a wider energy landscape and identification of configurations that are close to the global minimima - even when structural defects are present that a full-atom energy function would penalize. This fragment assembly is the basis for Rosetta *ab initio* structure predictions (*23*), and an essential step of the FunFolDes protocol described in the following chapters (*26*).

Following the coarse-grained exploration of the energy landscape, high-resolution atomic details are added to the protein structures to allow a more accurate energy calculation and optimization. The all-atom energy function is a linear combination of weighted score terms, including energy terms for van der Waals interactions (Lennard-Jones potential), electrostatic interactions, hydrogen bonds, solvation energy, backbone torsion angles and others (*27*).

The local sampling and refinement of backbone conformations is referred to as backbone minimization. Minimization aims to find the nearest local energy minimum by sampling backbone movements that are close in 3D space compared to the starting structure's conformation, achieved through local torsion angle perturbation. Optimizing this local conformational space is often performed simultaneously with side chain repacking, which attempts to find the lowest energy conformation of side chains on a fixed backbone. Together, minimization and repacking are referred to as a structure relaxation step.

The systematic sampling of side chains in all combinations and orientations is computationally expensive. Thus, Rosetta drastically reduces the search space through the use of discrete side chain conformations, known as *rotamers*, that are frequently observed in the PDB (*28*). These rotamers sample discrete combinations of side chain torsion angles dependent on backbone geometry, thereby reducing the number of conformations. Another strategy to reduce the search space is to restrict the amino acids sampled to hydrophobic residues for positions that are solvent-excluded in the designed backbone, and to hydrophilic side chains for solvent-exposed positions. With these constraints, the sequence design process is performed in a Monte Carlo simulated annealing search to guide the sampling towards an optimized score, as computed by the all-atom energy function (*27, 29*). While Monte Carlo sampling allows to explore the vast sequence-structure space in a reasonable time, it does not guarantee to reach the optimal solution, i.e. the global energy minimum (*30*).

Altogether, *de novo* design protocols tend to sample and optimize both backbone and side chain conformations simultaneously, aiming to find sequences that have their (predicted) lowest energy state close to the target structure. Depending on the degrees of freedom allowed, this heuristic sampling generates 10^4-10^6 decoys, a subset of which is selected for subsequent experimental validation of the computational models (*4*).

1.2 *De novo* protein design – state of the art

Historically, one of the first attempts of *de novo* protein design has been reported in 1979 by Gutte et al., who designed a 34-residue polypeptide based on rules obtained from known x-ray structures (*31*). The conceptual framework for the *de novo* protein design field was given by Eric Drexler, who reasoned that although predicting protein structure from sequence may be challenging, finding sequences that are compatible with a given structure may be more feasible (*16*), and Carl Pabo who defined the design of proteins as the inverse folding problem (*17*). Subsequent *de novo* protein design efforts were reported during the late 1980's with the design of relatively simple helical bundle structures (*32-34*). In 1997, the first computationally designed *de novo* protein was reported by Dahiyat et al., followed by the laboratory of David Baker, who used Rosetta to design the first *de novo* protein with a fold not observed in nature, known as Top7 (*35, 36*).

Since then, the *de novo* protein design field has rapidly evolved. Major efforts have been made in order to understand the rules that govern protein folding and translating them into a set of design principles. The groundwork towards this aim was laid by Koga et al., who defined a set of rules for the design of idealized protein structures (*37, 38*). The defined rules related local secondary structure elements to tertiary structures through defined loop lengths. Uncovering geometric and chemical principles from naturally occurring protein structures has subsequently enabled the *de novo* design of a wide variety of protein structures, such as alpha helical coiled coils (*39*), curved beta sheets (*40*) and others (*41, 42*). A notable achievement was the *de novo* design of a TIM-barrel, a fold commonly found in enzymes, consisting of eight alpha helices and eight beta strands that form a closed toroid (*42*). Altogether, from a structural perspective, it has become possible to design novel proteins which adopt the intended fold with high accuracy, and rules for the design of many protein folds and geometries have been defined.

Nonetheless, from a functional perspective, the problem of protein design is far from solved. The vast majority of studies has thus far focused on the design of highly idealized protein structures with the primary goal being structural accuracy, but deprived of biological function. Exceptions and first steps towards functional protein design are highlighted in the following section, where currently available methods for the *de novo* design of functional proteins are also discussed.

1.3 Towards functional protein design

Proteins have evolved a vast repertoire of molecular functions, which in turn control many different biological processes. While the molecular functions are very diverse (e.g. catalysis of chemical reactions, mechanical functions, recognition, etc.), a prerequisite for almost all protein functions is its physical interactions with other molecules, including but not limited to other proteins, small molecules, metals, nucleic acids, ions or lipids (*9*). The region of a protein that interacts with a ligand is called the binding site, which is formed by a particular arrangement of amino acids that form a network of non-covalent interactions with the ligand. In many cases, the physical interaction itself can be considered a function – examples include antibodies that bind and neutralize a pathogen, ligands that mediate receptor dimerization, and many others. In other cases, binding is only the trigger event for a more complex function – prominent examples are many enzymatic reactions, where substrate binding triggers conformational changes which then enable catalysis (*6*). Although binding and molecular recognition by themselves are not in all cases sufficient to mediate a biological function, it is a critical event that governs the vast majority of protein functions in biological systems.

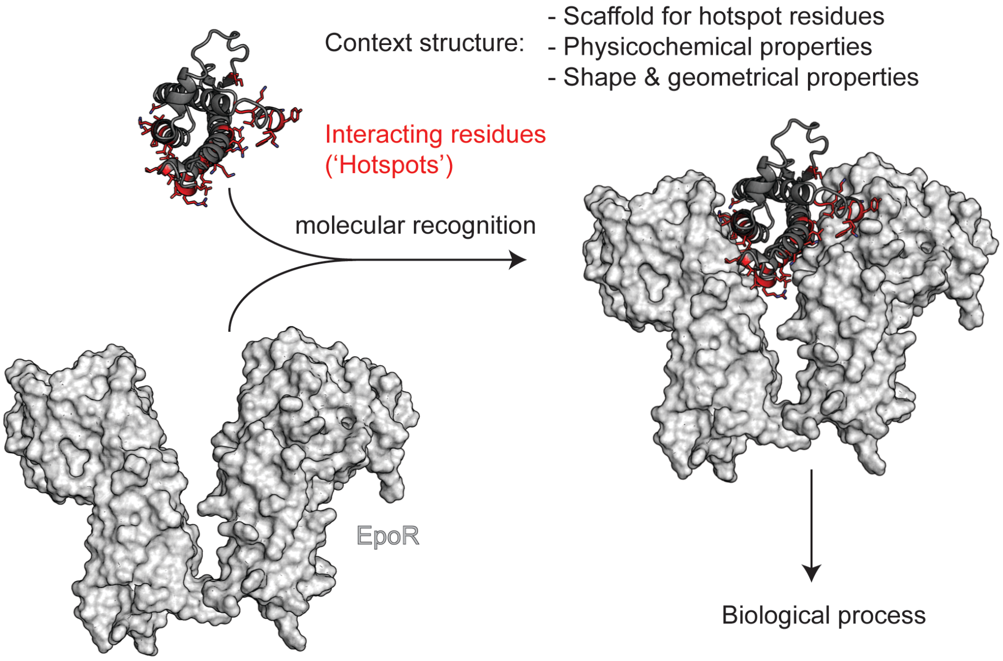

Context structure:
- Scaffold for hotspot residues
- Physicochemical properties
- Shape & geometrical properties
Interacting residues ('Hotspots')
molecular recognition
EpoR
Biological process

Figure 1.3: Structural determinants of protein function using the example of the Epo – EpoR recognition.
The example function shown is the molecular recognition between the EpoR (light grey surface) and its ligand Epo (dark grey cartoon), based on PDB entry 1EER. Side chains that are directly involved in the recognition of EpoR are colored red and represented as sticks. These hotspot residues are embedded in a context structure, which serves as a scaffold to stabilize their spatial configuration, and includes physicochemical and topological properties that impact its function.

Figure 1.3 illustrates a functional protein-protein interaction using the example of the erythropoietin (Epo) growth factor, an essential growth factor that stimulates red blood cell production, binding to its receptor (EpoR). The Epo-EpoR interface is formed by multiple helical and loop structures, where selected side chains ('hotspots') form a favourable network of polar and apolar interactions (*43*). All residues that are not in direct contact with the binding partner can be classified as scaffold or context structure, which has three major roles. First, it serves as a 'scaffold' to stabilize the interacting side chains in the correct spatial configuration.

Second, it carries a set of physicochemical properties that can impact its function, affecting for example stability, conformational dynamics, diffusion, cellular location, and more. The third role relates to its shape and geometry, which - in this example - is critical to mediate receptor homodimerization in a precise spatial orientation in order to trigger downstream signaling.

Together, the multi-layered complexity of structural, physicochemical and topological determinants of protein function poses an overarching challenge in design of functional proteins. At the same time, the ability to engineer and tune the different components opens a number of opportunities to improve our understanding of the structure-function relationship, and to design novel proteins with customized functions.

Broadly, two main approaches to design functional proteins can be employed: (1) re-design of existing functional proteins; and (2) functional site transplantation. The re-design of a functional proteins entails for example to alter or improve a desired property (e.g. ligand specificity, stability), generally achieved by altering the sequence in relevant regions, while maintaining the chemical and topological features that surround the functional site (*44*). In contrast, functional site transplantation refers to isolating a functional site from its native structural environment, and installing it in another, similar or entirely distinct, chemical and topological context.

The transplantation of functional sites presents a number of important opportunities, as further described throughout this thesis: first, it allows to engineer the complete protein predictably; second, through altering the physicochemical properties and the topological context, the protein may be functional under different conditions compared to the native protein; third, it is conceivable to engineer multi-functional proteins carrying multiple functional sites; lastly, it allows to test our understanding of the structure – function relationship for a given functional motif.

A common strategy for functional site transplantation is known as protein grafting (Approach 1, Fig. 1.4) (*45*). Briefly, grafting consists of two steps: First, the functional motif is isolated from its native context and used to search a database of known protein structures for proteins that have local structural similarity to the motif of interest. Once a protein with sufficient local similarity is identified, in the second step, either the relevant side chains or both backbone and side chains are transplanted. Grafting has been used successfully for the design of a variety novel functional proteins, including ligand binding proteins (*46*), various inhibitors (*47-52*) and immunogens (*53-56*). While some functions were limited to *in vitro* binding to a defined target (*54*), other studies have shown functionality *in vivo*, for example the induction of apoptosis in cells (*48, 49*), toxin neutralization (*51*) and anti-viral activity (*51, 52*). Importantly, some of the designs showed superior biochemical properties compared to antibodies, underlining the potential of engineering novel functional proteins with customized biochemical properties (*51*).

However, with only few exceptions (*54*), the grafted functional sites were regular helical segments - a common motif found in natural proteins. Thus, numerous grafting templates were available in the natural protein repertoire that showed high local structural similarity (<1 Å) to the motif, allowing for side chain-only grafting or, at most, limited backbone flexibility in the target structure (Fig 1.5). However, function in natural proteins is rarely contained within single, regular helical segments, but rather in the 3D arrangement of several, often irregular structural elements that are supported by defined topological features of the overall protein (*57, 58*). The difficulty of grafting these discontinuous motifs lies within identifying templates that have enough local structural similarity to multiple segments in their native orientation scales with the number of segments, as well as the structural complexity of each segment. Figure 1.5 illustrates that putative design templates with local structural similarity are widely available for regular, single segment motifs, but are sparse for irregular and discontinuous segments.

An alternative strategy to install functional sites in template structures was reported by Correia and colleagues in 2014 (*59*) (Approach 2, Fig 1.4). Conceptually, their design approach originated in protein grafting, but overcame its major limitation – the inability to handle structural motifs for which design templates with enough local structural similarity are unavailable in the PDB. The computational protocol, for which we recently published an upgraded version, was termed Rosetta Functional Folding and Design (FunFolDes), and is one of the important computational protein design methods used throughout this thesis (*26*).

The distinctive feature of FunFolDes compared to grafting is that it couples protein folding and design to bias the sampling towards backbones and sequences that allow stabilization of the transplanted motif, as illustrated in Fig 1.4. Briefly, the design process folds an initially extended protein chain around a given functional motif which is kept static, using Rosetta's fragment insertion machinery. This allows the scaffold's conformation to adapt to the three-dimensional constraints of the functional motif. Upon folding, sequences are designed that (a) stabilize the folded protein, and (b) maximize side chain and backbone contacts with the functional motif, allowing its accurate stabilization in the heterologous scaffold.

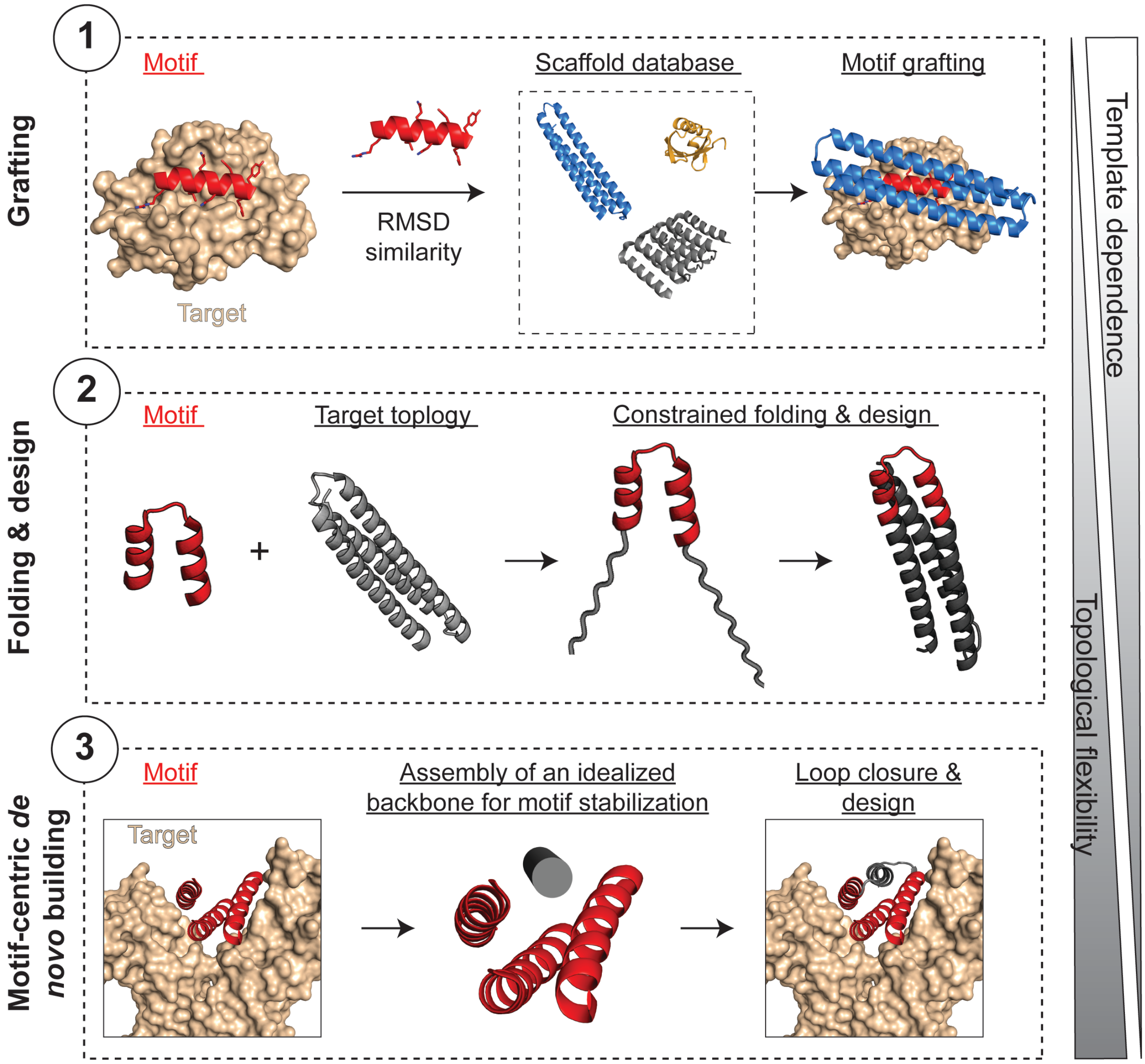

1
Grafting
Motif
Target
RMSD similarity
Scaffold database
Motif grafting
2
Folding & design
Motif
+
Target toplogy
Constrained folding & design
3
Motif-centric de novo building
Motif
Target
Assembly of an idealized backbone for motif stabilization
Loop closure & design
Template dependence
Topological flexibility

Figure 1.4: Strategies for functional site transplantation.
Approach 1: Motif grafting. A functional motif is transplanted onto a pre-existing protein structure that shows local structural similarity to the motif of interest. Approach 2: Folding and design. A target topology for a given functional motif is chosen and serves as template structure to guide *in silico* folding and design. See chapters 3 and 4 for more details. Approach 3: Motif-centric *de novo* building. The basis of approach 3 is to tailor a *de novo* protein to the functional motif of interest, rather than repurposing an existing template as in approaches 1 and 2.

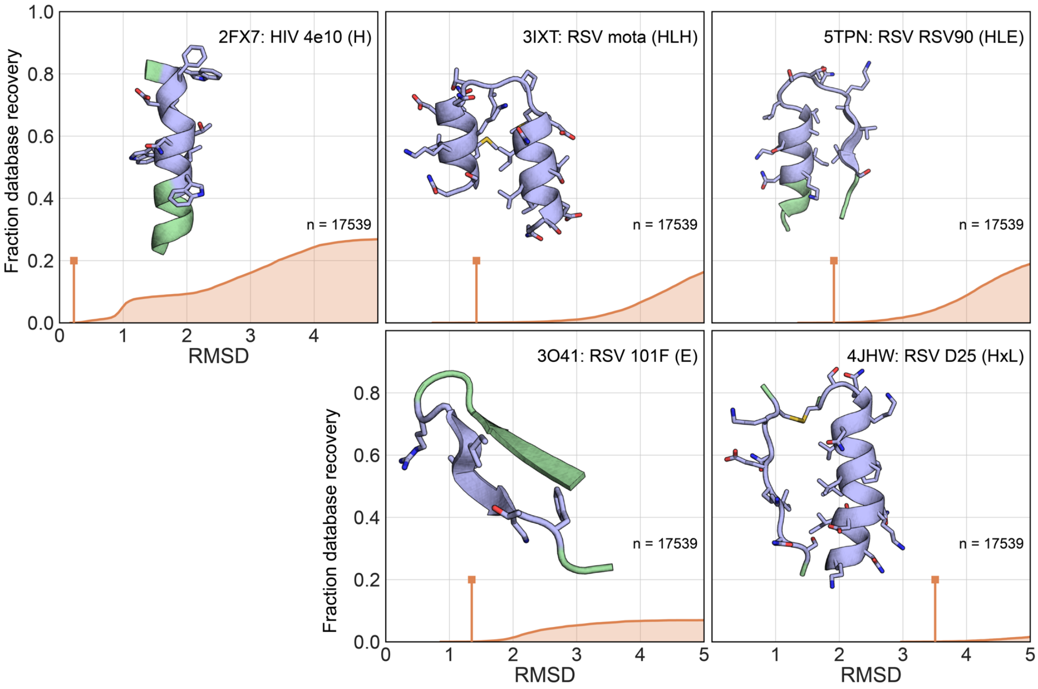


Figure 1.5: Available template structures for binding motifs of different structural complexity.
A MASTER search was performed over the nrPDB30 database (nonredundant subset of the PDB with a 30% sequence identity cutoff) containing a total of 17539 structures, querying the number of matches for different neutralization epitopes (colored in blue in each structure) of increasing structural complexity. Short stretches outside of the direct interface were included in the search (green). The fraction of the database recovered is plotted on the y-axis. Matches were filtered for protein size <180 residues and for matches that do not clash with the respective binder. The vertical line (orange) indicates the RMSD cutoff in Å for the first 10 scaffold identified. Secondary structure composition of the motifs is represented by: E - strand; L – Loop; H – helix; x – chain break.

Using the FunFolDes design strategy, Correia and colleagues designed an immunogen that accurately mimicked a neutralization epitope of the respiratory syncytial virus (RSV), a pathogen that currently lacks an effective vaccine (*59*). Strikingly, this computationally designed immunogen was shown to elicit RSV neutralizing antibodies in non-human primates, and thus served as a proof-of-principle for a new vaccine concept based on computationally designed proteins. Section 1.4 elaborates on the challenges of vaccine design, and points out how computationally designed immunogens, among other strategies, can aid the development of vaccines against pathogens for which more traditional vaccine development strategies have remained unsuccessful.

Beyond installing functional sites on pre-existing protein structures, recent studies have reported the stabilization of functional motifs in proteins built *de novo* (Approach 3, Fig 1.4). Although the exact protocols differ, the overall strategy involved building a *de novo* protein tailored to the functional motif or small molecule ligand. Collectively, we herein refer to this approach as motif-centric *de novo* building, as a protein topology is built around the functional motif, rather than adapting an already existing protein to the motif's structure (approaches 1 and 2). Successes using this approach included the design of a cofactor binding protein (*60*), and the design of an alpha helical barrel in a configuration that allowed for the installation of a functional catalytic triad (*61*). Notably, da Silva and colleagues have reported the design of an interleukin 2 (IL2) mimetic using this motif-centric design approach (*62*). The designed protein triggered IL2 receptor signaling, and had anti-tumor activity in mice with lower toxicity compared to the natural IL2 analog. Biophysically, the IL2 mimetic showed extreme thermostability (>120 °C), and structural studies have confirmed atomic level accuracy in mimicking the natural IL2 - IL2-receptor interface. This study serves as a powerful illustration of how *de novo* protein design can yield biologically active proteins with superior biophysical and functional properties.

Altogether, the transplantation of functional sites has allowed the design of functional proteins with controlled physicochemical and topological properties. Grafting remains the most conservative approach, as it is greatly dependent on a structurally similar design template, and allows only limited control over the topological context of the protein. Approach 2 is a much more generalistic approach, allowing to transplant motifs to virtually any target topology, irrespective of structural similarity. Lastly, approach 3 is the most flexible strategy, aiming to design functional proteins from scratch. It allows full control over the physicochemical and topological context, and thus designed proteins should be entirely free of any 'evolutionary baggage' that proteins designed using strategy 1, and to a lesser extent also using strategy 2, may carry.

In my thesis work, approaches 2 and 3 will be employed for the *de novo* of proteins carrying antibody binding sites ('epitopes') from viral proteins. Throughout chapter 2 and 3, the designed proteins will be tested as immunogens to elicit anti-viral antibody responses *in vivo*. In chapter 4, my work shows the potential of *de novo* designed proteins to serve as antibody biosensors. In the following two sections, the biological problem that the protein design efforts in my thesis work aim to address will be introduced.

1.4 Structure-based immunogen design strategies

As briefly introduced in the last section, recent advances in functional protein design have raised expectations for the design of novel immunogens as an alternative to current vaccine development strategies. The following sections will introduce the challenges faced by classical vaccine development strategies, and how structure-based immunogen design approaches may lead to the development of novel vaccines for pathogens that have eluded conventional vaccine development efforts. Different protein engineering strategies, some of which involve computational protein design, are described along with the immunological outcome that was achieved.

The content of this section has been published as review article in Curr Opin Struct Biol during 2018 (doi:10.1016/j.sbi.2018.06.002).

Authors and affiliations:

Fabian Sesterhenn[1,2], Jaume Bonet[1,2] & Bruno E. Correia[1,2].

[1]Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, CH-1015, Switzerland.
[2]Swiss Institute of Bioinformatics (SIB), Lausanne, CH-1015, Switzerland.

My contribution to this article:

I performed the literature research and wrote the manuscript, with input from all authors.

1.4.1 Abstract

Vaccines have been one of the most successful interventions in global health. However, traditional vaccine development has proven insufficient to deal with pathogens that elude the immune system through highly variable and non-functional epitopes. Emerging B cell technologies have yielded potent monoclonal antibodies targeting conserved epitopes, and their structural characterization has provided templates for rational immunogen design. Here, we review immunogen design strategies that leverage structural information to steer bulk immune responses towards the induction of precise antibody specificities targeting key antigenic sites. Immunogens designed to elicit well-defined antibody responses will become the basis of what we dubbed *precision vaccines*. Such immunogens have been used to tackle long-standing vaccine problems and have demonstrated their potential to seed the next generation of vaccines.

1.4.2 Introduction

Vaccination is likely the most efficacious prophylactic approach in modern medicine and has greatly reduced the burden of infectious diseases (*63*). Despite their numerous successes, classical vaccine strategies that rely on attenuated or inactivated pathogen formulations have failed to elicit neutralizing antibodies (nAbs) against a number of pathogens. Some notable examples are long-standing viral threats like human immunodeficiency virus 1 (HIV-1), respiratory syncytial virus (RSV), human metapneumovirus (hMPV) or dengue, as well as emerging pathogens such as zika, amongst others. Additionally, the constant antigenic drift of influenza requires formulation of seasonal vaccines, thus preventing the development of a universally protective vaccine (*64*).

Many of these pathogens have evolved strategies to evade targeted, neutralizing immune responses (*65*). The immunodominance of antigenic sites that do not confer broad and potent neutralization over those that can be targeted by potent nAbs is a poorly understood phenomenon; However, it is well established that non-neutralizing antibodies can facilitate virus entry into host cells between different serotypes of certain pathogens (e. g. dengue), thereby causing antibody-dependent disease enhancement (ADE) (*66, 67*).

To overcome current limitations in vaccine development, a rational vaccine strategy known as reverse vaccinology has been proposed (*68*). The aim of the reverse vaccinology concept is to focus the immune response onto epitopes where the pathogen is vulnerable to antibody-mediated neutralization. In essence, this strategy relies on the isolation of nAbs from human or animal repertoires, followed by the structural characterization of the nAb-antigen complex, and finally exploit the acquired atomic-level information to design novel immunogens. A key challenge for next-generation vaccines, then, is to place neutralizing epitopes in the immune system's spotlight for efficient recognition and enhanced epitope-specific antibody elicitation. We refer to such class of vaccines as *precision vaccines*, given their extremely well-defined epitope-directed antibody response. An essential requisite for precision vaccines is to encode the structural information of epitopes targeted by broad and potent nAbs in the designed immunogens.

Here, we review four structure-based immunogen design approaches that aim to elicit focused antibody responses (Fig 1.6): I) silencing non-neutralizing epitopes; II) conformational stabilization; III) germline targeting; IV) epitope scaffolding. A special emphasis is placed on linking the design strategy to the immunological outcomes for different pathogens.

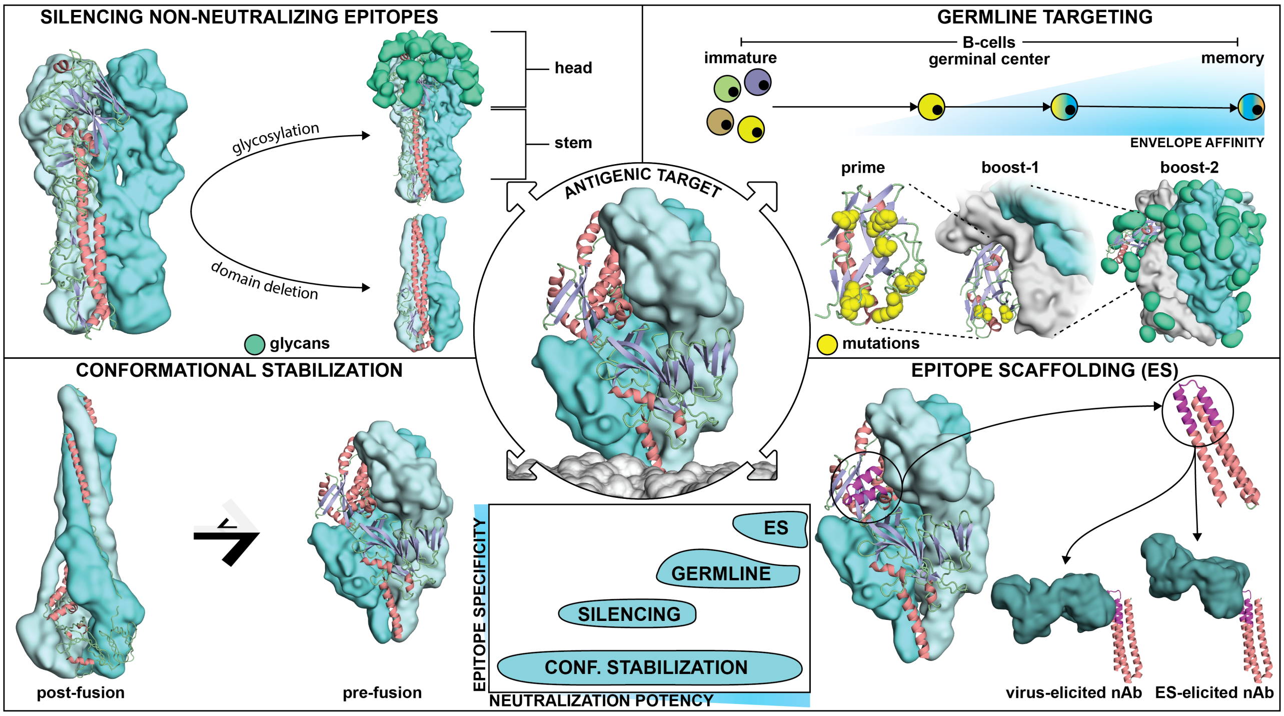

SILENCING NON-NEUTRALIZING EPITOPES
head
stem
glycosylation
domain deletion
glycans
GERMLINE TARGETING
B-cells
immature
germinal center
memory
ENVELOPE AFFINITY
prime
boost-1
boost-2
mutations
ANTIGENIC TARGET
CONFORMATIONAL STABILIZATION
post-fusion
pre-fusion
EPITOPE SCAFFOLDING (ES)
virus-elicited nAb
ES-elicited nAb
EPITOPE SPECIFICITY
NEUTRALIZATION POTENCY
ES
GERMLINE
SILENCING
CONF. STABILIZATION

Figure 1.6: Structure-based approaches for immunogen design.
Four main strategies have been used to design improved immunogens for the elicitation of nAbs that target precise conformational states, domains, epitopes or antibody lineages. **Center:** All these strategies were employed to stabilize viral fusion proteins or to enhance antibody responses towards neutralizing epitopes located in viral fusion proteins. **Center-bottom:** The structure-based approaches can be classified according to expected epitope specificity and neutralization potency of the elicited antibody responses. **Upper left:** Silencing non-neutralizing epitopes. Glycosylation or deletion of influenza's hemagglutinin head domain enables the focusing of antibody responses towards more conserved epitopes targeted by bnAbs. **Lower left:** Antigenic conformational stabilization. Respiratory syncytial virus fusion protein (RSVF) adopts two distinct conformations, a metastable prefusion state (right), and a stable postfusion (left) state. Through stabilization of the prefusion conformation, the elicitation of nAbs was greatly improved. **Upper Right:** Germline targeting. A domain of the HIV-1 envelope protein (Env) is modified to trigger the activation of inferred germline precursors of HIV-1 bnAbs. After priming, B-cell maturation is guided through consecutive boostings with increasingly native Env versions to select for antibodies that can bind the epitope in its native context. **Lower Right:** Epitope Scaffolding. A structurally characterized neutralization epitope is transplanted onto a heterologous, synthetic protein scaffold, which is used to elicit virus-reactive antibodies targeting the desired epitope.

1.4.3 Silencing non-neutralizing epitopes

The most intuitive strategy to focus antibody responses on immunologically subdominant epitopes is to remove or mask non-neutralizing epitopes. In this section, we describe two approaches to favor the immune recognition of conserved epitopes: I) excise antigen domains mainly targeted by non-neutralizing antibodies; II) reduce the accessibility of non-neutralizing epitopes through glycan masking.

On the influenza virus, the hemagglutinin (HA) surface protein is the main target of humoral responses, and is one of the most notorious examples where the antibody responses mostly target a highly sequence variable region (head domain) and yield mostly strain-specific responses (*69*). In contrast, several epitopes in the HA stem region are targeted by broadly neutralizing antibodies (bnAbs) (*70, 71*), providing a strong rationale for vaccines that aim to induce antibody responses towards the HA stem. Two independent studies reported the design of immunogens comprising only the HA-stem domain (*72, 73*). Several rounds of structure-based design were performed, including the removal of the entire head domain followed by the introduction of stabilizing core mutations and linker design. Upon vaccination with stem-derived antigens, mice, ferrets and cynomolgus monkeys mounted a cross-reactive antibody response and survived a challenge with a lethal dose of a highly pathogenic viral isolate.

In contrast to HA, the major target of respiratory syncytial virus (RSV) nAbs is the head domain of the fusion protein (RSVF) (*74*). Boyington et al. designed truncated versions of RSVF comprising only the head region. This head-only immunogen elicited comparable neutralization titers to those of the full length, prefusion RSVF (see the next section). However, mice primed with RSVF and boosted with a head-only immunogen showed enhanced antibody titers towards neutralization epitopes located on the apex of the head domain, indicating the value of heterologous prime-boost immunizations to focus antibody responses (*75*).

A structurally less aggressive approach to mask dominant epitopes is to hyper-glycosylate non-functional epitopes or domains. This strategy has been used by Eggink et al. to silence immunodominant epitopes in HA, generating hyper-glycosylated HA head domains (HGHD). In mice, the HGHDs yielded a heightened antibody response towards the subdominant stem domain and showed improved protection upon viral challenge (*76*).

Masking of non-neutralizing epitopes through glycosylation has also led to novel HIV-1 envelope (Env) immunogens with reduced antigenicity to non-neutralizing epitopes. Two main targets of non-nAbs are the flexible Env V3 loop, as well as the bottom of recombinant Env, which is inaccessible on membrane-bound Env spikes. The introduction of two glycans on V3 was shown to dampen the immune response against this site upon rabbit immunization (*77*). Similarly, glycosylation of the Env base showed reduced reactivity with sera of animals immunized with non-masked Env, indicating that glycan masking can render non-neutralizing epitopes inaccessible (*78*).

Overall, the domain deletion and masking of epitopes induces measurable immunological outcomes; Nevertheless, they don’t seem to radically transform the immune responses obtained. Perhaps most promising are the HA headless immunogens which elicited antibodies with increased reactivity across heterosubtypic influenza strains.

1.4.4 Antigenic conformational stabilization

Viruses are dependent on entering host cells for replication, a process mediated by viral surface proteins. Generally, these proteins tend to adopt a metastable prefusion state on the virion’s surface (e.g HIV, RSV, hMPV and others), undergoing large structural rearrangements to a stable postfusion conformation upon viral and host cell membrane fusion (*79*). Serum studies by Magro et al. have revealed that RSV nAbs mostly target the prefusion state of RSVF (*80*). This fueled the idea of promoting the elicitation of antibodies directed against the prefusion conformation rather than postfusion. Table 1.1 summarizes the design strategies used to accomplish the antigenic conformational stabilization.

The determination of the prefusion conformation of RSVF (*81*) provided a template for designing prefusion-stabilized versions of RSVF, achieved by introducing intra-protomer disulfide bonds and cavity-filling hydrophobic mutations (Ds-Cav1) (*82*). Similarly, Krarup et al. reported a stabilization strategy for RSVF, by introducing proline residues that prevent structural rearrangements required to adopt the postfusion conformation (*83*). In immunization studies, prefusion RSVF was eight times more potent in terms of neutralization than postfusion RSVF, as observed in mice, non-human primates and cotton rats. A large fraction of the elicited antibodies was directed against the apex of RSVF, a site that undergoes profound conformational changes in the postfusion state (*82*). In a follow up article, Joyce and colleagues further stabilized Ds-Cav1, resulting in an immunogen with improved physical stability in the prefusion state (RSVF-DS2). In mice, DS2 variants elicited up to four-fold more potent neutralizing responses as compared to Ds-Cav1, indicating that conformational stability correlated with immunogenicity (*84*). As a more relevant model for human RSV disease, a bovine RSVF-DS2 version elicited bovine RSV nAbs >100-fold higher than its postfusion counterpart in calves, and conferred full protection of upper and lower respiratory tract upon viral challenge (*85*).

Design strategy	**Structural stabilization effect**	**Examples**
Cavity filling mutations	Stabilization through improved hydrophobic core packing	RSVF DSCav1 (*82*) HIV Env (*86*)
Disulfides	Covalent linkage of residues/domains/protomers that are distant in sequence Increase thermostability	RSVF DsCav1 (*82*) HIV Env (*86*)
Substitution by prolines	Improve trimerization Disturb helical structure formation to prevent structural rearrangement Increase expression yield	HIV gp41 (*87*) RSVF (*83*)
Fusion of trimerization domains	Favor trimerization	T4 fibritin foldon (*76, 82*) GCN4 leucine zipper (*88*)
Structural deletions	Increase solubility through deletion of hydrophobic stretches Recombinant expression of cleavage independent antigens	HIV Env SOSIP.664 (*89*) HIV Env gp140 (*90*) RSVF DS2 (*84*)

Table 1.1: General design strategies for antigenic conformational stabilization of viral fusion proteins.

The structural similarity between RSV and MPV allowed similar design strategies for the stabilization of the MPVF prefusion state. Interestingly, immunization studies in mice have not yielded higher nAb responses than those elicited by the postfusion conformation, showing that although design strategies may be transferable to other pathogens' fusion proteins, the immunological outcome can differ (*91*).

The largest class I fusion proteins are found in human coronaviruses, and known as spike proteins (S). The structure of the S protein of a human coronavirus HKU1 has been solved (*92*), serving as template to stabilize a MERS-CoV S protein in its prefusion conformation. Immunogenicity studies in mice revealed that prefusion MERS-CoV S elicited nAbs with increased breadth and potency compared to postfusion S, showing that for this particular protein the antigenic conformational stabilization strategy yields a superior immunological outcome (*93*).

Using analogous design strategies, extensive efforts have been made to stabilize the HIV Env glycoprotein in a closed prefusion conformation. A particular challenge for Env engineering is to maintain its native conformation upon recombinant expression, and reduce the conformational "breathing" between the open and closed conformations that have been reported for the prefusion Env. The most widely used construct for native-like Env design is the SOSIP.664 trimer, which is stabilized by an intermolecular disulfide bond, an isoleucine-proline substitution within the fusion peptide and a truncation to remove the hydrophobic membrane proximal region. While a comprehensive analysis of the efforts to stabilize Env is beyond the scope of this article and have been the subject of several reviews (*94, 95*), we highlight two recent studies that employed rational stabilization strategies to suppress non-neutralizing antibody responses.

The Env V3 loop is a major target of non-neutralizing Abs, and is exposed in the Env open conformation. To reduce the V3 conformational dynamics, structure-based design was employed to strengthen hydrophobic packing, preventing V3 loop to adopt the open conformation (*96, 97*). Similarly, Kulp et al. employed computational design to replace a network of buried hydrophilic residues by hydrophobic amino acids, thereby limiting V3 exposure (*78*). Both stabilization strategies reduced reactivity to V3 directed antibodies, and immunogenicity studies in mice, rabbits and non-human primates confirmed dampened antibody responses against the V3 loop.

Another main target of non-neutralizing antibodies within Env are the CD4-induced epitopes. To prevent the CD4-induced conformational change while maintaining an epitope targeted by bnAbs, de Taeye et al. introduced two mutations that occur in HIV-1 strains unable to undergo CD4-induced conformational changes (*97*). Using a computational multi-state design protocol, Kulp et al. rationally designed mutations that abrogate CD4 binding to Env, while maintaining binding to bnAbs targeting the CD4 receptor binding site (CD4bs) (*78*). While these engineered Envs showed reduced binding to antibodies targeting CD4-induced epitopes, neither study could prove a direct impact of such modifications on antibody specificity or neutralization *in vivo*.

In summary, the antigenic conformational stabilization is clearly one of the most promising strategies for immunogen engineering, having shown the ability to dramatically transform the immunological outcome by presenting the most relevant antigenic conformation for the elicitation of functional antibodies. In this strategy, the precision character arises from the conformational specificity rather than epitope specificity which will be much more dominant in the following design strategies.

1.4.5 Germline targeting

Typical bnAbs against HIV-1 carry a large number of somatic hypermutations (*98-100*). Based on this observation, germline targeting has emerged as a novel strategy to prime selected antibody lineages. Germline targeting aims to engage the unmutated precursors of the bnAbs, and drive their maturation towards a bnAb by gradually acquiring the necessary somatic mutations to broadly neutralize HIV-1.

Using computational design and *in vitro* evolution, Jardine et al. developed an engineered outer domain (eOD) of the viral gp120 protein, and introduced mutations that enable binding of inferred germline precursors of the bnAb VRC01 (glVRC01) (*98, 101, 102*). Both in a glVRC01 heavy chain knock-in mouse model and in mice transgenic for human immunoglobulin loci, it was shown that the germline targeting immunogen (eOD-GT8) primed antibodies with characteristic features of VRC01-like antibodies (*103-105*). Primed B cells were shown to be recalled upon boosting immunizations with gradually more native Env versions, and somatic mutations were driven towards those found in VRC01-class bnAbs (*106*). Similarly, Medina-Ramirez et al. have engineered a SOSIP.664 trimer to bind germline precursors of CD4bs antibodies, and this engineered native-like trimer was shown to activate glVRC01 antibodies *in vitro* and *in vivo* (*107*).

In order to target another bnAb class binding to a different epitope on Env (PGT121-like bnAbs), Steichen et al. employed mammalian cell surface display to engineer germline targeting Env trimers, which were then shown to activate B cells carrying a PGT121 germline receptor *in vitro* and to prime PGT121-like antibody responses in knock-in mice (*108*).

Together, these studies established a first milestone in using structure-based design together with a stepwise vaccination protocol that targets and expands specific germline precursors. Albeit an important contribution, it remains to be shown what outcome the germline targeting strategy will have in terms of eliciting bnAbs in relevant animal models. Unlike the other described approaches, the precision aspect of this strategy is related to the specific antibody lineages that are deemed necessary to achieve broad HIV neutralizing activity.

1.4.6 Epitope scaffolding

The presentation of single epitopes in the context of scaffold proteins has gained traction in the last years with the aim of eliciting epitope-specific antibodies. Essentially, this approach uses structural information of the exact epitope conformation recognized by a nAb to design heterologous proteins that mimic the structure of the epitope and are structurally compatible with the antibody binding mode.

The first reports of epitope-scaffolds were presented in the context of HIV using as targets the neutralization epitopes 4E10 (*55*) and 2F5 (*56, 109*). In these two studies, the epitopes were grafted onto heterologous scaffolds using computational design approaches (*54*), and biochemical and structural characterization confirmed accurate epitope mimicry. Immunologically, the most remarkable result was that the epitope-scaffolds were able to induce antibodies with fine specificities similar to those of the parent antibodies (*55, 56, 109*), showing that the epitope-scaffolds achieved *in vivo* the immunological outcome for which they were primarily designed, although no serum neutralization was obtained.

A later effort by McLellan and colleagues applied the epitope scaffolding strategy to design immunogens that presented the antigenic site II of RSVF, the target of the FDA-approved monoclonal antibody Palivizumab. However, only the side chains of the epitope were grafted onto a scaffold with a similar backbone conformation as the epitope, yielding a scaffold with imperfect epitope mimicry. Much like the HIV epitope-scaffolds, in mouse immunizations, epitope-specific antibodies were elicited but no serum neutralization was achieved (*110*).

To improve the structural mimicry of site II in a synthetic scaffold, Correia et al. developed a new protein design algorithm, where protein folding and sequence design simulations were coupled to generate immunogens with the epitope structure stabilized in the exact native conformation. These designs showed an extremely high affinity to site II antibodies, and structural characterization revealed a perfect epitope mimicry. In mouse immunogenicity studies, the second-generation epitope-scaffold still did not elicit viral neutralization titers, nevertheless when used in non-human primates (NHP), low levels of neutralization were detected. Monoclonal antibodies isolated from the NHPs were site II-specific, and bound with high affinity to RSVF. Structural studies confirmed that the elicited antibodies recognized antigenic site II similarly to Palivizumab, and most importantly, neutralized RSV with superior potency compared to Palivizumab (*59*).

In summary, the epitope scaffolding approach is an efficient way of eliciting epitope-specific antibodies. Nevertheless, these are simplified immunogens that mostly encode the epitope binding motif surface and lack important molecular features regarding the tertiary and quaternary environments of the native viral proteins. This class of immunogens faces important challenges; Though their ability to induce very potent and epitope-specific antibodies has been shown, it remains to be seen how the fraction of functional antibodies within the overall repertoire can be increased.

1.4.7 Conclusions/Outlook

The field of vaccinology is facing incredible challenges to develop efficacious vaccines against sophisticated pathogens capable of evading immune responses in many different ways. Due to those escape mechanisms, mounted antibody responses against some of those pathogens have limited breadth and do not afford protection against antigenically drifted strains. Additionally, the induced neutralization titers are often low and decay over time allowing reinfection. The reverse vaccinology strategy was envisioned as an integrated approach to design immunogens that elicit or boost preexisting antibody responses focused on *bona fide* neutralization epitopes. In the age of systems approaches and big-data science, our understanding of complex biological phenomena, such as an immune response, has greatly expanded. Therefore, we are now able to define with exquisite accuracy the molecular determinants desirable for a particular vaccine development endeavor. Those determinants include but are not limited to antigen conformations, desired antibody lineages and key neutralization epitopes, which lead us to conclude that we have entered into a new age of precision vaccines. Structure-based approaches, as reviewed here, are essential to define the target problem and realize the full potential of precision vaccines. Despite the successes so far, none of the strategies described here have proven to be the "one-fits-all" solution (Table 1.2). While a prefusion stabilized RSVF was a far superior immunogen than the postfusion counterpart, the same approach for MPVF did not enhance the neutralizing responses. Similarly, the sole conformational stabilization of HIV Env is deemed unlikely to elicit bnAbs (*111*) and germline targeting approaches hold great promise to prime a bnAb response. Finally, while epitope-scaffolds elicited very potent and specific monoclonal antibodies in NHPs, the average neutralization titers were rather low. The strengths and weaknesses of each one of the presented strategies suggests a complex balance between the approach, the target pathogen and the expected immune response. Ultimately, a combination of the different strategies may be the best course of action to overcome their specific limitations and bring forth a new generation of rationally designed precision vaccines.

Approach	**Immunological outcome**	**Examples**
Silencing non-neutralizing epitopes	Immune response focused on HA stem region, conferring heterosubtypic protection upon challenge Boosting of RSVF-head directed antibodies	Influenza (*72, 73*) RSV (*75*)
Conformational stabilization	Increased serum neutralization Antibody responses focused on sites that are inaccessible in alternative conformation	RSV (*84*) HIV (*78, 97*)
Germline targeting	Activation of germline precursors in humanized mice Sequential boosting protocol selected productive somatic mutations	HIV (*104, 106*)
Epitope scaffolding	Elicitation of epitope specific antibodies Successful elicitation of RSV antibody with superior neutralization potency compared to clinically approved mAb	HIV (*55*) RSV (*59*)

Table 1.2: Overview of immunogen engineering approaches and their immunological outcomes.

1.5 Respiratory Syncytial Virus

As described during the previous section, RSV is one of the most important pathogens that has resisted vaccine development. As such, it has emerged as a model pathogen to develop and test structure-based immunogen design approaches. Multiple proof-of-concept studies were performed using RSV as a model pathogen (*59, 82*), and have subsequently catalysed the development of novel immunogens for other pathogens (*91, 112, 113*). On the basis of these studies, and due to the extensive structural knowledge available, RSV serves as a model for the development of an epitope-focused vaccination strategy using computationally designed immunogens throughout this thesis.

RSV is an enveloped negative-strand RNA virus, belonging to the family of *Pneumoviridae* (*114*). It is the cause of one of the most prevalent viral infections of childhood, and nearly every child will encounter at least one infection during the first two years of life (*115*). While most infections only cause cold-like symptoms, spreading of the infection to the lower respiratory tract can result in severe pneumonia. RSV infection is the leading cause of hospitalization of young infants (3.4 million annually), responsible for up to 200,000 deaths per year (*115*). Besides young infants, RSV also causes significant morbidity in the elderly and in immunocompromised patients (*116*).

To date, there is no licensed RSV vaccine. After its discovery in 1957, a vaccine trial was initiated in 1966 using a formalin-inactivated formulation of RSV isolates. Unfortunately, the trial had a disastrous outcome with 80% of the vaccinated group being hospitalized with severe adverse events and two infants dying due to disease-enhancing effects (*117*). Among other important factors (*118*), it is believed that the formalin-inactivated formulation induced non-neutralizing antibodies directed against non-protective epitopes, thus leading to an antibody-dependent disease enhancement (ADE) (*119, 120*). Briefly, the mechanism of ADE is that non-neutralizing antibodies enhance virus entry into monocytes or macrophages through interaction with Fc or complement receptors, and this phenomenon has also been observed in the context of other viral infections (*121*). While the precise contribution of ADE to the severe RSV disease has remained unclear, there is a clear need for the development of novel vaccination approaches that induce precisely controlled, protective antibody responses.

As described in the previous section, a detailed understanding of the antigen and its epitopes that are targeted by nAbs can inform the design of novel vaccine candidates. The main target of RSV nAbs is the RSV fusion protein (RSVF) (*74*). RSVF is synthesized as an F0 precursor protein followed by proteolytic processing through furin, resulting in an F1 and F2 subunit. The fusion protein forms a trimeric spike on the viral surface consisting of three protomers, with each protomer being a heterodimer of an F1 and F2 subunit. The resulting prefusion state is metastable, and RSVF spontaneously rearranges into the stable postfusion state after host cell attachment, thereby mediating the virus-host cell fusion (*122*).

There are six major antigenic sites that have been identified on the RSV fusion protein, known as antigenic sites 0, I, II, III, IV, and V (Fig 1.7). Multiple antibodies have been isolated for each antigenic site from either animal or human repertoires and structural information is available for at least one representative per site. It is important to note that antigenic sites 0, III and V are only present in the prefusion conformation of RSVF ('prefusion-specific epitopes'). Furthermore, antibodies targeting sites II and IV are generally conformation-independent, and antibodies targeting site I are specific for the postfusion conformation (*123*).

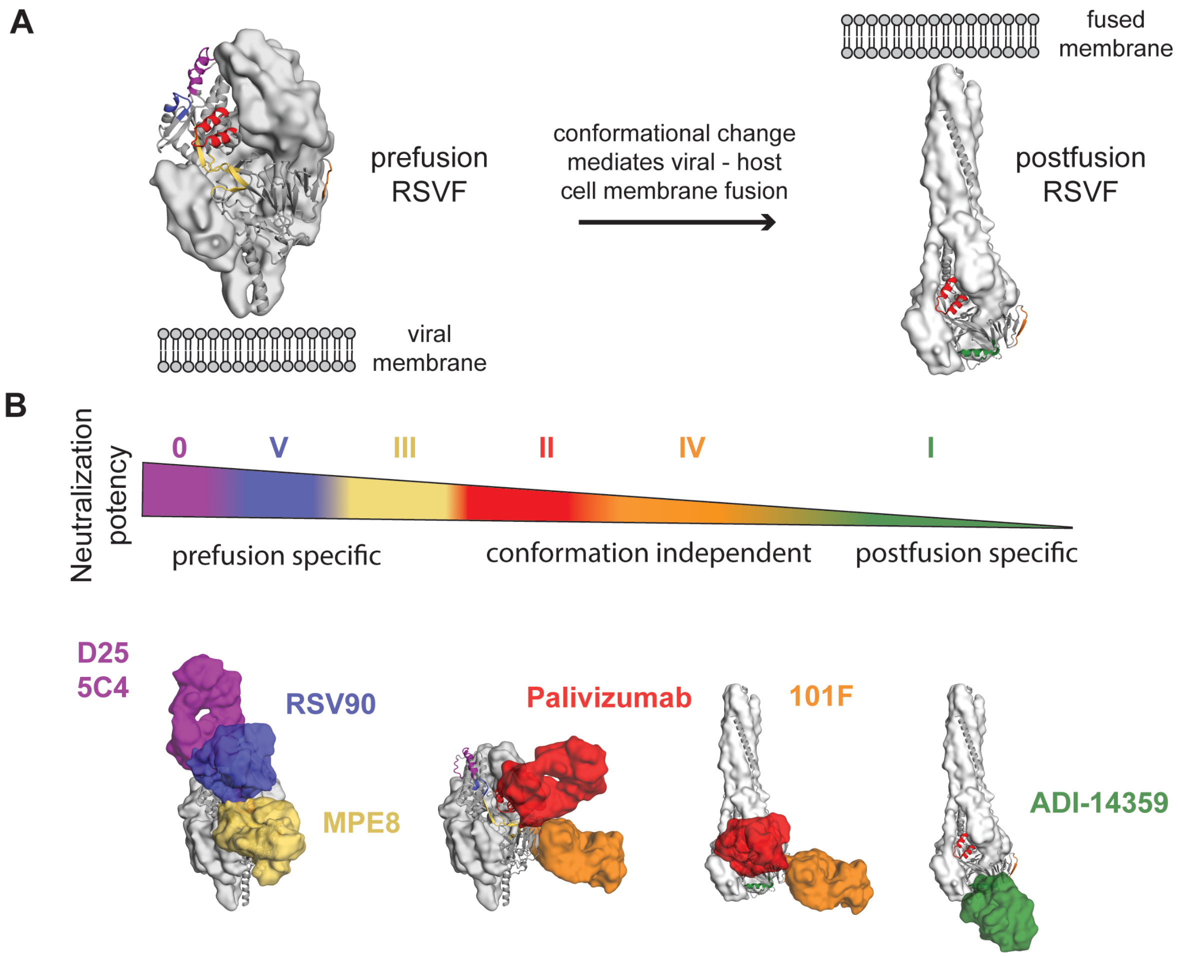

A
fused membrane
prefusion RSVF
conformational change mediates viral - host cell membrane fusion
postfusion RSVF
viral membrane
B
Neutralization potency
0
V
III
II
IV
I
prefusion specific
conformation independent
postfusion specific
D25
5C4
RSV90
MPE8
Palivizumab
101F
ADI-14359

Figure 1.7: Neutralization epitopes on pre- and postfusion RSVF.
(A) RSVF forms a diamond-like shape in its metastable prefusion conformation (left) that undergoes a drastic conformational change to a cone-like shape in its postfusion conformation (right), mediating the viral-host cell membrane fusion. The neutralization epitopes, named antigenic sites 0-V, are colored in both conformational states, if present. **(B)** Qualitative ranking of the antigenic sites with respect to the neutralization potency of their target antibodies (left to right, from high to low neutralization potency). Prefusion-specific antibodies, in particular the D25 and 5C4 antibodies that target antigenic site 0, are the most potent nAbs identified to date. Antigenic sites II and IV maintain their epitope structure both in prefusion and postfusion conformation, and are the target of Palivizumab and 101F, respectively. Postfusion-specific antibodies, such as ADI-14359, are generally non-neutralizing. Figure inspired by Graham et al., 2019 (*123*).

Multiple studies have shown that the most potent RSV nAbs target the prefusion conformation of RSVF (*1, 2, 74*). One of the most potent prefusion-specific nAbs identified is the D25 antibody, which targets antigenic site 0. Its potency stems from the fact that it locks RSVF in the prefusion conformation, thus preventing the conformational change necessary for host cell infection (*1*). As a general rule, antibodies that are prefusion-specific are often exceptionally potent for RSV neutralization, followed by antibodies that target both the pre- and postfusion conformation (Fig 1.7) (*2*). Palivizumab, which targets antigenic site II, has been one of the earliest RSV nAbs discovered, and is successfully used in the clinic to prevent RSV infection in high-risk infants (*124*). In contrast, antibodies specific for the postfusion conformation, such as the site I specific antibody ADI-14359, are generally non-neutralizing, and a role for such antibodies to mediate disease enhancement has been discussed (*120, 125*). Consequently, the primary target epitopes for RSV vaccine development are those that are present in the prefusion conformation, exclusively or non-exclusively, whereas the elicitation of postfusion-specific antibodies should be avoided.

Altogether, RSV has been particularly well studied structurally, and detailed human antibody profiling has provided a high-resolution antigenic map to inform vaccine design. This makes RSV an excellent model pathogen to test immunogen design strategies, for which the challenge is to elicit antibodies targeting defined neutralization epitopes. The following two chapters present how computationally designed, epitope-focused immunogens can be used to elicit antibody responses targeting precisely defined epitopes, providing fundamental steps towards an epitope-centric vaccination approach for RSV, and potentially other pathogens. Lastly, we discuss how the computational design methods developed for the design of RSV immunogens can be of use beyond immunogen design, opening the door for a general computational design pipeline for the *de novo* design of functional proteins.

Objectives

My thesis work originates at the interface of computational protein design and vaccine design. It aims to fulfil the following four main objectives:

Aim I - *De novo* design of protein scaffolds for structurally complex neutralization epitopes.

One of the biggest challenges for functional *de novo* protein design is to handle structurally complex (irregular, discontinuous) functional motifs. Towards this aim, chapter 3 showcases two design strategies for the *de novo* design of epitope-scaffolds that accurately present binding motifs of thus far unprecedented structural complexity. Within the scope of aim I, chapters 3 and 4 present an integrated pipeline of computational design, high-throughput experimental screening and next-generation sequencing, revealing how subtle topological differences and sequence features impact stability and function of computationally designed proteins.

Aim II - Reshaping pre-existing immunity through computationally designed epitope-focused immunogens.

Pre-existing immunity poses a major hurdle for vaccine development, mainly attributed to established immunodominance hierarchies where antibodies targeting *bona fide* neutralization epitopes are subdominant, i.e. below a functional threshold for broad and potent neutralization. Chapter 2 and 3 reveal the previously unknown potential of epitope-scaffolds as boosting immunogens to reshape pre-existing immunity, and boost functional antibodies targeting key RSV neutralization epitopes. Chapter 2 presents a proof-of-concept study toward this aim, using an optimized formulation of a previously developed site II immunogen. Chapter 3 presents additional work towards this aim, demonstrating how a cocktail formulation of multiple epitope-focused immunogens can focus immune responses in a prime-boost immunization study in non-human primates.

Aim III - Multi-site immunofocusing through cocktails of designed immunogens.

Viral pathogens have evolved strategies to obscure conserved, potent neutralization epitopes. Thus, extensive effort has been undertaken by the scientific community to develop immunogens that spotlight neutralization epitopes. With respect to this aim, we seek to advance an epitope-scaffold based vaccination strategy through the design of a multi-site cocktail vaccine. Results of this effort are presented in chapter 3, showing that a trivalent cocktail vaccine consistently induces neutralizing antibodies in mice and non-human primates directed against three non-overlapping neutralization epitopes.

Aim IV - Beyond immunogen design – a general bottom-up *de novo* protein design approach for the design of novel functional proteins.

Chapter 4 presents ongoing work towards aim IV, in which we generalize the design methodology introduced in chapter 3. We refer to this as a ‘bottom-up’ *de novo* protein design approach, allowing to build diverse protein folds from scratch to accommodate distinct functional binding motifs. On the functional side, we introduce a biosensor based on *de novo* designed proteins for the detection and quantification of epitope-specific antibody responses, potentially becoming a versatile tool to monitor RSV neutralizing antibody responses directly in human sera.

Altogether, the broad vision behind the work of my thesis is to leverage *de novo* protein design for the design of novel functional proteins, aiming to address biomedical challenges that have thus far been out of reach of natural proteins.

Chapter 2 Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen

This chapter is based on an article pulished in Plos Biology in 2019 (10.1371/journal.pbio.3000164).

Authors and affiliations:

Fabian Sesterhenn[1,2], Marie Galloux[3], Sabrina S. Vollers[1], Lucia Csepregi[4], Che Yang[1,2], Delphyne Descamps[3], Jaume Bonet[1,2], Simon Friedensohn[4], Pablo Gainza[1,2], Patricia Corthésy[1], Man Chen[5], Stéphane Rosset[1], Marie-Anne Rameix-Welti[6,7], Jean-François Éléouët[3], Sai T. Reddy[4], Barney S. Graham[5], Sabine Riffault[3] & Bruno E. Correia[1,2,*].

[1] Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland. [2] Swiss Institute of Bioinformatics (SIB), Lausanne CH-1015, Switzerland. [3] Unité de Virologie et Immunologie Moléculaires (UR892), INRA, Université Paris-Saclay, 78352, Jouy-en-Josas, France. [4] Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. [5] Viral Pathogenesis Laboratory, Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States. [6] UMR1173, INSERM, Université de Versailles St. Quentin, 78180 Montigny le Bretonneux, France. [7] AP-HP, Laboratoire de Microbiologie, Hôpital Ambroise Paré, 92104, Boulogne-Billancourt, France. * Corresponding author: bruno.correia@epfl.ch

Author contributions:

F.S. and B.E.C. conceived the work and designed the experiments. M.G. expressed and purified RSVN and NRM fusion protein. F.S., C.Y., S.S.V and P.C.H performed ELISAs. F.S. designed and performed SPR competition assay. F.S., C.Y and P.C.H performed ELISpot experiments, S.S.V, P.C.H and S.R performed mouse immunizations. L.C performed next generation sequencing analysis. F.S., J.B. and P.G. designed FFLM. F.S., P.C.H and M.C. performed RSV neutralization assay. F.S. and B.E.C wrote the paper. All authors commented on the manuscript.

2.1 Abstract

Throughout the last decades, vaccination has been key to prevent and eradicate infectious diseases. However, many pathogens (e.g. respiratory syncytial virus (RSV), influenza, dengue and others) have resisted vaccine development efforts, largely due to the failure to induce potent antibody responses targeting conserved epitopes. Deep profiling of human B-cells often reveals potent neutralizing antibodies that emerge from natural infection, but these specificities are generally subdominant (i.e., are present in low titers). A major challenge for next-generation vaccines is to overcome established immunodominance hierarchies and focus antibody responses on crucial neutralization epitopes. Here, we show that a computationally designed epitope-focused immunogen presenting a single RSV neutralization epitope elicits superior epitope-specific responses compared to the viral fusion protein. In addition, the epitope-focused immunogen efficiently boosts antibodies targeting the Palivizumab epitope, resulting in enhanced neutralization. Overall, we show that epitope-focused immunogens can boost subdominant neutralizing antibody responses *in vivo* and reshape established antibody hierarchies.

2.2 Introduction

The development of vaccines has proven to be one of the most successful medical interventions to reduce the burden of infectious diseases (*126*), and their correlate of protection is the induction of neutralizing antibodies (nAbs) that block infection (*127*).

In recent years, advances in high-throughput B-cell technologies have revealed a plethora of potent nAbs for different pathogens which have resisted the traditional means of vaccine development for several decades, including HIV-1 (*128*), influenza (*129*), respiratory syncytial virus (RSV) (*2, 130*), zika (*131, 132*), dengue (*133*) and others (*134-136*). A major target of these nAb responses is the pathogen's fusion protein, which drives the viral and host cell membrane fusion while undergoing a conformational rearrangement from a prefusion to a postfusion state (*79*). Many of these nAbs have been structurally characterized in complex with their target, unveiling the atomic details of neutralization epitopes (*1, 131, 137*). Together, these studies have provided comprehensive antigenic maps of the viral fusion proteins which delineate epitopes susceptible to antibody-mediated neutralization and provide a roadmap for rational and structure-based vaccine design approaches.

The conceptual framework to leverage nAb-defined epitopes for vaccine development is commonly referred to as reverse vaccinology (*68, 138, 139*). Although reverse vaccinology-inspired approaches have yielded a number of exciting advances in the last decade, the design of immunogens that elicit such focused antibody responses remains challenging. Successful examples of structure-based immunogen design approaches include conformational stabilization of RSVF in its prefusion state, which induces superior serum neutralization titers when compared to immunization with RSVF in the postfusion conformation (*82*). In the case of influenza, several epitopes targeted by broadly neutralizing antibodies (bnAbs) were identified within the hemagglutinin (HA) stem domain, and an HA stem-only immunogen elicited a broader nAb response than full-length HA (*72, 73*). Commonly, these approaches have aimed to focus antibody responses on specific conformations or subdomains of viral proteins. In a more aggressive approach, Correia et al. (*59*) computationally designed a synthetic immunogen presenting the RSV antigenic site II, and provided a proof-of-principle for the induction of site-specific, RSV nAbs using a synthetic immunogen.

The absence of a potent and long-lasting immune response upon natural infection is a major challenge associated with RSV, influenza virus and other pathogens. While a single exposure to pathogens like poliovirus confers life-long immunity, RSV, influenza and other pathogens have developed mechanisms to subvert the development of a durable and potent n response, thereby allowing such pathogens to infect humans repeatedly throughout their lives (*65*). One of the major factors hindering the induction of long-lasting protection after the first infection is related to the antibody specificities induced. Upon exposure to a pathogen, such as influenza, the human antibody responses predominantly target strain-specific antigenic sites, whereas potent bnAbs are subdominant (*140*). This phenomenon is generally referred to as B-cell immunodominance, which describes the unbalanced immunogenicity of certain antigenic sites within an antigen, favoring strain-specific, variable, non-neutralizing epitopes to the detriment of conserved, neutralization-sensitive epitopes (*69*). The factors that determine the antigenicity of specific epitopes remain unclear, making the categorization of immunodominant and subdominant epitopes an empirical classification based on serological analysis. Importantly, the presence of high levels of antibodies directed against immunodominant epitopes can sterically mask surrounding subdominant epitopes that may be targeted by bnAbs, preventing the immune system from mounting productive antibody responses against subdominant epitopes, and potentially limiting vaccination efficacy (*69, 140-142*).

The immunodominance hierarchy is established within the germinal center, where B-cells undergo a binding affinity-based competition for available antigen and subsequently initiate a clonal expansion stage, ultimately becoming long-lived plasma cells or memory B-cells (*143*). Controlling this competition and driving antibody responses towards the increased recognition of subdominant, neutralizing epitopes is of primary importance to enable development of novel vaccines against pathogens which have resisted traditional strategies. One of the few strategies to guide antibody maturation was tested in the HIV field and is referred to as germline targeting, which relies upon the activation and expansion of rare but specific B-cell lineages in naïve individuals (*103, 106*). In contrast, under conditions of pre-existing immunity acquired during natural infection or previous vaccination, the challenge is to manipulate already established B-cell immunodominance hierarchies and reshape serum antibody responses towards desired specificities. In an indirect approach towards increasing subdominant B-cell populations, Silva et al. (*144*) have shown that the targeted suppression of immunodominant clones during an active germinal center reaction can allow subdominant B-cell populations to overtake the germinal center response. Other approaches have used heterologous prime-boost immunization regimens with either alternative viral strains or rationally modified versions of the priming immunogen in order to steer antibody responses towards more conserved domains (*75, 145-147*). However, leveraging structural information of defined neutralization epitopes to guide bulk antibody responses towards specific, well-characterized single epitopes remains an unmet challenge.

Here, we investigate whether, under conditions of pre-existing immunity, a computationally designed immunogen presenting a single epitope is able to reshape serum antibody responses towards increased recognition of a specific neutralizing epitope. To mimic a scenario of pre-existing immunity against a relevant pathogen, we immunized mice with a prefusion-stabilized version of RSVF, and found that antibody titers against RSV antigenic site II were present in very low levels, i.e. a subdominant site II-specific response was elicited. Based on a previously developed epitope-focused immunogen for RSV site II (FFL_001) (*59*), we engineered an optimized nanoparticle presenting this immunogen, and investigated the potential of a rationally designed epitope-focused immunogen to boost these subdominant levels of site-specific antibodies.

We show that multivalent presentation of a designed epitope-focused immunogen elicits superior levels of epitope-specific antibodies compared to prefusion RSVF in naïve mice, indicating that the subdominance of a particular epitope can be altered through its presentation in a distinct molecular context. Repeated immunizations with RSVF failed to increase site II-specific antibodies, and instead further diluted site II-specific responses. In contrast, heterologous boosts with an epitope-scaffold nanoparticle enhanced the serum response towards the subdominant site II epitope, and the boosted antibodies neutralized RSV *in vitro*. For the first time, we provide compelling evidence that synthetic immunogens comprising a single epitope can efficiently redirect specificities in bulk antibody responses *in vivo* and enhance subdominant nAb responses. Such strategy may present an important alternative for pathogens where future vaccines are required to reshape pre-existing immunity and elicit finely tuned antibody specificities.

2.3 Results

2.3.1 Design of an RSV-based nanoparticle displaying a site II epitope-focused immunogen

In a previous study, a computationally-designed, RSV site II epitope-scaffold nanoparticle was shown to elicit serum neutralization activity in non-human primates (NHPs) (*59*). Despite the fact that very potent monoclonal antibodies were isolated from the immunized NHPs, the neutralization potency at the serum level was modest, indicating low titers of the potent antibodies. Therefore, our first aim was to take the best previously tested immunogen (FFL_001) and further optimize delivery and immunization conditions to maximize the induction of site II-specific antibodies. A comparative study of four different adjuvants revealed that alum, an adjuvant approved for human use, yielded highest overall immunogenicity and elicited antibodies cross-reactive with prefusion RSVF in four out of five mice (Fig S2.1).

Next, we sought to develop an improved, easily produced nanoparticle to multimerize the epitope-scaffold for efficient B-cell receptor crosslinking. Previously, Correia et al. (*59*) employed a chemical conjugation strategy of FFL_001 to a Hepatitis-B core antigen based nanoparticle, which resulted in a difficult construct with a laborious purification process. Recently, several studies have reported the use of the RSV nucleoprotein (RSVN) as a nanoparticle platform for immunogen presentation (*148, 149*). When expressed in *E. coli*, RSVN forms nanorings, 17 nm in diameter, containing 10 or 11 RSVN protomers (*150*). We reasoned that RSVN would be an ideal particle platform to multimerize an RSV epitope-scaffold, as RSVN contains strong, RSV-directed T-cell epitopes (*149*). However, our initial attempts to genetically fuse FFL_001 to RSVN yielded poorly soluble proteins that rapidly aggregated after purification. We therefore employed structure-based protein resurfacing (*151*), attempting to improve the solubility of this site II epitope-scaffold when arrayed in high density on RSVN. To guide our resurfacing design process, we leveraged information from a sequence homolog of the ribosomal recycling factor (PDB: 1ISE), the structural template originally used to design FFL_001. Based on a sequence alignment of the mouse homolog (NCBI reference: NP_080698.1) and FFL_001, we exchanged the FFL_001 amino acids for the mouse sequence homolog and used Rosetta Fixed Backbone Design (*152*) to ensure that the mutations were not energetically unfavorable, resulting in 38 amino acid substitutions (34.2% overall). We named this variant FFLM, whose expression yields in *E. coli* showed a five-fold increase when compared to FFL_001, and it was confirmed to be monomeric in solution (Fig S2.2).

To confirm that the resurfacing did not alter the epitope integrity, we measured the binding affinities of FFLM to Motavizumab, a high-affinity variant of Palivizumab (*153*), and to a panel of human site II nAbs previously isolated (*2*) using surface plasmon resonance (SPR). All antibodies bound with high affinity to FFLM, indicating broad reactivity of this immunogen with a diverse panel of human nAbs (Fig 2.1B and Fig S2.3). Interestingly, the tested nAbs showed approximately one order of magnitude higher affinity to the epitope-scaffold as compared to the latest version of prefusion RSVF, originally called DS2 (*84*), suggesting that the epitope is properly presented and likely further stabilized in a relevant conformation.

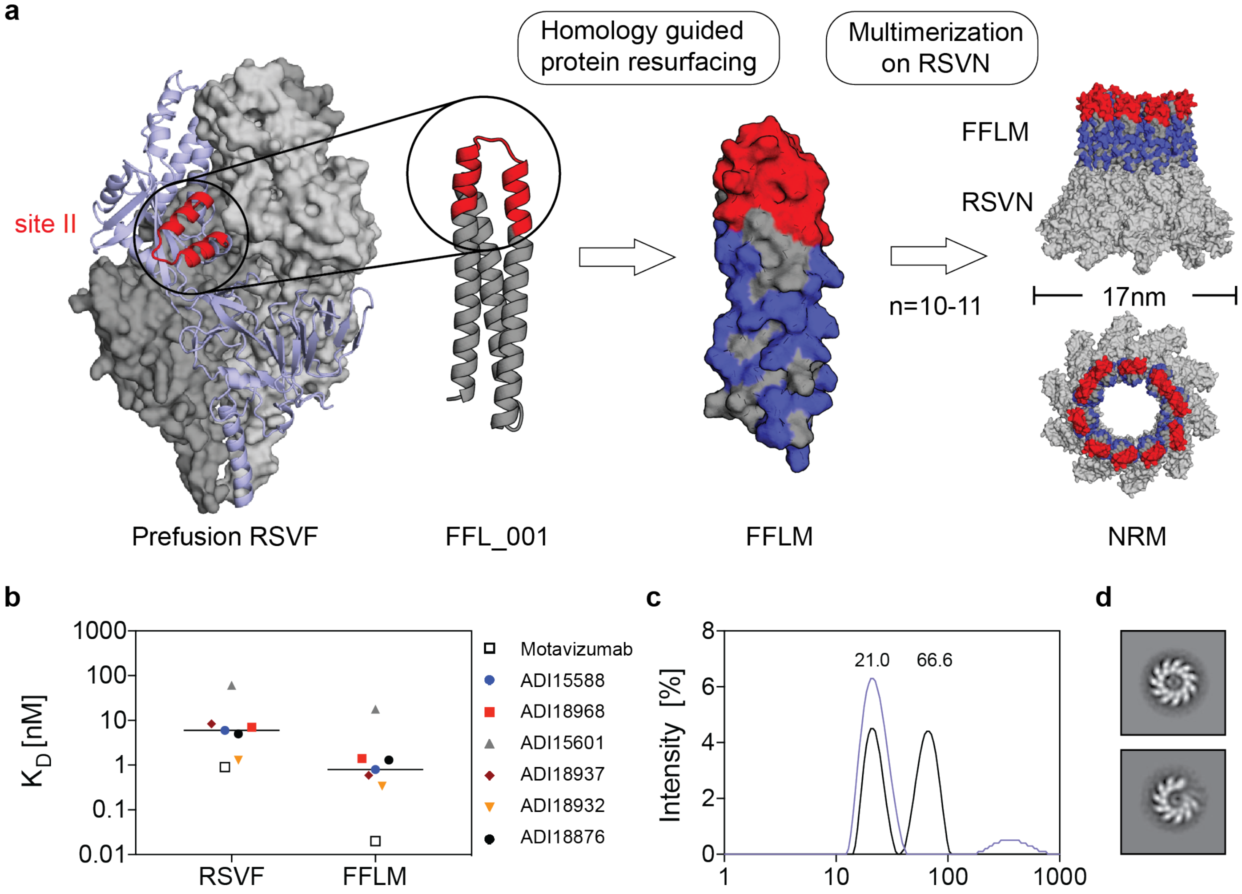

a
Homology guided protein resurfacing
Multimerization on RSVN
site II
FFLM
RSVN
n=10-11
17nm
Prefusion RSVF
FFL_001
FFLM
NRM
b
K_D [nM]
1000
100
10
1
0.1
0.01
RSVF
FFLM
Motavizumab
ADI15588
ADI18968
ADI15601
ADI18937
ADI18932
ADI18876
c
Intensity [%]
8
6
4
2
0
21.0
66.6
1
10
100
1000
d

Figure 2.1: Design of an RSV-based nanoparticle displaying a site II epitope-focused immunogen.
a) Structural model of the prefusion RSVF trimer (PDBID: 4JHW), with two subunits shown as a grey surface and one subunit shown as light blue cartoon representation with the epitope targeted by Palivizumab (antigenic site II) highlighted in red. FFL_001 was previously designed to present the site II epitope in a computationally designed scaffold. FFLM was designed by evolution-guided resurfacing, where changes in amino acid identity are highlighted in blue. FFLM was genetically fused to the N-terminus of the RSV nucleoprotein (RSVN), resulting in a high-density array of the epitope-scaffold, as shown by the structural model (based on PDBID: 2WJ8). **b)** Kinetic binding affinities of site II-specific human nAbs measured by SPR. K_Ds were measured with RSVF/FFLM immobilized as ligand and antibody fabs as analyte. Sensorgrams and fits are shown in Fig S2.3. **c)** Dynamic light scattering (DLS) profiles for FFL_001 and FFLM fused to RSVN. The FFL_001-RSVN fusion protein formed higher-order oligomers in solution (66.6 nm of median diameter), whereas the resurfaced FFLM-RSVN fusion protein (NRM) was monodisperse with a median diameter of 21 nm. **d)** Analysis of the NRM nanoparticles by negative stain electron microscopy. Shown are the 2D class averages of two representative classes.

Importantly, the FFLM-RSVN fusion protein expressed with high yields in *E. coli* (>10 mg/liter), forming a nanoring particle, dubbed NRM, that was monodisperse in solution with a diameter of approximately 21 nm (Fig 2.1C). Negative stain electron microcopy confirmed the ring-like structure as suggested by the model (Fig 2.1D). Although we cannot fully rationalize the factors that contributed to the solubility improvement upon multimerization, our strategy to transplant surface residues from a sequence homolog to synthetic proteins may prove useful to enhance the solubility of other computationally designed proteins.

2.3.2 NRM enhances the induction of site II-specific antibodies

We next tested the immunogenicity of NRM and its ability to elicit site II-specific antibodies. Three groups of ten mice were subjected to three immunizations with 10 μg of NRM, monomeric FFLM and prefusion RSVF (*84*), which is currently the leading immunogen for an RSV vaccine (Fig 2.2A). Based on the results of our adjuvant screen (Fig S2.1), all the immunogens were formulated in alum. As compared to FFLM, NRM showed a higher overall immunogenicity (directed both against RSVN and FFLM) (Fig 2.2B).

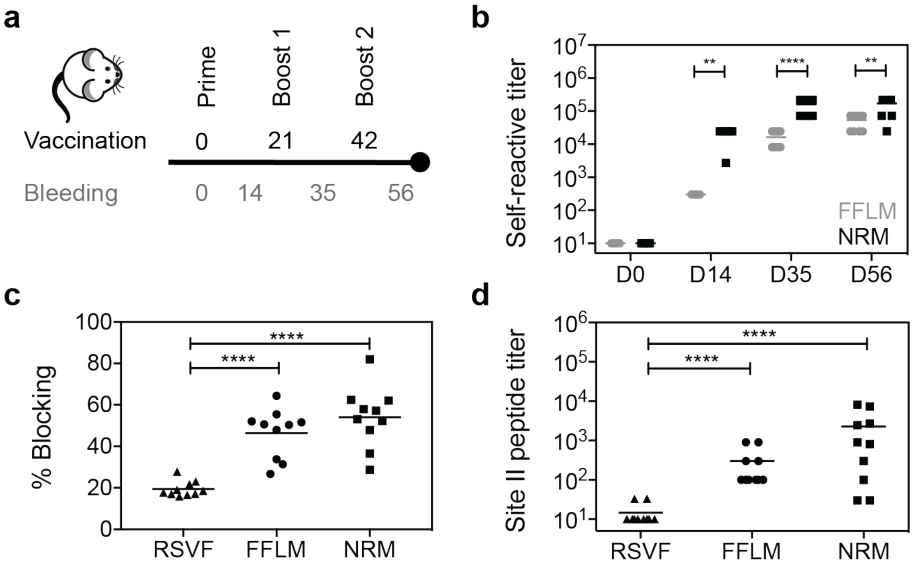


Figure 2.2: Immunogenicity and quantification of site II-specific antibody responses.
a) Immunization scheme. Balb/c mice were immunized three times on days 0, 21 and 42, and blood was drawn 14 days after each vaccination. **b)** Serum antibody titers elicited by FFLM and NRM at different timepoints measured by ELISA against the respective immunogen. NRM shows significantly increased immunogenicity at day 14, 35 and 56 relative to FFLM. **c)** SPR competition assay with Motavizumab. Day 56 sera of mice immunized with RSVF, FFLM or NRM was diluted 1:100 and SPR response units (RU) were measured on sensor chip surfaces containing the respective immunogen. Motavizumab binding sites were then blocked by saturating amounts of Motavizumab, and the residual serum response was measured to calculate the serum fraction competed by Motavizumab binding. Mice immunized with FFLM or NRM show significantly higher levels of serum antibodies that are competed by Motavizumab binding. **d)** Site II-specific serum titers at day 56 from mice immunized with RSVF, FFLM and NRM, measured by ELISA against site II peptide. Three immunizations with prefusion RSVF elicited low levels of site II-specific antibodies, whereas FFLM and NRM vaccinations yielded significantly higher peptide-specific serum titers. Data shown are derived from at least two independent experiments, each sample assayed in duplicate. Statistical comparisons were calculated using two-tailed Mann-Whitney U tests. ** indicates $p < 0.01$, *** indicates $p < 0.0001$, **** $p < 0.0001$.

A key aspect of epitope-focused vaccines is to understand how much of the antibody response targets the viral epitope presented to the immune system. Therefore, we sought to measure the site II-specific antibody titers elicited by NRM and FFLM and compare these epitope-specific antibody responses to those elicited by prefusion RSVF. We established an SPR competition assay (described in the methods and shown in Fig S2.4) to quantify the fraction of site II-specific antibodies elicited by each immunogen (FFLM, NRM or prefusion RSVF). Briefly, the respective antigen was immobilized on the sensor chip surface, and the fraction of the serum antibody response competed by Motavizumab was measured, serving as a proxy for site II-specific antibodies. We observed that NRM elicited site II-specific antibody responses superior to those elicited by RSVF (Fig 2.2C). This was surprising, given that the ratio of site II epitope surface area to overall immunogen surface is similar in both NRM and RSVF (Fig S2.2). To confirm this finding through a direct binding assay rather than a competitive format, we measured the binding levels of sera to the site II epitope in a peptide ELISA, where the site II peptide was immobilized on a streptavidin-coated surface. Peptides mimicking site II are known to be conformationally flexible (*110*) (Fig S2.5) but have been show to adopt a very similar conformation upon antibody binding to the one presented in the context of pre- and postfusion RSVF (*154*). Consistent with the previous experiment, we found that NRM elicited two orders of magnitude higher site II-specific responses than RSVF (Fig 2.2D). Together, we concluded that an epitope-focused immunogen, despite similar molecular surface area, can elicit substantially higher levels of site-specific antibodies compared to a viral fusion protein.

2.3.3 NRM induces low levels of RSVF cross-reactive antibodies with low neutralization potency

Given the substantial site II peptide-specific serum titers elicited by NRM in mice, we investigated whether these antibodies cross-reacted with prefusion RSVF and were sufficient to neutralize RSV *in vitro*.

Following three immunizations with NRM, all the mice (n=10) developed detectable serum cross-reactivity with prefusion RSVF (mean serum titer = 980) (Fig 2.3A). Sera also cross-reacted with the postfusion conformation of RSVF (mean serum titer = 380), but binding to virus-infected cell lysate was negligible for mice immunized with the epitope scaffold (Fig S2.6). Unsurprisingly, the overall quantity of RSVF cross-reactive antibodies elicited by immunization with an immunogen presenting a single epitope is more than two orders of magnitude lower than those of mice immunized with prefusion RSVF, which comprises at least six antigenic sites (*2*). Similarly, a B-cell ELISpot revealed that NRM-immunized mice presented prefusion RSVF-reactive antibody secreting cells, but their frequency was approximately one order of magnitude lower than upon immunization with prefusion RSVF (Fig 2.3C).

The major determinant for antibody specificity is attributed to the heavy chain CDR3 region (HCDR3) (*155*). While for certain classes of nAbs, the antibody lineages and their sequence features are well-defined (e.g. HIV neutralizing VRC01 class antibodies (*156*), or RSV neutralizing MPE8-like antibodies (*125*)), antibodies targeting RSV antigenic site II seem to be derived from diverse precursors and do not show HCDR3 sequence convergence in humans (*2*). While we did not expect to find dominant lineages or HCDR3 sequence patterns in mice, we used next-generation antibody repertoire sequencing (*157*) to ask whether NRM could elicit antibodies with similar sequence signatures to those elicited by prefusion RSVF. Indeed, we found 300 clonotypes, defined as antibodies derived from the same VH gene with the same HCDR3 length and 80% sequence similarity, that overlapped between NRM and the prefusion RSVF immunized cohort, suggesting that at the molecular level, relevant antibody lineages can be activated with the NRM immunogen (Fig S2.7). Notably, nine out of the 20 most expanded clonotypes in the NRM cohort were also present in mice immunized with prefusion RSVF, albeit not as expanded (Fig 2.3B). This finding might reflect the enrichment of site II-specific antibodies in the NRM cohort (Fig 2.2D).

We further investigated whether these low levels of prefusion RSVF-binding antibodies were sufficient to neutralize RSV *in vitro*. While three immunizations with prefusion RSVF elicited potent RSV-neutralizing serum titers (mean IC_{50}= 10,827), for NRM we only detected low levels of RSV-neutralizing serum activity in three out of ten mice (Fig 2.3D). This result is consistent with Correia *et al*. (*59*), who observed no serum neutralization in mice, but succeeded in inducing nAbs in NHPs with prior RSV seronegativity.

Altogether, we concluded that despite NRM's superior potential to induce high levels of site II-specific antibodies, the majority of antibodies activated from the naïve repertoire is not functional for RSV neutralization. A potential explanation, stemming from structural comparison between the epitope-focused immunogen and RSVF, is that these antibodies do not recognize the site II epitope in its native, quaternary environment in prefusion RSVF or on virions in sufficient amounts and with high enough affinity to potently neutralize RSV.

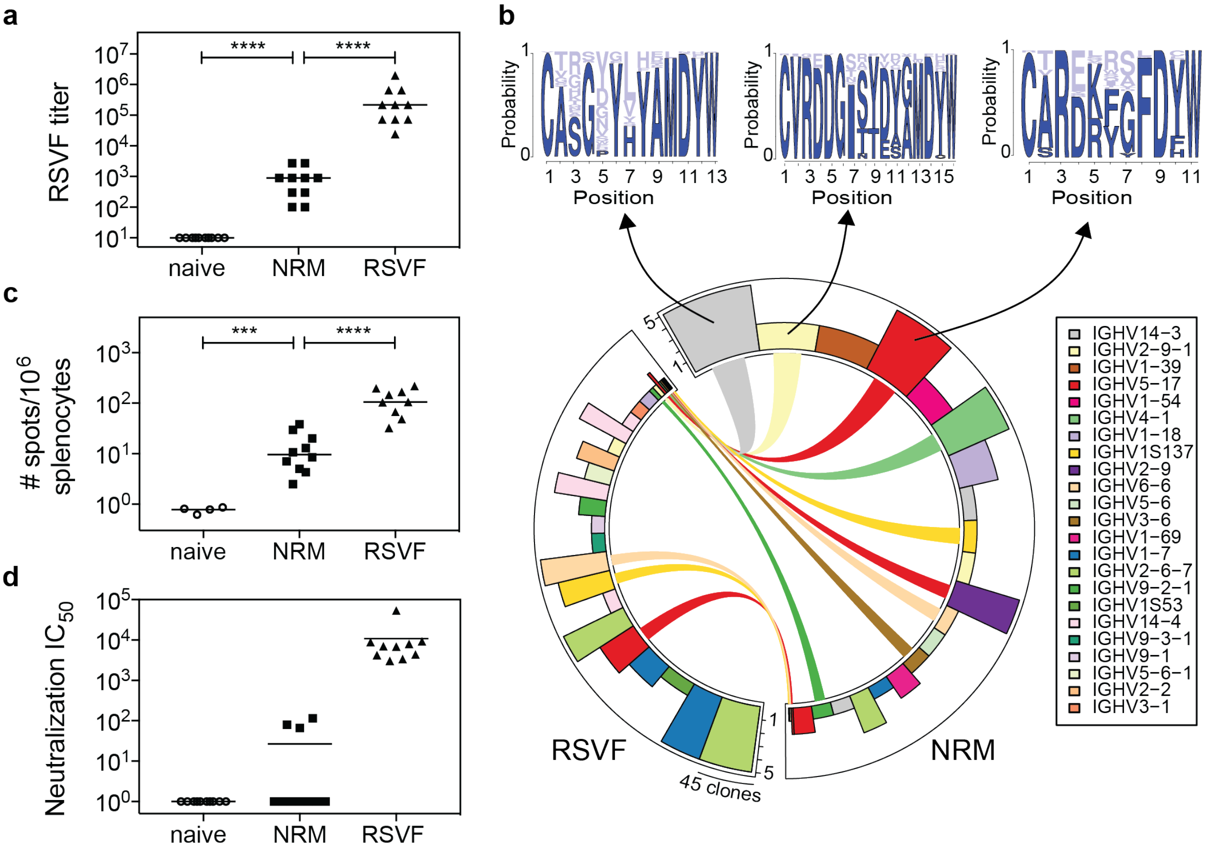


Figure 2.3: RSVF cross-reactivity and serum neutralization.
a) NRM elicits prefusion RSVF cross-reactive antibodies, which are two orders of magnitude lower compared to prefusion RSVF immunization. Mice immunized only with adjuvant (naïve) do not show RSVF cross-reactivity. **b)** Next-generation sequencing of antibody repertoire. Antibody variable heavy chains of mice immunized with RSVF or NRM (5 mice per cohort) were sequenced at day 56 and grouped into clonotypes. Circos plot showing the 20 most expanded clonotypes from both cohorts, with identical clonotypes connected. Height of bars indicates number of mice that showed the respective clonotype, width represents the clonal expansion within a clonotype (i.e. the number of clones grouped into the respective clonotype). Three clonotypes that occurred both in the RSVF and the NRM cohort, but were expanded within the NRM cohort were analyzed for their HCDR3 sequence profile, as shown by sequence logo plots (top). Dark blue color represents amino acid identities that occurred in RSVF cohort, light blue color represents amino acids uniquely found following NRM immunization. The frequency of each amino acid in the NRM cohort is indicated by the size of the letter. **c)** B-cell ELISpot of mouse splenocytes to quantify prefusion RSVF-specific antibody secreting cells (ASC). Number of ASCs per 10^6 splenocytes that secrete prefusion RSVF-specific antibodies following three immunizations with adjuvant only (naïve), NRM or prefusion RSVF. **d)** RSV neutralizing activity of mouse sera from day 56 shown as neutralization IC_{50}. Three out of ten mice immunized with NRM showed detectable RSV neutralizing activity, whereas all mice immunized with prefusion RSVF neutralized RSV (mean IC_{50} = 10,827). Data shown are from one out of two independent experiments. Statistical comparisons were calculated using two-tailed Mann-Whitney U tests. *** indicates $p < 0.001$, **** indicates $p < 0.0001$.

2.3.4 NRM boosts site II-specific antibodies under conditions of pre-existing immunity

While vaccination studies in naïve animal models are an important first step to validate novel immunogens, previous studies (*59*) and results presented here imply that epitope-scaffolds may not be able to elicit robust RSV neutralizing serum activity from a naïve antibody repertoire. However, given the high affinity of the epitope-scaffold towards a panel of site II-specific nAbs, together with the ability to elicit high titers of site II-specific antibodies *in vivo*, we hypothesized that such an epitope-focused immunogen could be efficient in recalling site II-specific B-cells in a scenario of pre-existing immunity, thereby achieving an enhanced site-specific neutralization response.

Our initial immunization studies with prefusion RSVF showed that site II-specific responses were subdominant (Fig 2.2C and 2.2D). Given that subdominance is a common immunological phenotype for many of the neutralization epitopes that are relevant for vaccine development (*158*), we sought to test if NRM could boost subdominant antibody lineages that should ultimately be functional and recognize the epitope in the quaternary environment of the viral protein. To test this hypothesis, we designed a mouse immunization experiment with three cohorts, as outlined in Fig 2.4A. Following a priming immunization with RSVF, cohort (1) was boosted with adjuvant only (“prime only”), cohort (2) received two boosting immunizations with prefusion RSVF (“homologous boost”), and cohort (3) received two boosts with NRM (“heterologous boost”).

A comparison between prefusion RSVF immunized groups prime only and homologous boost revealed that the two additional boosting immunizations with RSVF only slightly increased overall titers of prefusion RSVF-specific antibodies ($p = 0.02$), indicating that a single immunization with adjuvanted RSVF is sufficient to induce close to maximal serum titers against RSVF (Fig 2.4B). Following the heterologous boost with NRM, overall RSVF specific antibody titers remained statistically comparable to the prime only group ($p = 0.22$).

Next, we quantified the site II-specific endpoint serum titers in a peptide ELISA format (Fig 2.4C). Interestingly, the homologous boost with prefusion RSVF failed to increase site II-specific antibody levels, reducing the responses directed to site II to the lower limit of detection by ELISA. This result is yet another example of the underlying complexity inherent to the fine specificity of antibody responses elicited by immunogens and how important specificities can be dampened throughout the development of an antibody response. In contrast to the homologous boost, the heterologous boost with NRM significantly increased site II peptide-specific serum titers ($p < 0.0001$).

In order to understand whether this increase relied at least partially on an actual recall of antibodies primed by RSVF, or rather on an independent antibody response irrelevant for RSVF binding and RSV neutralization, we dissected the epitope specificity within the RSVF-specific serum response. In an SPR competition assay, a significantly higher fraction ($p = 0.02$) of prefusion RSVF-reactive antibodies were competed by Motavizumab in mouse sera primed with prefusion RSVF and boosted with NRM (mean % blocking = 37.5 ± 14.5%), as compared to mice immunized once or three times with prefusion RSVF (21.5% ± 12.1% or 19.5% ± 3.7%, respectively) (Fig 2.4D). Importantly, control mice immunized that were not primed with prefusion RSVF and instead were immunized three times with NRM did not yield detectable binding signals against prefusion RSVF in an SPR assay, despite detectable ELISA signals (Fig 2.3A). Thus, while the heterologous boost with NRM will also prime antibodies that do not bind RSVF, the increased fraction of prefusion RSVF binding, site II-specific antibodies is likely to arise from a recall of RSVF primed, site II-specific antibodies. Similarly, a competition ELISA revealed that a significantly larger fraction of overall RSVF reactivity was attributed to site II-specific antibodies upon heterologous boost, as compared to both control groups (mean % competition =

36.1% ± 2.5% versus 22.6% ± 9.1% or 14.4% ± 5.9%, respectively, $p = 0.002$ and $p < 0.0001$). In contrast, site II-specific antibodies were significantly higher in mice that received only one as opposed to three RSVF immunizations, indicating that RSVF boosting immunizations further diluted site II-specific antibody titers ($p = 0.03$) (Fig 2.4E).

Together, we have shown that the serum antibody specificity can be steered towards a well-defined antigenic site by boosting pre-existing, subdominant antibody levels with an epitope-focused immunogen. This is an important and distinctive feature of the epitope-focused immunogen compared to an immunogen based on a viral protein (prefusion RSVF), which was shown to decrease already subdominant antibody responses under the same conditions. These results may have broad implications on strategies to control antibody fine specificities in vaccination schemes, both for RSV and other pathogens.

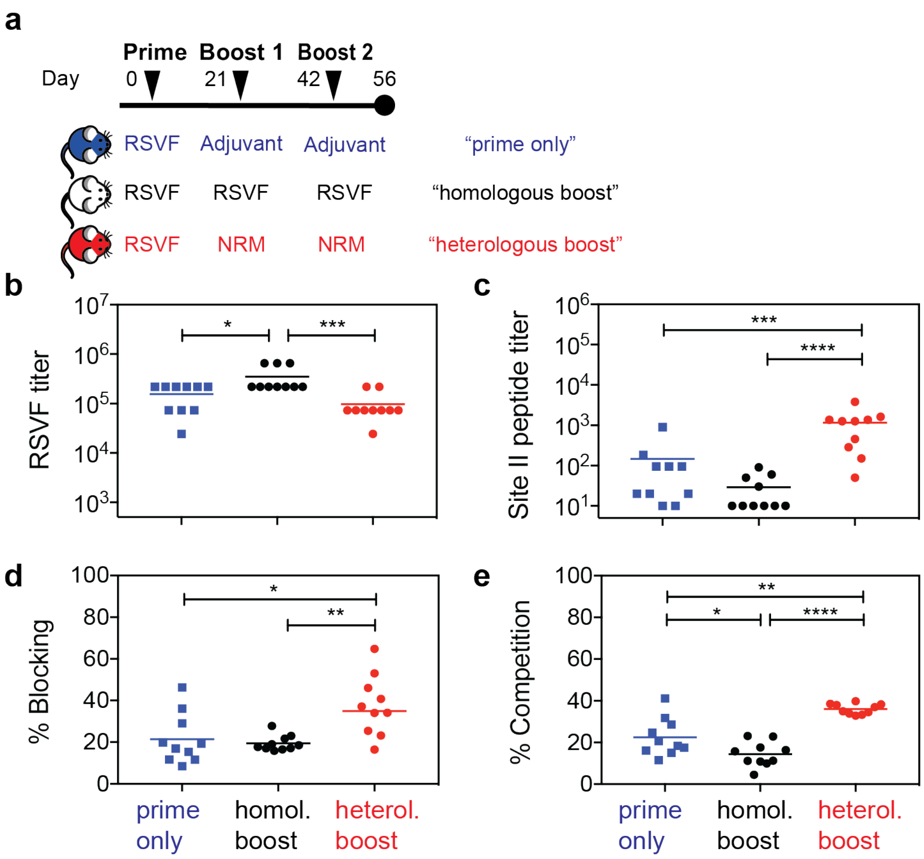


Figure 2.4: Heterologous prime boost reshapes antibody responses enhancing levels of site II-specific antibodies. a) Heterologous prime-boost study groups. Three mouse cohorts were immunized with either 1x RSVF ("prime only"), 3x RSVF ("homologous boost") or 1x RSVF followed by two boosts with NRM ("heterologous boost"). **b)** Antibody titers directed against prefusion RSVF. Mice receiving homologous boosting immunizations show slightly higher RSVF-specific serum titers compared to the prime only cohort, whereas heterologous boosting yielded statistically comparable titers to the prime only group. The difference between the homologous and heterologous boost cohorts was statistically significant. **c)** Site II-specific titers measured by ELISA showed that the heterologous boost significantly increases site II-specific titers compared to both prime and homologous boost groups. Although not statistically significant (p = 0.06), mice receiving a homologous boost had lower levels of site II-specific antibodies compared to prime only group. **d)** SPR competition assay with Motavizumab on a prefusion RSVF-coated sensor chip. Sera from indicated groups were diluted 1:100 and RSVF binding responses were quantified. Site II was then blocked with Motavizumab, and the remaining serum response quantified. The heterologous boost induced a significantly higher fraction of site II-directed antibodies that competed with Motavizumab for RSVF binding, as compared to both prime only and homologous boost groups. **e)** Quantification of site II-specific responses in a competition ELISA. Binding was measured against prefusion RSVF, and the area under the curve (AUC) was calculated in presence of NRM competitor, normalized to the AUC in the presence of RSVN as a control competitor. Compared to the prime only group, the homologous boost resulted in significantly lower site II-specific serum titers, confirming the trend observed in c). The heterologous boost increased the fraction of site II-targeting antibodies within the pool of prefusion RSVF-specific antibodies compared to both control groups. Data presented are from at least two independent experiments, with each sample assayed in duplicates. Statistical comparisons were calculated using two-tailed Mann-Whitney U tests. * indicates p < 0.05, ** indicates p < 0.01, *** indicates p < 0.0001, **** p < 0.0001.

2.3.5 Boosted antibodies neutralize RSV *in vitro*

The enhanced reactivity to site II observed in the heterologous prime-boost scheme led us to investigate if the antibodies boosted by a synthetic immunogen were functionally relevant for virus neutralization. In bulk sera, we observed 2.3-fold higher serum neutralization titers in mice receiving the heterologous boost (mean IC_{50}=7,654) compared to the prime only control group (mean IC_{50}=3,275) (Fig 2.5A). While this increase in serum neutralization was not statistically significant, we next assessed if this increase in neutralization was driven by increased levels of epitope-specific antibodies. We observed that site II-directed antibody levels correlated with overall serum neutralization titers in the heterologous prime boost group (r^2=0.76, p=0.0009) (Fig 2.5B), whereas the prime only (r^2=0.32, p=0.09) or the homologous boost cohorts showed no such correlation (r^2 = 0.18, p=0.22) (Fig S2.8).

To characterize the RSV neutralizing activity mediated by site II-specific antibodies, we pooled sera from each cohort, enriched site II-specific antibodies and measured viral neutralization (see methods and Fig S2.9). Briefly, we incubated pooled sera from each group with streptavidin beads conjugated to biotin-labeled antigenic site II peptide, and eluted bound antibodies. To control for the quality of the enrichment protocol, we verified by ELISA that the column flow-through was depleted of site II-specific antibodies (Fig S2.9). In agreement with the different site II peptide-specific serum levels shown in Fig 2.4, the overall quantity of site II-specific antibodies purified from equivalent amounts of sera differed between groups, with the heterologous boost and the 3x NRM groups showing the highest levels (Fig S2.9).

Next, we tested neutralization of this polyclonal pool of site II-specific antibodies. Strikingly, we found that the heterologous boost cohort showed a 15-fold greater site II-specific neutralization titer (IC_{50} = 99.6) as compared to the prime only and homologous boost cohorts (IC_{50} = 6.6 and IC_{50} = 4.9, respectively). It is important to note that albeit the homologous boost with RSVF showed increased serum neutralizing activity compared to the prime only control group, site II-mediated neutralization was similar. This finding is consistent with our observation that site II-specific antibodies do not increase with repeated immunizations of prefusion RSVF. Enriched antibodies from the 3x NRM group were non-neutralizing, despite the high concentration of site II-specific antibodies (Fig S2.9). Thus, NRM significantly enhances site II-mediated RSV neutralization, but requires the priming of a relevant subset of RSVF-binding antibodies. Finally, we addressed whether this increase in site II-mediated neutralization was due to higher amounts of site II-specific antibodies, or the intrinsic neutralization potency of the same antibodies. As shown in Fig 2.5D, antibodies from the prime only, heterologous boost and homologous boost cohorts exhibit similar neutralization potencies (IC_{50} ranging from 1.7 µg/ml to 4.9 µg/ml) of site II-specific antibodies. Consequently, the heterologous boosting scheme yielded higher amounts of site-specific, functional antibodies, rather than an increased potency of the same antibodies.

Altogether, we dissected the mode of action of the synthetic immunogens when used as heterologous boosters, where the observed enhanced neutralization resulted from the increase of sheer amounts of antibodies directed to site II.

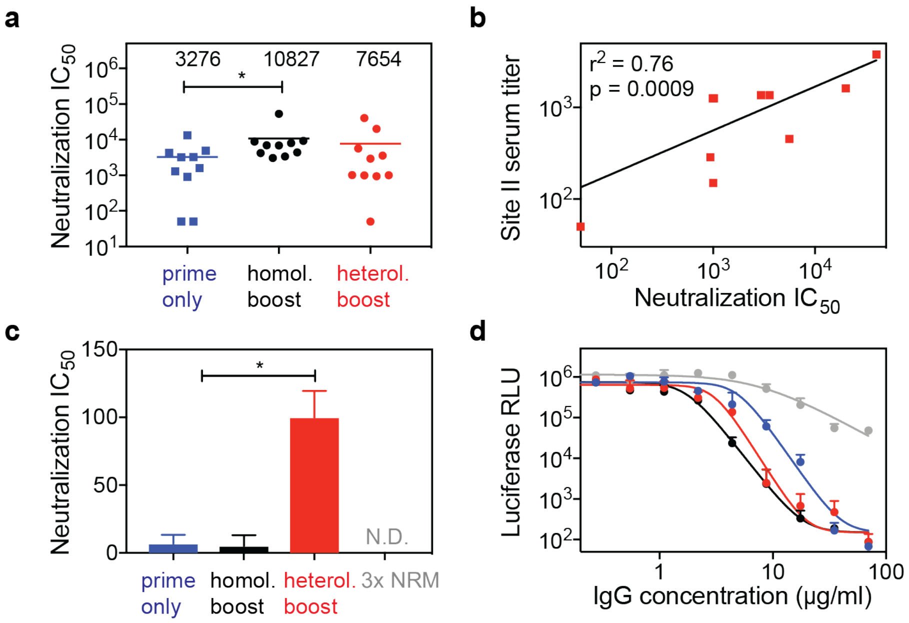


Figure 2.5: Boosted site II-specific antibodies are functional and mediate increased neutralization activity. **a)** *In vitro* RSV neutralization IC_{50} for each group. Compared to the prime only group, mice receiving a homologous boost showed increased RSV neutralization titers. On average, the heterologous boost yielded a 2.3-fold increase in serum neutralization titers compared to prime only, but these differences were statistically not significant when compared to either group. **b)** Correlation of site II-specific serum titer (measured by peptide ELISA) with RSV neutralization IC_{50} as determined for each mouse within the heterologous prime boost cohort. Correlations for control groups are shown in Fig S2.8. Data represent the mean of two independent experiments, each measured in duplicate. Pearson correlation coefficient (r^2) and p-value were calculated in GraphPad Prism. **c)** Serum fractionation revealed increased levels of site II-mediated neutralization. Site II-specific antibodies from mouse sera were enriched in an affinity purification as described in the methods and shown in Fig S2.9. The IC_{50} values are the dilution factor of site II-specific antibodies eluted and quantify the total site II-mediated neutralization. Antibodies purified from the heterologous boost group showed a 15-fold increase in RSV neutralizing activity compared to the prime only and homologous boost control groups. No site II-mediated neutralization was detected for mice receiving three immunizations of NRM (N.D. = non-detectable). **d)** RSV neutralization potency of affinity-purified site II-specific antibodies. Purified antibodies from each group were diluted to a concentration of 70 μg/ml to measure their RSV neutralization potencies. Site II antibodies from prime only, heterologous boost and homologous boost exhibited similar neutralization potencies (IC_{50} = 4.9 μg/ml, IC_{50} = 3.0 μg/ml, and IC_{50} = 1.7 μg/ml, respectively). Data are presented from two independent experiments, and each sample was assayed in duplicate with additional controls shown in Fig S2.9. Statistical comparisons were calculated using two-tailed Mann-Whitney U tests. * indicates $p < 0.05$.

2.4 Discussion

Despite a rapid increase in our atomic-level understanding of antibody-antigen interactions for various pathogens, the translation of structural information into efficacious immunogens that elicit antibody responses specific to *bona fide* epitopes remains a key challenge for next-generation vaccine development.

Multiple strategies have been investigated to focus nAb responses on defined neutralization epitopes (*159*). Among them, epitope-scaffolds have been shown to elicit RSV site II-specific, nAb responses in naïve non-human primates. While the overall serum neutralization was modest, a monoclonal antibody induced by vaccination, showed superior neutralization potency to that of Palivizumab (*59*). However, a major limitation of epitope-scaffold immunogens (*56, 110, 160*) is that the quaternary environment of the epitope presented in the native viral protein is lost. Thus, the binding mode of a significant fraction of the elicited antibodies is likely incompatible with the epitope in its native environment. This observation is reinforced by our finding that although NRM elicited high serum levels of site II peptide-specific antibodies, only low levels were cross-reactive with RSVF and neutralizing activity was residual. This finding is consistent with previous studies using epitope-scaffolds (*55, 110, 161*). Together, these results highlight the limitations of synthetic scaffolds in an epitope-focused vaccine approach in naïve individuals.

However, our finding that an epitope which is subdominant (site II) in its native environment (prefusion RSVF) is readily targeted by the immune system when presented in a distinct molecular context (NRM), supported the potential use of synthetic immunogens to reshape antibody responses towards such well-defined antigenic sites. Pre-existing immunity against a viral protein (RSVF, influenza HA or others), in which certain antibody specificities are subdominant, is a common scenario in humans that have encountered repeated natural infections throughout their life (*120, 142, 162, 163*). Therefore, a major challenge for vaccine development is to boost pre-existing, subdominant antibodies and enhance site-specific neutralization.

To date, boosting nAbs targeting specific epitopes under conditions of pre-existing immunity has been challenging. For instance, strong antibody responses against immunodominant epitopes can sterically mask the neutralization epitope, preventing the induction of a potent antibody response targeting the subdominant site (*140-142, 164*). Overcoming these established immunodominance hierarchies is complex, as such hierarchies seem to be impacted by multiple factors including serological antibody levels, their specificity, memory B-cell counts, adjuvants, and the immunization or infection route (*69*).

Heterologous prime-boost schemes are a promising strategy to guide the fine specificity of antibody responses and to focus these responses on vulnerable antigenic sites. Several vaccine studies have been conducted for influenza (*146, 147*), RSV (*75*) and HIV (*106*), where the heterologous immunogens were alternative strains or modified viral fusion proteins, but yet not as heterologous as a computationally designed epitope-scaffold. Importantly, it is possible that immunogens based on modified viral proteins retain immunodominant signatures that steer antibody responses away from the target epitopes. While this scenario may not be fully absent in synthetic epitope-scaffolds, it is at least mitigated by the fact that the protein has not evolved under the pressure of escaping the immune system.

Our study demonstrates that a heterologous boosting immunogen that optimally presents a single neutralization epitope can boost pre-existing, subdominant antibody responses targeting this epitope, yielding increased epitope-mediated neutralization. The ability to narrowly focus antibody responses to a single epitope that mediates clinical protection underlines the potential of rationally designed immunogens for vaccine development against elusive pathogens. In particular, our results demonstrate that albeit single-epitope immunogens may not be the most powerful to select functional antibodies from a naïve repertoire, they have a unique ability to boost neutralizing epitope-specific antibodies primed by a viral protein. Further studies in more relevant animal models will reveal if nAbs primed by natural infection with RSV can also be boosted mimicking a more realistic vaccination scenario.

Given that the approach presented here is generalizable and that epitope-scaffold nanoparticles can be proven successful in boosting nAbs specific for other sites, this strategy holds great potential to tune levels of antibody specificities through heterologous prime boost vaccination schemes which are now frequently used in for challenging pathogens (*106, 146, 165*).

The original antigenic sin theory in the influenza field describes that the first viral exposure permanently shapes the antibody response, which causes individuals to respond to seasonal vaccines dependent on their immune history (*140, 166*). Seasonal vaccines generally fail to boost antibodies targeting broadly neutralization epitopes on the hemagglutinin stem region (*140*). Focusing antibody responses on these defined epitopes may remove the need for annual vaccine reformulation, and may also protect against emerging pandemic strains (*137, 158, 167, 168*). The influenza vaccine challenge seems particularly well suited to our approach considering that the human population has pre-existing immunity to influenza, including some subdominant bnAbs that seasonal vaccines fail to stimulate (*140*).

Lastly, vaccine development against antigenically related viruses such as zika and dengue could benefit of the approach presented here, as antibodies mounted against the envelope protein of a dengue subtype can facilitate infection with zika (*169*) or other dengue subtypes (*170*). A site conserved between all four dengue subtypes and zika envelope protein has been structurally characterized and suggested for the development of an epitope-focused immunogen (*131*).

When seeking to apply an immunofocusing strategy to other antigenic sites and pathogens, one challenge is the development of epitope-scaffolds stably presenting the epitope in a synthetic immunogen that is compatible with antibody binding. While the RSV antigenic site II is a structurally simple helix-turn-helix motif, many other identified neutralization epitopes comprise multiple, discontinuous segments. However, continuous advances in rational protein design techniques (*26*) will allow the design of more complex protein scaffolds to stabilize increasingly complex epitopes.

Altogether, we have shown how an optimized presentation of a computationally designed immunogen in an RSVN-based nanoparticle can reshape bulk serum responses and boost subdominant, nAb responses *in vivo.* This is a distinctive feature compared to using prefusion RSVF as a boosting immunogen, and underscores how subdominant epitopes can be converted to immunodominant epitopes when presented in a different environment. We foresee the great promise of this strategy to overcome the challenge of boosting and focusing pre-existing immunity towards defined neutralization epitopes, potentially applicable to multiple pathogens.

2.5 Methods

2.5.1 Protein design, expression and purification

Homology-guided protein resurfacing

The previously published RSV site II epitope-scaffold ("FFL_001") (*59*) was designed based on a crystal structure of a mutant of ribosome recycling factor from *E. coli* (PDB entry 1ISE. Using BLAST, we identified sequence homologs of 1ISE from eukaryotic organisms and created a multiple sequence alignment with clustal omega (CLUSTALO (1.2.1)) (*171*) of the mouse homolog sequence (NCBI reference NP_080698.1), 1ISE and FFL_001. Surface-exposed residues of FFL_001 were then mutated to the respective residue of the mouse homolog using the Rosetta fixed backbone design application (*152*), resulting in 38 surface mutations. Amino acid changes were verified to not impact overall Rosetta energy score term.

Protein expression and purification

FFLM

DNA sequences of the epitope-scaffold designs were purchased from Genscript and cloned in pET29b, in frame with a C-terminal 6x His tag. The plasmid was transformed in *E. coli* BL21 (DE3) and grown in Terrific Broth supplemented with Kanamycin (50 µg/ml). Cultures were inoculated to an OD_{600} of 0.1 from an overnight culture and incubated at 37 °C. After reaching OD_{600} of 0.6, expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside (IPTG) and cells were incubated for further 4-5h at 37 °C. Cell pellets were resuspended in lysis buffer (50 mM TRIS, pH 7.5, 500 mM NaCl, 5% Glycerol, 1 mg/ml lysozyme, 1 mM PMSF, 1 µg/ml DNase) and sonicated on ice for a total of 12 minutes, in intervals of 15 seconds sonication followed by a 45 seconds pause. Lysates were clarified by centrifugation (18,000 rpm, 20 minutes), sterile-filtered and purified using a His-Trap FF column on an Äkta pure system (GE healthcare). Bound proteins were eluted in buffer containing 50 mM Tris, 500 mM NaCl and 300 mM imidazole, pH 7.5. Concentrated proteins were further purified by size exclusion chromatography on a Superdex^{TM} 75 300/10 (GE Healthcare) in PBS. Protein concentrations were determined by measuring the absorbance at 280 nm on a Nanodrop (Thermo Scientific). Proteins were concentrated by centrifugation (Millipore, #UFC900324) to 1 mg/ml, snap frozen in liquid nitrogen and stored at -80 °C.

NRM

The full-length N gene (sequence derived from the human RSV strain Long, ATCC VR-26; GenBank accession number AY911262.1) was PCR amplified using the Phusion DNA polymerase (Thermo Scientific) and cloned into pET28a+ at NcoI-XhoI sites to obtain the pET-N plasmid. The sequence of FFLM was then PCR amplified and cloned into pET-N at NcoI site to the pET-NRM plasmid. *E. coli* BL21 (DE3) bacteria were co-transformed with pGEX-PCT (*172*) and pET-FFLM-N plasmids and grown in LB medium containing ampicillin (100 µg/ml) and kanamycin (50 µg/ml). The same volume of LB medium was then added, and protein expression was induced by the addition of 0.33 mM IPTG to the medium. Bacteria were incubated for 15 h at 28 °C and then harvested by centrifugation. For protein purification, bacterial pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.8, 60 mM NaCl, 1 mM EDTA, 2 mM dithiothreitol, 0.2% Triton X-100, 1 mg/ml lysozyme) supplemented with a complete protease inhibitor cocktail (Roche), incubated for one hour on ice, and disrupted by sonication. The soluble fraction was collected by centrifugation at 4 °C for 30 min at 10,000 x g. Glutathione-Sepharose 4B beads (GE Healthcare) were added to clarify supernatants and incubated at 4 °C for 15h. The beads were then washed one time in lysis buffer and two times in 20 mM Tris pH 8.5, 150 mM NaCl. To isolate NRM, beads containing bound complex were incubated with thrombine for 16 h at 20 °C. After cleavage of the GST tag, the supernantant was loaded onto a Sephacryl S-200 HR 16/30 column (GE Healthcare) and eluted in 20 mM Tris-HCl, 150 mM NaCl, pH 8.5.

Antibody variable fragments (Fabs)

For Fab expression, heavy and light chain DNA sequences were purchased from Twist Biosciences and cloned separately into the pHLSec mammalian expression vector (Addgene, #99845) using AgeI and XhoI restriction sites. Expression plasmids were pre-mixed in a 1:1 stoichiometric ratio, co-transfected into HEK293-F cells and cultured in FreeStyle^{TM} medium (Gibco, #12338018). Supernatants were harvested after one week by centrifugation and purified using a kappa-select column (GE Healthcare). Elution of bound proteins was conducted using 0.1 M glycine buffer (pH 2.7) and eluates were immediately neutralized by the addition of 1 M Tris ethylamine (pH 9), followed by buffer exchange to PBS pH 7.4.

Respiratory Syncytial Virus Fusion protein (prefusion RSVF)

Protein sequence of prefusion RSVF corresponds to the sc9-10 DS-Cav1 A149C Y458C S46G E92D S215P K465Q variant designed by Joyce et al. (*84*), which we refer to as RSVF DS2. RSVF DS2 was codon optimized for mammalian expression and cloned into the pHCMV-1 vector together with two C-terminal Strep-Tag II and one 8x His tag. Plasmids were transfected in HEK293-F cells and cultured in FreeStyle^{TM} medium. Supernatants were harvested one week after transfection and purified via Ni-NTA affinity chromatography. Bound protein was eluted using buffer containing 10 mM Tris, 500 mM NaCl and 300 mM Imidazole (pH 7.5), and eluate was further purified on a StrepTrap HP affinity column (GE Healthcare). Bound protein was eluted in 10mM Tris, 150 mM NaCl and 20 mM Desthiobiotin (Sigma), pH 8, and size excluded in PBS, pH 7.4, on a Superdex 200 Increase 10/300 GL column (GE Healthcare) to obtain trimeric RSVF.

2.5.2 Mouse immunizations and serum analysis

Ethics statement

All animal experiments were approved by the Veterinary Authority of the Canton of Vaud (Switzerland) according to Swiss regulations of animal welfare (animal protocol number 3074).

Mouse immunizations

Six-week-old, female Balb/c mice were ordered from Janvier labs and acclimatized for one week. Immunogens were thawed on ice and diluted in PBS pH 7.4 to a concentration of 0.2 mg/ml. The immunogens were then mixed with an equal volume of 2% Alhydrogel (Invivogen), resulting in a final Alhydrogel concentration of 1%. Other adjuvants were formulated according to manufacturer's instructions. After mixing immunogens and adjuvants for one hour at 4°C, each mouse was injected with 100 μl, corresponding to 10 μg immunogen adsorbed to Alhydrogel. All immunizations were done subcutaneously, with no visible irritation around the injection site. Immunizations were performed on day 0, 21 and 42. 100-200 μl blood were drawn on day 0, 14, 35, and the maximum amount of blood (200-1000μl) was taken by cardiac puncture at day 56, when mice were sacrificed.

Antigen ELISA

Nunc Medisorp plates (Thermo Scientific, # 467320) were coated overnight at 4°C with 100 μl of antigen (recombinant RSVF, FFLM and NRM) diluted in coating buffer (100 mM sodium bicarbonate, pH 9) at a final concentration of 0.5 μg/ml. For blocking, plates were incubated for two hours at room temperature with blocking buffer (PBS + 0.05% Tween 20 (PBST) supplemented with 5% skim milk powder (Sigma, #70166)). Mouse sera were serially diluted in blocking buffer and incubated for one hour at room temperature. Plates were washed five times with PBST before adding 100 μl of anti-mouse HRP-conjugated secondary antibody diluted at 1:1500 in blocking buffer (abcam, #ab99617). An additional five washes were performed before adding Pierce TMB substrate (Thermo Scientific, # 34021). The reaction was stopped by adding 100 μl of 2M sulfuric acid, and absorbance at 450 nm was measured on a Tecan Safire 2 plate reader. Each plate contained a standard curve of Motavizumab to normalize signals between different plates and experiments. Normalization was done in GraphPad Prism. The mean value was plotted for each cohort and statistical analysis was performed using GraphPad Prism.

Whole-virus ELISA

Nunc MaxiSorp ELISA plates (Thermo Scientific, # 44-2404-21) were coated with heat inactivated, frozen-thawed cell lysates from Hep2 cells which were infected for 48 hours with RSV (*149*). A control lysate was prepared from uninfected Hep2 cells to subtract background signals. ELISA was performed as described for the antigen ELISA.

Competition ELISA

Prior to incubation with a coated antigen plate, sera were serially diluted in the presence of 100 μg/ml competitor antigen and incubated overnight at 4°C. ELISA curves of a positive control, Motavizumab, are shown Fig S2.10. Curves were plotted using GraphPad Prism, and the area under the curve (AUC) was calculated for the specific (NRM) and control (RSVN) competitor. % competition was calculated using the following formula (*78*):

$$\text{\% competition} = \left(1 - \left(\frac{AUC(specific\ competitor\ (NRM))}{AUC(control\ competitor\ (NR))}\right)\right) * 100$$

Peptide sandwich ELISA

The antigenic site II was synthesized as peptide by JPT Peptide Technologies, Germany. The following sequence was synthesized and biotinylated at the N-terminus:

MLTNSELLSKINDMPITNDQKKLMSNNVQI

For ELISA analysis of peptide-reactive serum antibodies, Nunc MediSorp plates were coated with 5 μg/ml streptavidin (Thermo Scientific, #21122) for one hour at 37°C. Subsequently, ELISA plates were blocked as indicated above, followed by the addition of 2.4 μg/ml of the biotinylated site II peptide. Coupling was performed for one hour at room temperature. The subsequent steps were performed as described for the antigen ELISA.

Serum competition using Surface Plasmon Resonance

Approximately 300 RU of antigen were immobilized via amine coupling on a CM5 chip. Mouse sera were diluted 1:100 in HBS-EP+ running buffer and flowed as analyte with a contact time of 120 seconds to obtain an initial response unit ($RU_{\text{non-blocked surface}}$). The surface was regenerated using 50 mM NaOH. Sequentially, Motavizumab was injected four times at a concentration of 2 μM, leading to complete blocking of Motavizumab binding sites as confirmed by signal saturation. The same serum dilution was reinjected to determine the remaining response ($RU_{\text{blocked surface}}$). The delta serum response ($\Delta\, SR$) corresponds to the baseline-subtracted, maximum signal of the injected sera.

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

Percent blocking was calculated as follows:

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

A schematic representation of the SPR experiment is shown in Fig S2.4.

Enzyme-linked immunospot assay (ELISPOT)

B-cell ELISPOT assays were performed using the Mouse IgG ELISpot HRP kit (Mabtech, #3825-2H) according to the manufacturer’s instructions. Briefly, mouse spleens were isolated, and pressed through a cell strainer (Corning, #352350) to obtain a single cell suspension. Splenocytes were resuspended in RPMI media (Gibco, #11875093) supplemented with 10% FBS (Gibco), Penicillin/Streptomycin (Gibco), 0.01 µg/ml IL2, 1 µg/ml R848 (Mabtech, #3825-2H) and 50 µM β-mercaptoethanol (Sigma) for ~60 hours stimulation at 37 °C, 5% CO_2. ELISpot plates (PVDF 96-well plates, Millipore, #MSIPS4510) were coated overnight with 15 µg/ml antigen diluted in PBS, followed by careful washing and blocking using RPMI + 10% FBS. Live splenocytes were counted and the cell number was adjusted to $1x10^7$ cells/ml. Serial dilutions of splenocytes were plated in duplicates and incubated overnight with coated plates. After several wash steps with PBS buffer, plates were incubated for two hours with biotinylated anti-mouse total IgG (Mabtech, # 3825-6-250) in PBS, followed by incubation with streptavidin-conjugated to HRP (Mabtech, #3310-9) for one hour. Spots were revealed using tetramethylbenzidine (TMB, Mabtech, #3651-10) and counted with an automatic reader (Bioreader 2000; BioSys GmbH). Results were represented as number of spots per 10^6 splenocytes.

RSV neutralization assay

The RSV A2 strain carrying a luciferase gene (RSV-Luc) was a kind gift of Marie-Anne Rameix-Welti, UFR des Sciences et de la Santé, Paris. Hep2 cells were seeded in Corning 96-well tissue culture plates (Sigma, #CLS3595) at a density of 40,000 cells/well in 100 μl of Minimum Essential Medium (MEM, Gibco, #11095-080) supplemented with 10% FBS (Gibco, 10500-084), L-glutamine 2 mM (Gibco, #25030-081) and penicillin-streptomycin (Gibco, #15140-122), and grown overnight at 37 °C with 5% CO2. Sera were heat-inactivated for 30 minutes at 56 °C. Serial two-fold dilutions were prepared in an untreated 96-well plate using MEM without phenol red (M0, Life Technologies, #51200-038) containing 2 mM L-glutamine, penicillin + streptomycin, and mixed with 800 pfu/well RSV-Luc (corresponding to a final MOI of 0.01). After incubating diluted sera and virus for one hour at 37 °C, growth media was removed from the Hep2 cell layer and 100 μl/well of the serum-virus mixture added. After 48 hours, cells were lysed in 100 µl buffer containing 32 mM Tris pH 7.9, 10 mM $MgCl_2$, 1.25% Triton X-100, 18.75% glycerol and 1 mM DTT. 50 µl lysate were transferred to a 96-well plate with white background (Sigma, # CLS3912). 50 µl of lysis buffer supplemented with 1 µg/ml luciferin (Sigma, #L-6882) and 2 mM ATP (Sigma, #A3377) were added to each well immediately before reading luminescence signal on a Tecan Infinite 500 plate reader. On each plate, a Palivizumab dilution series was included to ensure comparability of neutralization data. In our assay, we determined IC_{50} values for Palivizumab of

0.32 µg/ml, which is similar to what other groups have reported (*153*). The neutralization curve was plotted and fitted using the GraphPad variable slope fitting model, weighted by $1/Y^2$.

Serum fractionation

400 µl of streptavidin agarose beads (Thermo Scientific, #20347) were pelleted at 13,000 rpm for 2 minutes in a table top centrifuge and washed with phosphate buffered saline (PBS). 200 µg of biotinylated site II peptide were incubated for 2 hours at room temperature to allow coupling of biotinylated peptide to streptavidin beads. Beads were washed three times with 1 ml PBS to remove excess of peptide and resuspended to a total volume of 500 µl bead slurry. Mouse sera from the same cohort (n=10) were pooled (4 µl each, 40 µl total) in a total volume of 200 µl PBS, and 90 µl diluted sera were mixed with 150 µl of bead slurry, followed by an overnight incubation at 4 °C. Beads were pelleted by centrifugation and the supernatant carefully removed by pipetting. Beads were then washed twice with 200 µl PBS and the wash fractions were discarded. To elute site II-specific antibodies, beads were resuspended in 200 µl elution buffer (0.1 M glycine, pH 2.7) and incubated for 1 minute before centrifugation. Supernatant was removed, neutralized with 40 µl neutralization buffer (1 M Tris pH 7.5, 300 mM NaCl), and stored at -20 °C for subsequent testing for RSV neutralization. As a control, unconjugated streptavidin was used for each sample to account for non-specific binding.

2.5.3 Biophysical characterization

Affinity determination using Surface Plasmon Resonance

Surface Plasmon Resonance experiments were performed on a Biacore 8K at room temperature with HBS-EP+ running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20) (GE Healthcare). Approximately 100 response units (RU) of FFLM were immobilized via amine coupling on a CM5 sensor chip (GE Healthcare). Serial dilutions of site II-specific antibody variable fragments (fabs) were injected as analyte at a flow rate of 30 µl/min with 120 seconds contact time. Following each injection cycle, ligand regeneration was performed using 0.1 M glycine, pH 2. If not stated otherwise, data analysis was performed using 1:1 Langmuir binding kinetic fits within the Biacore evaluation software (GE Healthcare).

Negative-stained sample preparation, data acquisition, and image processing

NRM was size excluded in PBS on a Superose 6 column (GE Healthcare) and diluted to a concentration of 0.015 mg/ml. The sample was adsorbed to a glow-discharged carbon-coated copper grid (EMS, Hatfield, PA, USA) washed with deionized water and stained with a solution of uranyl formate 0.75%. Observation was made using an F20 electron microscope (Thermo Fisher, Hillsboro, USA) operated at 200 kV. Digital images were collected using a direct detector camera Falcon III (Thermo Fisher, Hillsboro, USA) 4098 X 4098 pixels. Automatic data collection was performed using EPU software (Thermo Fisher, Hillsboro, USA) at a nominal magnification of 50,000X, corresponding to a pixel size of 2 Å, and defocus range from -1 µm to -2 µm.

Contrast transfer function for each image was estimated using CTFFIND4 (*173*). 1,000 particles of nanorings were picked using XMIPP manual-picking utility within SCIPION framework (*174*). Manually picked particles were used as input into XMIPP auto-picking utility, resulted in 13,861 particles. Particles were extracted and binned to have the box size of 100 pixels, corresponding to the pixel size of 4 Å, phase-flipped and subjected for three rounds of reference-free 2D classification without CTF correction in RELION-3.0 Beta (*175*).

2.5.4 Next-generation antibody repertoire sequencing (NGS)

RNA isolation

Mouse bone marrow was isolated from femurs and re-suspended in 1.5 ml Trizol (Life Technologies, #15596) and stored at -80˚C until further processing. RNA extraction was performed using the PureLink RNA Mini Kit (Life Technologies, #12183018A) following the manufacturer guidelines.

Antibody sequencing library preparation

Library preparation for antibody variable heavy chain regions was performed using a protocol that incorporates unique molecular identifier (UID) tagging, as previously described in Khan *et al.* (*176*). Briefly, first-strand cDNA synthesis was performed by using Maxima reverse transcriptase (Life Technologies, #EP0742) following the manufacturer instructions, using 5 µg RNA with 20 pmol of IgG gene-specific primers (binding IgG1, IgG2a, IgG2b, IgG2c, and IgG3) with an overhang of a reverse UID (RID). After cDNA synthesis, samples were subjected to a left-hand sided SPRIselect bead (Beckman Coulter, #B23318) cleanup at 0.8X. Quantification of target-specific cDNA by a digital droplet (dd)PCR assay allowed exact input of 135000 copies into the next PCR step. Reaction mixtures contained a forward multiplex primer set that was specific for variable heavy region framework 1 and possessed forward UID (FID), a 3' Illumina adapter specific reverse primer, and 1X KAPA HIFI HotStart Uracil+ ReadyMix (KAPA Biosystems, #KK2802). PCR reactions were then left-hand side SPRIselect bead cleaned as before and quantified using ddPCR assay. Finally, an Illumina adaptor-extension PCR step was carried out using 820000 copies of the previous PCR product. Following 2nd-step adaptor-extension PCR, reactions were cleaned using a double-sided SPRIselect bead cleanup process (0.5X-0.8X) and eluted in TE buffer.

NGS with Illumina MiSeq (2 x 300 bp)

After library preparation, individual NGS libraries were characterized for quality and quantified by capillary electrophoresis using a Fragment Analyzer (Advanced Analytical DNF-473 Standard Sensitivity). Samples were then pooled and NGS was performed on the Illumina MiSeq platform with a MiSeq Reagent Kit V3, 2 x 300 bp paired-end (Illumina, #MS-102-3003), using an input concentration of 10 pM with 10% PhiX.

Error and bias correction

Error and bias correction were performed using molecular amplification fingerprinting pipeline, as previously described (*176, 177*).

Bioinformatic preprocessing

Paired-end FASTQ files obtained from Illumina MiSeq were imported into CLC Genomics Workbench 10 on the ETH Zurich Euler High Performance Computing (HPC) cluster. A preprocessing workflow was run containing the following steps: trimming of low-quality reads, merging of paired-end reads, removal of sequences not aligning to mouse IGH constant sequences, and length filtering.

Error correction by consensus building

After pre-processing all datasets were downsampled to contain the same amount of sequencing reads as the dataset with the lowest overall number of reads (361749 sequencing reads). For error correction, a custom Python script was used to perform consensus building on the sequences for which at least three reads per UID were required. VDJ annotation and frequency calculation was then performed by our in-house aligner

(*176, 177*). The complete error-correction and alignment pipeline is available under https://gitlab.ethz.ch/reddy/MAF.

Sequence analysis and data visualization

Data analysis was done by customized scripts in R. For the identification of clonotypes hierarchical clustering (*177*) was utilized to group CDR3 sequences together. The following parameters were used: identical IGHV and IGHJ gene segment usage, identical CDR3 length, and at least 80% CDR3 amino acid similarity to one other sequence in the given clonotype (single linkage). The overlap of clonotypes between both cohorts was analyzed by extracting the 20 most expanded clonotypes from each cohort and visualizing their size, occurrence, and Vgene usage by a circos plot using R software circlize (*178*). CDR3 sequence similarities between overlapping clonotypes were represented graphically with the R software motifStack (*179*). All scripts are available upon request.

2.6 Supplementary Information

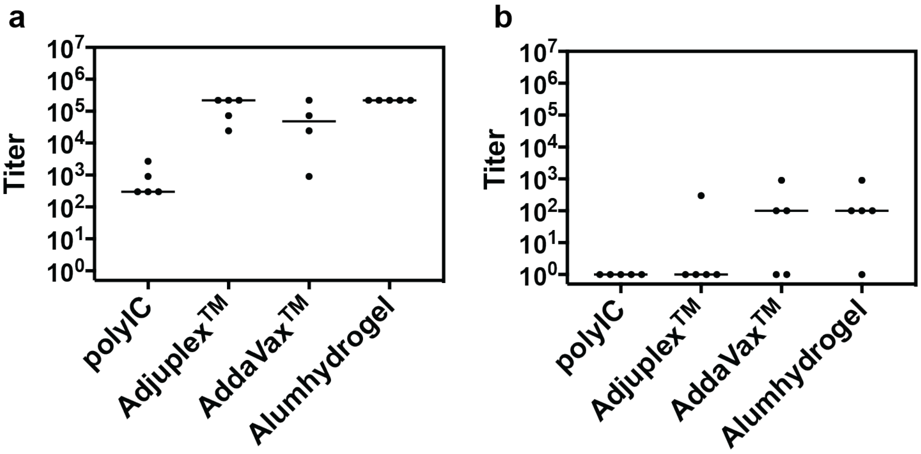


Figure S 2.1: Adjuvant screen for FFL_001 immunogen.
Female Balb/c mice (five animals/group) were immunized three times (day 0, 21, 42) with 10 µg FFL_001 monomer adsorbed to different adjuvants and serum was analysed on day 56. **a)** Immunogenicity of FFL_001 formulated in different adjuvants. Serum titers were determined against FFL_001 at day 56 of the immunization protocol. FFL_001 emulsed in Alhydrogel® showed highest overall immunogenicity. **b)** Prefusion RSVF cross-reactivity of FFL_001 immunized mice after three immunizations. 4/5 mice immunized with FFL_001 formulated in Alhydrogel® showed serum cross-reactivity with prefusion RSVF.

a

```
CLUSTAL O(1.2.4) multiple sequence alignment


FFL_001      GSRSDMRKDAERRFDKFVEAAKNKFDKFKAALRKGDIKEERRKDMKKLARKEAEQARRAV    60
FFLM         ASREDMREEADEDFKSFVEAAKDNFNKFKARLRKGKITREHREMMKKLAKQNANKAKEAV    60
             .**.***::*:. *..******::*:**** ****.*..*:*: *****::*::*:.**

FFL_001      RNRLSELLSKINDMPITNDQKKLMSNDVLKFAAEAEKKIEALAADAEDKFTQ 112
FFLM         RKRLSELLSKINDMPITNDQKKLMSNQVLQFADDAEAEIDQLAADATKEFTG 112
             *:************************:**:** :** :*: ***** .:**
```

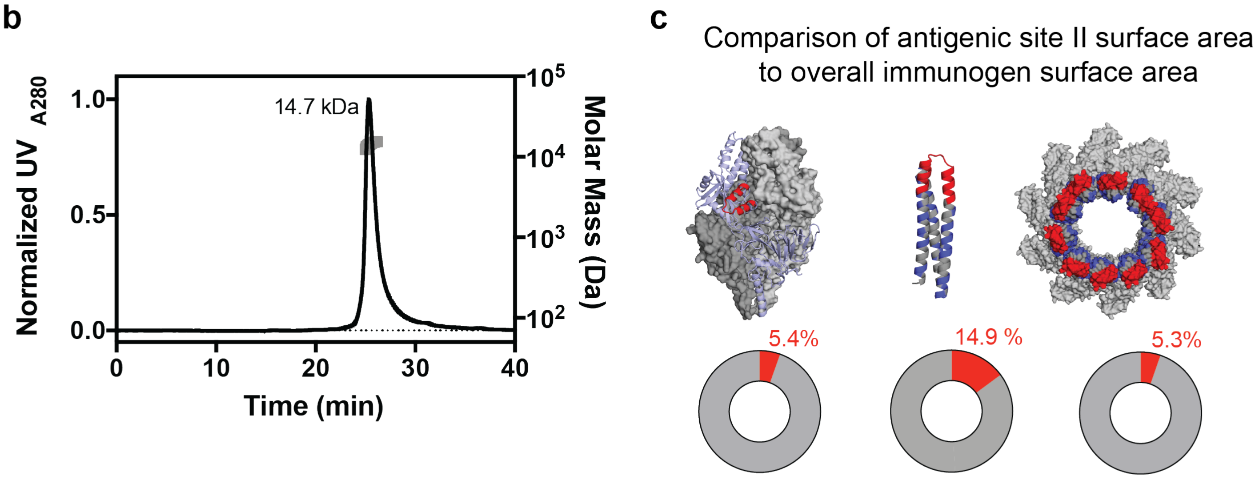


Figure S 2.2: Homology-guided resurfacing of FFL_001.

a) Sequence alignment of FFL_001 and FFLM. **b)** Resurfaced variant FFLM is monomeric in solution as assessed by size-exclusion coupled to an on-line multi-angle-light scattering detector. Determined mass in solution is 14.7 kDa ± 3.5 %, which is close to the theoretical molecular weight of 14.4 kDa. **c)** Relative surface area of RSVF antigenic site II in pre-fusion RSVF (PDBID 4JHW), FFLM and NRM (model based on RSVN structure with PDBID 2WJ8). The Motavizumab epitope is highlighted in red, blue patches indicate sequence changes of FFLM compared to FFL_001, and piecharts show the fraction of antigenic site II surface area compared to overall immunogen surface area. Solvent accessible surface area (SASA) was computed in PyMol in presence and absence of Motavizumab. % SASA of antigenic site II is nearly identical when comparing RSVF and NRM, whereas the FFLM monomer shows approximately three-fold greater relative surface area of antigenic site II, due to its small size.

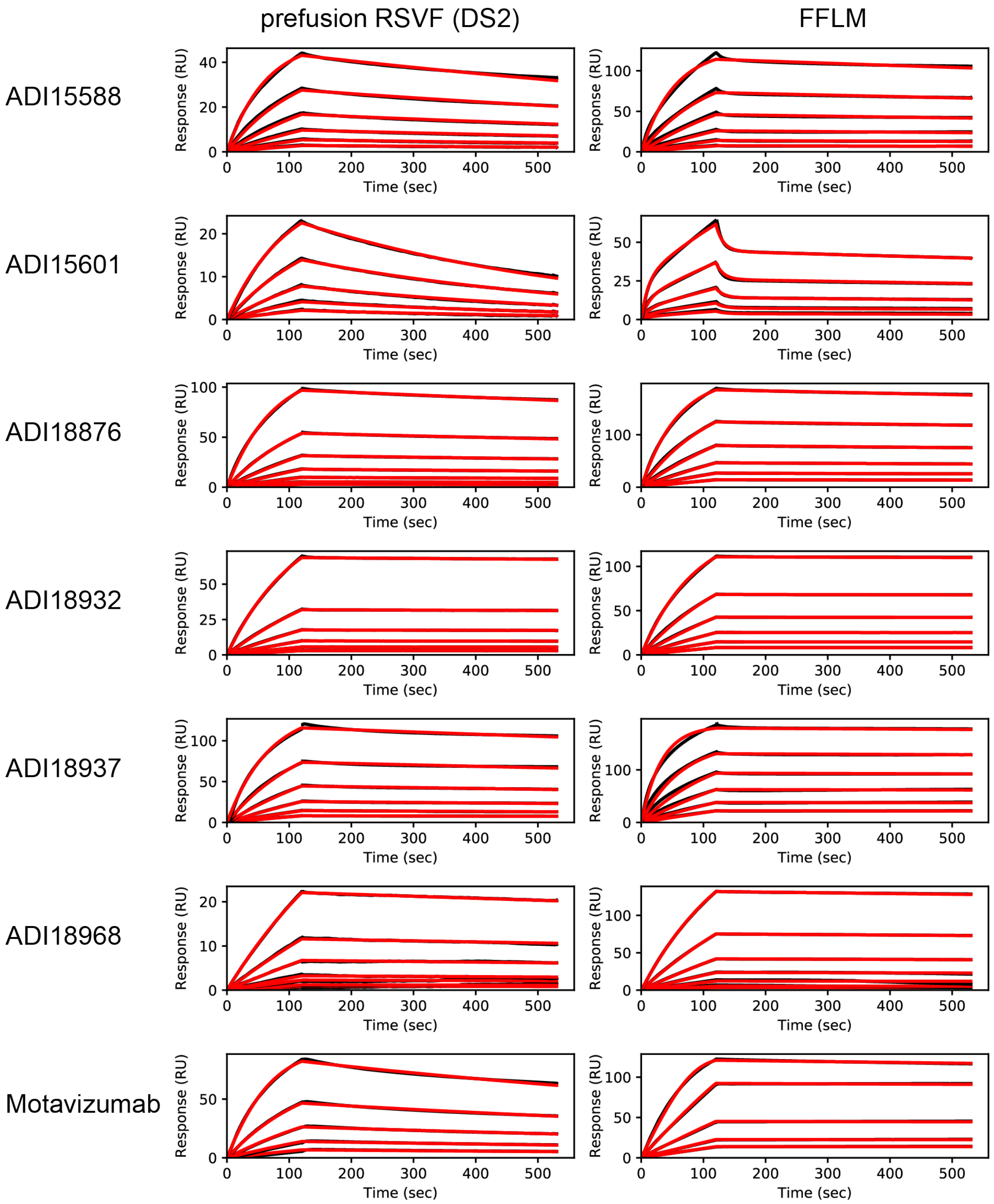


Figure S 2.3: SPR sensorgrams for site II nAbs.

Prefusion RSVF or FFLM were immobilized on the sensor chip surface via amine coupling. Serial dilutions of site II-specific antibody variable fragments (fabs) were injected as analyte. With the exception of ADI15601, which was fitted to a two-state reaction model for binding to FFLM, all data fitted to a 1:1 Langmuir model within the Biacore evaluation software (GE Healthcare).

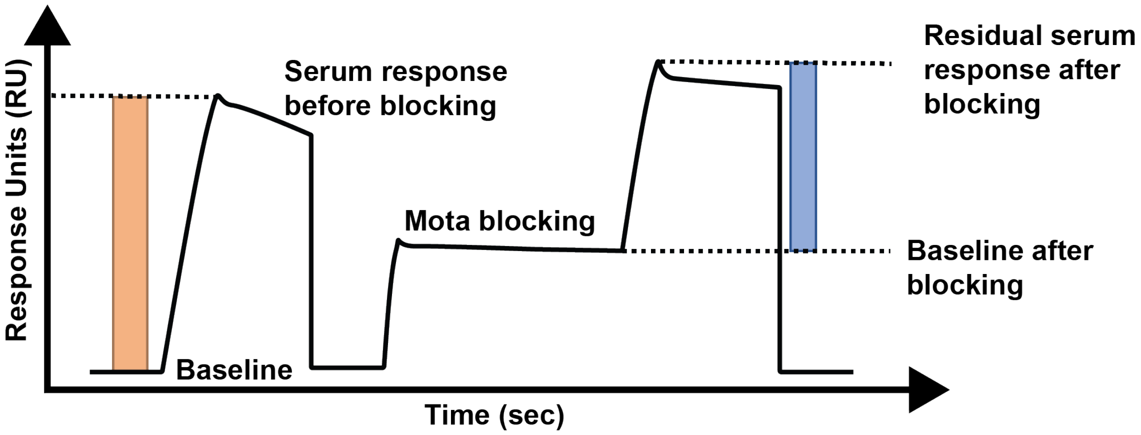

Response Units (RU)
Serum response before blocking
Residual serum response after blocking
Mota blocking
Baseline after blocking
Baseline
Time (sec)

Figure S 2.4: Schematic representation of the SPR serum competition assay.
Mouse sera were injected on an antigen coated sensor chip surface to measure initial response (orange). Following regeneration, Motavizumab binding sites were blocked with saturating amounts of Motavizumab. Residual serum response was determined on a blocked surface (blue). For data analysis, response units at indicated timepoints were extracted, and % competition was calculated as described in methods.

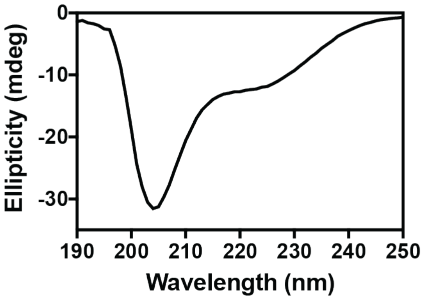


Figure S 2.5: Far-ultraviolet circular dichroism spectrum of antigenic site II peptide.
The site II peptide adopts a flexible conformation in solution, measured in phosphate-buffered saline buffer at 25 °C.

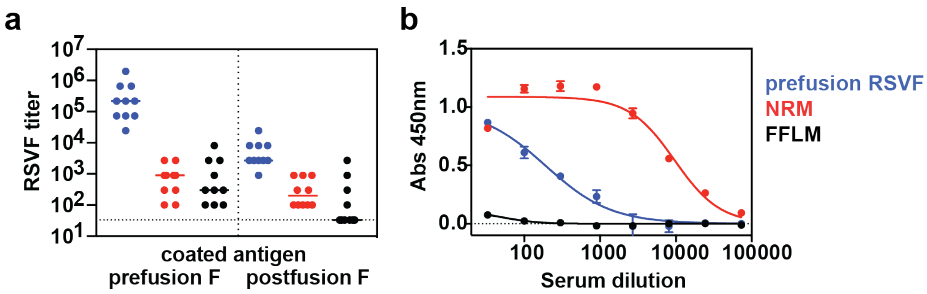


Figure S 2.6: Serum reactivity with RSVF and viral lysate.
Mice immunized with synthetic immunogen show low levels of cross-reactivity with recombinant RSVF but negligible binding to viral lysate. Mice were immunized three times with prefusion RSVF, NRM or FFLM as shown in Figure 2.2. **a)** Sera from day 56 was analyzed by ELISA for binding to prefusion and postfusion RSVF. Prefusion RSVF immunized mice showed lower reactivity to postfusion RSVF than to the prefusion form. FFLM and NRM immunized mice showed low levels of cross-reactivity with pre- and postfusion RSVF. Data shown are from one out of three independent experiments. **b)** Day 56 sera from ten mice were pooled, and tested for binding to lysate of Hep2 cells which had been infected for 48 hours with RSV. As background control, non-infected Hep2 cell lysate was prepared, and curves shown were background-subtracted. NRM immunized mouse sera strongly react with viral lysate, whereas mice immunized with FFLM only show negligible binding to viral lysate. The strong reactivity of NRM immunized mice derives from antibodies raised against the RSVN carrier protein. Sera from prefusion RSVF immunized mice are shown as control. Data shown are from one experiment performed in triplicates.

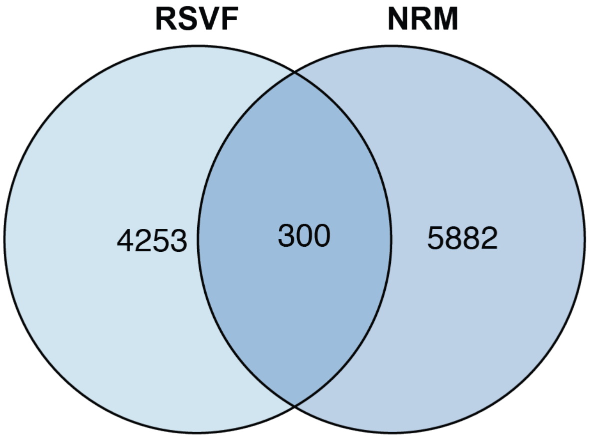


Figure S 2.7: Overlapping antibody clonotypes of mice immunized with RSVF and NRM.
When comparing antibody clonotypes, defined as the same VH gene and 80% sequence similarity in the HCDR3, NRM and RSVF immunizations yield 300 overlapping clonotypes.

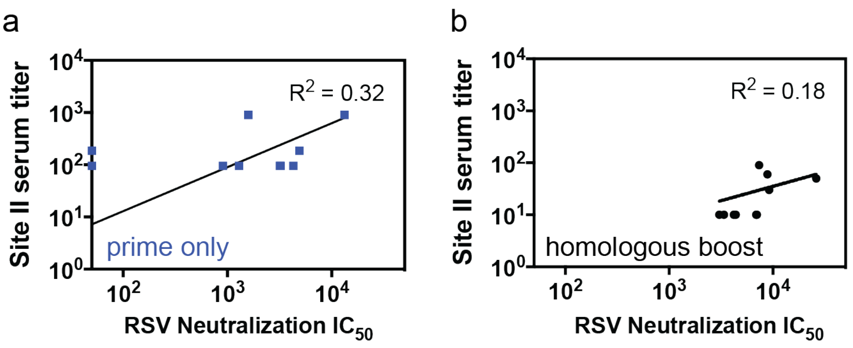


Figure S 2.8: Correlation of site II peptide specific serum titer with RSV neutralization IC_{50}.
Correlations for the prime only mouse cohort (a) and the homologous boost cohort (b). Data represent the mean of two independent experiments, each measured in duplicates. Pearson correlation coefficients (r^2) and p-values were calculated in GraphPad Prism.

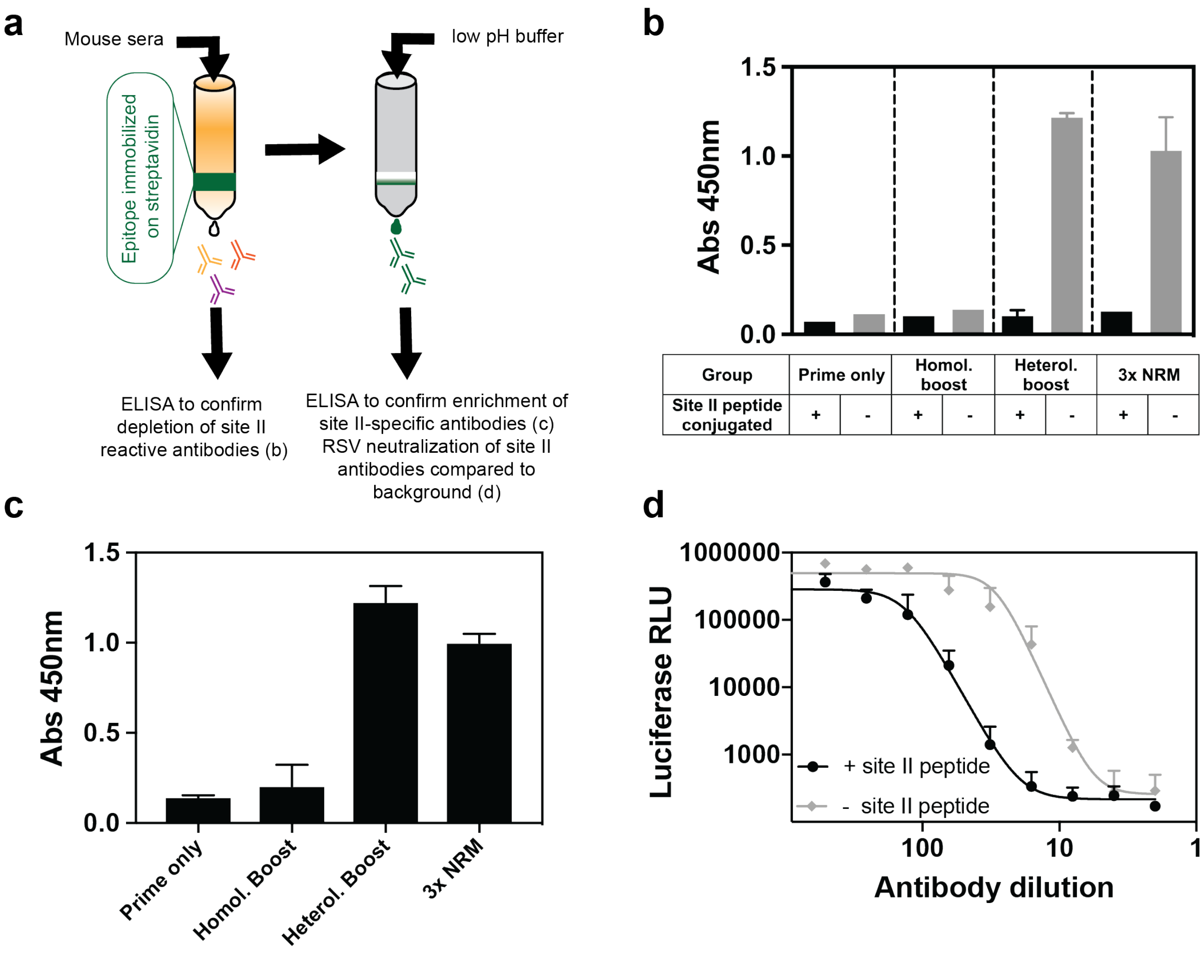


Figure S 2.9: Enrichment of site II-specific antibodies from mouse sera.
a) Experimental setup. Streptavidin agarose beads were conjugated to biotinylated antigenic site II peptide. As control, unconjugated streptavidin beads were prepared. Sera from ten mice within each cohort were pooled and mixed with conjugated and unconjugated beads. Column flow-through and elution fractions were analyzed by ELISA (b,c), and eluted site II-specific antibodies were analyzed in an RSV neutralization assay (d). **b)** Analysis of column flow-through for site II peptide reactivity by ELISA. Immunization groups as described in Figure 2.4 (prime only, homologous boost, heterologous boost and 3x NRM). ELISA signal (OD at 450 nm) for site II peptide reactivity is shown for column flow-

through from serum fractionation as depicted in (a). Streptavidin beads that were not coupled to antigenic site II peptide were used as controls, and did not deplete site II reactivity in the flow-through. Data and error bars presented are averaged from two independent experiments. **c)** ELISA against antigenic site II peptide of the elution fractions as shown in (a). Antibodies eluted bound specifically to the antigenic site II peptide. **d)** Example RSV neutralization assay curves from elution fractions, obtained from site II conjugated (black) or unconjugated streptavidin beads (grey). Luciferase signal is plotted on the y-axis, and is a measure for RSV replication. The dilution factor of purified antibodies is indicated on the x-axis. Data shown are from one experiment performed in duplicates.

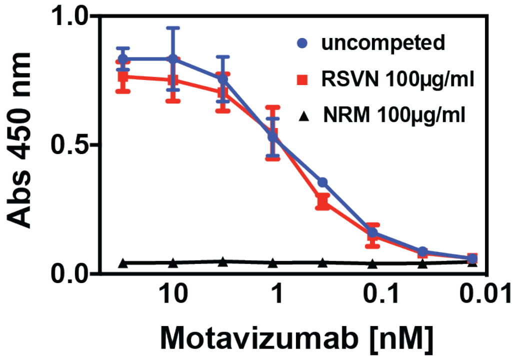


Figure S 2.10: Competition ELISA with Motavizumab antibody control.

Plates were coated with prefusion RSVF as described in methods. Three-fold serial dilutions of Motavizumab (initial concentration = 30 nM) were prepared in presence of different competitors (RSVN, NRM or none). Following overnight competition at 4°C, binding of Motavizumab to RSVF was measured. As expected, RSVN competition did not affect RSVF binding of Motavizumab. In contrast, NRM efficiently competed with RSVF for Motavizumab binding at the indicated competitor concentration. Data shown are from one experiment, with error bars derived from technical duplicates.

Chapter 3 Trivalent cocktail of *de novo* designed immunogens enables the robust induction and focusing of functional antibodies *in vivo*

This chapter is based on an article that is currently under peer review. A preprint is available in bioRxiv (doi: https://doi.org/10.1101/685867).

Authors

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

***These authors contributed equally.**

Affiliations:

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

Author contributions:

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

3.1 Abstract

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

3.2 Introduction

Efforts to design novel proteins from first principles have revealed a variety of rules to control the structural features of *de novo* proteins (*37, 40, 180, 181*). However, the *de novo* design of functional proteins has been far more challenging (*19*). A commonly used strategy to design *de novo* functional proteins is to transplant the binding motifs found in existing protein structures to pre-existing or *de novo* templates. In nearly all cases, the binding motifs transplanted were commonly found in existing protein structures, such as linear helical segments, allowing the grafting of such motifs without extensive backbone adjustments (*48, 49, 51*). Most protein functional sites, however, are not contained within a single, regular segment in protein structures but arise from the three-dimensional arrangement of several, often irregular, structural elements that are supported by defined topological features of the overall structure (*57, 58, 182*). As such, it is of utmost importance for the field to develop computational approaches to endow *de novo* designed proteins with irregular and multi-segment complex structural motifs that can perform the desired functions.

Functional protein design has raised expectations in the domain of immune response modulation; in particular, on the induction of neutralizing antibodies (nAbs) *in vivo* (*59*). Inducing nAbs targeting defined epitopes remains an overarching challenge for vaccine development (*68*). Our increasing structural understanding of many nAb-antigen interactions has provided templates for the rational design of immunogens for respiratory syncytial virus (RSV), influenza, HIV, dengue and others (*130, 183, 184*). Despite this extensive structural knowledge, these and other pathogens are still lacking efficacious vaccines, highlighting the need for next-generation vaccines to efficiently guide antibody responses towards key neutralization epitopes in both naïve and pre-exposed immune systems. The elicitation of antibody-responses with defined epitope specificities has been an enduring challenge for immunogens derived from modified viral proteins (*159*).

Recently, Correia and colleagues (*59*) have shown that computationally designed immunogens could elicit epitope-specific responses. The RSVF antigenic site II, a linear helix-turn-helix motif, was transplanted onto a heterologous protein scaffold, which elicited nAbs in non-human-primates (NHPs) after repeated boosting immunizations. Despite being a proof-of-principle for the induction of functional antibodies using a computationally designed immunogen, several major caveats emerged; namely, the lack of applicability of the computational approach to structurally complex epitopes, and the inconsistent neutralization titers observed in the immunogenicity studies.

To address these limitations, here we designed epitope-focused immunogens mimicking irregular and discontinuous RSV neutralization epitopes (site 0 (*1*) and IV (*3*), Fig 3.1) and showcase two computational design methodologies that enable the presentation of these structurally challenging motifs in *de novo* designed proteins. *In vivo*, cocktail formulations including an optimized site II immunogen (*185*) yielded consistent neutralization levels above the protective threshold directed against all three epitopes. The design strategies presented provide a blueprint to engineer proteins stabilizing irregular and discontinuous binding sites, applicable to vaccine design for pathogens that require fine control over the antibody specificities induced, and more generally for the design of *de novo* proteins displaying complex functional motifs.

3.3 Results

3.3.1 *De novo* design of immunogens with structurally complex epitopes

Designing proteins with structurally complex functional sites has remained a largely unmet challenge in the field of computational protein design (*19*). We sought to design accurate mimetics of RSV neutralization epitopes, which have been structurally well characterized, and evaluate their functionality in immunization studies. We chose antigenic sites 0 and IV (Fig 3.1), which are both targeted by potent nAbs, and are structurally distinct from functional motifs that have previously been handled by computational protein design algorithms. The antigenic site 0 presents a structurally complex and discontinuous epitope consisting of a kinked 17-residue alpha helix and a disordered loop of 7 residues, targeted by nAbs D25 and 5C4 (*1, 186*), while site IV presents an irregular 6-residue bulged beta-strand and is targeted by nAb 101F (*3*).

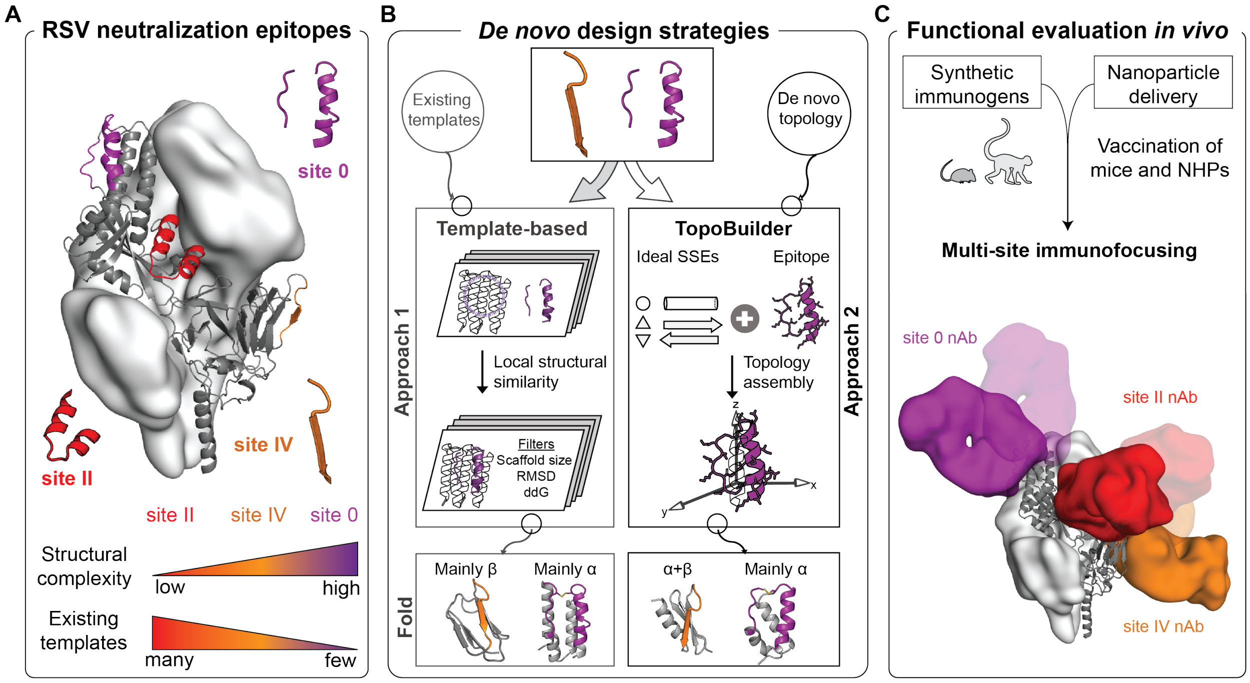

A
RSV neutralization epitopes
site 0
site IV
site II
site II
site IV
site 0
Structural complexity
low
high
Existing templates
many
few
B
De novo design strategies
Existing templates
De novo topology
Template-based
Local structural similarity
Filters
Scaffold size
RMSD
ddG
TopoBuilder
Ideal SSEs
Epitope
Topology assembly
Approach 1
Approach 2
Fold
Mainly β
Mainly α
α+β
Mainly α
C
Functional evaluation in vivo
Synthetic immunogens
Nanoparticle delivery
Vaccination of mice and NHPs
Multi-site immunofocusing
site 0 nAb
site II nAb
site IV nAb

Figure 3.1: Computational design of immunogens to elicit RSV nAbs focused on three distal epitopes. **(A)** Prefusion RSVF structure (PDB 4JHW) with sites 0, II and IV highlighted. An immunogen for site II was previously reported (Chapter 2). **(B)** Computational protein design strategies. Approach 1: Design templates were identified in the PDB based on loose structural similarity to site 0/IV, followed by *in silico* folding and design, and sequence optimization through directed evolution. Approach 2: A motif-centric *de novo* design approach was developed ("TopoBuilder") to tailor the protein topology to the motif's structural constraints. Bottom: Computational models of designed immunogens using different approaches. **(C)** Cocktail formulations of three synthetic immunogen nanoparticles to elicit nAbs focused on three non-overlapping epitopes.

The computational design of proteins mimicking structural motifs has previously been performed by first identifying compatible protein scaffolds, either from naturally occurring structures or built *de novo*, which then serve as design templates to graft the motif (*49, 51, 52, 54, 110*). Given the structural complexity of sites 0 and IV, this approach did not provide any promising matches, even using loose structural criteria (Fig 1.5).

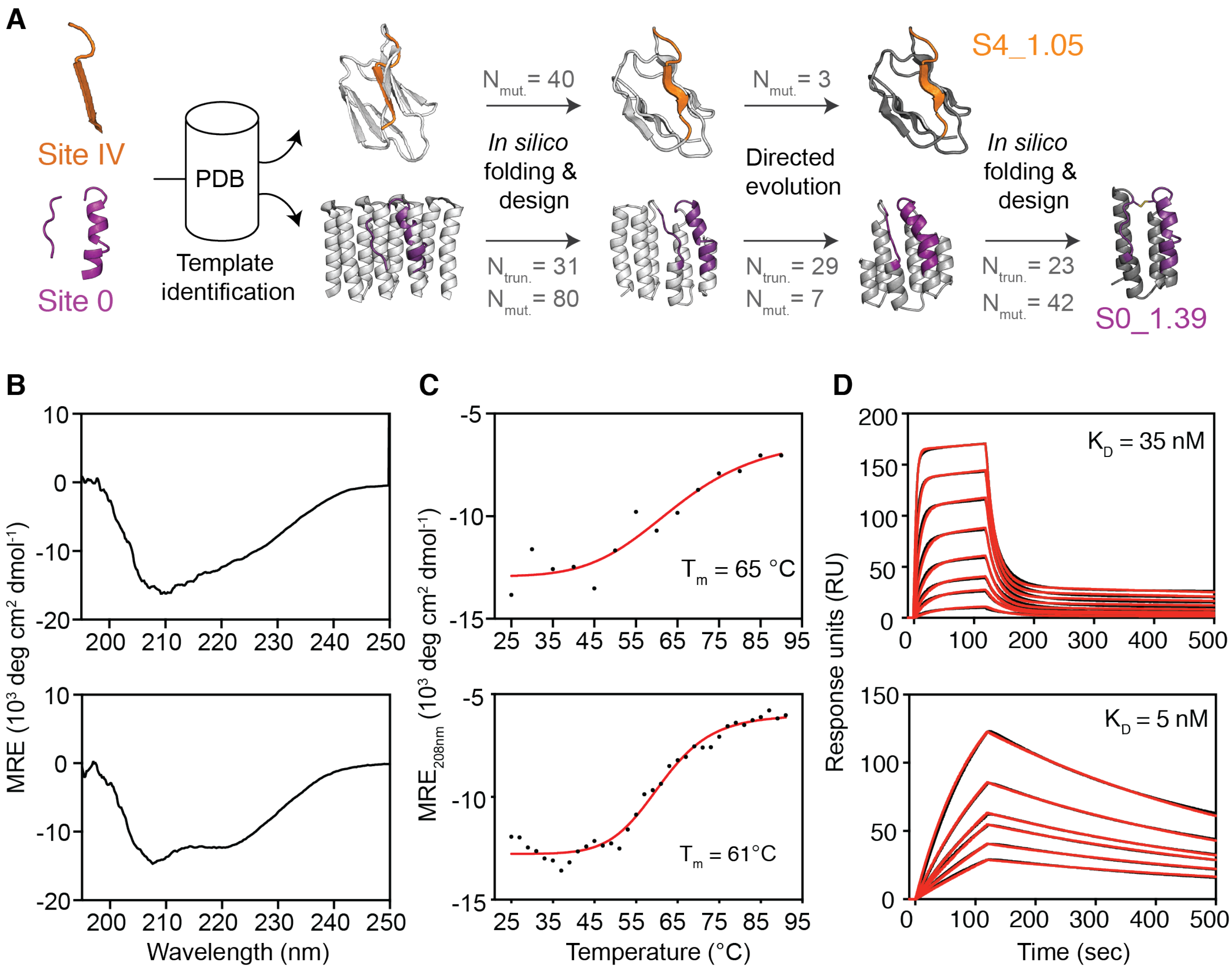


Figure 3.2: Templated computational design and biophysical characterization of synthetic immunogens.
(A) Protein design strategy - templates with structural similarity to sites IV and 0 were identified by native domain excision or loose structural matching, followed by *in silico* folding, design and directed evolution. An additional *in silico* folding and design step was necessary to install site 0 on a truncated template sequence revealed by directed evolution. Computational models of intermediates and final designs (S4_1.5 and S0_1.39) are shown, and the number of mutations (N_{mut}) and truncated residues (N_{trun}) are indicated for each step. **(B)** CD spectra measured at 20 °C of S4_1.5 (top) and S0_1.39 (bottom), are in agreement with the expected secondary structure content of the design model. **(C)** Thermal melting curves measured by CD at 208 nm in presence of 5 mM TCEP reducing agent. **(D)** Binding affinity measured by SPR against target antibodies 101F (top) and D25 (bottom). Sensorgrams are shown in black and fits in red lines. CD - Circular dichroism, T_m - melting temperature, SPR - Surface plasmon resonance.

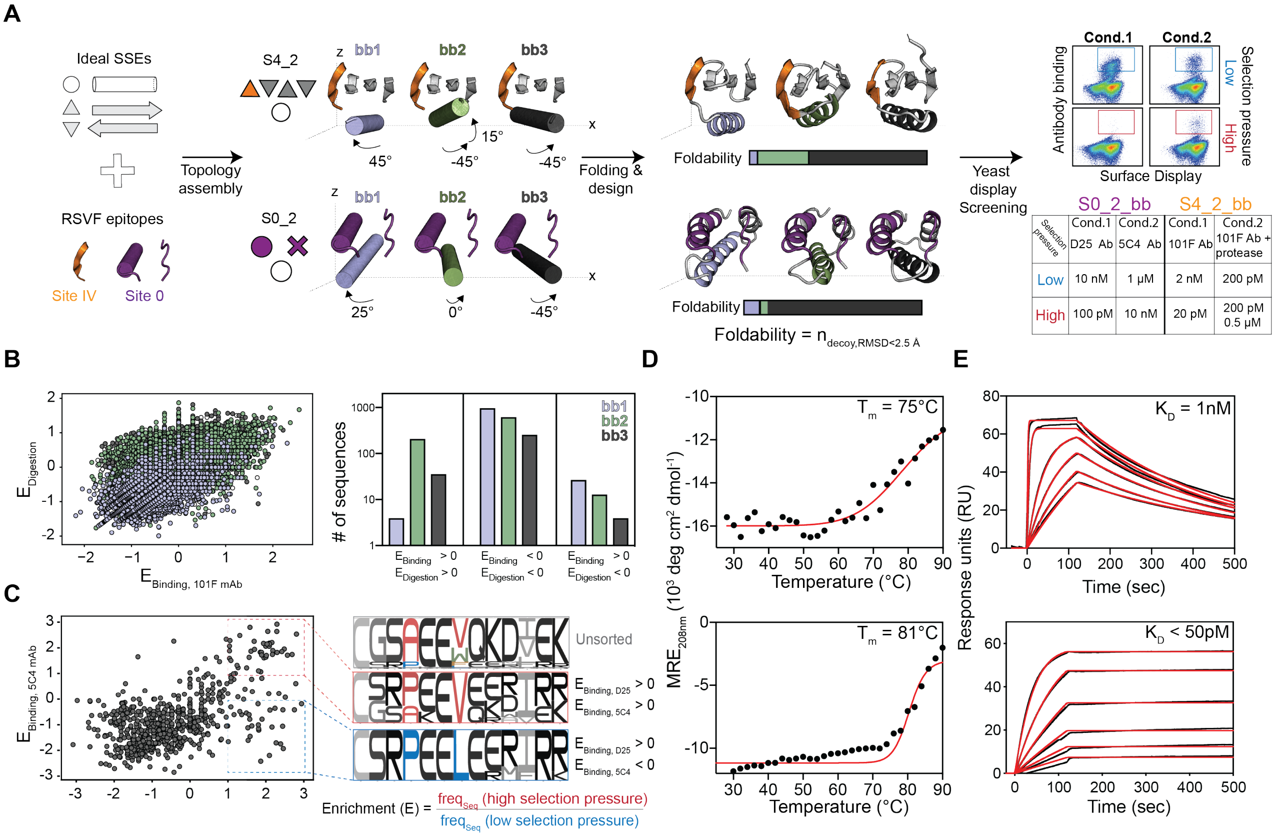
Thus, for site IV, we noticed that a small structural domain that resembles an immunoglobulin fold containing the epitope could be excised from the prefusion RSVF (preRSVF) structure. We hypothesized this would be a conservative approach to maintain its native, distorted epitope structure (Fig S3.1). We optimized the sequence for stability and epitope mimicry using Rosetta FunFolDes (*26*) (Fig 3.2a), and our best computational design (S4_1.1) bound with a K_D > 85 µM to the 101F target antibody. To improve binding affinity, we performed deep mutational scanning followed by next-generation sequencing, as previously described (*187*) (Fig S3.1). We tested combinations of enriched positions in recombinantly expressed proteins for antibody binding, obtaining a double mutant (S4_1.5) that bound with a K_D of 35 nM to the 101F target antibody, showed a circular dichroism (CD) spectrum of a folded protein, and was thermostable up to 65 °C (Fig 3.2b-d and Fig S3.2).

The discontinuous structure of site 0 was not amenable for domain excision and stabilization. Thus, we searched for template structures that mimicked the helical segment of the epitope, and simultaneously allowed grafting the loop segment, ultimately selecting a designed helical repeat protein as template (PDB 5cwj) (Fig 3.2a and Fig S3.3) (*41*). In order to avoid steric clashes with the target antibody D25, we truncated the N-terminal 31 residues of the 5cwj template, and performed *in silico* folding and design simulations to sample local and global changes on the scaffold to allow the presentation of the site 0 epitope (Fig 3.2a). Out of 9 sequences tested, the best design (S0_1.1) bound with a K_D of 1.4 µM to the D25 antibody (Fig S3.4), which is four orders of magnitude lower than the affinity of D25 to preRSVF (*82*). Following multiple rounds of directed evolution using yeast display, we found an enriched sequence that was C-terminally truncated by 29 residues, and showed a ~5-fold increased affinity towards D25 (Fig S3.3-S3.4). We used the truncated structure as a new template for *in silico* folding and design. Ultimately, this multi-stage process yielded S0_1.39, a design further truncated by additional 23 residues, which bound with 5 nM to D25 (Fig 3.2d). S0_1.39 was also recognized by the 5C4 antibody (Fig S3.4), which has been shown to engage site 0 in a different orientation compared to D25 (*186*), with an affinity of 5 nM, identical to that of the 5C4-preRSVF interaction (*82*).

The primary goals for the designs were achieved in terms of the stabilization of irregular and complex binding motifs in a conformation relevant for antibody binding, however, the overall strategy presented important limitations with respect to its general utility. Despite the large number of structures available to serve as design templates, the fraction of those that are practically useful for the design of functional proteins becomes increasingly limited with the structural complexity of the motif (Fig 1.5). As described above, suboptimal design templates require extensive backbone flexibility on the design process and multiple rounds of directed evolution until a sequence with high-affinity binding is identified. Additionally, the starting topology determines the overall shape of the designed protein, which may be suboptimal for the accurate stabilization of the motif, and may impose unwanted tertiary steric constraints that interfere with the designed function. In particular, for immunogen design it is desired to preserve native-like accessibility of the epitope in the context of the designed immunogen, thereby maximizing the induction of antibodies that can cross-react with the native antigen presented by the pathogen. An illustrative example on how a template-based design approach can fail to fulfil these criteria is the comparison between the quaternary environment of the site 0 epitope in preRSVF and S0_1.39 showing that this topology does not mimic such environment, albeit allowing the binding of several monoclonal antibodies (Fig S3.5).

To overcome these limitations, we developed a template-free design protocol - the TopoBuilder - that generates tailor-made topologies to stabilize complex functional motifs. Within the TopoBuilder, we parametrically sample the placement of idealized secondary structure elements which are then connected by loop segments, to assemble topologies that can stabilize the desired conformation of the structural motif. Next, these topologies are diversified to enhance structural and sequence diversity with a folding and design stage using Rosetta FunFolDes (see Fig S3.6 and methods for details). For this approach, we defined two new design objectives which were unmet by our previous template-based designs: (1) building stable *de novo* topologies that stabilize the epitope, while mimicking their native quaternary environment; (2) fine-tuning the topology's secondary structure arrangements to maximize the fold stability and optimize epitope presentation for high affinity antibody binding.

To present antigenic site IV, we designed a fold composed of a β-sheet with 4 antiparallel strands and one helix (Fig 3.3a), referred to as S4_2 fold. Within the S4_2 topology, we generated three structural variants (S4_2_bb1-3), by parametrically sampling three distinct helical secondary structural elements, varying both orientations and lengths to maximize the packing interaction against the ß-sheet. Sequences generated from two structural variants (S4_2_bb2 and S4_2_bb3) showed a strong propensity to recover the designed structures in Rosetta *ab initio* simulations (Fig 3.3a and Fig S3.7).



Selection pressure	S0_2_bb Cond.1 D25 Ab	S0_2_bb Cond.2 5C4 Ab	S4_2_bb Cond.1 101F Ab	S4_2_bb Cond.2 101F Ab + protease
Low	10 nM	1 µM	2 nM	200 pM
High	100 pM	10 nM	20 pM	200 pM 0.5 µM

Figure 3.3: Motif-centric *de novo* design of epitope-focused immunogens.
(A) Ideal secondary structure elements (SSEs) are assembled around RSVF epitopes, sampling different orientations within the same topology, followed by a single round of *in silico* folding and design. Rosetta *ab initio* simulations are performed for designs of each topology to assess its propensity to fold into the designed structures, returning a foldability score. Selected designs are then displayed on yeast surface and sorted under two different selection pressures for subsequent deep sequencing. **(B)** All three designed topological variants were screened for high affinity binding and resistance to chymotrypsin to select stably folded proteins. Enrichment analysis revealed a strong preference for one of the designed helix orientations (S4_2_bb2, green) to resist protease digestion and bind with high affinity to 101F. **(C)** Enrichment analysis of sorted populations under high and low selective pressures. Sequences highly enriched for both D25 and 5C4 binding show convergent sequence features in critical core positions of the site 0 scaffold. **(D)** Thermal melting curves measured by CD for best designs (S4_2.45 (top) and S0_2.126 (bottom)) showing high thermostability. (E) Dissociation constants (K_D) of S4_2.45 to 101F (top) and S0_2.126 to D25 (bottom) antibodies measured by SPR.

We screened a defined set of computationally designed sequences using yeast display and applied two selective pressures – binding to 101F and resistance to the nonspecific protease chymotrypsin, an effective method to digest partially unfolded proteins (*51, 188, 189*). Deep sequencing of populations sorted under different conditions revealed that S4_2_bb2-based designs were strongly enriched under stringent selection conditions for folding and 101F binding, showing that subtle topological differences in the design template can have substantial impact on function and stability. We expressed 15 S4_2_bb2 design variants and successfully purified and biochemically characterized 14. The designs showed mixed alpha/beta CD spectra and bound to 101F with affinities ranging from 1 nM to 200 nM (Fig S3.8). The best variant, S4_2.45 (K_D = 1 nM), was well folded and thermostable according to CD and NMR with a T_m of 75 °C (Fig 3.3d and Fig S3.9).

Similarly, we built a minimal *de novo* topology to present the tertiary structure of the site 0 epitope. The choice for this topology was motivated by the fact that site 0, in its native preRSVF environment, is accessible for antibody binding from diverse angles (*186*), in contrast to the S0_39 natural template which topologically constrained site 0 accessibility (Fig S3.5). By building a template *de novo,* we attempted to respect site 0's native quaternary constraints, while stabilizing both irregular epitope segments with high accuracy.

We explored the topological space within the shape constraints of preRSVF and built three different helical orientations (S0_2_bb1-3) that support both epitope segments. Evaluation of the designed sequences with Rosetta *ab initio* showed that only sequences generated based on one topology (S0_2_bb3) presented a funnel-shaped energy landscape (Fig S3.10). A set of computationally designed sequences based on S0_2_bb3 was screened in yeast under the selective pressure of two site 0-specific antibodies (D25 and 5C4) to ensure the presentation of the native epitope's conformation. Deep sequencing of the double-enriched clones and subsequent sequence analysis revealed that a valine at position 28 is critical to retain a cavity formed between the two epitope segments, ensuring binding to both antibodies (Fig 3.3b). We selected five sequences, differing from 3 to 21 mutations, for further biochemical characterization (Fig S3.11). The design with best solution behaviour (S0_02.126) showed a CD spectrum of a mostly helical protein, with extremely high thermostability even under reducing conditions (T_m = 81 °C, Fig 3.3d) and a well-dispersed HSQC NMR spectrum (Fig S3.9). Strikingly, S0_2.126 bound with ~50 pM affinity to D25, similar to that of the preRSVF-D25 interaction (~150 pM), and with a K_D = 4 nM to 5C4 (Fig 3.3e and Fig S3.12).

Overall, the properties of the designs generated by topological assembly with the TopoBuilder showed improved binding affinities and thermal stabilities as compared to those using available structural templates. To investigate whether this design and screening procedure yielded scaffolds that better mimicked the viral epitope presented, or rather revealed sequences with a highly optimized interface towards the antibodies used during the selection, we determined the affinities of S4_2.45 and S0_2.126 against a panel of human site-specific antibodies. Compared to the first-generation designs, S4_2.45 and S0_2.126 showed large affinity improvements across the antibody panels, exhibiting a geometric mean affinity closely resembling that of the antibodies to preRSVF (Fig S3.12). In the light of such results, we concluded that the topologically designed immunogens were superior mimetics of sites IV and 0 relative to the template-based designs.

3.3.2 *De novo* designed topologies adopt the predicted structures with high accuracy

To evaluate the structural accuracy of the computational design approach, we solved the crystal structure of S4_2.45 in complex with 101F at 2.6 Å resolution. The structure closely matched our design model, with a full-atom RMSD of 1.5 Å (Fig 3.4a). The epitope was mimicked with an RMSD of 0.135 Å, and retained all essential interactions with 101F (Fig 3.4d,e). Importantly, the structural data confirmed that we presented an irregular beta strand, a common motif found in many protein-protein interactions (*190*), in a fully *de novo* designed protein with sub-angstrom accuracy.

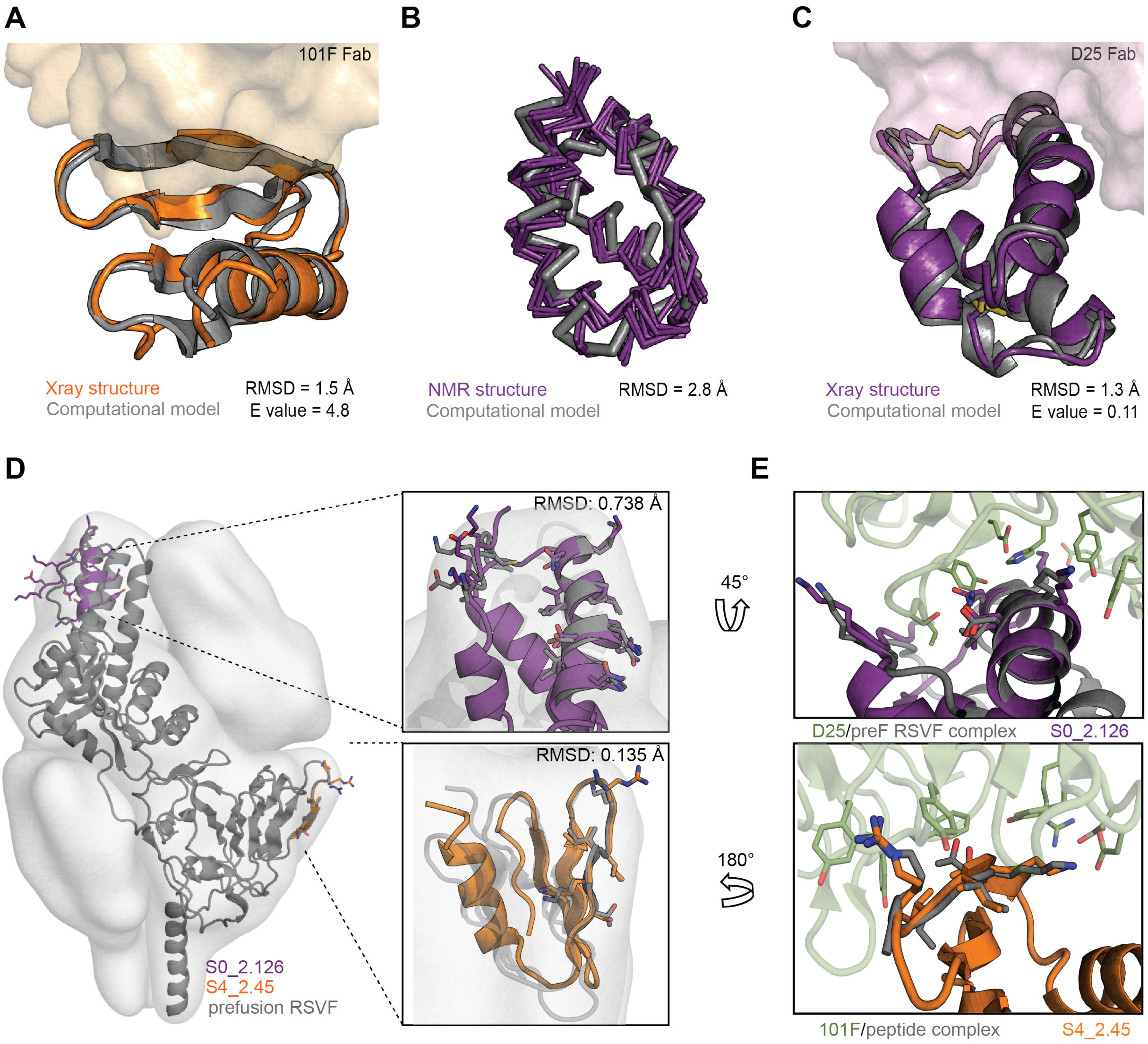

A
101F Fab
Xray structure
Computational model
RMSD = 1.5 Å
E value = 4.8
B
NMR structure
Computational model
RMSD = 2.8 Å
C
D25 Fab
Xray structure
Computational model
RMSD = 1.3 Å
E value = 0.11
D
RMSD: 0.738 Å
RMSD: 0.135 Å
S0_2.126
S4_2.45
prefusion RSVF
45°
180°
E
D25/preF RSVF complex
S0_2.126
101F/peptide complex
S4_2.45

Figure 3.4: Structural characterization of *de novo* designed immunogens.
(A) Crystal structure of S4_2.45 (orange) bound to 101F Fab closely matches the design model (grey, RMSD = 1.5 Å). **(B)** NMR structural ensemble of S0_2.126 (purple) superimposed to the computational model (grey). The NMR structure is in agreement with the design model (backbone RMSD of 2.8 Å). **(C)** Crystal structure of S0_2.126 (purple) bound to D25 Fab closely resembles the design model (grey, RMSD = 1.3 Å). **(D)** Superposition of the preRSVF sites 0/IV and designed immunogens shows sub-angstrom mimicry of the epitopes. Designed scaffolds are compatible with the shape constraints of preRSVF (surface representation). **(E)** Close-up view of the interfacial side-chain interactions between D25 (top) and 101F (bottom) with designed immunogens as compared to the starting epitope structures.

Next, we solved an unbound structure of S0_2.126 by NMR, confirming the accuracy of the designed fold with a backbone RMSD between the average structure and the model of 2.8 Å (Fig 3.4b). Additionally, we solved a crystal structure of S0_2.126 bound to D25 at a resolution of 3.0 Å. The structure showed an overall RMSD of 1.5 Å to the design model, and an RMSD of 0.9 Å over the discontinuous epitope compared to preRSVF (Fig 3.4c-e). To the best of our knowledge, this is the first computationally *de novo* designed protein that presents a two-segment, structurally irregular, binding motif with atomic-level accuracy. In comparison with native proteins, S0_2.126 showed exceptionally low packing due to a large core cavity (Fig S3.13), but retained a very high thermal stability. The core cavity was essential for antibody binding and highlights the potential of *de novo* approaches to design small proteins hosting structurally challenging motifs and preserving cavities required for function (*40*). Notably, due to the level of control and precision of the TopoBuilder, both designed antigens respected the shape constraints of the respective epitope in their native environment within preRSVF, a structural feature that may be important for the improved elicitation of functional antibodies (Fig S3.5).

3.3.3 Cocktails of designed immunogens elicit neutralizing antibodies *in vivo*

Lastly, we sought to evaluate the designed antigens for their ability to elicit focused nAb responses *in vivo*. Our rationale for combining site 0, II and IV immunogens in a cocktail formulation is that all three sites are non-overlapping, as verified by electron microscopy analysis (Fig S3.14), and thus could induce a more potent antibody response *in vivo*. To increase immunogenicity, each immunogen was multimerized on self-assembling protein nanoparticles. We chose the RSV nucleoprotein (RSVN), a self-assembling ring-like structure of 10-11 subunits, previously shown to be an effective carrier for the site II immunogen (*185*), and formulated a trivalent immunogen cocktail containing equimolar amounts of S0_1.39, S4_1.05 and S2_1.2 immunogen nanoparticles (“Trivax1”, Fig S3.15). The fusion of S0_2.126 and S4_2.45 to RSVN yielded poorly soluble nanoparticles, prompting us to use ferritin particles for multimerization, with a 50% occupancy (~12 copies), creating a second cocktail comprising S2_1.2 in RSVN and the remaining immunogens in ferritin (“Trivax2”, Fig S3.16).

In mice, Trivax1 elicited low levels of RSVF cross-reactive antibodies, and sera did not show RSV neutralizing activity in most animals (Fig S3.17). In contrast, Trivax2 induced robust levels of RSVF cross-reactive serum levels, and the response was balanced against all three epitopes (Fig 3.5a,b). Strikingly, Trivax2 immunization yielded RSV neutralizing activity above the protective threshold in 6/10 mice (Fig 3.5c). These results show that vaccine candidates composed of *de novo* designed proteins mimicking viral neutralization epitopes can induce robust antibody responses *in vivo*, targeting multiple specificities. This is an important finding given that mice have been a traditionally difficult model to induce nAbs with scaffold-based immunogens (*59, 110*).

In parallel, we sought to test the potential of a trivalent immunogen cocktail in NHPs. The previously designed site II immunogen showed promise in NHPs, but the induced neutralizing titers were low and inconsistent across animals, requiring up to five immunizations to elicit nAbs in 2/4 animals (*59*). We immunized seven RSV-naïve NHPs with Trivax1, as detailed in Fig 3.5d. In contrast to mice, NHPs developed robust levels of RSVF cross-reactive serum titer in all animals (Fig 3.5e), and antibodies induced were directed against all three epitopes (Fig 3.5f). Strikingly, we found that 6/7 NHPs showed RSV neutralizing serum levels above the protective threshold after a single boosting immunization (median IC_{50} = 312) (Fig 3.5g). Neutralization titers were maximal at day 84 (median IC_{50} = 408), four-fold above the protective threshold (*82*), and measurements were confirmed by an independent laboratory (Fig S3.18).

Immunization studies in naïve animals are essential to test the capability of the designed immunogens to induce functional antibodies. However, an overarching challenge for vaccine development to target pathogens such as RSV, influenza, dengue and others is to focus or reshape pre-existing immunity of broad specificity on defined neutralizing epitopes that can confer long-lasting protection (*131, 140, 185*). To mimic a serum response of broad specificity towards RSV, we immunized 13 NHPs with preRSVF. All animals developed strong preRSVF-specific titers and cross-reactivity with all the epitope-focused immunogens, indicating that epitope-specific antibodies were primed and recognize the designed immunogens (Fig S3.19). Group 2 (6 animals) subsequently served as control group to follow the dynamics of epitope-specific antibodies over time, and group 3 (7 animals) was boosted three times with Trivax1 (Fig 3.5d). PreRSVF-specific antibody and neutralization titers maximized at day 28 and were maintained up to day 119 in both groups (Fig 3.5h,i). Analysis of the site-specific antibody levels showed that site 0, II and IV responses were dynamic in the control group, with site II dropping from 37% to 13% and site 0 from 17% to 4% at day 28 and 91, respectively (Fig 3.5j). In contrast, site IV specific responses increased from 13% to 43% over the same time span. Although Trivax1 boosting immunizations did not significantly change the magnitude of the preRSVF-specific serum response, they reshaped the serum specificities in primed animals. Site II specific titers were 6.5-fold higher (day 91) compared to the non-boosted control group (84% vs 13%, p = 0.02, Mann-Whitney), and unlike the rapid drop of site 0-specific antibodies in the non-boosted group, these antibodies were maintained upon Trivax1 boosting (25% vs 4%, p=0.02, Mann-Whitney) (Fig 3.5j). In contrast, site IV specific responses increased to similar levels in both groups, 43% and 40% in group 2 and 3, respectively. Strikingly, upon depletion of site 0, II and IV specific antibodies from pooled sera, we observed a 60% drop in neutralizing activity in group 3 as compared to only a 7% drop in the non-boosted control group, indicating that Trivax1 boosting reshaped a serum response of broad specificity towards a more focused response that predominantly relies on site 0, II and IV-specific antibodies for RSV neutralization (Fig 3.5k).

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

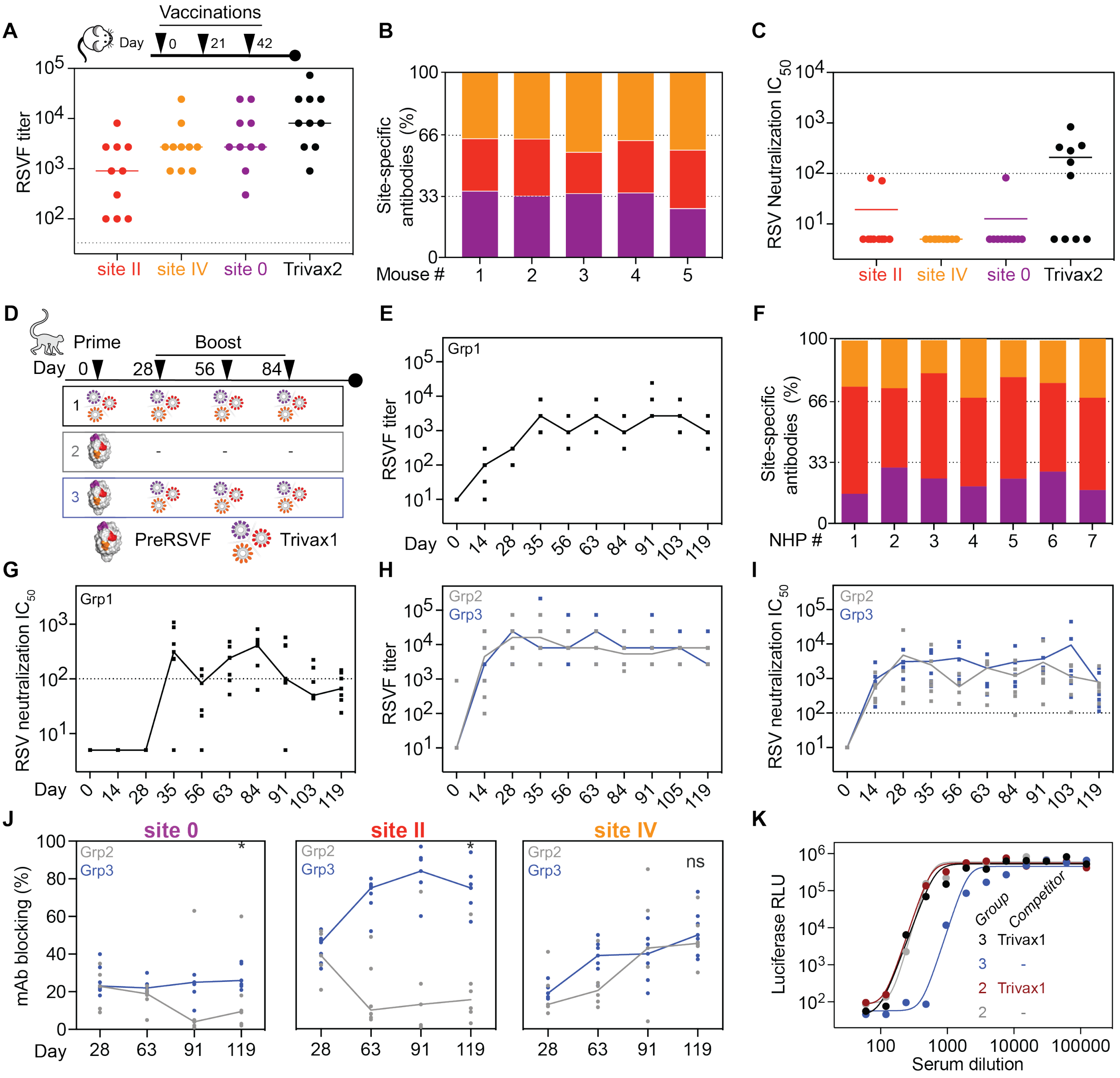


Figure 3.5: Synthetic immunogens elicit neutralizing serum responses in mice and NHPs, and focus pre-existing immunity on sites 0 and II.

(A-C) Trivax2 immunization study in mice. **(A)** PreRSVF cross-reactive serum levels following three immunizations with single immunogens or Trivax2 cocktail (day 56). **(B)** Serum specificity shown for 5 representative mice immunized with Trivax2, as measured by an SPR competition assay with D25, Motavizumab and 101F IgGs as competitors, exhibiting an equally balanced response towards all sites. **(C)** RSV neutralization titer of mice at day 56, immunized with Trivax2 components individually and as cocktail. Dotted line (IC_{50} = 100) indicates protective threshold as defined by protective level of Palivizumab. (D-K) Trivax1 immunization study in NHPs. **(D)** NHP immunization scheme. **(E)** PreRSVF cross-reactive serum levels for group 1. **(F)** Serum antibodies target all three antigenic sites in all 7 animals as measured by an SPR competition assay. **(G)** RSV neutralization titers of group 1. **(H)** PreRSVF titer in group 2 (grey) and 3 (blue). **(I)** RSV neutralization titer of group 2 and 3. **(J)** Site-specific antibody levels measured by SPR competition assay. Site 0 and site II-specific titers were significantly higher in group 3 compared to 2 following Trivax1 boosting ($p < 0.05$, Mann-Whitney U test). **(K)** RSV neutralization curves upon depletion of day 91 sera with site 0, II, IV-specific scaffolds. 60% of the neutralizing activity is competed in group 3, whereas no significant decrease is observed in the control group 2.

3.4 Discussion and conclusions

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

To overcome these limitations, we developed the TopoBuilder, a motif-centric design approach that tailors a protein fold directly to the functional site of interest. Compared to previously employed functional *de novo* design protocols, in which a stable scaffold topology was constructed first and endowed with binding motifs in a second step (*51*), our method has significant advantages for structurally complex motifs. First, it tailors the topology to the structural requirements of the functional motif from the start of the design process, rather than through the adaptation (and often destabilization) of a stable protein to accommodate the functional site. Second, the topological assembly and fine-tuning allowed to select for optimal backbone orientations and sequences that stably folded and bound with high affinity in a single screening round, without further optimization through directed evolution, as often necessary in computational protein design efforts (*49, 51, 62, 191*). Together, our approach enabled the computational design of *de novo* proteins presenting irregular and discontinuous structural motifs that are typically required to endow proteins with diverse biochemical functions (e.g. binding or catalysis), thus providing a new means for the *de novo* design of functional proteins.

As to the functional aspect of our design work, we showed *in vivo* that these immunogens consistently elicited neutralizing serum levels in mice and NHPs as cocktail formulations. The elicitation of focused nAb responses by vaccination remains the central goal for vaccines against pathogens that have frustrated conventional vaccine development efforts (*68*). Using RSV as a model system, we have shown that cocktails of computationally designed antigens can robustly elicit neutralizing serum levels in naïve animals. These neutralization levels were much superior to any previous report on epitope-focused immunogens (*59*) and provide a strong rationale for an epitope-focused vaccination strategy involving multiple, non-overlapping epitopes. Also, their capability to dramatically reshape the nature of non-naïve repertoires in NHPs addresses an important challenge for many next-generation vaccines to target pathogens for which efficacious vaccines are needed (*131, 164, 185*). An important pathogen from this category is influenza, where the challenge is to overcome immunodominance hierarchies (*69*), which have been established during repeated natural infections, and that favour strain-specific antibody specificities, rather than cross-protecting nAbs found in the hemagglutinin stem region (*137*). The ability to selectively boost subdominant nAbs targeting defined, broadly protective epitopes that are surrounded by strain-specific epitopes could overcome long-standing challenges in vaccine development, given that cross-neutralizing antibodies can persist for years once elicited (*192*). A tantalizing future application for epitope-focused immunogens could marry this technology with engineered components of the immune system, which could be used to stimulate antibody production of adoptively transferred, engineered B-cells that express monoclonal therapeutic antibodies *in vivo* (*193*).

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

3.5 Methods

3.5.1 Computational design of template-based epitope-focused immunogens

All code used for the computational design and analysis is available through a public github repository: https://github.com/lpdi-epfl/trivalent_cocktail. It contains the TopoBuilder source code, RosettaScripts used for the design, analysis scripts and detailed information on how designs were selected. Tables with the amino acid sequences of the experimentally characterized sequences are available in the supplementary material of a publicly available preprint of the manuscript, found under: https://doi.org/10.1101/685867.

Site 0

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

The MASTER software (*194*) was used to perform structural searches over the Protein Data Bank (PDB, from August 2018), containing 141,920 protein structures, to select template scaffolds with local structural similarities to the site 0 motif. A first search with a Cα RMSD threshold below 2.5 Å did not produce any usable structural matches both in terms of local mimicry as well as global topology features. A second search was performed, where extra structural elements that support the epitope in its native environment were included as part of the query motif to bias the search towards matches that favoured motif-compatible topologies rather than those with close local similarities. The extra structural elements included were the two buried helices that directly contact the site 0 in the preRSVF structure (4JHW residues 70-88 and 212-229). The search yielded initially 7,600 matches under 5 Å of backbone RMSD, which were subsequently filtered for proteins with a length between 50 and 160 residues, high secondary structure content, as well as for accessibility of the epitope for antibody binding. Remaining matches were manually inspected to select template-scaffolds suitable to present the native conformation of antigenic site 0. Subsequently, we selected a computationally designed, highly stable, helical repeat protein (*41*) consisting of 8 regular helices (PDB ID: 5CWJ) with an RMSD of 4.4 Å to the query (2.82 Å for site 0 segments only). To avoid steric clashes with the D25 antibody, we truncated the 5CWJ template structure at the N-terminus by 31 residues, resulting in a structural topology composed of 7 helices.

Using Rosetta FunFolDes (*26*) the truncated 5CWJ topology was folded and designed to stabilized the grafted site 0 epitope recognized by D25. We generated 25,000 designs and selected the top 300 by Rosetta energy score (RE), designed backbones that presented obvious flaws, such as low packing scores, distorted secondary structural elements and buried unsatisfied atoms were discarded. From the top 300 designs, 3 were retained for follow-up iterative cycles of structural relaxation and design using Rosetta FastDesign (*195*), generating a total of 100 designed sequences.

The best 9 designs by Rosetta energy score were recombinantly expressed in E. coli. 2 designed sequences derived from the same backbone, were successfully expressed and purified. The best variant was named S0_1.1, and subjected to experimental optimization using yeast surface display (Fig S3.3-S3.4). In one of the libraries, we found a truncated sequence (S0_1.17) enriched for expression and binding, which served as template for a second round of computational design (Fig S3.3-S3.4). We performed 25,000 folding and design simulations using Rosetta FunFolDes (*26*). The best 300 decoys by total Rosetta energy score were extracted, and relaxed using the Rosetta Relax application (*196*). We computed the mean total RE, and selected designs that showed a lower energy score than the mean of the design population (RE = -155.2), RMSD drift of the epitope after relaxing of less than 0.7 Å, and a cavity volume <60 Å^3. We selected one of the best 5 scoring decoys, truncated the C-terminal 29 and N-terminal 23 residues which did not contribute to epitope stabilization, and introduced a disulfide bond between residue 1 and 43. Four sequences were experimentally tested (S0_1.37-40). The best variant according to binding, S0_1.39, bound with 5 nM affinity to antibody D25, and, importantly, also gained binding to the 5C4 antibody (K_D = 5 nM).

Site IV

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

The crystallized peptide-epitope corresponds to the residues 429-434 of the RSVF protein. Structurally the 101F-bound peptide-epitope adopts a bulged strand and several studies suggest that 101F recognition extends beyond the linear β-strand, contacting other residues located in antigenic site IV (*197*). Despite the apparent structural simplicity of the epitope, structural searches for designable scaffolds failed to yield promising starting templates. However, we noticed that the antigenic site IV of RSVF is self-contained within an individual domain that could potentially be excised and designed as a soluble folded protein. To maximize these contacts, we first truncated the seemingly self-contained region from RSVF pre-fusion structure (PDB ID: 4JHW, residue: 402-459) forming a β-sandwich and containing site IV. We used Rosetta FastDesign to optimize the core positions of this minimal topology, obtaining our initial design: S4_wt. However, S4_wt did not show a funnel-shaped energy landscape in Rosetta *ab initio* simulations, and we were unable to obtain expression in *E.coli*.

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

3.5.2 TopoBuilder - Motif-centric *de novo* design

Given the limited availability of suitable starting templates to host structurally complex motifs such as site 0 and site IV, we developed a template-free design protocol, which we named TopoBuilder (see Fig S3.6). In contrast to adapting an existing topology to accommodate the epitope, the design goal is to build protein scaffolds around the epitope from scratch, using idealized secondary structures (beta strands and alpha helices). The length, orientation and 3D-positioning are defined by the user for each secondary structure with respect to the epitope, which is extracted from its native environment. The topologies built were designed to meet the following criteria: (1) Small, globular proteins with a high contact order between secondary structures and the epitope, to allow for stable folding and accurate stabilization of the epitope in its native conformation; (2) Context mimicry, i.e. respecting shape constraints of the epitope in its native context (Fig S3.5). For assembling the topology, the default distances between alpha helices was set to 11 Å and for adjacent beta-strands was 5 Å. For each discontinuous structural sketch, a connectivity between the secondary structural elements was defined and loop lengths were selected to connect the secondary structure elements with the minimal number of residues that can cover a given distance, while maintaining proper backbone geometries.

For site 0, the short helix of S0_1.39 preceding the epitope loop segment was kept, and a third helix was placed on the backside of the epitope to: (1) provide a core to the protein and (2) allow for the proper connectivity between the secondary structures. A total of three different orientations (45°, 0° and -45° degrees to the plane formed by site 0) where tested for the designed supporting alpha helix (Fig. 3.3).

In the case of site IV, the known binding region to 101F (residues 428F-434F) was extracted from prefusion RSVF (PDB 4JWH). Three antiparallel beta strands, pairing with the epitope, plus an alpha helix on the buried side, were assembled around the 101F epitope. Three different configurations (45°, (-45°,0°,10°) and -45° degrees with respect to the β-sheet) were sampled parametrically for the alpha helix (Fig. 3.3 and Fig S3.7).

The structural sketches were used to generate Cα distance constraints to guide Rosetta FunFolDes (*26*) folding trajectories. Around 25,000 trajectories were generated for each sketch. The newly generated backbones were further subjected to layer-based FastDesign (*195*), meaning that each amino acid position was assigned a layer (combining 'core', 'boundary', 'surface' and 'sheet' or 'helix') on the basis of its exposure and secondary structure type, that dictated the allowed amino acid types at that position.

After iterative cycles of sequence design, unconstraint FastRelax (*198*) (i.e sidechain repacking and backbone minimization) was applied over the designs to evaluate their conformational stability of the epitope region. After each relax cycle, structural changes of the epitope region were evaluated (epitope RMSD drift). Designs with epitope RMSD drifts higher than 1.2 Å were discarded. Designs were also ranked and selected according hydrophobic core packing (packstat score), with a cutoff of 0.5 for site 0 and 0.6 for the site IV design series, and a cavity volume of < 50 Å^3. Between 1,000 and 10,000 of the designed sequences were generated from this computational protocol. We evaluated sequence profiles for the designs, and encoded the critical positions combinatorially by assembling overlapping oligos. Upon PCR assembly, libraries were transformed in yeast and screened for antibody binding and stability as assessed by protease digestion assays (*51, 188, 189*).

3.5.3 Immunization studies

Mouse immunizations

All animal experiments were approved by the Veterinary Authority of the Canton of Vaud (Switzerland) according to Swiss regulations of animal welfare (animal protocol number 3074). Female Balb/c mice (6-week old) were purchased from Janvier labs.

Immunogens were thawed on ice, mixed with equal volumes of adjuvant (2 % Alhydrogel, Invivogen or Sigma Adjuvant System, Sigma) and incubated for 30 minutes. Mice were injected subcutaneously with 100 μl vaccine formulation, containing in total 10 μg of immunogen (equimolar ratios of each immunogen for Trivax immunizations). Immunizations were performed on day 0, 21 and 42. 100-200 μl blood were drawn on day 0, 14 and 35. Mice were euthanized at day 56 and blood was taken by cardiac puncture.

NHP immunizations

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

3.5.4 Serum analysis

For the following methods please refer to the method section in chapter 2: ELISA and RSV neutralization assay.

Serum fractionation

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

Dissection of serum antibody specificities by SPR

To quantify the epitope-specific antibody responses in bulk serum from immunized animals, we performed an SPR competition assay with the monoclonal antibodies (D25, Motavizumab and 101F) as described previously (*185*). Briefly, approximately 400 RU of prefusion RSVF were immobilized on a CM5 chip via amine coupling, and serum diluted 1:10 in running buffer was injected to measure the total response ($RU_{\text{non-blocked surface}}$). After chip regeneration using 50 mM NaOH, the site 0/II/IV epitopes were blocked by injecting saturating amounts of either D25, Motavizumab, or 101F IgG, and serum was injected again to quantify residual response ($RU_{\text{blocked surface}}$). The delta serum response (ΔSR) was calculated as follows:

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

Percent blocking was calculated for each site as:

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

3.5.5 *In vitro* evolution

Yeast surface display

Libraries of linear DNA fragments encoding variants of the designed proteins were transformed together with linearized pCTcon2 vector (Addgene #41843) based on the protocol previously described by Chao and colleagues (*199*). Transformation procedures generally yielded ~10^7 transformants. The transformed cells were passaged twice in SDCAA medium before induction. To induce cell surface expression, cells were centrifuged at 7,000 r.p.m. for 1 min, washed with induction media (SGCAA) and resuspended in 100 ml SGCAA with a cell density of 1 x 10^7 cells/ml SGCAA. Cells were grown overnight at 30 °C in SGCAA medium. Induced cells were washed in cold wash buffer (PBS + 0.05% BSA) and labelled with various concentration of target IgG or Fab (101F, D25, and 5C4) at 4°C. After one hour of incubation, cells were washed twice with wash buffer and then incubated with FITC-conjugated anti-cMyc antibody and PE-conjugated anti-human Fc (BioLegend, #342303) or PE-conjugated anti-Fab (Thermo Scientific, #MA1-10377) for an additional 30 min. Cells were washed and sorted using a SONY SH800 flow cytometer in 'ultra-purity' mode. The sorted cells were recovered in SDCAA medium, and grown for 1-2 days at 30 °C.

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

Site saturation mutagenesis library (SSM)

A SSM library was assembled by overhang PCR, in which 11 selected positions surrounding the epitope in the S4_1.1 design model were allowed to mutate to all 20 amino acids, with one mutation allowed at a time. Each of the 11 libraries was assembled by primers containing the degenerate codon 'NNK' at the selected position. All 11 libraries were pooled, and transformed into EBY-100 yeast strain with a transformation efficiency of 1 x 10^6 transformants.

Combinatorial library

Combinatorial sequence libraries were constructed by assembling multiple overlapping primers (Table S2) containing degenerate codons at selected positions for combinatorial sampling of hydrophobic amino acids in the protein core. The theoretical diversity was between 1 x 10^6 and 5 x 10^6. Primers were mixed (10 µM each), and assembled in a PCR reaction (55 °C annealing for 30 sec, 72 °C extension time for 1 min, 25 cycles). To amplify full-length assembled products, a second PCR reaction was performed, with forward and reverse primers specific for the full-length product. The PCR product was desalted, and transformed into EBY-100 yeast strain with a transformation efficiency of at least 1 x 10^7 transformants (*199*).

Next-generation sequencing of design pools

After sorting, yeast cells were grown overnight, pelleted and plasmid DNA was extracted using Zymoprep Yeast Plasmid Miniprep II (Zymo Research) following the manufacturer's instructions. The coding sequence of the designed variants was amplified using vector-specific primer pairs, Illumina sequencing adapters were attached using overhang PCR, and PCR products were desalted (Qiaquick PCR purification kit, Qiagen). Next generation sequencing was performed using an Illumina MiSeq 2 x 150bp paired end sequencing (300 cycles), yielding between 0.45 - 0.58 million reads/sample.

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

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

The high selective pressure corresponds to low labelling concentration of the respective target antibodies (100 pM D25, 10 nM 5C4 or 20 pM 101F, as shown in Fig. 3.3), or a higher concentration of chymotrypsin protease (0.5 µM). The low selective pressure corresponds to a high labelling concentration with antibodies (10 nM D25, 1 µM 5C4 or 2 nM 101F), or no protease digestion, as indicated in Fig. 3.3. Only sequences that had at least one count in both sorting conditions were included in the analysis.

3.5.6 Protein expression and purification

For the following methods, please refer to the method section in chapter 2: Expression of antibodies and Fabs, prefusion RSVF and nanoring-based immunogens.

Designed scaffolds

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

Ferritin-based immunogens

The gene encoding Helicobacter pylori ferritin (GenBank ID: QAB33511.1) was cloned into the pHLsec vector for mammalian expression, with an N-terminal 6x His Tag. The sequence of the designed immunogens (S0_2.126 and S4_2.45) were cloned upstream of the ferritin gene, spaced by a GGGGS linker. Ferritin particulate immunogens were produced by co-transfecting a 1:1 stochiometric ratio of “naked” ferritin and immunogen-ferritin in HEK-293F cells, as previously described for other immunogen-nanoparticle fusion constructs (*106*). The supernatant was collected 7-days post transfection and purified via Ni-NTA affinity chromatography and size exclusion on a Superose 6 increase 10/300 GL column (GE).

3.5.7 Negative-stain transmission electron microscopy

Sample preparation

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

Data acquisition

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

Image processing

CTFFIND4 program (*173*) was used to estimate the contrast transfer function for each collected image. Around 1000 particles were manually selected using the installed package XMIPP within SCIPION framework (*174*). Manually picked particles were served as input for XMIPP auto-picking utility, resulting in at least

10,000 particles. Selected particles were extracted with the box size of 100 pixels and subjected for three rounds of reference-free 2D classification without CTF correction using RELION-3.0 Beta suite (*175*).

RSVF-Fabs complex formation and negative stain EM

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

Purified complexes of RSVF and Fabs were deposited at approximately 0.02 mg/ml onto carbon-coated copper grids and stained with 2% uranyl formate. Images were collected with a field-emission FEI Tecnai F20 electron microscope operating at 200 kV. Images were acquired with an Orius charge-coupled device (CCD) camera (Gatan Inc.) at a calibrated magnification of ×34,483, resulting in a pixel size of 2.71 Å. For the complexes of RSVF with a single Fab, approximately 2,000 particles were manually selected with Cryosparc2 (*200*). Two rounds of 2D classification of particle images were performed with 20 classes allowed. For the complexes of RSVF with D25, Motavizumab and 101F Fabs, approximately 330,000 particles were picked using Relion 3.0 (*175*) and subsequently imported to Cryosparc2 for two rounds of 2D classification with 50 classes allowed.

3.5.8 Biophysical characterization of designed proteins

SEC-MALS

Size exclusion chromatography with an online multi-angle light scattering (MALS) device (miniDAWN TREOS, Wyatt) was used to determine the oligomeric state and molecular weight for the protein in solution. Purified proteins were concentrated to 1 mg/ml in PBS (pH 7.4), and 100 µl of sample was injected into a Superdex 75 300/10 GL column (GE Healthcare) with a flow rate of 0.5 ml/min, and UV_{280} and light scattering signals were recorded. Molecular weight was determined using the ASTRA software (version 6.1, Wyatt).

Circular Dichroism

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

Determining binding affinities by Surface plasmon resonance (SPR)

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

3.5.9 Structural characterization by NMR and x-ray crystallography

NMR

Protein samples for NMR were prepared in 10 mM sodium phosphate buffer, 50 mM sodium chloride at pH 7.4 with the protein concentration of 500 µM. All NMR experiments were carried out in a 18.8T (800 MHz proton Larmor frequency) Bruker spectrometer equipped with a CPTC ^{1}H,^{13}C,^{15}N 5 mm cryoprobe and an Avance III console. Experiments for backbone resonance assignment consisted in standard triple resonance spectra HNCA, HN(CO)CA, HNCO, HN(CO)CA, CBCA(CO)NH and HNCACB acquired on a 0.5 mM sample doubly labelled with ^{13}C and ^{15}N (*201*). Sidechain assignments were obtained from HCCH-TOCSY experiments acquired on the same sample plus HNHA, NOESY-^{15}N-HSQC and TOCSY-^{15}N-HSQC acquired on a ^{15}N-labeled sample. The NOESY-^{15}N-HSQC was used together with a 2D NOESY collected on an unlabelled sample for structure calculations. Spectra for backbone assignments were acquired with 40 increments in the ^{15}N dimension and 128 increments in the ^{13}C dimension, and processed with 128 and 256 points by using linear prediction. HCCH-TOCSY were recorded with 64-128 increments in the ^{13}C dimensions and processed with twice the number of points. ^{15}N-resolved NOESY and TOCSY spectra were acquired with 64 increments in ^{15}N dimension and 128 in the indirect ^{1}H dimension, and processed with twice the number of points. ^{1}H-^{1}H 2D-NOESY and 2D TOCSY spectra were acquired with 256 increments in the indirect dimension, processed with 512 points. Mixing times for NOESY spectra were 100 ms and TOCSY spin locks were 60 ms. Heteronuclear ^{1}H-^{15}N NOE was measured with 128 ^{15}N increments processed with 256 points, using 64 scans and a saturation time of 6 seconds. All samples were prepared in 20 mM phosphate buffer pH 7, with 10% $^{2}H_2O$ and 0.2% sodium azide to prevent sample degradation.

All spectra were acquired and processed with Bruker's TopSpin 3.0 (acquisition with standard pulse programs) and analyzed manually with the program CARA (http://cara.nmr.ch/doku.php/home) to obtain backbone and sidechain resonance assignments. Peak picking and assignment of NOESY spectra (a ^{15}N-resolved NOESY and a 2D NOESY) were performed automatically with the program UNIO-ATNOS/CANDID (*202, 203*) coupled to Cyana 2.1 (*204*), using standard settings in both programs. The run was complemented with dihedral angles derived from chemical shifts with Talos-n (*205*). The solution NMR structure of S0_2.126 has been deposited in the Protein Data Bank under accession code 6S28.

NMR restraints	
Total NOEs from Unio [a]	306
Intraresidual	124
Interresidual	182
Sequential (i – j = 1)	112
Medium-range (1 < i – j < 5)	47
Long-range (i – j ≥ 5)	23
Dihedral Angles from Talos-n [b]	88
φ	43
ψ	45
Structural statistics	
Violations [c]	
Distance restraints (Å)	0.0254 ± 0.009
Dihedral angle constraints (°)	6.8 ± 0.12
Ramachandran plot (all residues/ordered residues) [d]	
Most favored (%)	84.7 / 95.8
Additionally allowed (%)	14.3 / 4.5
Generously allowed (%)	0.98 / 0.1
Disallowed (%)	0 / 0
Average pairwise RMSD (Å) [e]	
Heavy	3.3 / 1.8
Backbone	2.8 / 1.2
Structure Quality Factors (raw score/z-score) [e]	
Procheck G-factor (phi/psi)	0.15 / 0.9
Procheck G-factor (all)	-0.48 / -2.84

[a] From UNIO-ATNOS/CANDID’s last cycle (cycle 7)

[b] Obtained from chemical shifts with Talos-N server

[c] From Cyana in Unio’s last cycle

[d] All residues from Cyana un Unio’s last cycle; ordered residues (5-22,26-57) from the Protein Structure Validation Suite at http://psvs-1_5-dev.nesg.org/results/testbc/OUTPUT.html

[e] From the Protein Structure Validation Suite

Table S 3.1: Refinement statistics of the S0_2.126 NMR structure.

Co-crystallization of complex D25 Fab with S0_2.126

After overnight incubation at 4 °C, the S0_2.126/D25 Fab complex was purified by size exclusion chromatography using a Superdex200 26 600 (GE Healthcare) equilibrated in 10 mM Tris pH 8, 100 mM NaCl and subsequently concentrated to ~10 mg/ml (Amicon Ultra-15, MWCO 3,000). Crystals were grown at 291K using the sitting-drop vapor-diffusion method in drops containing 1 µl purified protein mixed with 1 µl reservoir solution containing 10% PEG 8000, 100 mM HEPES pH 7.5, and 200 mM calcium acetate. For cryo protection, crystals were briefly swished through mother liquor containing 20% ethylene glycol.

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

Diffraction data was recorded at ESRF beamline ID30B. Data integration was performed by XDS (*206*) and a high-resolution cut at I/σ=1 was applied. The dataset contained a strong off-origin peak in the Patterson function (88% height rel. to origin) corresponding to a pseudo translational symmetry of ½, 0, ½. The structure was determined by the molecular replacement method using PHASER (*207*) using the D25 structure (*1*) (PDB ID 4JHW) as a search model. Manual model building was performed using Coot (*208*), and automated refinement in Phenix (*209*). After several rounds of automated refinement and manual building, paired refinement (*210*) determined the resolution cut-off for final refinement (Table S3.2).

D25 S0_2.126	
Wavelength	0.9763
Resolution range	49.09-3.0 (3.107-3.0)
Space group	P 21 21 21
Unit cell	126.3 127.0 156.1 90 90 90
Total reflections	700184 (72248)
Unique reflections	50740 (5000)
Multiplicity	13.8 (14.4)
Completeness (%)	98.76 (99.22)
Mean I/sigma(I)	12.63 (2.00)
Wilson B-factor	74.78
R-merge	0.1622 (1.484)
R-meas	0.1684 (1.538)
R-pim	0.04506 (0.4019)
CC1/2	0.999 (0.893)
CC*	1 (0.971)
Reflections used in refinement	50284 (4971)
Reflections used for R-free	2519 (249)
R-work	0.2699 (0.3677)
R-free	0.2936 (0.3972)
CC(work)	0.949 (0.817)
CC(free)	0.958 (0.793)
Number of non-hydrogen atoms	14453
macromolecules	14452
Protein residues	1921
RMS(bonds)	0.004
RMS(angles)	1.02
Ramachandran favored (%)	94.45
Ramachandran allowed (%)	5.07
Ramachandran outliers (%)	0.48
Rotamer outliers (%)	0.00
Clashscore	7.35
Average B-factor	97.74
macromolecules	97.74
solvent	59.33
Number of TLS groups	12

Table S 3.2: X-ray data collection and refinement statistics of S0_2.126 crystal structure.

Co-crystallization of complex 101F Fab with S4_2.45

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

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

Diffraction data were collected at SSRL facility, BL9-2 beamline at the SLAC National Accelerator Laboratory. The crystals belonged to space group P3221. The diffraction data were initially processed to 2.6 Å with X-ray Detector Software (XDS) (Table S3.3).

Molecular replacement searches were conducted with the program PHENIX PHASER using 101F Fab model (PDB ID: 3O41) and S4_2.45/101F Fab computational model generated from superimposing epitope region of S4_2.45 with the peptide-bound structure (PDB ID: 3O41), and yielded clear molecular replacement solutions. Initial refinement provided a R_{free} of 42.43% and R_{work} of 32.25% and a complex structure was refined using Phenix Refine, followed by manual rebuilding with the program COOT. The final refinement statistics, native data and phasing statistics are summarized in Table S3.3.

101F S4_2.45	
Wavelength	0.98
Resolution range	38.49 - 2.6 (2.693 - 2.6)
Space group	P 32 2 1
Unit cell	148.224 148.224 45.046 90 90 120
Total reflections	113069 (7302)
Unique reflections	17464 (1567)
Multiplicity	6.5 (4.7)
Completeness (%)	98.57 (89.58)
Mean I/sigma(I)	17.03 (1.66)
Wilson B-factor	56.09
R-merge	0.06712 (0.8361)
R-meas	0.07282 (0.9424)
R-pim	0.02776 (0.4231)
CC1/2	0.999 (0.635)
CC*	1 (0.881)
Reflections used in refinement	17455 (1565)
Reflections used for R-free	1748 (166)
R-work	0.2298 (0.3682)
R-free	0.2736 (0.3503)
CC(work)	0.462 (0.203)
CC(free)	0.353 (0.190)
Number of non-hydrogen atoms	3794
macromolecules	3686
solvent	108
Protein residues	485
RMS(bonds)	0.010
RMS(angles)	1.46
Ramachandran favored (%)	93.53
Ramachandran allowed (%)	5.64
Ramachandran outliers (%)	0.84
Rotamer outliers (%)	0.96
Clashscore	2.19
Average B-factor	38.90
macromolecules	38.37
solvent	56.78
Number of TLS groups	3

Table S 3.3: X-ray data collection and refinement statistics of S4_2.45 crystal structure.

3.6 Supplementary information

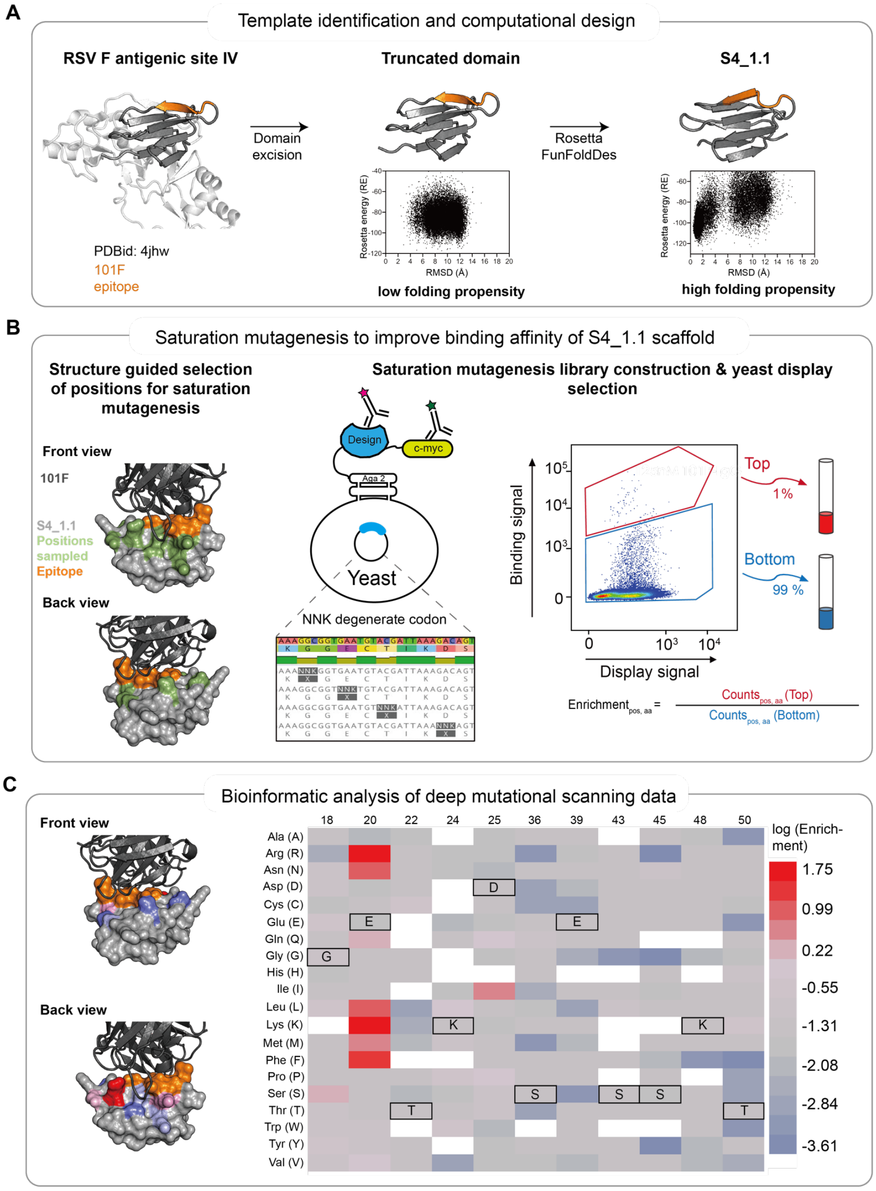

A
Template identification and computational design
RSV F antigenic site IV
Truncated domain
S4_1.1
Domain excision
Rosetta FunFoldDes
PDBid: 4jhw
101F epitope
Rosetta energy (RE)
RMSD (Å)
low folding propensity
high folding propensity
B
Saturation mutagenesis to improve binding affinity of S4_1.1 scaffold
Structure guided selection of positions for saturation mutagenesis
Saturation mutagenesis library construction & yeast display selection
Front view
101F
S4_1.1
Positions sampled
Epitope
Back view
Design
c-myc
Aga 2
Yeast
NNK degenerate codon
Binding signal
Display signal
Top
1%
Bottom
99 %
Enrichment$_{pos, aa}$ = Counts$_{pos, aa}$ (Top) / Counts$_{pos, aa}$ (Bottom)
C
Bioinformatic analysis of deep mutational scanning data
Front view
Back view
18 20 22 24 25 36 39 43 45 48 50
Ala (A)
Arg (R)
Asn (N)
Asp (D)
Cys (C)
Glu (E)
Gln (Q)
Gly (G)
His (H)
Ile (I)
Leu (L)
Lys (K)
Met (M)
Phe (F)
Pro (P)
Ser (S)
Thr (T)
Trp (W)
Tyr (Y)
Val (V)
log (Enrichment)
1.75
0.99
0.22
-0.55
-1.31
-2.08
-2.84
-3.61

Figure S 3.1: Computational design and experimental optimization of S4_1 design series.
(A) Template identification and computational design of S4_1.1. RSVF antigenic site IV is located in a small contained domain of preRSVF. This excised domain failed to show a folding funnel in Rosetta *ab initio* predictions, and failed to express recombinantly in *E.coli*. Using the excised domain as template, we folded and sequence-designed this topology using Rosetta FunFolDes, yielding design S4_1.1 which showed a strong funnel-shape energy landscape in *ab initio* folding simulation. **(B)** Experimental optimization of S4_1.1 through saturation mutagenesis. A saturation mutagenesis library was constructed using overhang PCR for 11 positions proximal (green) to the site IV epitope (orange), allowing one position at a time to mutate to any of the 20 amino acids, encoded by the degenerate codon 'NNK'. The library (size 11 positions x 32 codons = 352) was transformed in yeast, and designs were displayed on the cell surface. The selection was done by labeling the cells with 125 nM of 101F antibody. The top 1 % of clones binding with high affinity to 101F antibody were then sorted, as well as the bottom 99 % as shown. Following next-generation sequencing of the two populations, the enrichment values were computed for each sequence variant, corresponding to the relative abundance of each variant in the top versus bottom gate. **(C)** Bioinformatic analysis of deep mutational scanning data. The log(enrichment) is shown as heatmap (right) for each sequence variant, and mapped to the structure (left). White indicates missing data. Position 20 showed the highest enrichment for arginine and lysine, together with other less pronounced enrichments seen for other positions.

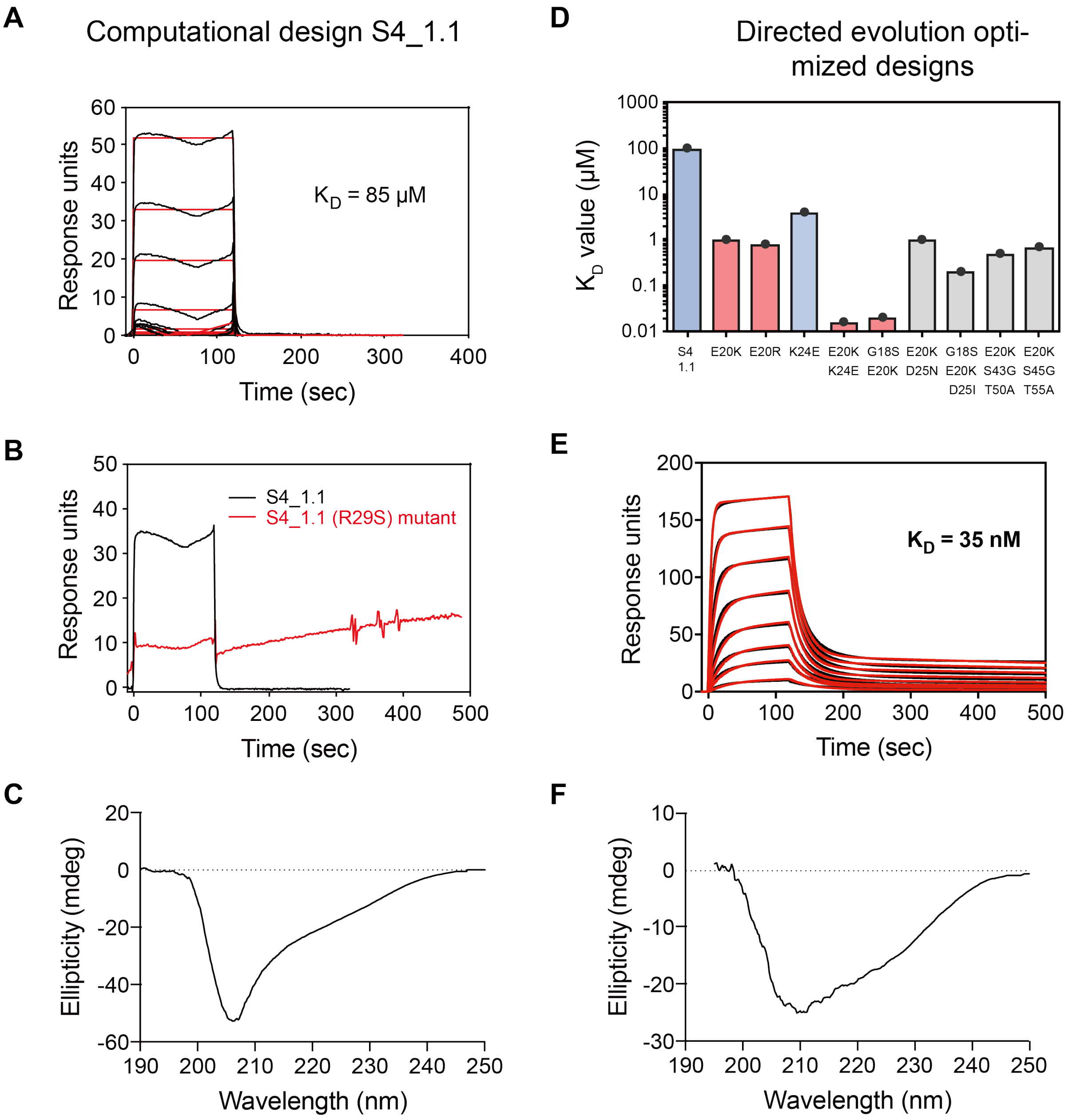


Figure S 3.2: Experimental characterization of S4_1 design series.
(A) Surface plasmon resonance measurement for the initial computational design S4_1.1 against 101F antibody revealed a dissociation constant of > 85 µM. **(B)** Despite low affinity, an R29S mutant revealed that binding was specific to the epitope of interest. **(C)** Circular dichroism spectrum of S4_1.1 at 20 °C. **(D)** Dissociation constants (K_D) for single and combined mutations of S4_1.1 that were identified in the deep mutational scanning screen. E20K/K24E double mutant (named S4_1.5) showed a binding affinity of 35 nM. **(E)** SPR sensorgram of S4_1.5 against 101F. **(F)** Circular dichroism spectrum of S4_1.5 at 20 °C.

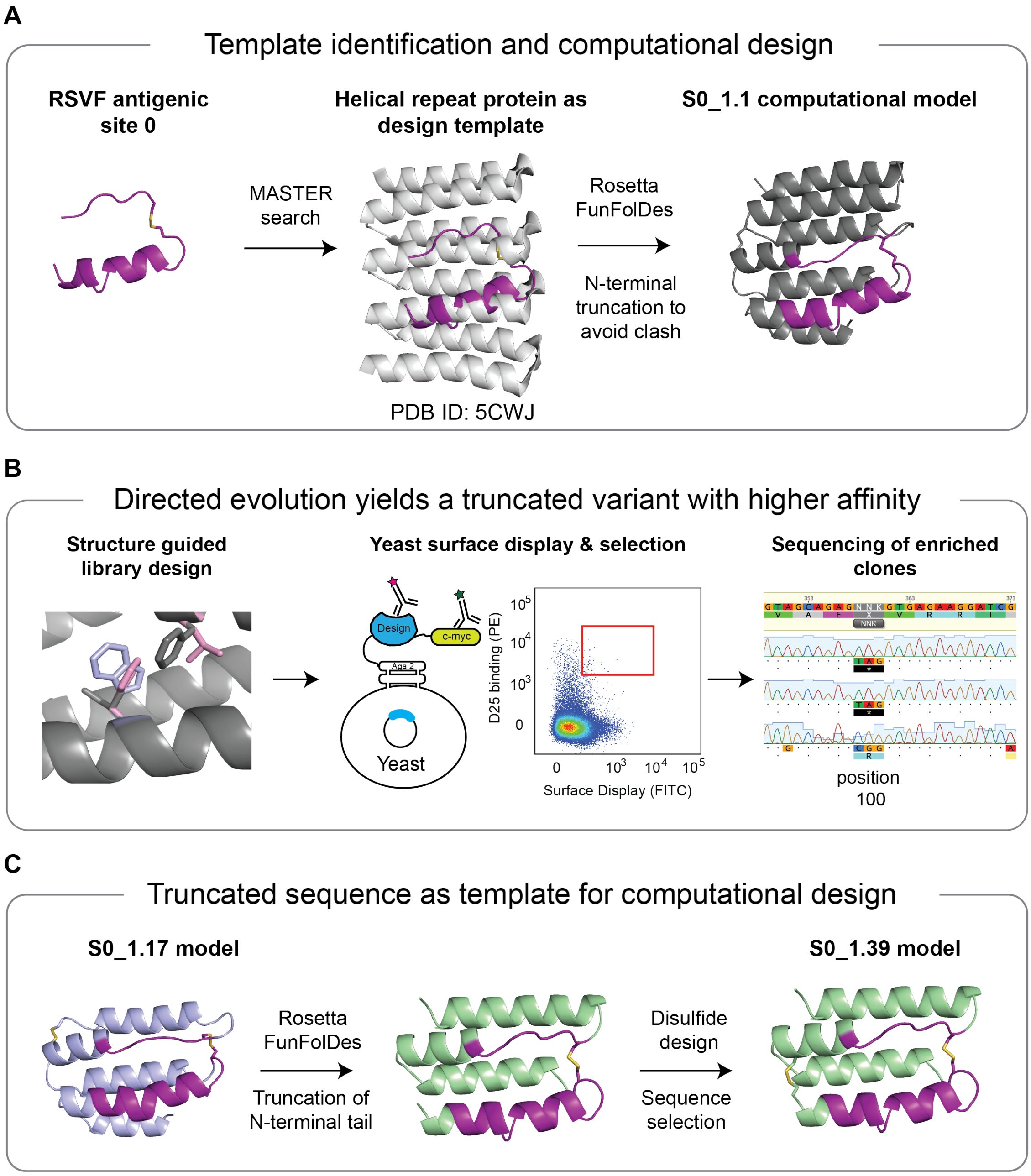

A
Template identification and computational design
RSVF antigenic site 0
MASTER search
Helical repeat protein as design template
Rosetta FunFolDes
N-terminal truncation to avoid clash
S0_1.1 computational model
PDB ID: 5CWJ
B
Directed evolution yields a truncated variant with higher affinity
Structure guided library design
Yeast surface display & selection
Design
c-myc
Aga 2
Yeast
D25 binding (PE)
Surface Display (FITC)
0
10^3
10^4
10^5
Sequencing of enriched clones
NNK
position 100
C
Truncated sequence as template for computational design
S0_1.17 model
Rosetta FunFolDes
Truncation of N-terminal tail
Disulfide design
Sequence selection
S0_1.39 model

Figure S 3.3: Computational design and experimental optimization of S0_1 design series.
(A) Template identification and design. Using MASTER, we identified a designed helical repeat protein (PDB ID: 5CWJ) to serve as design template to present and stabilize antigenic site 0 (see methods for details). The N-terminal 29 residues were truncated to avoid clashing with the D25 antibody, and Rosetta FunFolDes was used to design S0_1.1. See methods for details on the design process. **(B)** Based on S0_1.1, a combinatorial sequence library was constructed and screened using yeast surface display. After three consecutive sorts of high-affinity binding clones, individual colonies were sequenced. Position 100 was frequently found to be mutated to a stop codon, leading to a truncated variant with increased expression yield, and a ~5-fold improved binding affinity to D25 (Fig S3.4). **(C)** A model of the truncated variant served as template for a second round *of in silico* folding and design. We truncated the template further by the N terminal 14 residues, and introduced a disulfide bond between residues 1 and 43, leading to S0_1.39. See methods for full details on the design selection process.

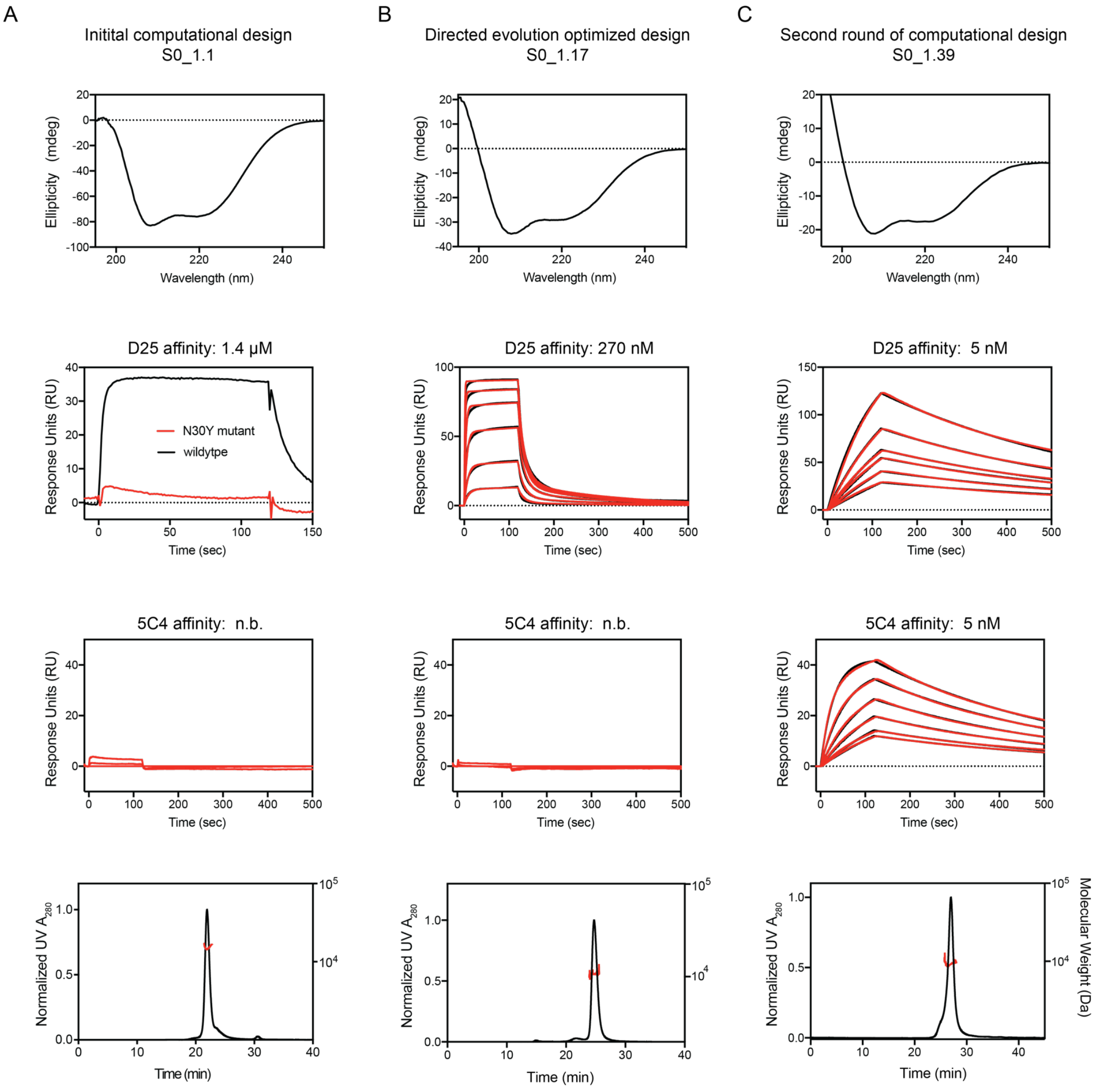


Figure S 3.4: Biophysical characterization of the S0_1 design series.

Top: Circular dichroism spectra at 20 °C. Middle: Surface plasmon resonance measurements against D25 and 5C4. Bottom: Multi-angle light scattering coupled to size exclusion chromatography. **(A)** S0_1.1 bound with a K_D of 1.4 µM to D25 and no detectable binding to 5C4. To verify that the binding interaction was specific to the epitope we generated a knockout mutant (N30Y) and observed that the binding interaction was absent. **(B)** S0_1.17 showed a K_D of 270 nM to D25 and no binding to 5C4. **(C)** SPR sensorgrams of S0_1.39 binding to D25 and 5C4 antibodies. D25 or 5C4 IgG was immobilized as ligand on the sensor chip surface, and S0_1.39 was flown as analyte. All designs showed CD spectra typical of helical proteins and behaved as monomers is solution (Top and bottom rows).

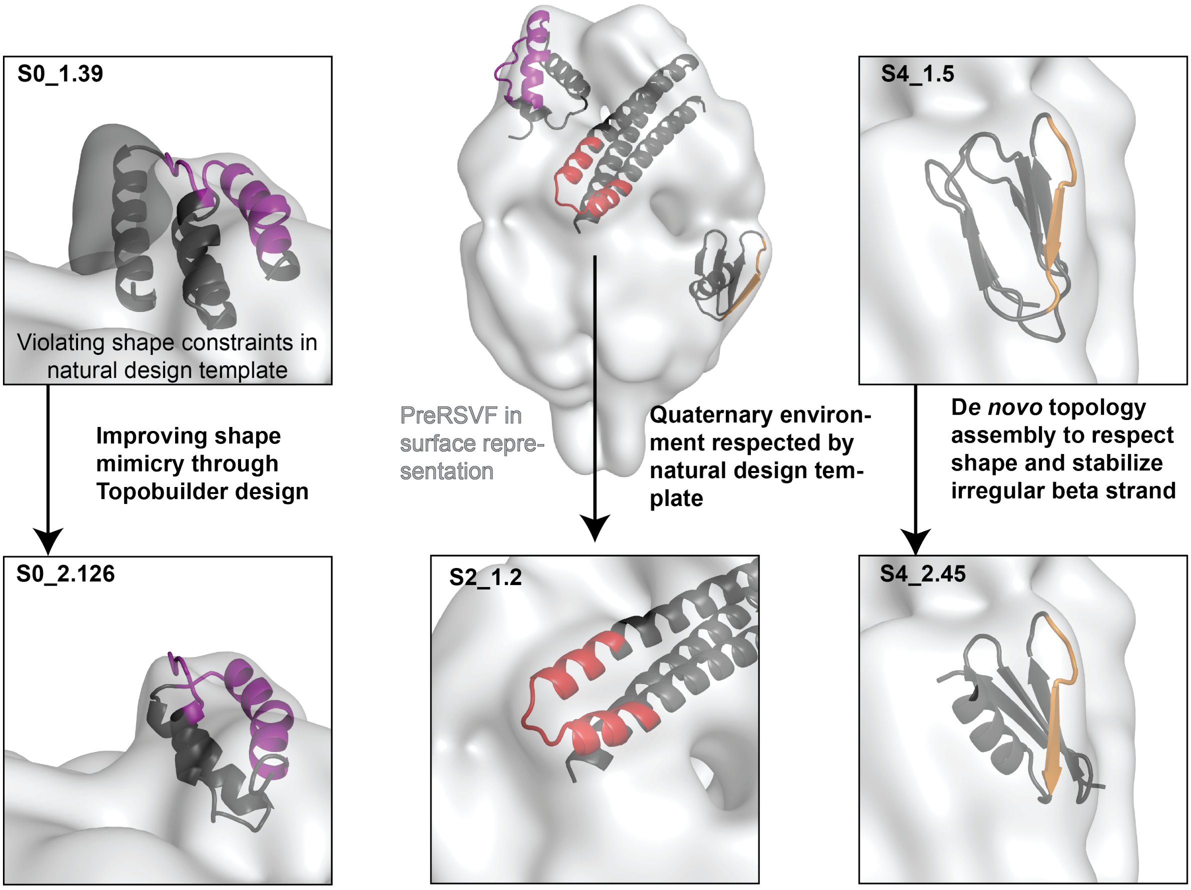

S0_1.39
Violating shape constraints in natural design template
Improving shape mimicry through Topobuilder design
S0_2.126
PreRSVF in surface representation
Quaternary environment respected by natural design template
S2_1.2
S4_1.5
De *novo* topology assembly to respect shape and stabilize irregular beta strand
S4_2.45

Figure S 3.5: Shape mimicry of computationally designed immunogens compared to prefusion RSVF.
Prefusion RSVF is shown in surface representation (light grey), with designed immunogens superimposed. Close-up views are shown for template-based designs (S0_1.39 and S4_1.5, top row). While site 0 is freely accessible for antibody binding in preRSVF, the C-terminal helix of S0_1.39 constrains its accessibility (dark grey surface). Through defined backbone assembly using TopoBuilder, S0_2.126 was designed, mimicking the native quaternary environment of site 0 (bottom left). RSVF antigenic site II, which is a structurally simple helix-turn-helix motif frequently found in natural proteins, was previously designed based on a design template that respects the quaternary constraints of site II in its native environment (S2_1.2, bottom middle). For site IV, a topology was assembled (S4_2.45) that respects the shape constraints while improving the stabilization of the irregular, bulged beta strand compared to the S4_1.5 design (right).

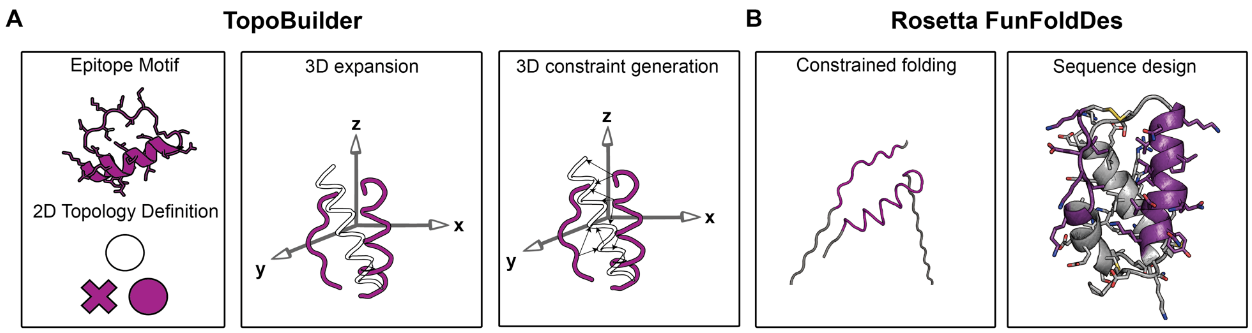

A
TopoBuilder
B
Rosetta FunFoldDes
Epitope Motif
2D Topology Definition
3D expansion
z
x
y
3D constraint generation
z
x
y
Constrained folding
Sequence design

Figure S 3.6: TopoBuilder design strategy.
(A) The motif of interest is extracted from its native environment, and a 2D form is generated that allows to connect the discontinuous epitope segments. The 2D form is then expanded to the 3D space, applying user defined rotations and translations along x,y and z coordinates. From the 3D sketch, Cα constraints are generated to guide the folding process. **(B)** Rosetta FunFolDes is used to fold the idealized 3D sketch (using fragment insertions of sizes 3 and 9), and to build connecting loops between the secondary structures. A sequence that stabilizes the folded pose is designed in a last step using Rosetta FastDesign. Further details on the design process, the TopoBuilder code and scripts used for folding and design are available in the online repository.

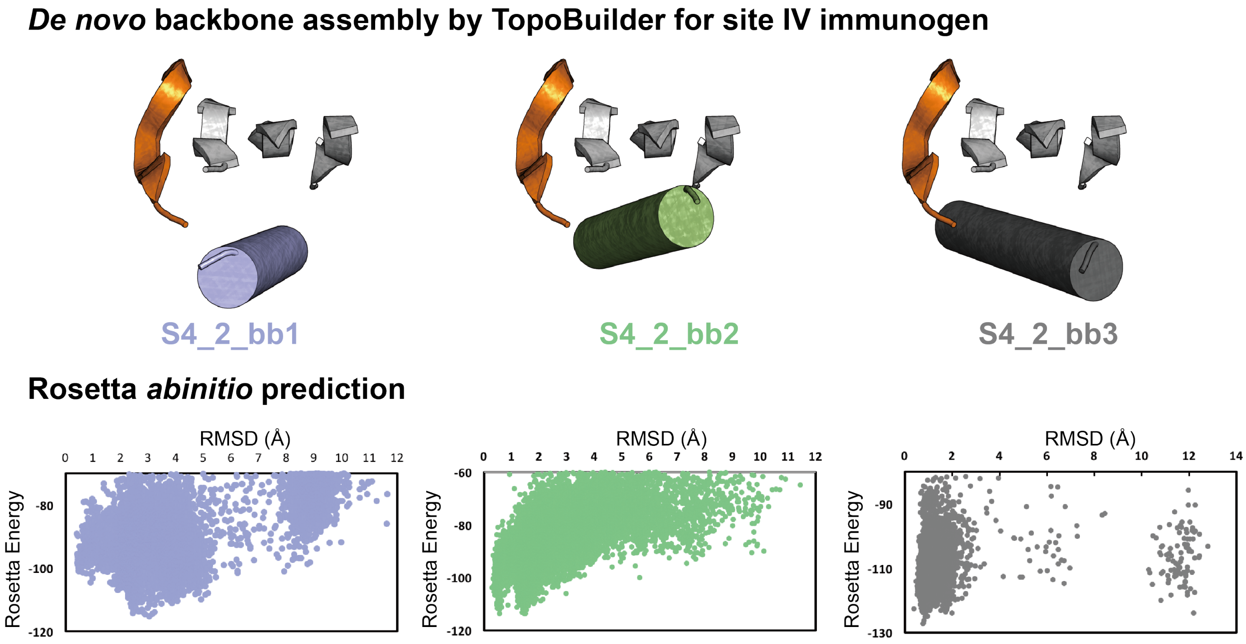


Figure S 3.7: *De novo* backbone assembly for site IV immunogen.
The site IV epitope was stabilized with three antiparallel beta strands built *de novo*, and a helix packing in various orientations against this beta sheet (bb1-bb3). Each backbone was simulated in Rosetta *ab initio* simulations for its ability to fold into a low energy state that is close to the design model, indicating that S4_2_bb2 and bb3 have a stronger tendency to converge into the designed fold.

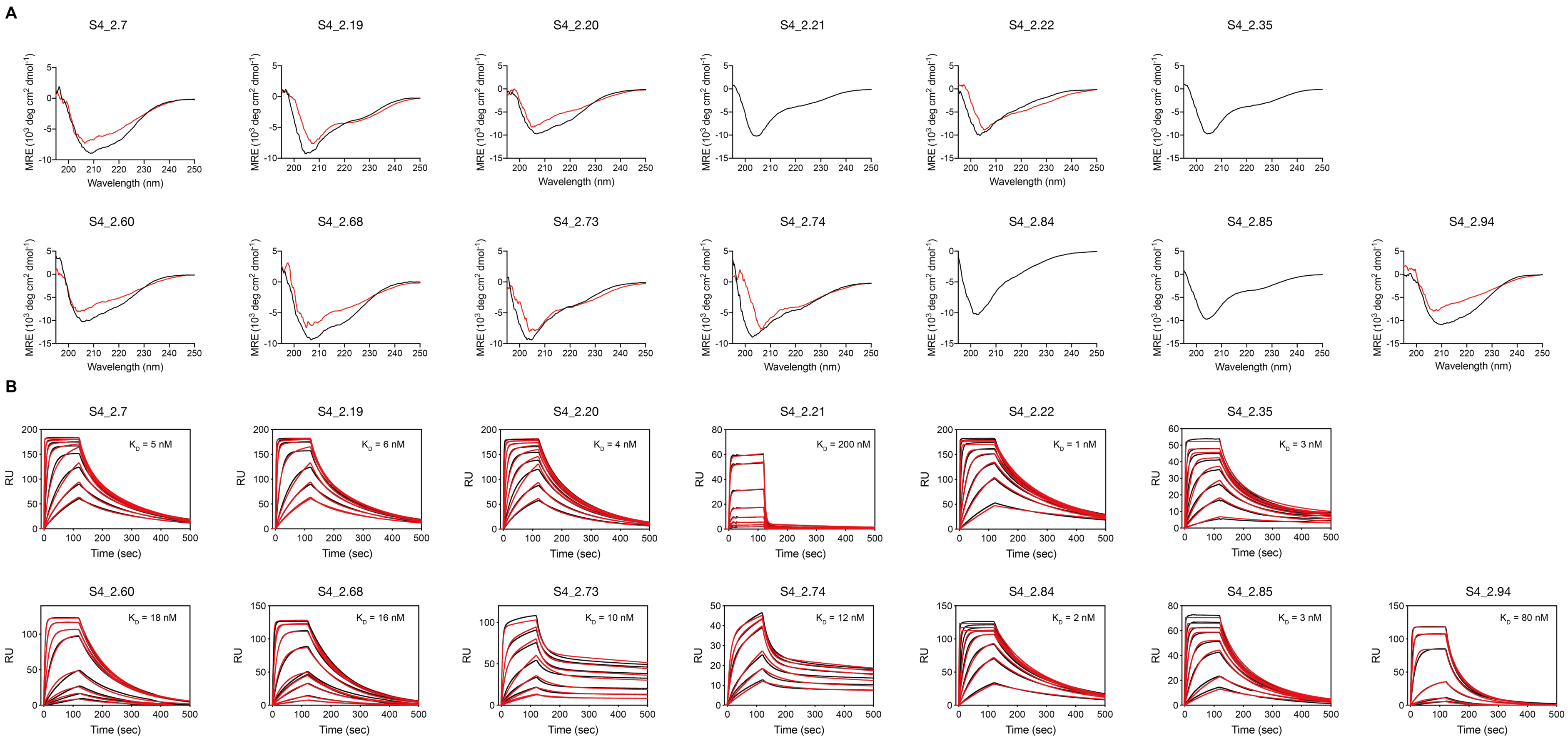


Figure S 3.8: Biophysical characterization of S4_2 design series.

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

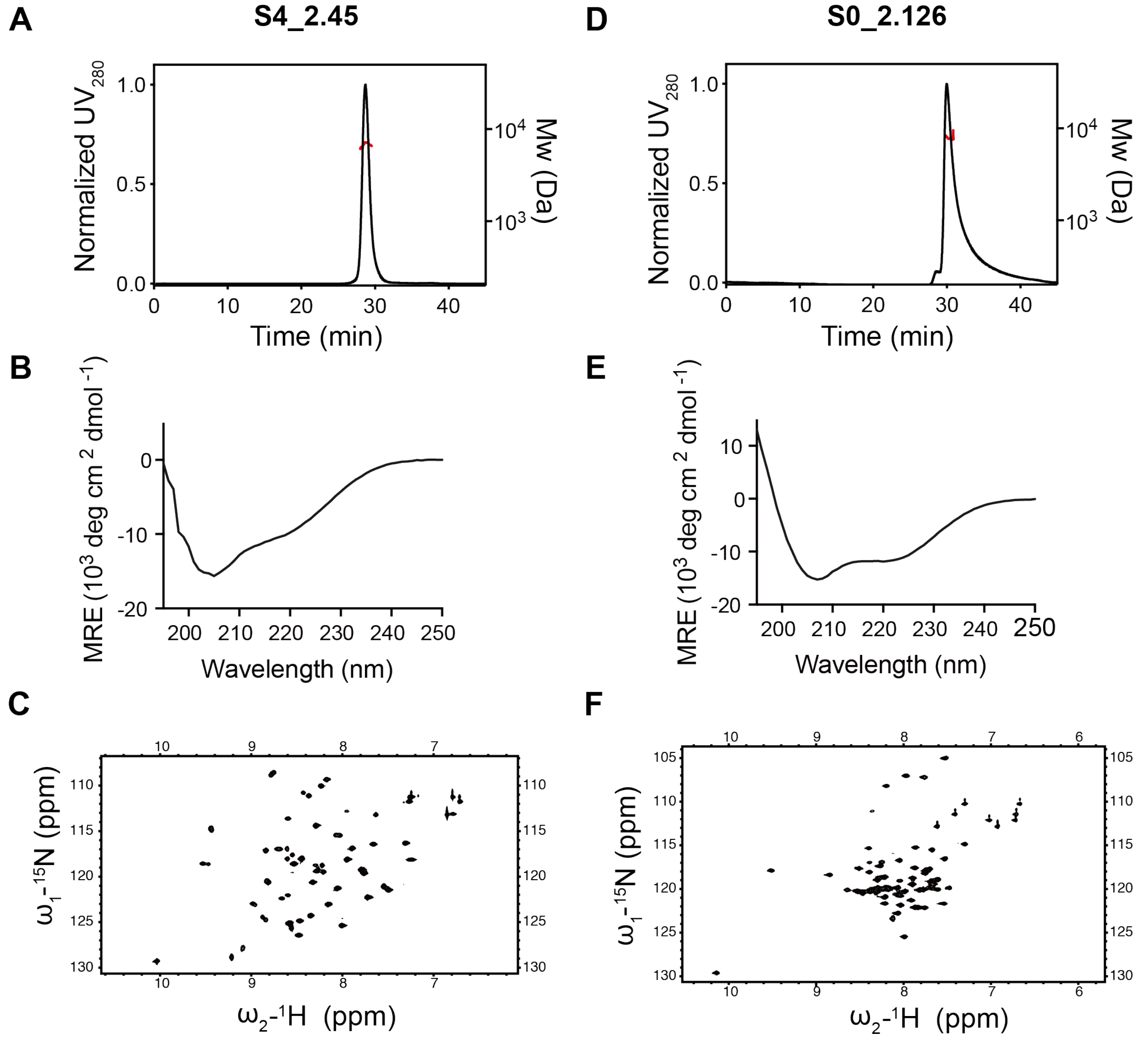


Figure S 3.9: Biophysical characterization of S4_2.45 and S0_2.126.
A,D: S4_2.45 (A) and S0_2.126 (D) are monomeric in solution as shown by SEC-MALS profile. **B,E:** Circular dichroism spectra at 25 °C. **C,F:** 2D NMR of ^{15}N HSQC spectra for S4_2.45 (C) and S0_2.126 (F) are well dispersed, confirming that the designs are well folded in solution. Mw: Molecular Weight.

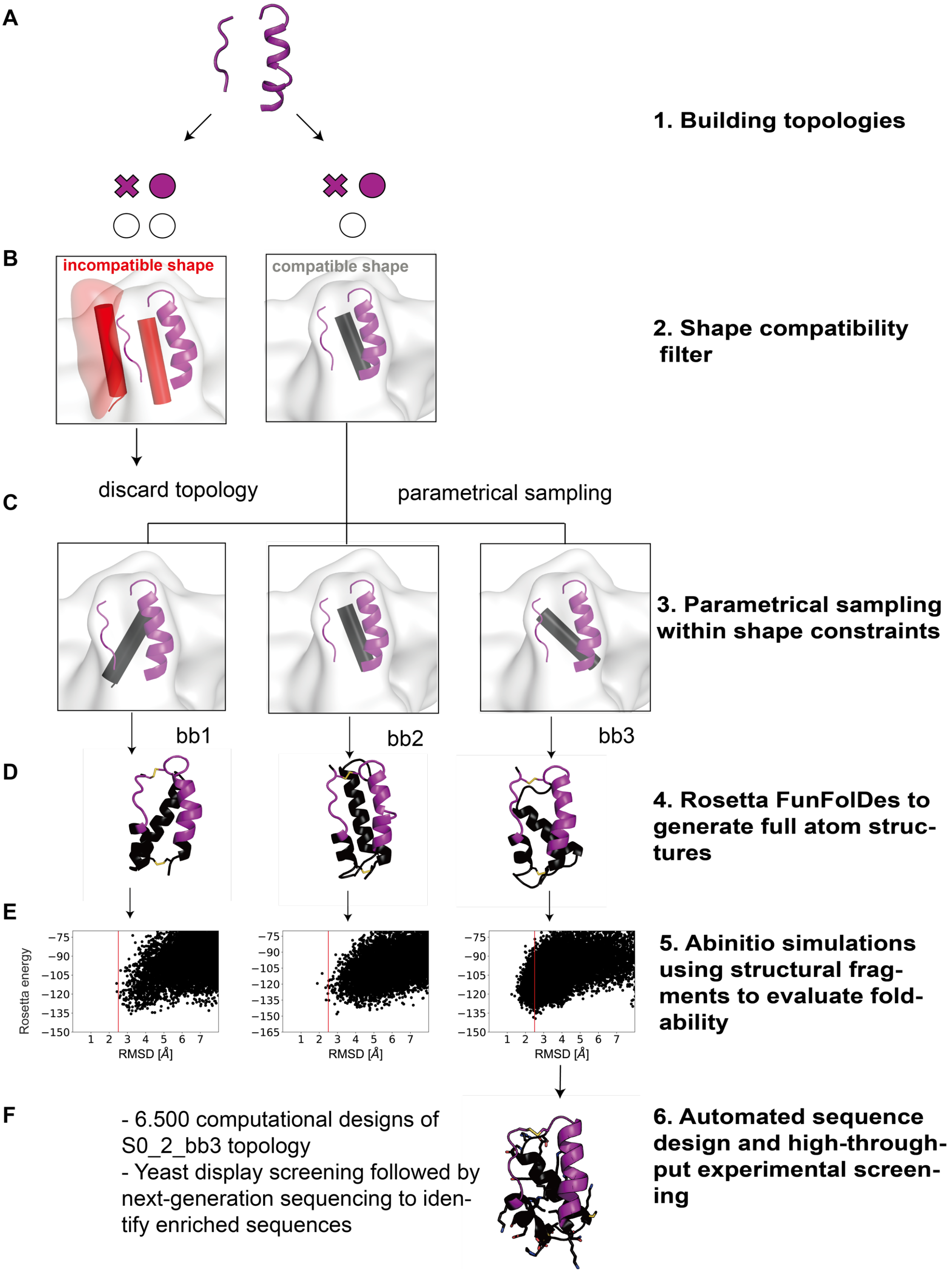

A
1. Building topologies
B
incompatible shape
compatible shape
2. Shape compatibility filter
discard topology
C
parametrical sampling
3. Parametrical sampling within shape constraints
bb1
bb2
bb3
D
4. Rosetta FunFolDes to generate full atom structures
E
Rosetta energy
−75
−90
−105
−120
−135
−150
−165
1 2 3 4 5 6 7
RMSD [Å]
5. Abinitio simulations using structural fragments to evaluate foldability
F
- 6.500 computational designs of S0_2_bb3 topology
- Yeast display screening followed by next-generation sequencing to identify enriched sequences
6. Automated sequence design and high-throughput experimental screening

Figure S 3.10: *De novo* topology assembly to stabilize site 0 using TopoBuilder.
(A) Topologies that allow to connect the discontinuous site 0 motif are built as 2D form, and then translated into a 3D sketch. **(B)** Generated sketches are evaluated for their compatibility with the shape constraints of prefusion RSVF. A compatible shape (right) respects the quaternary constraints of the motif in its native environment, whereas an incompatible shape (left) is one that sterically constrains the motif's accessibility in a non-native manner (red surface). **(C)** Once a compatible topology is identified, the secondary structure is parametrically sampled (3 different helical orientations, S0_2_bb1-bb3). **(D)** Three customized helical orientations were assembled (S0_2_bb1-bb3) to support site 0 epitope, and evaluated for their ability to fold into the designed topology in Rosetta *ab initio* simulations **(E)**. S0_2_bb3 showed a funnel-shaped energy landscape, and was selected for subsequent automated sequence design and experimental screening **(F)**.

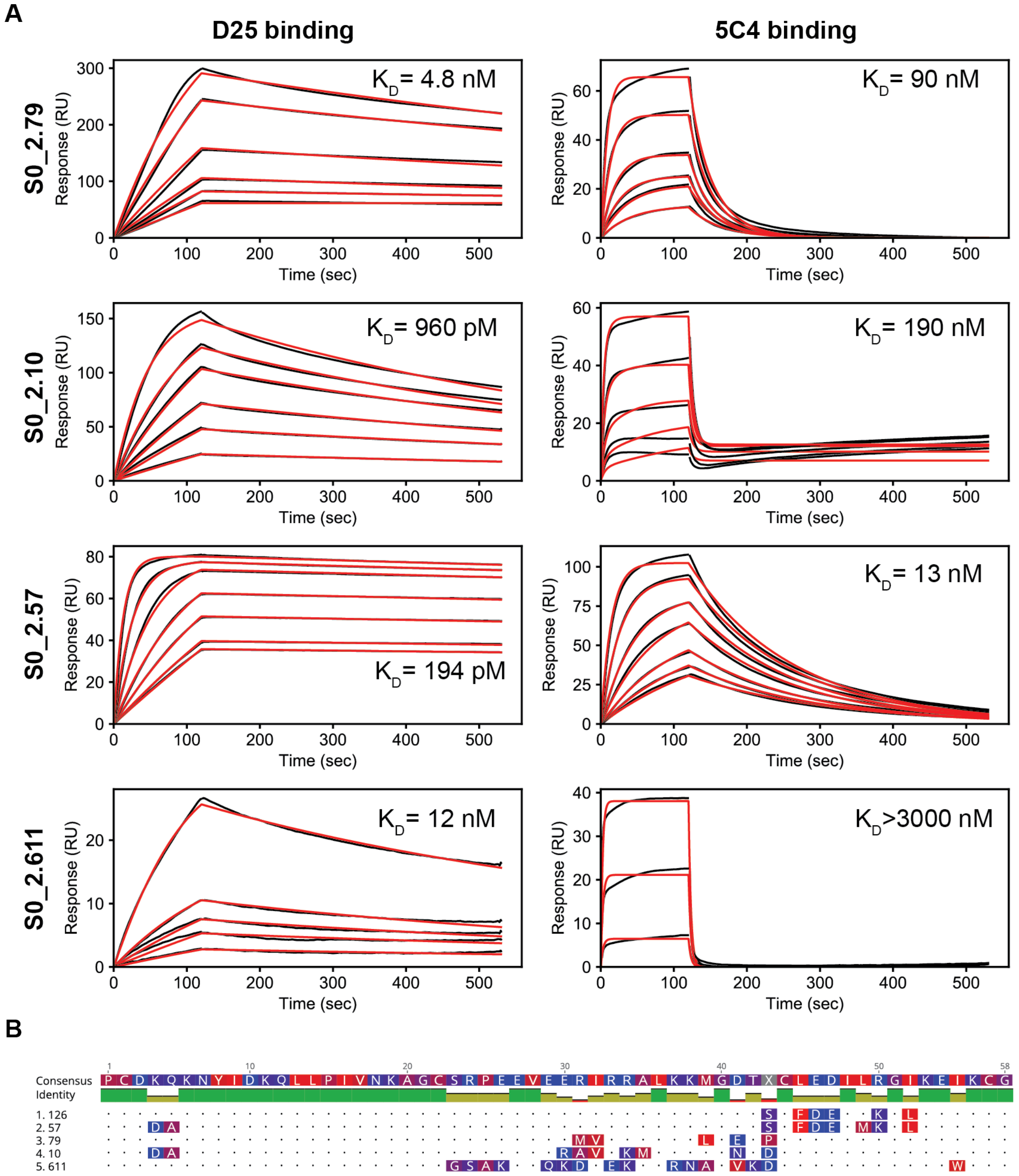


Figure S 3.11: Biophysical characterization of S0_2 design series.
(A) Binding affinity measurement for D25 and 5C4 binding of *de novo* site 0 scaffolds. Shown are the SPR sensorgrams of enriched designs that were successfully expressed and purified after the yeast display selection. D25 or 5C4 IgG were immobilized on the sensor chip surface, and scaffolds were injected as analyte. Dissociation constants shown are kinetic fits using a 1:1 Langmuir model. **(B)** Sequence alignment of experimentally characterized sequences, in comparison to S0_2.126. The closest sequence homolog to S0_2.126 is S0_2.57, differing in 3 amino acids.

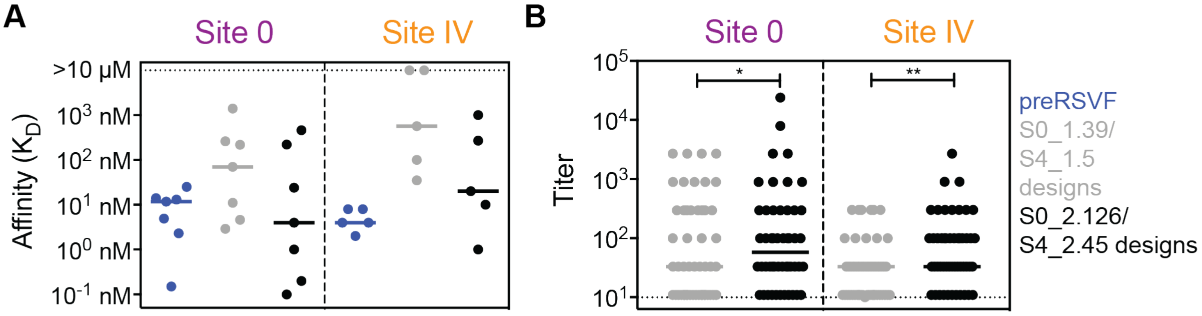


Figure S 3.12: Binding affinity towards panels of human neutralizing antibodies and human sera.

(A) Binding affinity (K_D, determined by SPR flowing Fabs as analyte) of S0_1.39 (grey) and S0_2.126 (black) towards a diverse panel of site-specific neutralizing antibodies, in comparison to prefusion RSVF (blue). Antibodies shown for site 0 are 5C4, D25 (*1*), ADI-14496, ADI-18916, ADI-15602, ADI-18900 and ADI-19009 (*2*). For site IV, the binding affinity was tested against 101F (*3*), ADI-15600 (*2*), 17E10, 6F18 and 2N6 (*5*), comparing S4_1.5 (grey) and S4_2.45 (black) to prefusion RSVF. The higher binding affinity of the second-generation designs (S0_2.126 and S4_2.45) compared to the first-generation and to prefusion RSVF indicates a greatly improved, near-native epitope mimicry of the respective antigenic sites in the designed immunogens. **(B)** ELISA reactivity of designed immunogens with sera obtained from 50 healthy human adults that were seropositive for prefusion RSVF. Both S0_2.126 and S4_2.45 showed significantly increased reactivity compared to the first-generation designs, confirming an improved epitope-mimicry on the serum level (* $p < 0.05$ and ** $p < 0.01$, Wilcoxon test). Data are representative from one out of two independent experiments.

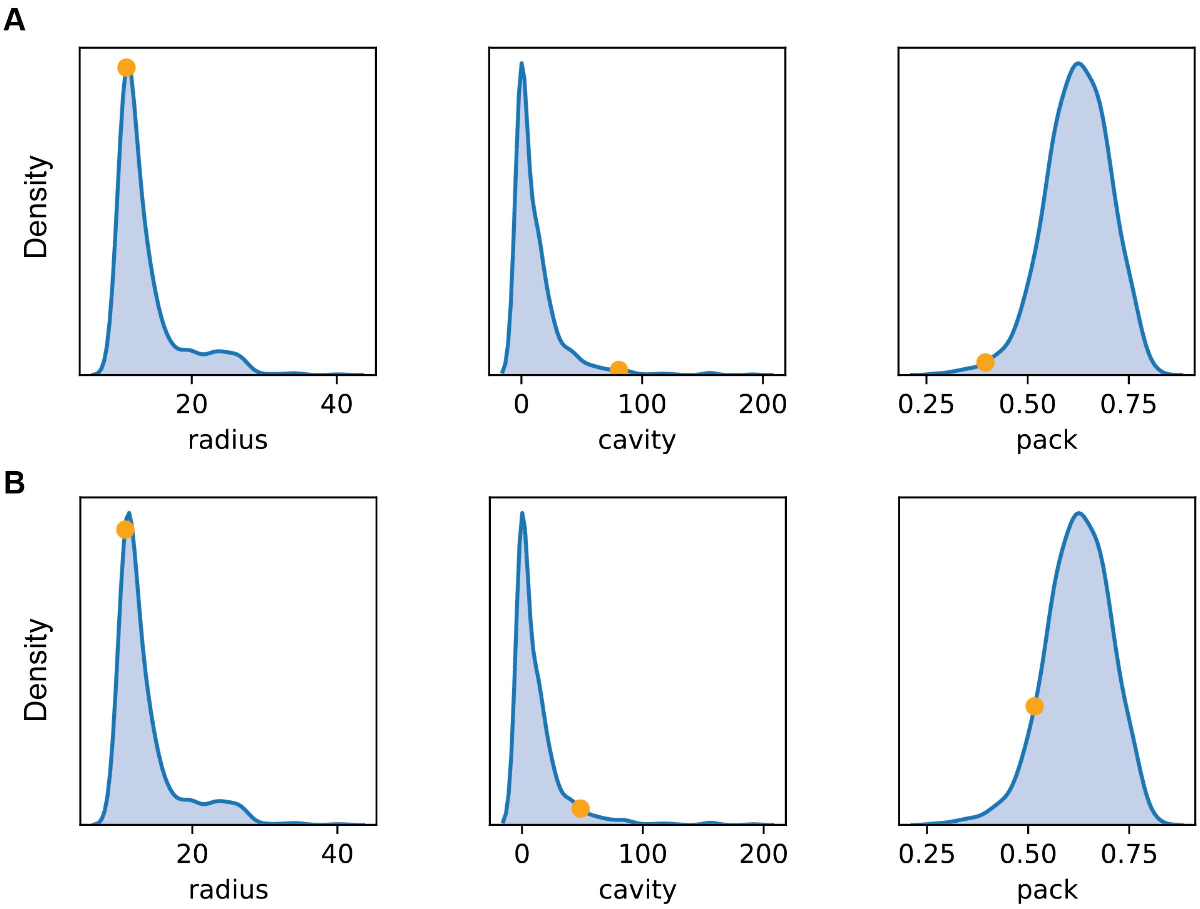


Figure S 3.13: Comparison of S0_2.126 Rosetta scores against natural proteins of similar size.
Protein structures within the same size as S0_2.126 (57 +/- 5 residues) were downloaded from the CATH database and filtered by 70 % sequence homology, yielding a representative database of natural proteins with similar size as S0_2.126 (n = 1,013 structures). Proteins were then minimized and scored by Rosetta to compute their radius of gyration, intra-protein cavities (cavity) and core packing (packstat). Plotted is the distribution for these score terms in 1,013 natural proteins (blue density plot), and the same scores for S0_2.126 are shown in orange. The NMR structure of S0_2.126 is shown in **(A)**, the computational model of S0_2.126 is shown in **(B)**, indicating that, despite similar radius of gyration, S0_2.126 shows a substantial cavity volume as well as a very low core packing compared to natural proteins of similar size. CATH database and scores were pre-calculated, loaded and visualized using the rstoolbox python library (*4*).

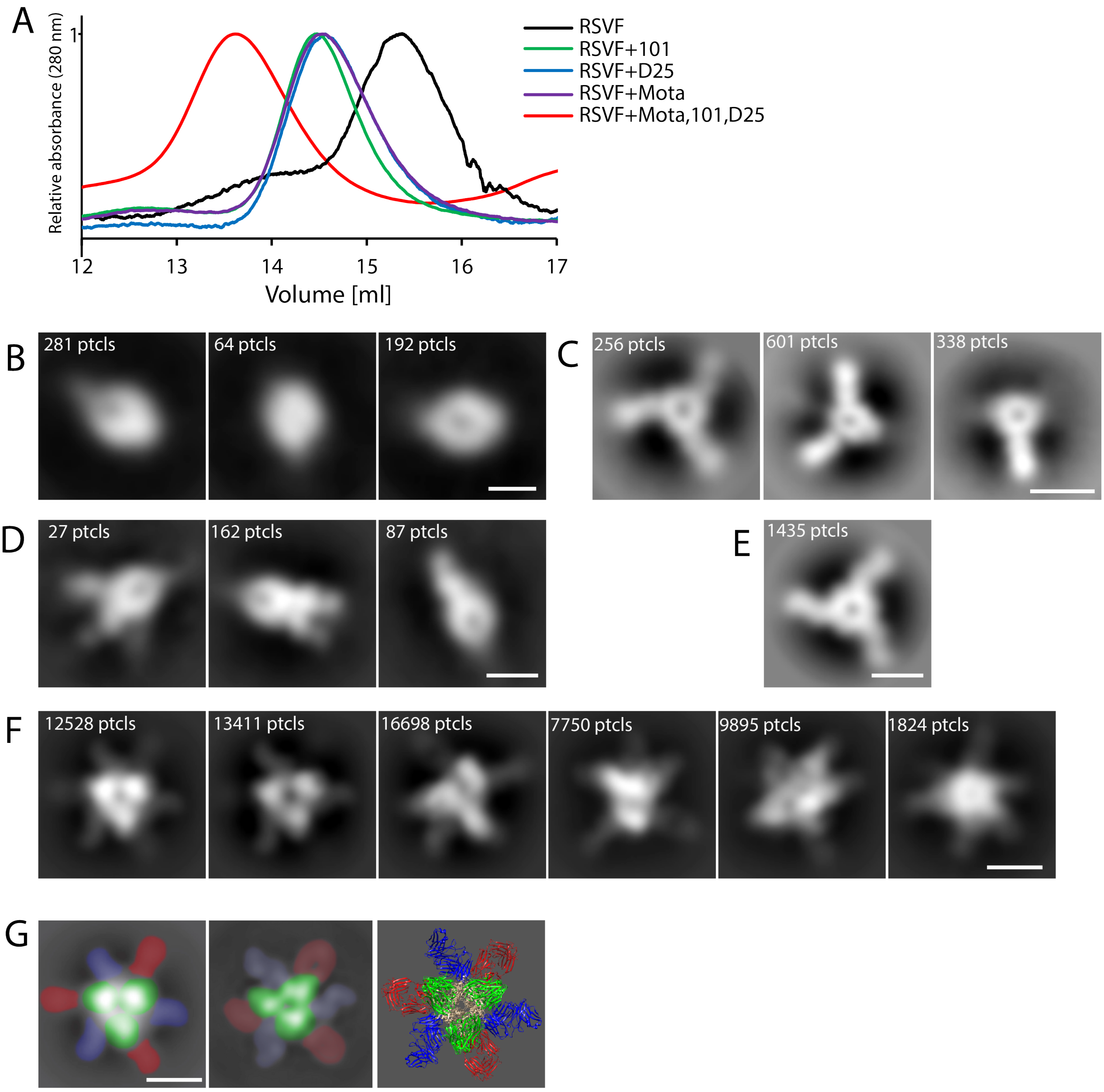

A
Relative absorbance (280 nm)
1
12
13
14
15
16
17
Volume [ml]
RSVF
RSVF+101
RSVF+D25
RSVF+Mota
RSVF+Mota,101,D25
B
281 ptcls
64 ptcls
192 ptcls
C
256 ptcls
601 ptcls
338 ptcls
D
27 ptcls
162 ptcls
87 ptcls
E
1435 ptcls
F
12528 ptcls
13411 ptcls
16698 ptcls
7750 ptcls
9895 ptcls
1824 ptcls
G

Figure S 3.14: Electron microscopy analysis of site-specific antibodies in complex with prefusion RSVF.
(A) Superposed size-exclusion profiles of unliganded RSVF (black line) and RSVF in complex with 101F (green line), D25 (blue line), Mota (purple line) and all three (101F, D25, Mota - red line) Fabs. **(B-F)** Representative reference-free 2D class averages of the unliganded RSVF trimer (B) and RSVF in complex with 101F (C), D25 (D), Mota (E) or all three (101F, D25, Mota (F)) Fabs. Fully-saturated RSVF trimers bound by Fabs are observed, as well as sub-stoichiometric classes. **(G)** Left panel: reference-free 2D class average of RSVF trimer with three copies of 101F, D25 and Mota Fabs visibly bound. The predicted structure of RSVF in complex with 101F, D25 and Mota was used to simulate 2D class averages in Cryosparc2, and simulated 2D class average with all three types of Fabs is shown in the middle panel. Right panel: predicted structure of RSVF trimer with bound 101F, D25 and Mota Fabs based on the existing structures of RSVF with individual Fabs (PDB ID 4JHW, 3QWO and 3O45). Fabs are colored as follow: red - 101F; blue - Mota; green - D25. Scale bar - 100 Å.

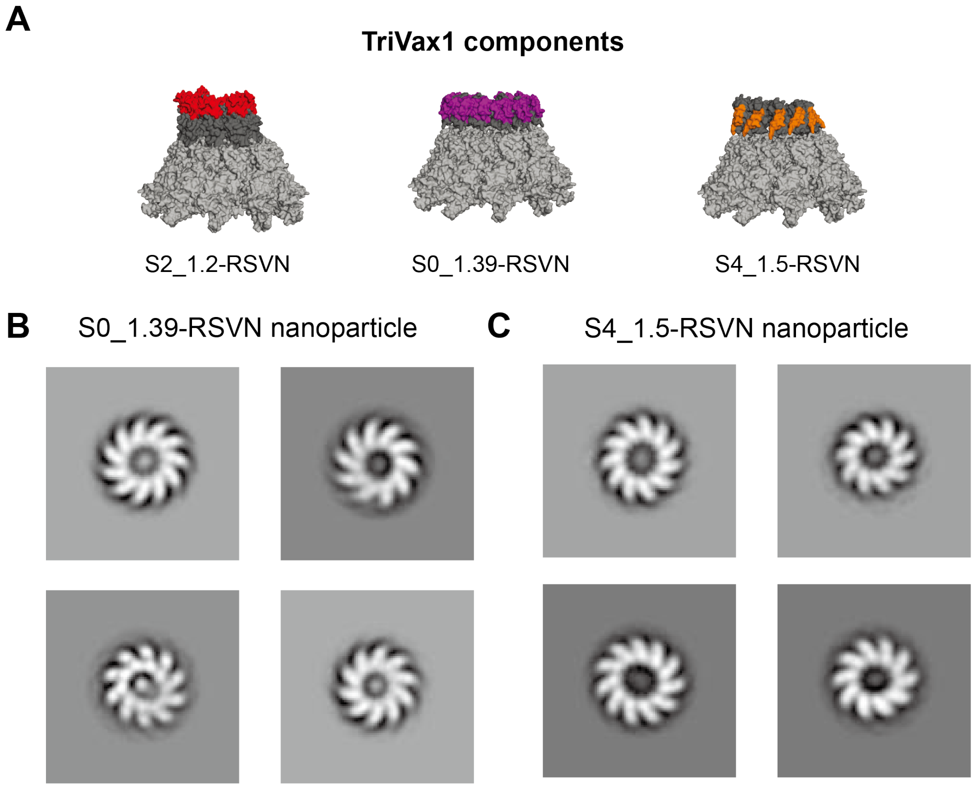

A
TriVax1 components
S2_1.2-RSVN
S0_1.39-RSVN
S4_1.5-RSVN
B
S0_1.39-RSVN nanoparticle
C
S4_1.5-RSVN nanoparticle

Figure S 3.15: Composition and EM analysis of Trivax1 RSVN nanoparticles.
(A) Trivax1 contains equimolar amounts of site II, 0 and IV epitope focused immunogens fused to the self-assembling RSVN nanoparticle with a ring-like structure (n = 10 - 11 subunits). The site II-RSVN nanoparticle has been described previously (*185*). Shown are the computational models for the nanoparticles-immunogen fusion proteins. **(B,C)** Negative stain electron microscopy for S0_1.39-RSVN and S4_1.5-RSVN nanoparticles confirms that the ring-like structure is maintained upon fusion of the designed immunogens.

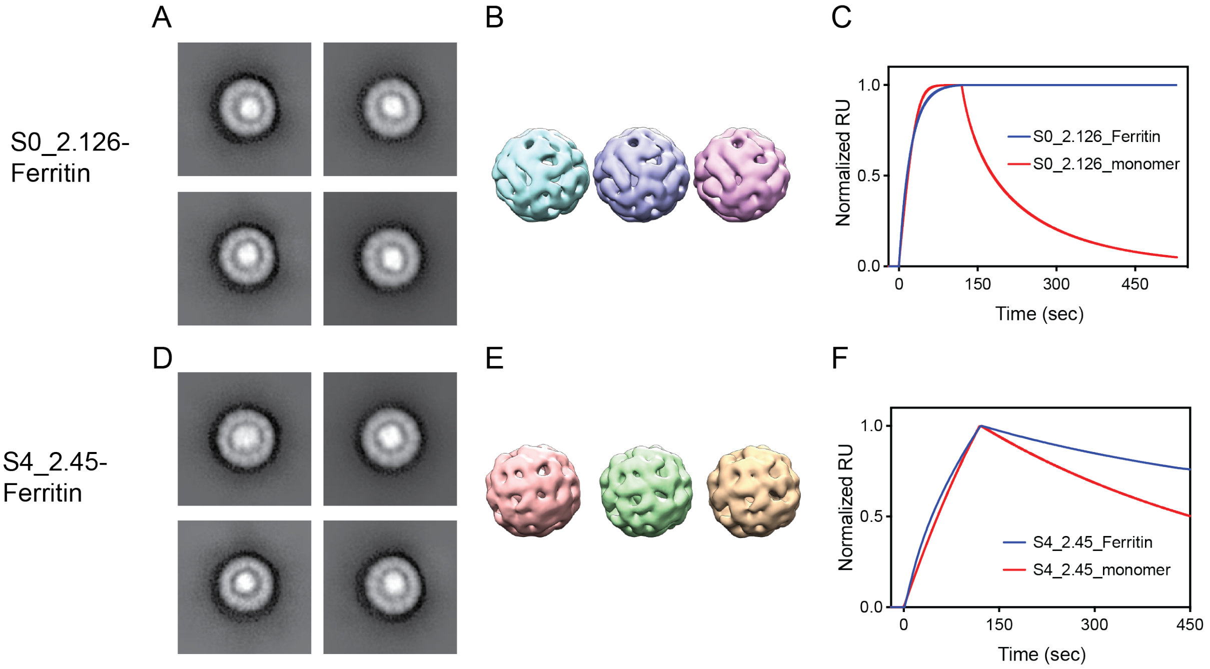


Figure S 3.16: EM analysis of Trivax2 ferritin nanoparticles.

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

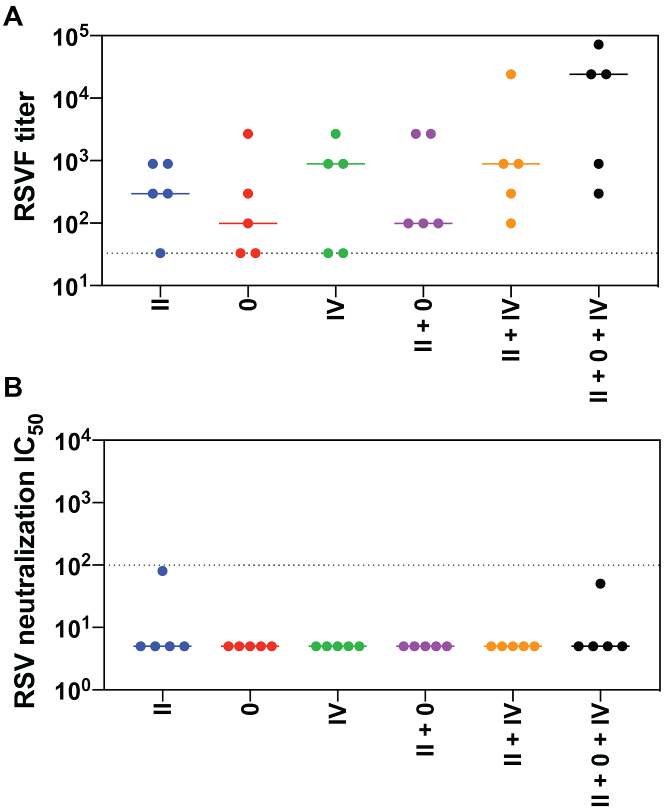


Figure S 3.17: Mouse immunization studies with Trivax1.
(A) RSVF cross-reactivity of epitope-focused immunogens formulated individually, as cocktail of two, and three (Trivax1). **(B)** RSV neutralizing serum titer of mice immunized with designed immunogens and combinations thereof.

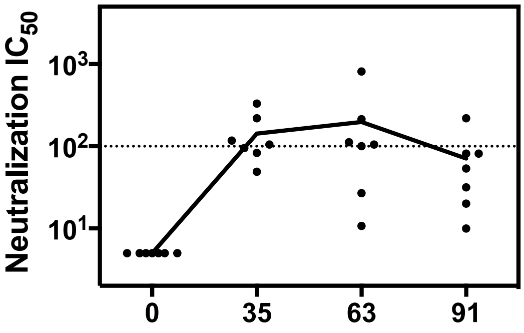


Figure S 3.18: NHP neutralization titer measured by an independent laboratory.
Sera from indicated timepoints were tested for RSV neutralization by an independent laboratory in a different RSV neutralization assay, using a Vero-118 cell line and a GFP readout.

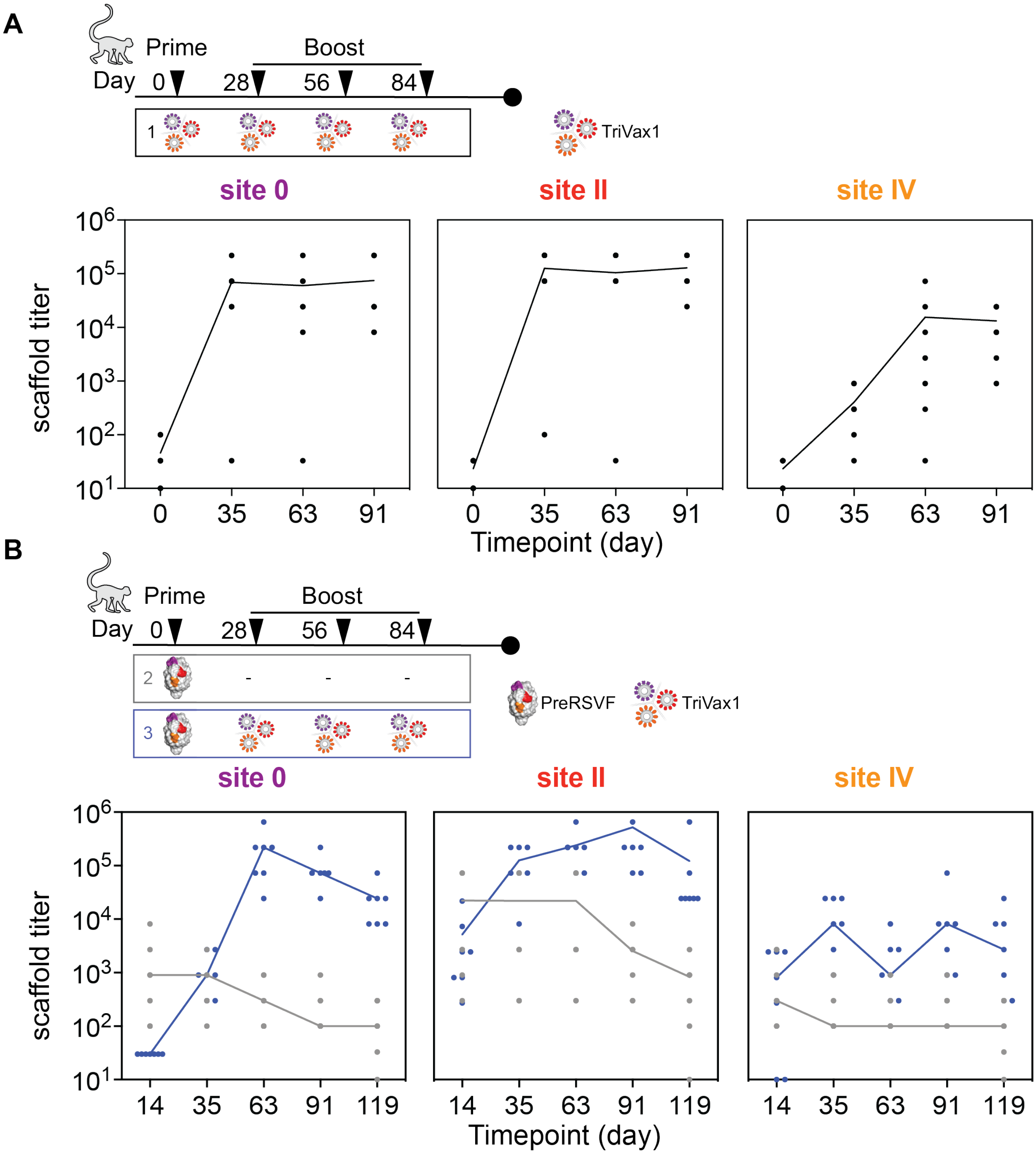


Figure S 3.19: NHP serum reactivity with designed immunogens.

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

Chapter 4 A bottom-up design approach for the *de novo* design of functional proteins

This chapter presents ongoing work towards a more general framework for the *de novo* design of proteins carrying functional binding sites. Using the same RSV epitopes as in chapter 3, this section presents a large-scale protein design effort to design multiple protein topologies accommodating the same binding motifs. Beyond their application as immunogens shown in previous chapters, this work uses the *de novo* proteins in a biosensor platform for the detection and quantification of RSVF epitope-specific antibodies. Although further validation is needed, we foresee a great potential for the designed scaffolds as versatile tools for monitoring human serum responses with high resolution, i.e. on the single epitope level.

A manuscript based on this work is currently in preparation. My contribution to this work is the design and characterization of the presented site 0 scaffolds, and I conceived the work related to the biosensors. Che Yang designed and characterized the site IV scaffolds and the bi-functional scaffold, and solved x-ray structures. Jaume Bonet developed the TopoBuilder algorithm. The experimental work related to the biosensors was performed by a master's student (Eva van Aalen) in our lab who worked under my supervision, and is now continuing to work on this project in Maarten Merkx's lab. In collaboration with Leo Scheller and Martin Fussenegger, we are currently exploring the bi-functional scaffold for synthetic biology applications.

Preliminary author list:

Yang C[1,2*], **Sesterhenn F[1,2*]**, Aalen E[3], Scheller L[4], Bonet J[1,2], Cramer JT[5], Wen X[6], Abriata LA[1,2], Rosset S[1,2], Georgeon S[1,2], Vollers SS[1,2], Jardetzky T[6], Krey T[5,7], Fussenegger M[4], Merkx M[3] & Correia BE[1,2].

***These authors contributed equally.**

1 Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland. 2 Swiss Institute of Bioinformatics (SIB), Lausanne CH-1015, Switzerland. 3 Laboratory of Chemical Biology and Institute for Complex Molecular Systems, Department of Biomedical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands. 4 Department of Biosystems Science and Engineering, ETH Zurich, Basel CH-4058, Switzerland. 5 Institute of Virology, Hannover Medical School, Germany. 6 Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305, USA; 7 German Center for Infection Research (DZIF), Hannover, Germany.

4.1 Introduction

De novo protein design has emerged as a powerful approach to delineate the rules governing protein folding, and translate these rules into the design of novel proteins with defined structures (*211, 212*). Numerous successes have been reported for the *de novo* design of diverse protein folds, with the primary design goal for the vast majority of these studies being structural accuracy (*37, 39, 42, 213*). In contrast, the design of *de novo* functional proteins is still in its infancy (*214*). Nonetheless, successes reported illustrate the potential of *de novo* design to transform multiple arenas of biology and biotechnology, including the design of vaccine candidates (*59, 185, 215*), synthetic signaling molecules (*62*), antivirals (*51*), pH-responsive carriers (*216*) and others (*217-219*).

Central to numerous protein functions is the ability to specifically interact with other molecules, including proteins, small molecules, nucleic acids, metals and others (*9*). Thus, albeit often not the only determinant of protein function, a central requirement for *de novo* design of functional proteins is to endow *de novo* proteins with structural motifs that mediate binding interactions. Beyond the difficulty of designing structurally accurate *de novo* proteins, functional design requires an exceptional level of accuracy, which is further complicated by the high structural complexity of many binding sites.

A widely used approach to design functional proteins is to transplant functional sites from one protein to another, known as protein grafting (*220*). A prerequisite for protein grafting is the availability of template structures with enough local structural similarity to accommodate the transplanted functional site. Thus, with few exceptions (*54*), grafting has been largely limited to single, regular secondary structures that are frequently found in natural proteins (*48, 49, 53*). However, the vast majority of functional sites is not contained in single, regular helical segments, but rather in multiple, often irregular structural segments that are stabilized by the overall protein structure (*57, 58, 182*).

Beyond nature's repertoire, *de novo* protein design allows the creation of novel structural templates that are deprived of any evolutionary baggage. In a subsequent step, the *de novo* proteins can be empowered with a functional motif, such as was recently shown for miniproteins endowed with influenza stem binding motifs (*51, 212*). We refer to this two-step approach, consisting of first building a stable, functionless scaffold, which subsequently serves as template for grafting, as a 'top-down' approach. There are several important limitations of a 'top-down' approach for functional protein design: first, most *de novo* proteins are built with a high proportion of regular secondary structures, high contact order and minimal loops (*37*). While these structures are generally highly stable, *de novo* proteins that were built 'function-agnostic' are unlikely to show enough local structural similarity to an irregular, multi-segment functional motif that would make them amenable to grafting. Second, beyond local structural similarity, the overall topology must be compatible with the designed function, as shown for *de novo* designed receptor agonists and epitope mimetics (*62, 215*).

In contrast to this 'top-down' approach, a 'bottom-up' strategy would consider the local structural and global topological requirements of the functional motif, irrespective of its complexity, and build supporting secondary structure elements to stabilize the functional site in its native conformation (*221*). A handful of studies have shown such a function-centric design strategy (*60-62, 217*), but the field is still lacking a general approach that allows the systematic building of *de novo* proteins with embedded functional motifs, controlled connectivity and precise control over the spatial positioning of each motif and secondary structure element (SSE).

Here, we describe a general ‘bottom-up’ design strategy to build and design *de novo* functional proteins. We demonstrate the power of this approach by designing various protein folds to accommodate irregular and discontinuous binding motifs. The chosen motifs are antibody binding motifs of different structural complexity, found on the respiratory syncytial virus fusion (RSVF) and glycoprotein (RSVG). Upon scaffolding these epitopes in their native conformation, we show their utility as biosensors to detect and quantify epitope-specific antibodies, which is potentially useful to inform vaccine design. Altogether, we present a versatile strategy for the *de novo* design of customized proteins carrying structurally complex binding motifs applicable to a wide range of functional protein design challenges.

4.2 Results

4.2.1 Computational design and high-throughput screening

We previously reported a prototype computational protocol for the *de novo* design of functional proteins, the TopoBuilder, and showcased its utility for the *de novo* design of epitope-focused immunogens (chapter 3). Here, we used an upgraded version of the TopoBuilder and present a general, bottom-up protein design approach for the *de novo* design of proteins endowed with functional binding motifs from other proteins (see methods and Fig S4.1).

As functional motifs, we chose four structurally well characterized antibody binding motifs of the respiratory syncytial virus fusion (RSVF) and glycoprotein (RSVG), differing in secondary structure type and complexity (Fig 4.1): 1) RSVF site IV, a linear, irregular beta strand (*3*); 2) RSVF site 0, a discontinuous epitope consisting of a kinked alpha helix and a disordered loop (*1*); 3) RSVF site II, a linear helix-turn-helix motif (*154*); and 4) RSVG 2D10 epitope, a linear helix-loop segment that is constrained by two disulfide bonds (*222*).

There are three major advantages to using the TopoBuilder for functional protein design, as shown in Fig 4.1: first, it enables a modular assembly of user-defined protein topologies with respect to the functional motif's structural requirements. As such, idealized secondary structure elements are positioned element by element to support the motif, enabling the building and systematic exploration of multiple topologies which accommodate the same binding motif. Leveraging this feature, we constructed three different topologies to present the site IV epitope: 3E2H (three-strand β-sheet plus two helices), 4E1H (four-strand β-sheet plus one helix) and 4E2H (four-strand β-sheet plus two helices). Second, it allows full control over the protein's connectivity, as shown by the design of two connectivities for a 3H1L topology (three helices with a cross-over loop) to host the discontinuous site 0 epitope. Third, we showcase the utility of the TopoBuilder to assemble a bi-functional protein and parametrically sample the placement of secondary structures and binding motifs. In one layer, the 4H topology presents the site II epitope, whereas the second helical layer contains the 2D10 epitope. To ensure optimal presentation of both functional motifs, we sampled two configurations (4H_01 - 4H_02), where the first layer is tilted by 20° compared to the second layer (Fig 4.1).

Each topology was initially defined as a two-dimensional form which was subsequently projected in the three-dimensional space, with respect to the coordinates of the functional motif extracted from its native context (prefusion RSVF or RSVG). Secondary structures were placed with a distance of 10 - 11 Å between the center of mass of alpha helices, and 4.5 Å between adjacent beta strands. From these idealized 3D sketches, distance constraints were derived to guide constrained *ab initio* folding coupled to sequence design using Rosetta FunFolDes (*26*) (Fig 4.1 and Fig S4.1).

Between 10,000 - 20,0000 designs were generated for each topology, and filtered for decoys that formed a stable and well-packed hydrophobic core, as well as for backbones with a funnel-shaped energy landscape in Rosetta *ab initio* simulations. Critical hydrophobic core positions based on the design simulations were selected, and encoded combinatorially for subsequent high-throughput yeast display screening (10^5 - 10^6 variants per topology).

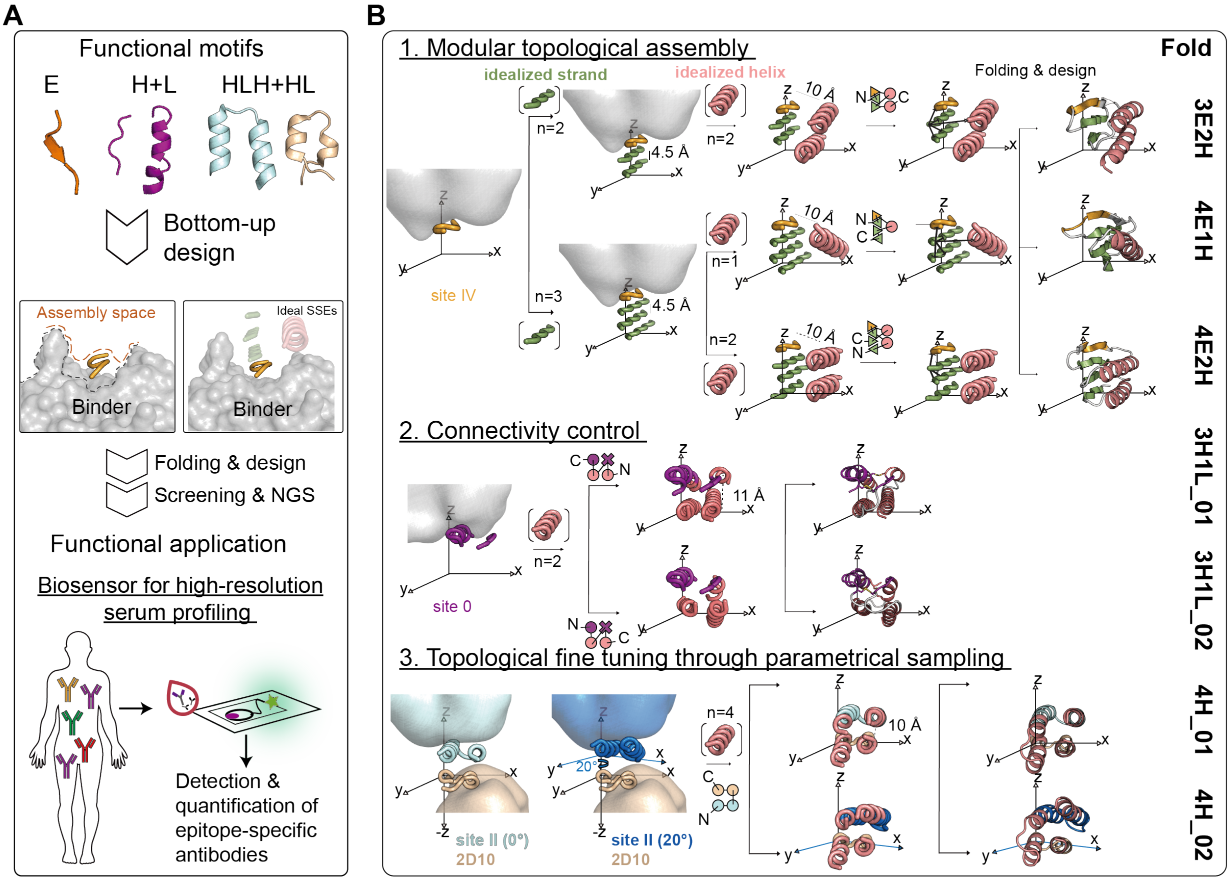

A
Functional motifs
E
H+L
HLH+HL
Bottom-up design
Assembly space
Binder
Ideal SSEs
Binder
Folding & design
Screening & NGS
Functional application
Biosensor for high-resolution serum profiling
Detection & quantification of epitope-specific antibodies
B
1. Modular topological assembly
idealized strand
idealized helix
site IV
n=2
n=3
4.5 Å
4.5 Å
n=2
n=1
n=2
10 Å
10 Å
10 Å
N
C
Folding & design
Fold
3E2H
4E1H
4E2H
2. Connectivity control
site 0
n=2
C
N
11 Å
N
C
3H1L_01
3H1L_02
3. Topological fine tuning through parametrical sampling
site II (0°)
2D10
site II (20°)
2D10
20°
n=4
C
N
10 Å
4H_01
4H_02
x
y
z
-z

Figure 4.1: A bottom-up *de novo* protein design strategy for the design of functional proteins.
(A) *De novo* proteins containing binding motifs of different secondary structure types and structural complexity were built using a bottom-up design approach. Idealized secondary structure elements were assembled around the binding motifs, respecting spatial constraints opposed by the binder. To identify lead candidates for functional characterization, designed topologies were screened using yeast display, followed by next-generation sequencing (NGS). As functional application, we showcase the utility of the designed proteins as biosensors for the detection of epitope-specific antibodies. **(B)** *In silico* building and design of multiple topologies for four different binding motifs. The TopoBuilder enables a modular topological assembly of virtually any topology that can be described in layers, allowing the building of both beta strands and alpha helices of defined length and spatial positioning. Leveraging this feature, three different folds were assembled to present the site IV epitope (3E2H, 4E1H, 4E2H). In addition, the connectivity of the assembled secondary structure elements is controllable, as illustrated by the design of two 3H1L topologies to stabilize site 0, differing in connectivity. Lastly, the TopoBuilder enables the design of bi-functional proteins (carrying both the site II and the 2D10 epitopes), and fine tuning of the spatial positioning of the motifs with respect to each other. E: beta strand, H: alpha helix, L: loop structure.

The libraries were sorted under two selective pressures – binding to their target antibody and resistance to treatment with the nonspecific protease chymotrypsin for the site IV motif, and binding to two different antibodies for the discontinuous site 0 and for the bi-functional 4H topology. Sorted populations were bulk-sequenced using next-generation sequencing (see chapter 3), revealing sequence profiles and structural determinants for stability and accurate motif presentation of the different design series (Fig S4.2 and S4.5).

4.2.2 Selected designs are well-folded and bind with high affinities to their target antibodies

For each topology, 10 - 20 designs that showed strong enrichment under two selection pressures in the yeast screening were selected for recombinant expression in *E.coli* and biophysical characterization.

Out of the three topologies designed to accommodate site IV, two design series (4E1H and 4E2H) yielded well-folded, stable proteins that bound the 101F antibody with affinities ranging from 22 to 200 nM. The best designs, 4E1H_95 and 4E2H_210, were thermostable up to 58 °C and 84 °C, respectively, and bound the 101F antibody with dissociation constants (K_D) of 22 nM and 180 nM (Fig 4.2). In contrast, the 3E2H topology mostly yielded insoluble proteins upon recombinant expression, underscoring the value of a general design method that allows the exploration of a wide range of topologies to increase chances of success.

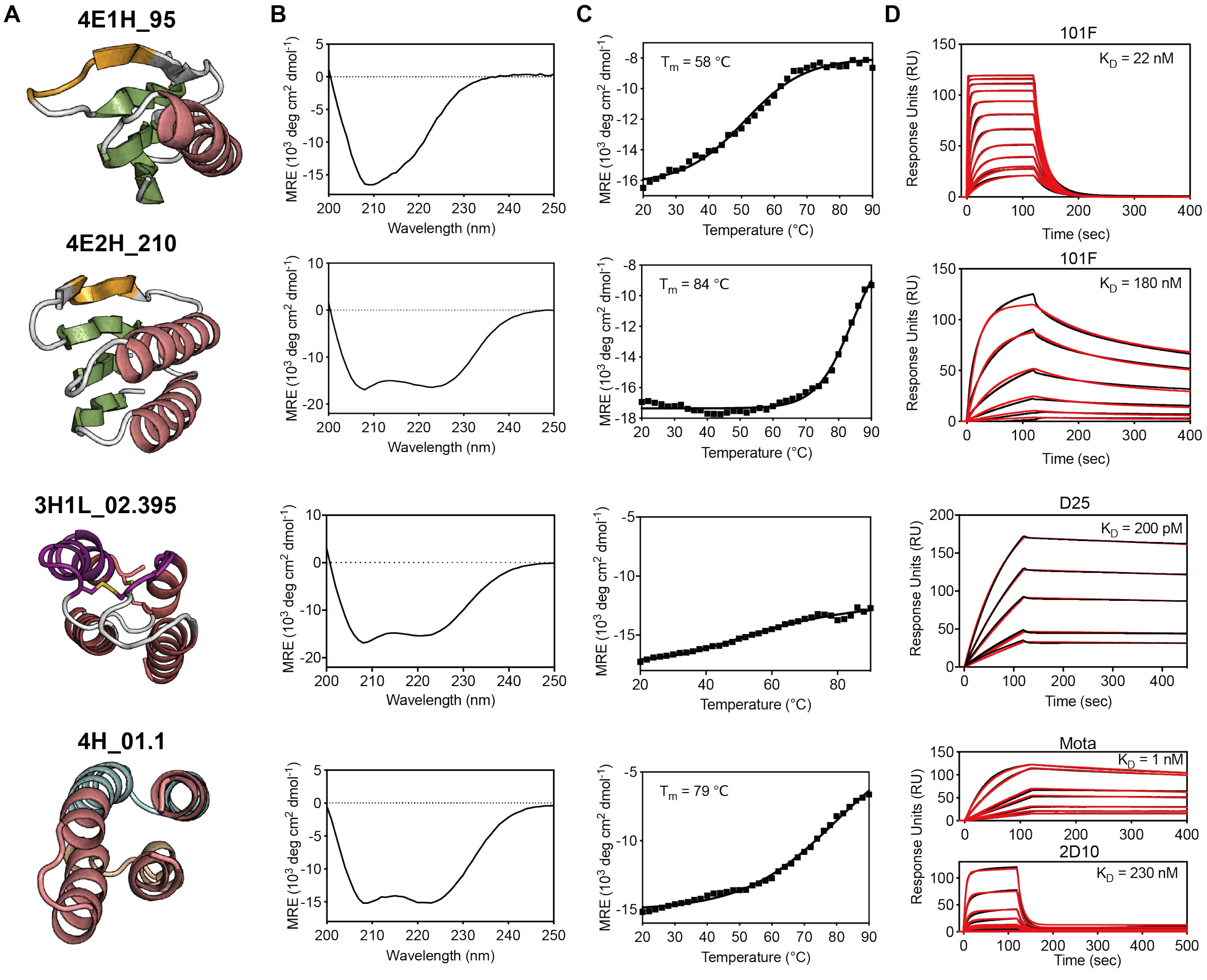


Figure 4.2: Biophysical characterization of lead variants from each topology.
(A) Computational models and names of lead variants. **(B)** CD spectra at 25 °C are in agreement with the secondary structure content designed. **(C)** Thermal melting curves as measured by CD at 208 nm. **(D)** Binding affinities to target antibodies determined by SPR. Raw data are shown in black, fitted curves (1:1 Langmuir kinetic fit) are shown in red. Affinities of 4E1H_95 and 4E2H_210 were measured against immobilized 101F IgG, 3H1L_02.395 against D25 IgG, and 4H_01.1 against both Motavizumab and 2D10 IgG.

With respect to the connectivity control in the TopoBuilder, both design series of the 3H1L topology, that presented the discontinuous site 0 epitope, were recombinantly expressed and successfully bound to their target antibody D25. However, all sequences tested deriving from the 3H1L_01 design series were dimeric in solution, and failed to bind to the 5C4 antibody (Fig S4.3). The 5C4 antibody was shown to engage site 0 from a different angle compared to D25 (*186*), and binding to both antibodies can thus serve as a probe for the integrity of the discontinuous site 0. In contrast, the best design of the 3H1L_02 series (3H1L_02.395) showed subnanomolar binding affinity to D25 (K_D = 200 pM) (Fig 4.2), and bound to 5C4 with an affinity of 25 nM (Fig S4.4). Both binding affinities closely match the reference binding affinity of the respective antibodies to prefusion RSVF, indicating that this discontinuous motif was mimicked accurately in a *de novo* designed protein (*1*). Also, 3H1L_02.395 was monomeric in solution, thermostable with only partial unfolding at 95 °C, and showed a well-dispersed HSQC NMR spectrum (Fig 4.2 and Fig S4.4).

High-throughput screening and next generation sequencing of the two 4H topological variants (4H_01 and 4H_02) revealed that sequences with strong enrichment for both Motavizumab and 2D10 binding were exclusively derived from the 4H_01 design series (Fig S4.5). We expressed 20 sequence variants from the 4H_01 design series and successfully biochemically characterized 14 of these. The designs showed CD spectra typical of alpha-helical proteins and bound to both target antibodies (Motavizumab and 2D10) with a binding affinity ranging from K_D = 1 - 10 nM to Motavizumab and K_D = 230 - 400 nM to 2D10. The best variant, 4H_01.1, was well-folded and thermostable according to CD with a T_m of 79 °C, and bound with K_D = 1 nM to Motavizumab and K_D = 230 nM to 2D10 (Fig 4.2).

4.2.3 Crystal structure of 4E1H_95 confirms accuracy of the design model

We next solved a crystal structure of 4E1H_95 in complex with the target antibody 101F at 2.9 Å resolution. The structure is in close agreement with the computational design model, with a full-atom RMSD of 1.8 Å (Fig 4.3a). The backbone hydrogen bonding network was accurately recapitulated in the structure compared to the design model, and the RSV site IV epitope was mimicked with sub-angstrom accuracy (Fig 4.3b).

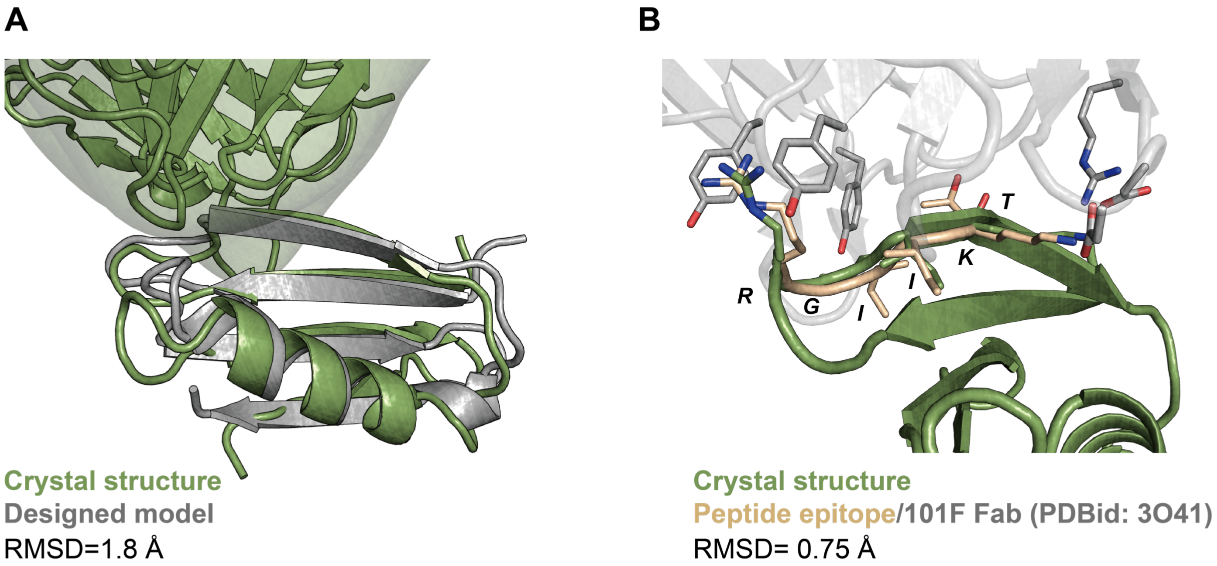

A
B
R
G
I
I
K
T
Crystal structure
Designed model
RMSD=1.8 Å
Crystal structure
Peptide epitope/101F Fab (PDBid: 3O41)
RMSD= 0.75 Å

Figure 4.3: Structural characterization of 4E1H_95.
(A) A crystal structure of 4E1H_95 bound to 101F Fab (green surface) was solved with a 2.9 Å resolution, closely matching the design model (grey, RMSD 1.8 Å). **(B)** Superposition of 4E1H_95 crystal structure (green) with the site IV peptide epitope (lightorange, PDB 3O41), showing accurate mimicry with an RSMD of 0.75 Å.

4.2.4 Detecting and quantifying RSV epitope-specific antibodies using a biosensor based on *de novo* proteins

Serological assays that allow for the detection of antibodies are widely used to diagnose disease states, including autoimmune diseases, allergies and infectious diseases (*223, 224*). In addition, serological readouts frequently serve as a measure for vaccination success, either inferred from the level of bulk serum titers or, in more laborious assays, by measuring pathogen neutralization (*127*). Commonly, these assays measure bulk serum antibody responses and do not allow the quantification of antibody responses with single epitope resolution. However, with our increasing structural understanding of epitopes targeted by protective versus non-protective antibodies, assays that enable the rapid dissection of serum responses on the epitope level would be highly informative.

To address this need, we sought to develop an antibody biosensor for the detection of epitope-specific antibody responses based on the *de novo* designed proteins, which stabilize major RSV antigenic sites in their native conformation. We designed a biosensor based on the recently developed Lumabs platform, as shown in Fig. 4.4a (*225*). In its closed conformation, bioluminescence resonance energy transfer (BRET) occurs between the nanoluciferase donor and the mNeonGreen acceptor. In presence of antibodies specific for the presented epitope, the sensor adopts an open state, as measured by a decrease in BRET ratio (Fig 4.4b).

As a proof-of-concept, we have successfully expressed and characterized one sensor for site IV and one for site 0. The site IV sensor is based on the 4E2H_210 design, and the site 0 sensor is based on the S0_2.126 scaffold presented in chapter 3. Ongoing efforts are directed towards characterizing sensors for the 3H1L design series, as well as the 4E1H_95 design.

Sensors were robustly expressed in *E.coli*, and showed a concentration-dependent decrease in BRET signal upon adding either 101F or D25 antibody, respectively (Fig 4.4c,d). Sensors were not responsive to an irrelevant IgG (Fig 4.4e,f).

The sensitivity of the different sensors was directly linked to the affinity of the respective scaffold to its monoclonal antibody. D25-like responses were detected at subnanomolar concentration (< 0.1 µg/ml), whereas 101F-like responses were detected above 1 µg/ml. This highlights the value of designing a large pool of scaffolds with different affinities, allowing the tailoring of the scaffold affinity to the desired detection range. It is noteworthy that a Lumabs sensor containing the site IV peptide epitope showed an approximately 500-fold lower sensitivity, which in most cases would be insufficient to detect physiologically relevant concentrations of epitope-specific antibodies.

Future efforts will be directed towards characterizing the sensors in serum samples. Ultimately, we aim to quantify epitope-specific responses in human sera from individuals that were infected with RSV, as well as from animals immunized with pre- and postfusion RSVF. Such high-resolution profiling may allow, for the first time, a detailed picture of epitope-specific antibodies elicited by viral infection and different vaccination strategies, and thus may become a versatile tool for immune monitoring.

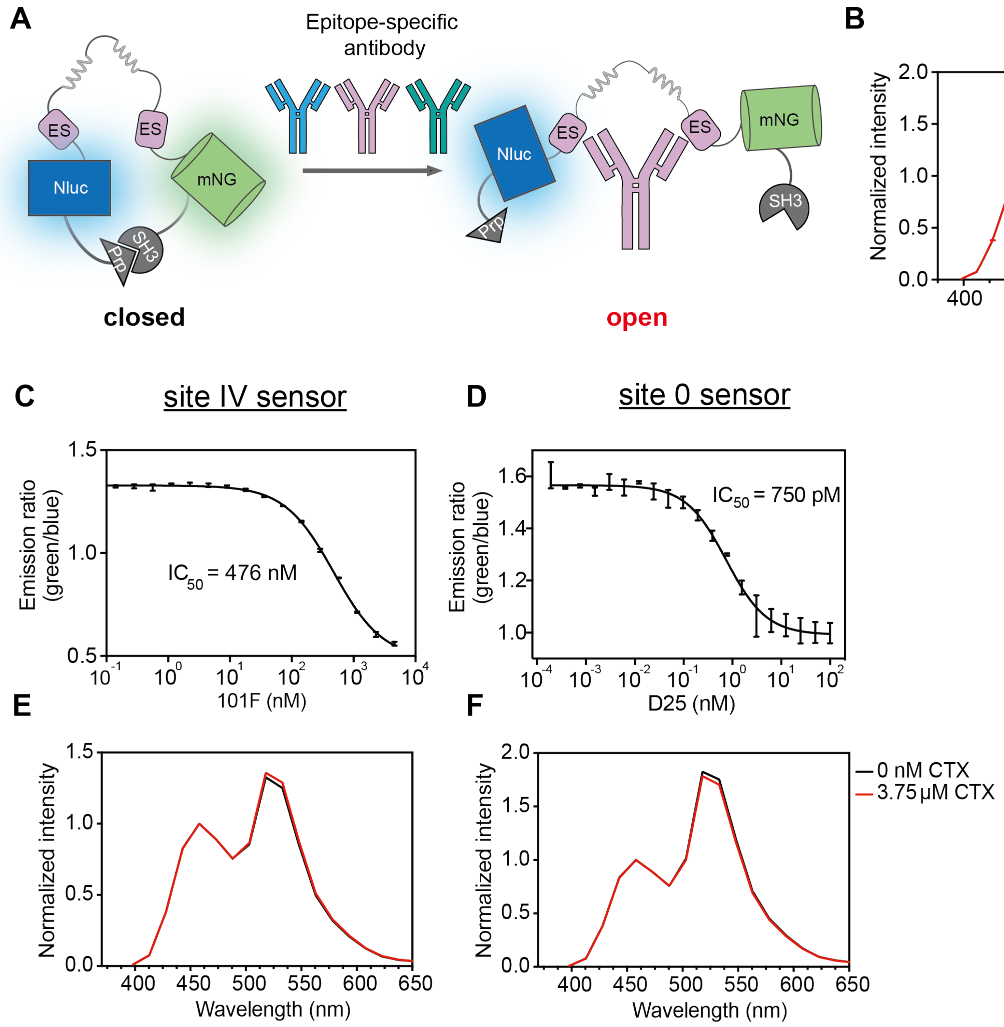


Figure 4.4: Biosensor for the detection of antibody responses based on *de novo* designed proteins.
(A) Schematic representation of the Lumabs sensor. NanoLuc luciferase (Nluc) and mNeonGreen (mNG) are held in close proximity by two helper domains (SH3 domain and a proline-rich-peptide, Prp), thus allowing efficient BRET. The fluorescence signals were normalized to the emission intensity of Nluc. Antibodies binding to the epitope-scaffold (ES) disrupt the weak Prp-SH3 interaction, opening the sensor and decreasing BRET efficiency. **(B)** Luminescence spectrum of the sensor in a closed (black) and open (red) conformation. **(C)** Site IV sensor characterization based on 4E2H_210 scaffold. The ratio between 518 and 458 nm emission upon titration of different 101F antibody concentrations is plotted. The IC_{50} of the 4E2H_210_Lumabs was determined as IC_{50} = 476 nM, which is in close agreement with the binding affinity of 4E2H_210 to 101F (180nM). **(D)** Site 0 sensor characterization based on S0_2.126 scaffold from chapter 3 upon titration of D25 IgG. The IC_{50} was determined as 750 pM, in agreement with the high S0_2.126 affinity for D25 (~50pM), allowing detection of site 0 antibodies at less than 0.1 µg/ml. **(E,F)** Both the site IV and site 0 sensors were irresponsive to a micromolar concentration of an irrelevant IgG (Cetuximab, CTX). Data points are plotted as mean ± SEM.

4.3 Discussion

Here, we have presented a general strategy for the ‘bottom-up’ design of diverse protein topologies endowed with binding motifs and showcased a relevant functional application to monitor RSV epitope-specific antibody responses in serum.

From a protein design perspective, the presented ‘motif-centric’ approach is of general utility for the *de novo* design of functional proteins. In contrast to previously employed ‘top-down’ approaches, the TopoBuilder accounts for the structural and sequence constraints of functional motifs irrespective of their structural complexity, and allows the building of customized scaffolds for their accurate stabilization.

A common problem for *de novo* functional protein design efforts is to select a protein fold that fulfills both the biophysical and functional requirements. The best topology for functional purposes may be difficult to predict *in silico*, hence requiring experimental testing of a diverse set of topologies. To address this problem, the TopoBuilder allows one to systematically build and fine-tune multiple protein topologies that are compatible with a given functional motif. While previous reports have shown a conceptually similar ‘bottom-up’ building and design of functional proteins, only one topology per functional site was reported, and topologies were structurally limited to helical bundles (*60-62, 217, 226*).

Thus, the design strategy presented here has significant advantages with regard to its general utility for *de novo* functional design, permitting the systematic exploration of a wide variety of possible topologies tailored to a given functional motif, control over their connectivity, and fine tuning of the secondary structure arrangement through parametrical sampling. To validate the methodology, we have attempted to design six different topologies, four of which have yielded well-folded, stable proteins. Structurally, the functional sites were accurately presented, as shown by high affinity binding to their target antibodies. Importantly, a crystal structure of one of the designs has confirmed the atomic level accuracy of the design model.

With respect to the functional aspect of this work, one of the most attractive domains for *de novo* design is to create proteins with functional and biophysical properties that are out of reach of natural proteins. Here, the designed RSV epitope scaffolds proved functional as diagnostic tools to detect epitope-specific antibodies when embedded in a biosensor, which represents an important and yet unmet need to profile immune responses and further inform vaccine design. With our rapidly growing understanding of protective anti-viral antibody responses, monitoring levels of epitope-specific antibody responses on a serum level is of utmost importance to assess the efficacy of novel vaccine candidates beyond bulk serum titers and *in vitro* virus neutralization assays.

The current state-of-the art to detect key antibody specificities is to quantify palivizumab-competing antibodies (PCA) in an ELISA-based competition assay, which is routinely included as a measure for vaccine-elicited immunity in RSV human clinical trials (*227, 228*). However, such competition assays are laborious due to many liquid handling steps and only yield a vague picture of the antibody specificities. One of the major confounding factors of competition assays, such as the PCA assay, is that they measure any palivizumab competing antibody, among them non-neutralizing site I antibodies, and are thus unable to exclusively quantify antibodies with palivizumab-like epitope-specificities (*229*).

To overcome these limitations, we present a proof-of-principle for *de novo* proteins as versatile tools for immune monitoring. The sensors enable the detection and reliable quantification of epitope-specific antibodies targeting site 0 and IV in an easy, fast and scalable assay, requiring only a very limited set of equipment and reagents. To the best of our knowledge, there is no alternative strategy to accomplish this task using naturally occurring proteins or modifications of thereof, underscoring the potential for functional *de novo* protein design to address challenges that would otherwise be intractable. Future developments will include the design of sensors for additional neutralization epitopes of RSV and their integration in a paper-based format, potentially allowing the dissection of serum specificities with an unprecedented level of resolution from a single drop of blood (*225, 230*). Beyond RSV, the development of similar sensors for conserved influenza epitopes may enable to predict protection against antigenically drifted strains, even in the absence of measurable neutralizing activity (*73*). Also, sensors for epitopes conserved among different dengue subtypes may be useful to distinguish protective from disease-enhancing antibodies, which poses a major concern for the development of a safe and effective dengue vaccine (*131, 170*).

Lastly, we show the utility of the TopoBuilder for the *de novo* design of proteins accommodating multiple binding sites. Carrying more than one binding site is a common scenario in many natural proteins that mediate important cellular functions, for example signal transducing adaptor proteins and ligands that mediate receptor heterodimerization (*9*). Thus, a tantalizing future route of exploration will be the design of novel receptor ligands to control cellular behavior in synthetic biology applications. Given that the precise spatial configuration of ligand-induced receptor dimerization is a critical determinant for efficient signaling (*43*), the TopoBuilder should prove as a versatile tool in the design of topologically controlled ligands to tune cellular behavior.

Altogether, the presented motif-centric design strategy represents a step forward for the design of functional proteins. While the functions herein were limited to antibody binding motifs, the approach should be widely applicable to any known functional motif, and, ultimately, the ability to precisely control topological and physicochemical features may allow the design of novel functions absent in natural proteins.

4.4 Methods

The experimental and computational methods used in chapter 4 are largely identical to those in chapter 3. Additional methods and variations of methods explained in chapter 3 are listed below, for all other methods refer to chapter 3.

4.4.1 *De novo* protein design

Template building using TopoBuilder

The initial step of TopoBuilder is to create a 2D representation of secondary structure placements (forms) around a given functional motif. The form is built layer-by-layer, thereby allowing a modular assembly and simple definition of a user-defined protein fold. The topologies built were selected to yield small, globular proteins with a high contact order between secondary structures and the functional motif, and a user-defined connectivity between the secondary structure elements.

The 2D forms are subsequently expanded into idealized 3D structures (C-alpha only topologies, referred to as sketches). For each sketch, secondary structure type, length, and spatial positioning in relation to the fixed structural motif are defined by the user (Fig S4.1 and Table S4.1).

A shift is referred to as a translation along the x,y or z axes, whereas a tilt is referred to as a rotation along the x,y,z axes (Fig S4.1).

Finally, distance and atom-pair constraints are derived from this sketch to guide folding trajectories using Rosetta FunFolDes (*26*), followed by iterative design-relax cycles to generate large ensembles of structures and sequences for the target topologies.

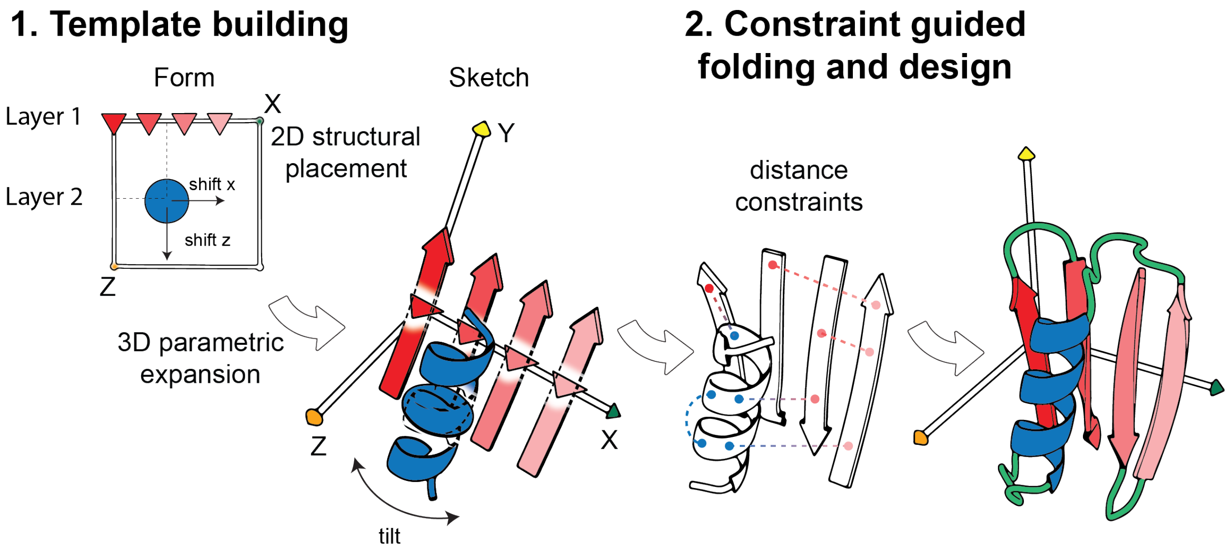


Figure S 4.1: TopoBuilder workflow.
In a first step, a 2D form is defined, followed by a 3D parametric expansion to obtain a sketch. Parameters that define the secondary structure type, length and spatial positioning are given in Table S4.1. Constraints are derived from the sketch to guide the folding and design protocol, which includes building the connecting loops.

Layer	SS-type	Length (# residues)	Shift [x,y,z] Å	Tilt [x,y,z] °
3H1L_01				
1	H	20	0,(-6),0	0,0,0
1	H	20	11,(-6),0	0,0,0
2	Site 0 epitope			
3H1L_02				
1	H	16	3,2,1.5	0,0,(-30)
1	H	20	12,0,2	(-20),0,(-20)
2	Site 0 epitope			
3E2H				
1	E	7	-1,-2,13	0,90,0
1	H	16	9,-2,10	-5,0,0
1	E	7	-1,-2, 6.8	0,-95,0
2	Site IV epitope			
2	H	15	11.5, 0, -3	-5,0,0
4E1H				
1	E	8	4.5,0,20.8	0,-75,2.5
2	E	8	2.7, 0,13.8	0,-92,0.5
2	H	17	12, 0, 17.6	-15, 35,0
3	E	8	1.8,-1,6.8	0,-85,3
4	Site IV epitope			
4E2H				
1	E	7	2,(-2),20	0,100,0
1	E	7	1.3,(-2),13.5	0,90,0
1	H	15	11.5,(-2), 10.5	0,45,0
1	E	7	1,(-2), 6.5	0,-90,0
2	Site IV epitope			
2	H	15	12, 0, -2.5	-5, -180, 0

4H_01				
1	Site II epitope		3, 0,12	0, 20,0
2	2D10 epitope			
4H_02				
1	Site II epitope		3, 0,12	0,0,0
2	2D10 epitope			

Table S 4.1: Fold definition and spatial positioning of each SSE in the designed topologies.

Sequence design, selection and high-throughput screening

Between 10,000 and 20,000 designs were generated using Rosetta FunFolDes. Decoys were filtered according to multiple parameters, such as core packing, RMSD drift of the epitope after relaxing, and sequence-structure compatibility. A subset of sequences was evaluated in Rosetta *ab initio* simulations for their tendency to recover the designed fold. Chapter 3 describes the sequence design and selection protocol in further detail.

We evaluated sequence profiles for the designs and encoded critical hydrophobic core positions in a combinatorial yeast library. Sequences were screened for binding to their target antibodies and for resistance to various concentrations of the non-specific protease chymotrypsin. The top 1% of clones was sorted and sequenced using next-generation sequencing, as described in chapter 3. For each sequence, an enrichment score was computed based on its frequency under high-stringency selection versus low-stringency selection. Between 10 and 20 sequences that were strongly enriched for two different selection pressures, such as high-affinity binding to two different antibodies or, alternatively, high-affinity binding to one antibody and chymotrypsin resistance, were selected for recombinant expression and biophysical characterization.

4.4.2 Biosensors

Lumabs expression and purification

Genes encoding Lumabs were synthesized by Genscript and cloned into a pET28 *E.coli* expression vector. For expression, constructs were transformed in *E.coli* BL21 (DE3) cells, grown until the OD reached 0.6, and induced with 1mM isopropyl β-d-1-thiogalactopyranoside (IPTG). After overnight expression at 20 °C, cells were harvested, lysed and purified using NiNTA affinity chromatography, followed by size exclusion chromatography on a Superdex200 increase column (GE Healthcare) using PBS as running buffer. Sensors were stored frozen at −80 °C until use.

Sensor characterization

The titrations with antibody were performed in a total reaction volume of 100 µL, in white polystyrene 96 wells plates (Thermo Fisher Scientific-Nunc F96 MicroWell, 136101), on a Tecan Spark 10M. A sensor concentration of 50 pM was used in phosphate buffered saline with 1 mg/mL bovine serum albumin, at pH = 7.4. After an incubation step of at least 40 minutes, NanoGlo assay substrate (Promega, N1110) was added (2000x diluted). Emission was subsequently recorded using a luminescence scan between 398 nm and 653 nm, with an integration time of 300 ms.

4.5 Supplementary Information

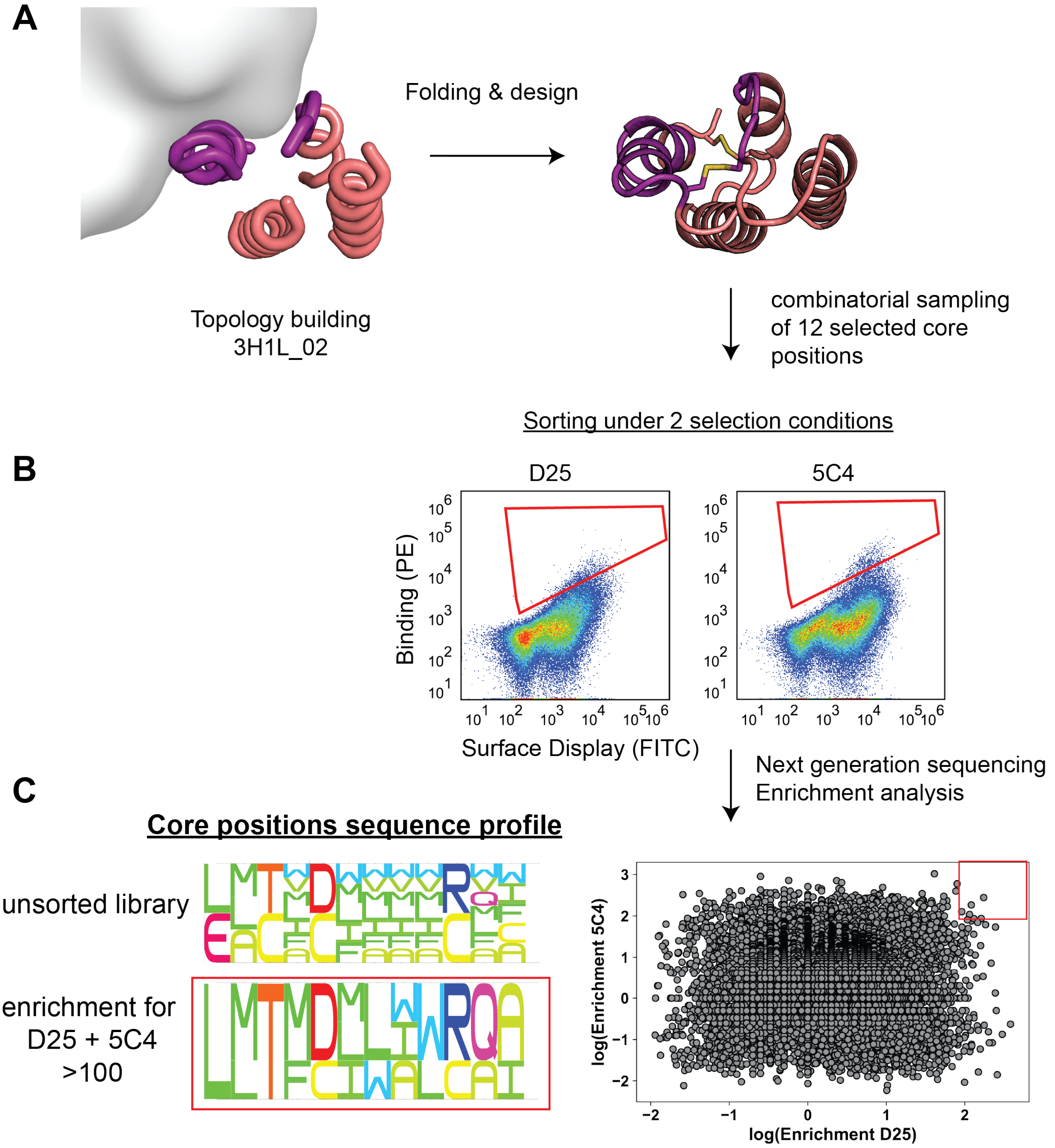


Figure S 4.2: Design and high-throughput screening for 3H1L_02 topology.
(A) The topology was assembled using the TopoBuilder, followed by folding and design using Rosetta FunFolDes. The site 0 epitope is shown in purple; the secondary structure elements that were built by the TopoBuilder are colored salmon. From the ensemble of the designed sequences, 12 critical core positions were selected and encoded in a combinatorial sequence library (theoretical size ~10^7). **(B)** The library was transformed into yeast, and clones with high-affinity binding to D25 and 5C4 were sorted (red sorting gate). **(C)** Right: the sorted populations were sequenced using next-generation sequencing, and enrichment scores were computed for each sequence binding to 5C4 and D25. Sequences that were strongly enriched for D25 and 5C4 (top right corner, red) were selected for recombinant expression and biophysical characterization. Left: Sequence profile of the 12 selected core positions in the unsorted library, and of sequences that were >100 fold enriched for D25 and 5C4 binding.

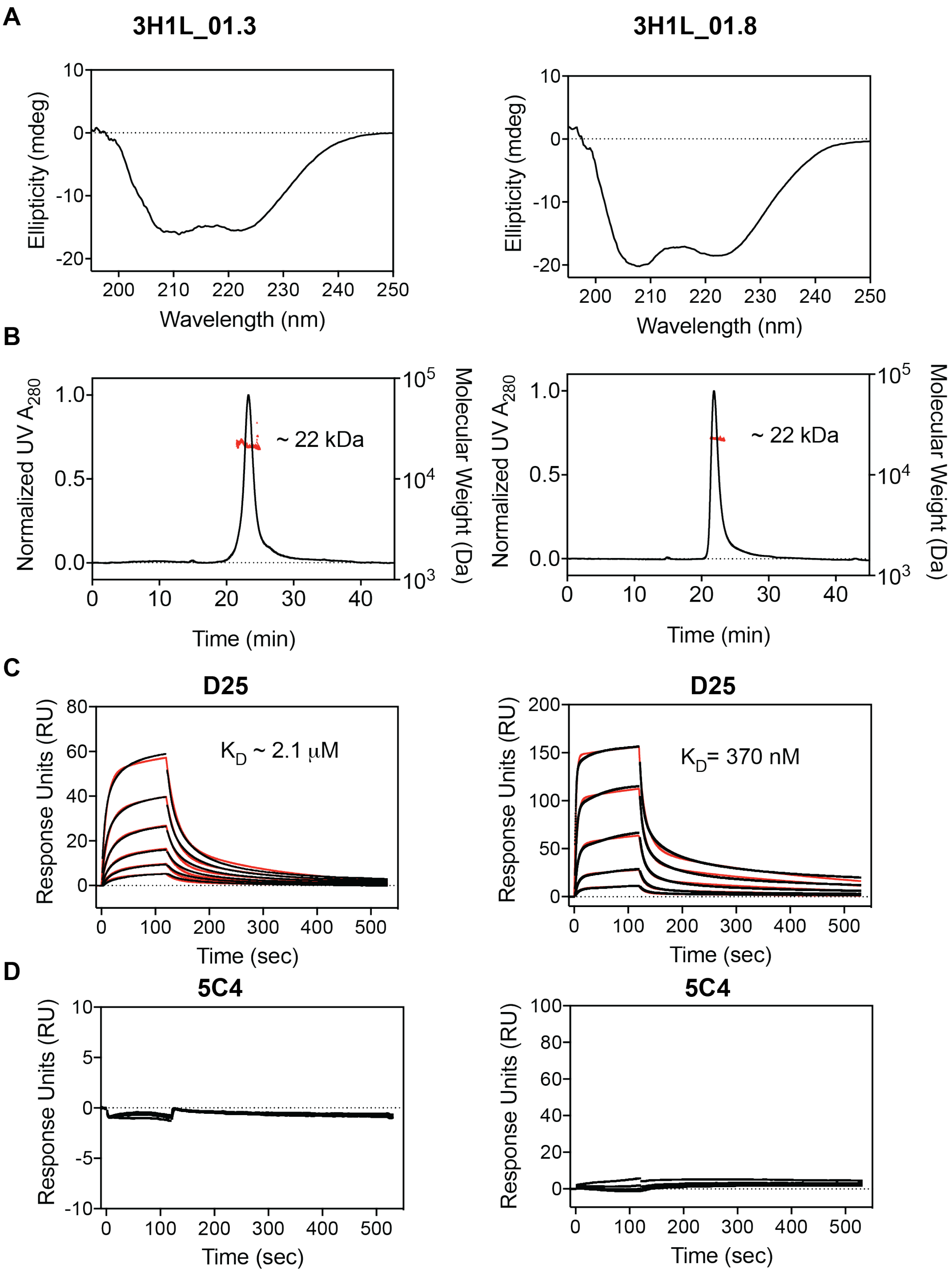


Figure S 4.3: Biophysical characterization of best 3H1L_01 variants.
Following yeast display screening, 10 sequences were selected for recombinant expression and biophysical characterization. Data are shown for the two best sequences (3H1L_01.3 and 3H1L_01.8) according to their D25 binding affinity. **(A)** Circular dichroism spectra at 25 °C are in agreement with a helical protein. **(B)** Size-exclusion chromatography with on-line multi-angle light scattering (SEC-MALS) shows that both designs are dimeric in solution. The expected molecular weight of monomeric 3H1L_01 designs is ~11kDa. **(C,D)** Measurement of binding affinity to D25 and 5C4 by SPR. 3H1L_01 designs were immobilized on the sensor chip surface, and D25 or 5C4 Fab were injected as analyte in various concentrations. Kinetic dissociation constants based on fitting with a 1:1 Langmuir model are indicated, with no detectable binding to 5C4.

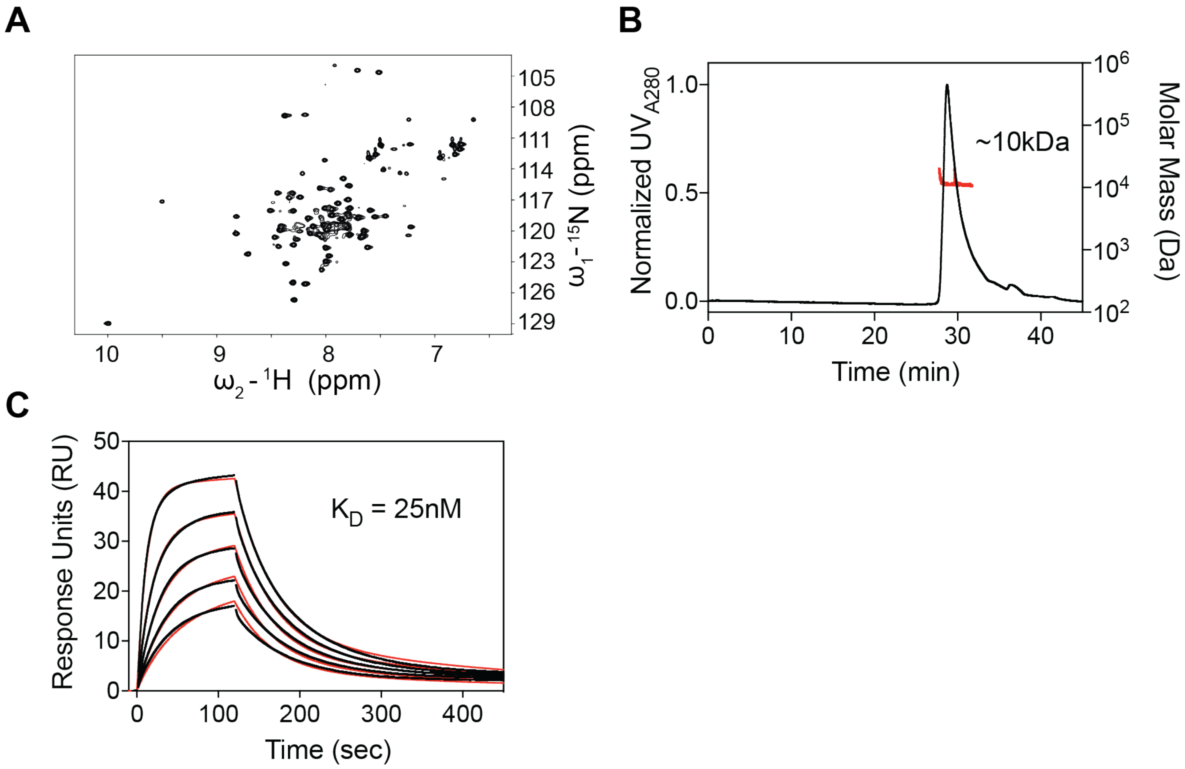


Figure S 4.4: Extended biophysical characterization of 3H1L_02.395.

(A) The 2D NMR ^{15}N HSQC spectrum for 3H1L_02.395 is well dispersed, confirming that it is well-folded in solution. **(B)** 3H1L_02.395 is monomeric in solution, with the determined molecular weight of 10 kDa closely matching its theoretical molecular weight of 10.5 kDa. **(C)** 3H1L_02.395 binds with a K_D of 25 nM to 5C4, as determined by SPR.

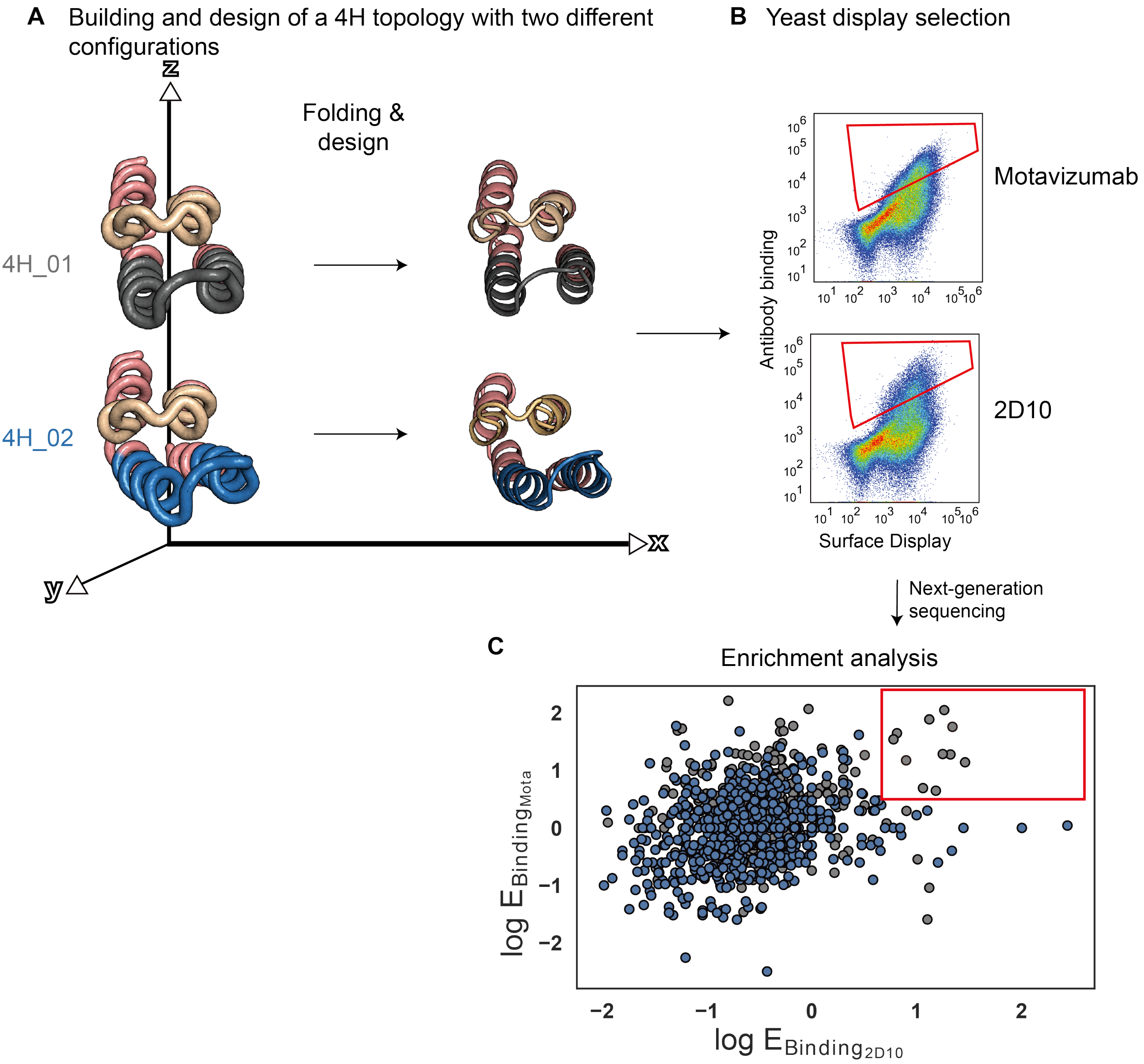


Figure S 4.5: Topological tuning of the 4H topology for optimal presentation of both binding motifs.
(A) Two different configurations of the 4H topology were built. For 4H_02, the helical layer containing the site II epitope was tilted by 20° relative to the helical layer presenting the 2D10 binding motif. **(B)** Designs from both templates were screened in a single yeast library, and clones with high affinity binding to Motavizumab or 2D10 were sorted and sequenced. **(C)** Enrichment analysis of the sorted populations; each dot represents a unique sequence. Sequences from the 4H_01 design series are colored grey; 4H_02-based designs are colored blue. For each sequence, an enrichment score (E) was computed based on its frequency under high selective pressure versus low selective pressure (see chapter 3 methods for details). The enrichment was computed for binding to Motavizumab and to 2D10. Sequences that are strongly enriched for binding to both antibodies (red square) uniquely derive from the 4H_01 design series (indicated in grey).

Chapter 5 Conclusions & Perspectives

The work presented in my thesis details significant progress within two domains of molecular bioengineering: (I) the structure-based immunogen design field; and (II) the *de novo* design of functional proteins. Since October 2015, when my thesis work was started, both fields have progressed at an enormous pace, and the underlying work has contributed to these scientific advances in multiple ways. With regard to the overall scientific progress, the following conclusions and future implications arise from this work:

5.1 Structure-based vaccine design

Throughout the last few years, high-throughput antibody isolation campaigns have yielded a plethora of potent nAbs for RSV and other pathogens. As an illustrative example, there were approximately a dozen RSV nAbs known and characterized before 2014, whereas between 2015 and 2019 more than 800 new human RSV nAbs were reported (*2, 125*). This massive increase in the number of antibodies has greatly improved our understanding, at the molecular level, of what constitutes an effective RSV antibody response. Crystal structures of a subset of these have unveiled the atomic details of antibody-mediated RSV neutralization, providing templates for rational immunogen design (*123*). Despite this rapid progress, a general challenge for the reverse vaccinology approach remains in translating this structural knowledge into immunogens that effectively elicit targeted nAbs *in vivo*.

An important breakthrough in the RSV field was the development of prefusion RSVF in 2013, which induces potent RSV nAbs in naïve subjects, and its antigenic properties were further improved in 2016 (*82, 84*). Nevertheless, it has remained an open question whether antigens other than those based on whole viral proteins could emerge as an alternative approach to induce protective immunity; and in particular, induce a more focused antibody response towards key neutralization epitopes. Building upon the proof-of-principle work by Correia and colleagues from 2014, this work has investigated strategies to design epitope-focused immunogens for structurally challenging epitopes, and reveal their potential as antigens in different immunological scenarios.

Conclusion 1: Synthetic immunogens boost subdominant neutralizing antibodies under conditions of pre-existing immunity.

My thesis work has established that epitope-focused immunogens can efficiently boost subdominant populations of nAbs under conditions of pre-existing immunity, which represents a highly relevant scenario for vaccine development against RSV, influenza and dengue. Chapter 2 described how the use of a site II epitope-focused immunogen results in steering serum responses of broad specificity towards a site II-focused response. Strikingly, this led to a marked increase of site II-mediated serum neutralization, indicating that boosted antibodies were functional and relevant for viral neutralization. This ability was unique to the synthetic immunogen with its defined epitope specificity, as the whole viral antigen has been unable to focus antibody responses on site II (Fig 5.1).

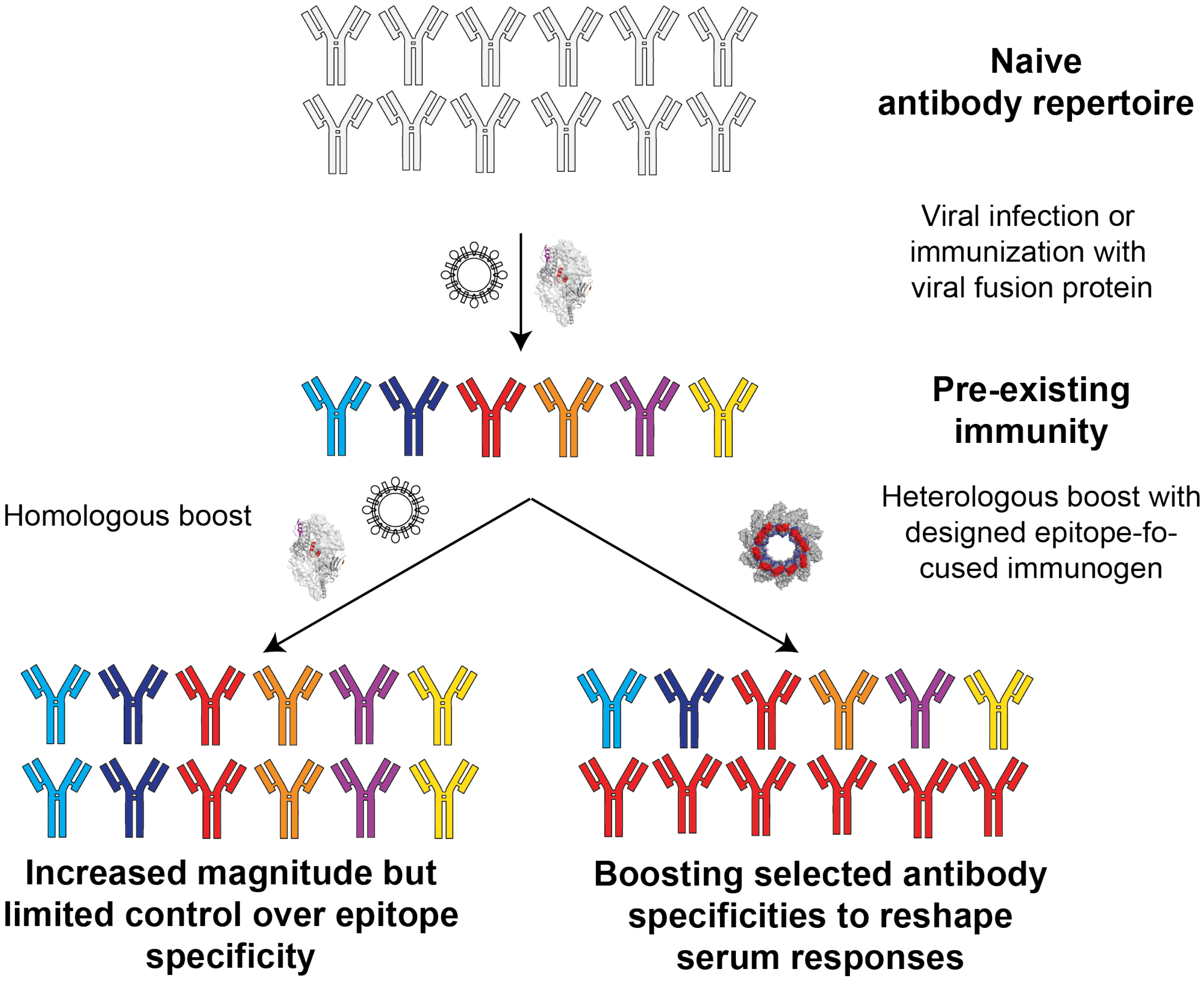

Naive
antibody repertoire
Viral infection or
immunization with
viral fusion protein
Pre-existing
immunity
Homologous boost
Heterologous boost with
designed epitope-focused immunogen
Increased magnitude but
limited control over epitope
specificity
Boosting selected antibody
specificities to reshape
serum responses

Figure 5.1: Synthetic immunogens as boosting immunogens under conditions of pre-existing immunity.
As shown throughout chapter 2 and 3, scaffold-based immunogens bear unique potential to reshape serum responses of broad specificity towards increased levels of site-specific antibodies (red), resulting in increased viral neutralization through boosted antibodies. In contrast, boosting immunizations with a viral fusion protein were unable to focus the response on the respective epitope.

Thus, an important conclusion arising from this work is that such *'precision immunogens'* are effective to steer pre-existing antibody responses towards a selected epitope *in vivo*, highlighting an important application for this class of immunogens that had not been uncovered previously. In chapter 3, we confirmed this potential application in NHPs. Boosting immunizations of non-naïve NHPs with a cocktail formulation of epitope-focused immunogens led to a significant increase of site II and site 0 antibodies, resulting in a drastic dependency of the serum neutralization on the boosted epitope specificities.

These findings are of broad interest for the vaccine development community. In particular, one of the major challenges that has hindered the development of a universal influenza vaccine is the immunodominance of antibodies targeting strain-specific epitopes over those targeting broadly protective, conserved epitopes (*69*). Despite repeated natural infections and seasonal vaccination, serum antibodies that target conserved stem epitopes are transient and rarely reach a functional threshold (*140*). Thus, the ability to effectively increase the fraction of functional, pre-existing antibodies specific for broadly protective (stem) epitopes would be a significant step forward towards the development of a universal influenza vaccine (*231*).

Given that *de novo* proteins have not evolved under the pressure of the immune system and do not carry the immunodominance signatures of whole viral antigens, scaffold-based antigens present an alternative approach that may emerge as a viable strategy for influenza vaccine development. Future efforts will be devoted to the development of scaffold-based immunogens for influenza, and immunization studies will reveal whether antibodies that target cross-neutralizing epitopes can be boosted beyond their functional threshold.

Conclusion 2: Cocktail formulations of *de novo* designed immunogens consistently elicit neutralizing antibodies *in vivo*.

My thesis work presents, for the first time, the development and *in vivo* testing of a cocktail formulation of epitope-focused immunogens. Results presented throughout chapter 3 prove the ability of a designed trivalent cocktail vaccine to elicit antibody responses targeting three defined, non-overlapping epitopes. In contrast to a previous study with a single epitope-focused immunogen (*59*), the Trivax1 cocktail consistently elicits neutralizing serum levels above the protective threshold in NHPs, providing a starting point for further optimization to increase the magnitude and durability of the response (Fig 5.2).

To optimize the designed antigens, we have developed a novel computational protein design algorithm (‘TopoBuilder’), yielding second generation antigens (‘Trivax2’) with greatly improved epitope mimicry. The novelty of the computational design procedure and its utility in designing epitope mimetics stems from the fact that it is generally applicable to motifs of high structural complexity, and allows the encoding of global structural similarity features in addition to accurate local epitope mimicry. While a head-to-head comparison in an NHP immunization study is still missing, the fact that Trivax2 elicited neutralizing serum levels in mice underscores the potential of this new design strategy as a powerful means for the design of epitope-focused immunogen cocktails for pathogens beyond RSV.

Nonetheless, in comparison to prefusion RSVF, the neutralizing serum levels induced by Trivax remain inferior by at least one order of magnitude. Further optimizations of antigenic presentation, delivery and formulation would be needed and could potentially improve the magnitude and durability of the antibody response. From a translational perspective, however, epitope-scaffold based immunogens are still far from becoming a true competitor of prefusion RSVF for vaccination of naïve subjects. Nevertheless, the potential of epitope-scaffold immunogens to dial selected antibody specificities is unique, and is out of reach of immunogens based on whole viral proteins.

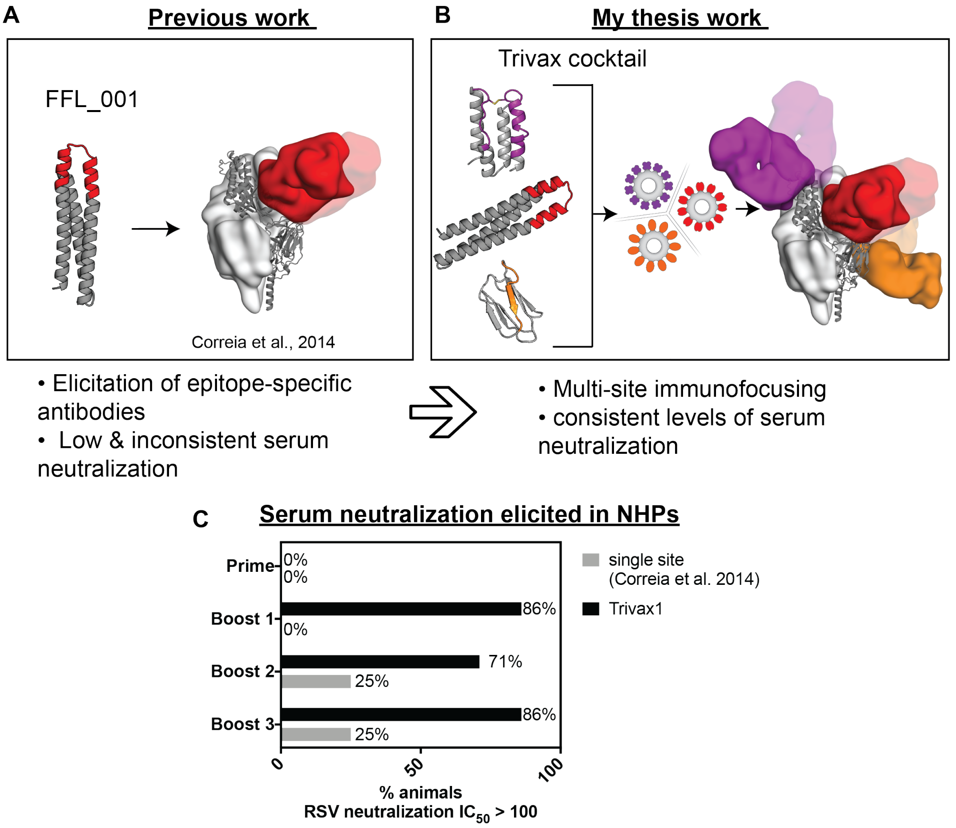

A
Previous work
FFL_001
Correia et al., 2014
• Elicitation of epitope-specific antibodies
• Low & inconsistent serum neutralization
B
My thesis work
Trivax cocktail
• Multi-site immunofocusing
• consistent levels of serum neutralization
C
Serum neutralization elicited in NHPs
Prime
Boost 1
Boost 2
Boost 3
0%
0%
86%
0%
71%
25%
86%
25%
0
50
100
% animals
RSV neutralization IC50 > 100
single site (Correia et al. 2014)
Trivax1

Figure 5.2: Single- versus multi-site immunofocusing using synthetic immunogens.
(A) In a previous proof-of-principle study by Correia and colleagues, it has been shown that a synthetic immunogen can elicit targeted nAbs in naïve NHPs, but neutralizing serum levels were low and inconsistent. **(B)** In my thesis work, the development of a trivalent cocktail has allowed to focus nAbs on three defined epitopes, leading to consistent neutralizing serum responses. **(C)** Number of animals with detectable neutralizing serum titers (IC_{50} > 100) using a single site (Correia et al. 2014) versus Trivax1 in NHPs.

Projecting beyond RSV, priming selected antibody specificities in naïve subjects may prove useful for pathogens that display only a limited number of conserved neutralization epitopes (e.g. HIV-1), and for which alternative approaches have proven ineffective. The most promising application of epitope-focused immunogens in HIV-1 vaccination protocols would likely be as priming immunogen, followed by boosting immunizations with more native-like envelope proteins to guide affinity maturation towards bnAbs. The feasibility of this approach is underlined by a recent study that primed epitope-focused responses using the HIV-1 fusion peptide, the target of several bnAbs, conjugated to an immunogenic carrier (*165*). Subsequent boosting immunizations with prefusion-stabilized envelope trimers yielded serum nAbs with notable breadth, whereas a homologous prime-boost scheme with envelope trimers only did not induce heterologous neutralizing responses. Given that a single antibody specificity is unlikely to mediate long-term protection for a highly variable virus like HIV-1 (*232*), the proven ability to prime epitope-focused responses targeting multiple, non-overlapping epitopes with scaffold-based immunogens may be a promising route of future investigation.

Altogether, these RSV-based studies have led to the development of a general methodological framework for the accurate design of synthetic immunogens presenting structurally challenging epitopes, and provide a proof-of-concept for the consistent elicitation of RSV nAbs targeting multiple sites using a cocktail vaccine.

5.2 Protein design field

The protein design field has been rapidly evolving during the last few years. This tremendous progress was catalyzed by novel high-throughput, low-cost DNA oligo synthesis technologies, which now allow the testing of computationally designed sequences on a much larger scale than was recently possible. This work has greatly benefited from these technologies, permitting the testing and optimization of the computational design procedure, which was further informed by experimental feedback. In parallel, the Rosetta protein design community has been constantly growing, providing new protocols and features that enable higher precision modeling (*24*), as well as tools to ease the analysis of large-scale computational protein design data (*4*). Leveraging these advances, a major contribution of my thesis work is the *de novo* design of protein scaffolds presenting irregular and discontinuous functional motifs, which has been largely unachieved previously.

Conclusion 3: Tailored design of *de novo* protein scaffolds enables the stabilization of structurally complex functional motifs.

The vast majority of protein functional sites contain irregular (loop) structures, and are formed by the tertiary configuration of side chains stabilized in discontinuous structural segments (*57, 58, 182*). Transplanting such functional sites onto other proteins is an overarching challenge for the protein design field, as it requires extraordinary precision of the design process to maintain the target structure and function outside its native structural environment. To date, most attempts have focused on those functional sites that are contained in linear, structurally regular motifs - especially short helical stretches - which could be grafted to pre-existing or *de novo* structural templates with similar local conformations (RMSD <1.5 Å) (*49, 51, 52, 55, 56, 59, 110*).

Using iterative rounds of constrained protein folding and directed evolution optimization, this work has shown that a discontinuous motif can be installed on a design template with unprecedently low structural similarity (2.8 Å for site 0), even including the replacement of a regular helical structure with a disordered loop structure. However, we noticed that this ‘top-down’ approach, which relies on a pre-existing template that needs extensive remodeling to adapt to a given functional motif, has substantial limitations with regard to its general utility: (1) the number of design templates with sufficient local similarity to a motif becomes sparse with increasing structural complexity of the motif (Fig 1.5); (2) the starting topology determines the overall shape of the designed protein, which may be suboptimal to accurately stabilize the motif, and may impose steric constraints that interfere with the designed function (Fig S3.5); (3) it is a labor-intensive approach, illustrated by the fact that multiple rounds of computational design and directed evolution were needed to reach the design goal (Fig S3.1 - S3.4).

To overcome these limitations, we have introduced a ‘bottom-up’ protein design approach, allowing the assembly of secondary structure elements from scratch with respect to the functional motif’s requirements. Conceptually, the presented ‘bottom-up’ design strategy is a combination of a previously reported, template-based folding and design algorithm (*26*), and other template-free design strategies that have been used for specific design challenges (*60-62, 217*) (Fig 5.3). However, the topology building step is newly implemented, allowing the building of both helices and beta strands, and positioning them with respect to a given functional motif. Our bottom-up design approach is based on a low-resolution definition of an intended protein fold (‘Form definition’ by Taylor (*233*)), which is subsequently projected into the three-dimensional space, and the derived constraints are used to guide the folding and design process (Fig 5.3). This strategy has significant advantages with respect to its general utility, and as compared to previously reported strategies for the *de novo* design of functional proteins (approaches 2 and 3, Fig 1.4 and Fig 5.3): First, it allows the building of any protein topology with one or multiple embedded functional motifs, with virtually any arrangement of secondary structures as defined by the user. Second, in order to extensively sample the defined structural space, our bottom-up design strategy allows a coarse-grained exploration of the structural space through parametric sampling of secondary structure elements. Subsequently, it leverages the capability of the folding and design protocol for the exploration of a wider structural and sequence space as compared to methods with no or limited backbone flexibility (*26*).

We have validated this approach by designing a variety of different topologies for irregular and discontinuous binding motifs throughout chapters 3 and 4, showcasing its utility to build tailored topologies with customized biophysical and structural properties. We found that subtle topological differences sampled by the TopoBuilder can substantially impact the fold stability and the stabilization of the site IV neutralization epitope, as revealed by integrating high-throughput design, testing and next-generation sequencing. Beyond topological differences, this integrated approach has also uncovered the sequence determinants required for S0_2.126 to maintain a substantial core cavity, which was required for its function but is rarely found in natural proteins of similar size. Ultimately, we could prove that the TopoBuilder designs mimicked neutralization epitopes accurately and induced functional antibodies *in vivo*.

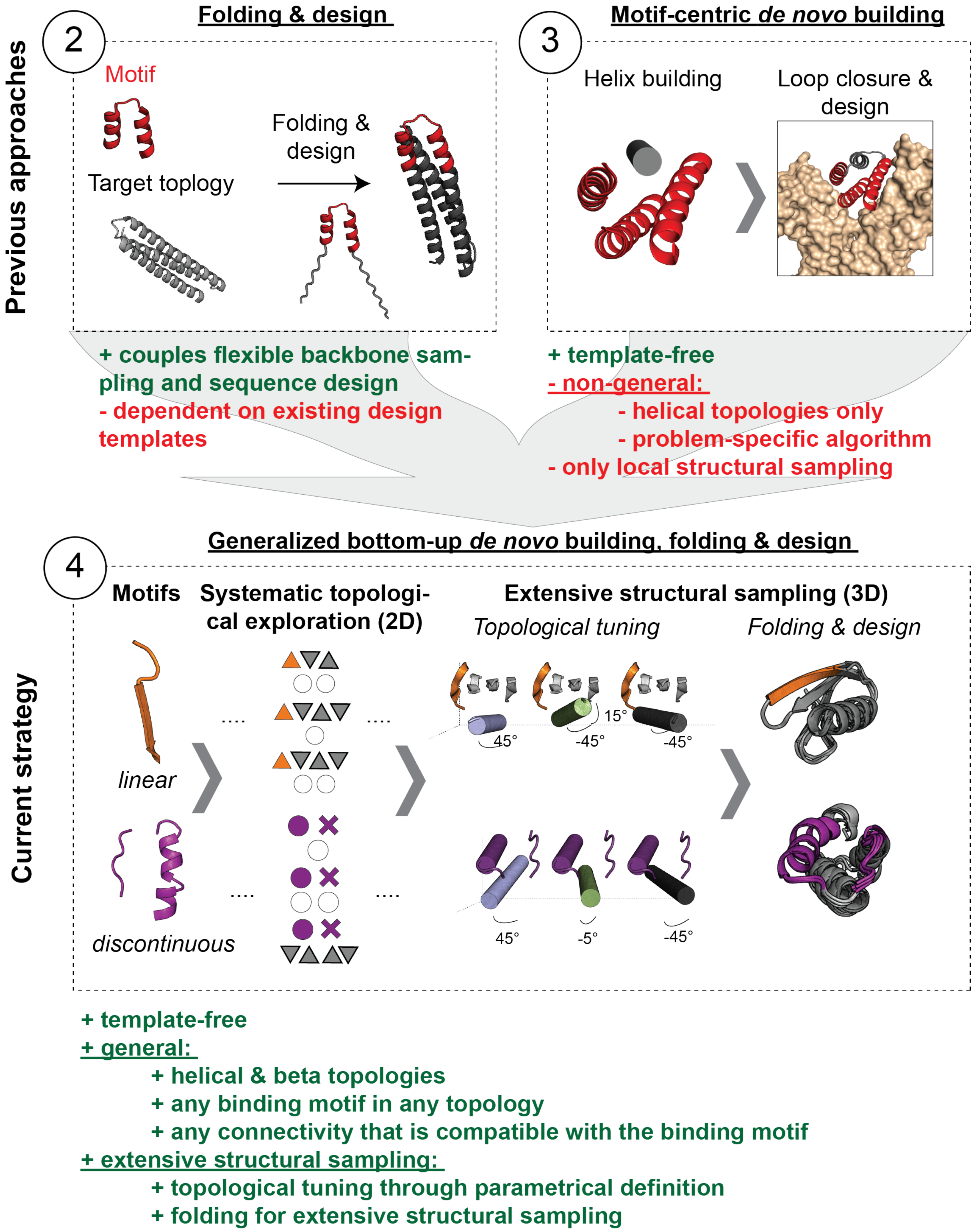

Previous approaches
2
Folding & design
Motif
Target toplogy
Folding & design
3
Motif-centric *de novo* building
Helix building
Loop closure & design
+ couples flexible backbone sampling and sequence design
- dependent on existing design templates
+ template-free
- non-general:
- helical topologies only
- problem-specific algorithm
- only local structural sampling
Current strategy
4
Generalized bottom-up *de novo* building, folding & design
Motifs
Systematic topological exploration (2D)
Extensive structural sampling (3D)
Topological tuning
Folding & design
linear
discontinuous
....
....
....
....
45°
-45°
-45°
15°
45°
-5°
-45°
+ template-free
+ general:
+ helical & beta topologies
+ any binding motif in any topology
+ any connectivity that is compatible with the binding motif
+ extensive structural sampling:
+ topological tuning through parametrical definition
+ folding for extensive structural sampling

Figure 5.3: Bottom-up *de novo* design of functional proteins.
Previous approaches for the *de novo* design of functional approaches included approaches 2 and 3, as introduced in Fig 1.4. The strategy presented throughout chapter 3 and 4 is a combination of both approaches, leading to a generalized strategy for the bottom-up building, folding and design of *de novo* proteins with embedded functional motifs.

Taking a glimpse into the future, the *de novo* protein design strategies showcased throughout my thesis open the door to address a number of biomedical challenges that seem intractable with natural proteins or modified versions of thereof. Beyond immunogens, another application described in this thesis was the development of novel biosensors for the detection and quantification of epitope-specific antibodies. These may become valuable tools for the in-depth analysis and quantification of epitope-specific antibodies, allowing the profiling of antibody specificities in human sera following natural infection or vaccination in a robust, fast and easily scalable assay.

Other future opportunities for the presented design pipeline include the *de novo* design of receptor-ligand mimetics to program cellular behavior and for synthetic biology applications. Cytokine receptors, for example, are activated by ligand-induced receptor homo- or heterodimerization (*234*), and their signaling efficiency is greatly dependent on their spatial orientation upon dimerization (*43*). The ability to design topologically controllable proteins, allowing control and manipulation of epitope distance, their orientation and affinity, could be leveraged to design receptor ligands with customized topological and biophysical properties. Together with the ability to stabilize complex binding sites, this will allow the design of an unprecedent variety of tunable, *de novo* receptor agonists and antagonists, as depicted in Fig 5.4.

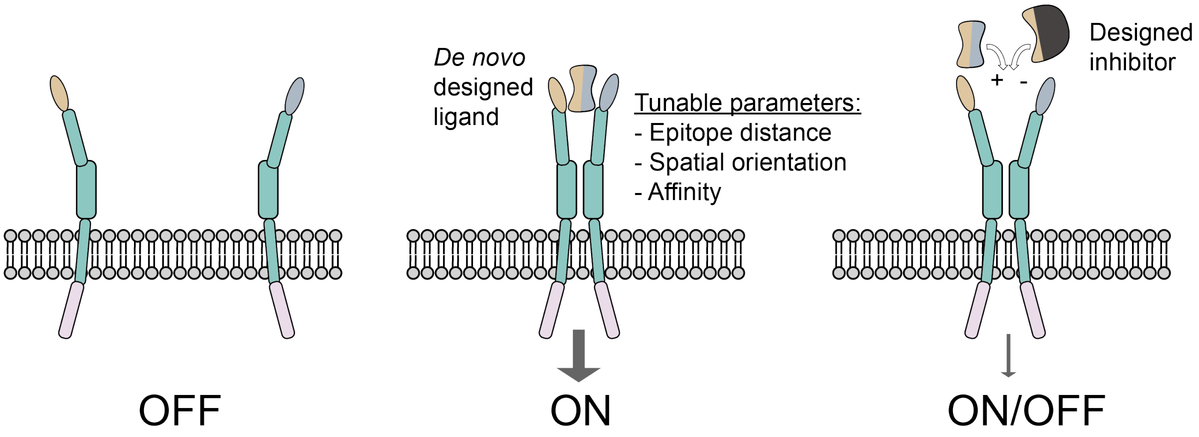


Figure 5.4: Potential future application of *de novo* proteins for synthetic biology applications.
Shown is a schematic representation of a cell surface receptor, that is active upon heterodimerization. A *de novo* designed ligand for both subunits could serve as a trigger to activate downstream signaling. Leveraging the TopoBuilder design strategy allows tuning of the distance between the two binding sites, their spatial orientation as well as their affinity, thus enabling the design of a wide range of ligands with defined structural and functional properties. Similarly, a designed inhibitor that has only one binding site could be used to turn off signaling.

Overall, it appears that the protein design field is now departing from designing proteins for the sole reason of structural accuracy, and has entered into an era where *de novo* proteins can reveal their potential as novel drugs, vaccines and diagnostic tools, and will enable us to study and engineer biological systems with designer molecules that used to be out of reach. This work contributes to this transition by introducing and showcasing a ‘bottom-up’ design approach, and its proven ability to accommodate multi-segment functional motifs in fully *de novo* proteins. There are exciting times ahead of us to see how far *de novo* protein design can improve our understanding of biology, and to tackle some of the outstanding challenges we are facing in the 21st century.

References

1. J. S. McLellan *et al.*, Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113-1117 (2013).
2. M. S. Gilman *et al.*, Rapid profiling of RSV antibody repertoires from the memory B cells of naturally infected adult donors. *Sci Immunol* **1**, (2016).
3. J. S. McLellan *et al.*, Structure of a major antigenic site on the respiratory syncytial virus fusion glycoprotein in complex with neutralizing antibody 101F. *J Virol* **84**, 12236-12244 (2010).
4. J. Bonet, Z. Harteveld, F. Sesterhenn, A. Scheck, B. E. Correia, rstoolbox - a Python library for large-scale analysis of computational protein design data and structural bioinformatics. *BMC Bioinformatics* **20**, 240 (2019).
5. J. J. Mousa *et al.*, Human antibody recognition of antigenic site IV on Pneumovirus fusion proteins. *PLoS Pathog* **14**, e1006837 (2018).
6. V. J. Voet D, Biochemistry, 4th Edition. *Wiley*, (2011).
7. F. Crick, Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).
8. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).
9. B. Alberts, Molecular biology of the cell. (2015).
10. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).
11. I. Sillitoe *et al.*, New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res* **41**, D490-498 (2013).
12. A. Andreeva *et al.*, SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res* **32**, D226-229 (2004).
13. R. Kolodny, L. Pereyaslavets, A. O. Samson, M. Levitt, On the universe of protein folds. *Annu Rev Biophys* **42**, 559-582 (2013).
14. R. K. Wierenga, The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett* **492**, 193-198 (2001).
15. P. Bork, L. Holm, C. Sander, The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* **242**, 309-320 (1994).
16. K. E. Drexler, Molecular engineering: An approach to the development of general capabilities for molecular manipulation. *Proc Natl Acad Sci U S A* **78**, 5275-5278 (1981).
17. C. Pabo, Molecular technology. Designing proteins and peptides. *Nature* **301**, 200 (1983).
18. M. S. Packer, D. R. Liu, Methods for the directed evolution of proteins. *Nat Rev Genet* **16**, 379-394 (2015).
19. P. S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature* **537**, 320-327 (2016).
20. N. K. Fox, S. E. Brenner, J. M. Chandonia, SCOPe: Structural Classification of Proteins--extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res* **42**, D304-309 (2014).
21. K. O. Kopec, A. N. Lupas, beta-Propeller blades as ancestral peptides in protein evolution. *PLoS One* **8**, e77074 (2013).
22. C. A. Orengo, J. M. Thornton, Protein families and their evolution-a structural perspective. *Annu Rev Biochem* **74**, 867-900 (2005).
23. C. A. Rohl, C. E. Strauss, K. M. Misura, D. Baker, Protein structure prediction using Rosetta. *Methods Enzymol* **383**, 66-93 (2004).
24. J. W. Koehler Leman, B.D.; Lewis, S.M.; Consortium, R.; Bonneau, R., Macromolecular Modeling and Design in Rosetta: New Methods and Frameworks. *Preprints*, (2019).
25. K. T. Simons, C. Kooperberg, E. Huang, D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* **268**, 209-225 (1997).
26. J. Bonet *et al.*, Rosetta FunFolDes - A general framework for the computational design of functional proteins. *PLoS Comput Biol* **14**, e1006623 (2018).
27. R. F. Alford *et al.*, The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* **13**, 3031-3048 (2017).
28. R. L. Dunbrack, Jr., M. Karplus, Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol* **230**, 543-574 (1993).

29. Z. Li, H. A. Scheraga, Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc Natl Acad Sci U S A* **84**, 6611-6615 (1987).

30. P. Gainza, H. M. Nisonoff, B. R. Donald, Algorithms for protein design. *Curr Opin Struct Biol* **39**, 16-26 (2016).

31. B. Gutte, M. Daumigen, E. Wittschieber, Design, synthesis and characterisation of a 34-residue polypeptide that interacts with nucleic acids. *Nature* **281**, 650-655 (1979).

32. L. Regan, W. F. DeGrado, Characterization of a helical protein designed from first principles. *Science* **241**, 976-978 (1988).

33. S. P. Ho, W. F. Degrado, Design of a 4-Helix Bundle Protein - Synthesis of Peptides Which Self-Associate into a Helical Protein. *Journal of the American Chemical Society* **109**, 6751-6758 (1987).

34. M. H. Hecht, J. S. Richardson, D. C. Richardson, R. C. Ogden, De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* **249**, 884-891 (1990).

35. B. I. Dahiyat, S. L. Mayo, De novo protein design: fully automated sequence selection. *Science* **278**, 82-87 (1997).

36. B. Kuhlman *et al.*, Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-1368 (2003).

37. N. Koga *et al.*, Principles for designing ideal protein structures. *Nature* **491**, 222-227 (2012).

38. Y. R. Lin *et al.*, Control over overall shape and size in de novo designed proteins. *Proc Natl Acad Sci U S A* **112**, E5478-5485 (2015).

39. A. R. Thomson *et al.*, Computational design of water-soluble alpha-helical barrels. *Science* **346**, 485-488 (2014).

40. E. Marcos *et al.*, Principles for designing proteins with cavities formed by curved beta sheets. *Science* **355**, 201-206 (2017).

41. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580-584 (2015).

42. P. S. Huang *et al.*, De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat Chem Biol* **12**, 29-34 (2016).

43. R. S. Syed *et al.*, Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* **395**, 511-516 (1998).

44. L. L. Looger, M. A. Dwyer, J. J. Smith, H. W. Hellinga, Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185-190 (2003).

45. J. Fernandez-Carneado *et al.*, Surface grafting onto template-assembled synthetic protein scaffolds in molecular recognition. *Biopolymers* **55**, 451-458 (2000).

46. U. Scheib, S. Shanmugaratnam, J. A. Farias-Rico, B. Hocker, Change in protein-ligand specificity through binding pocket grafting. *J Struct Biol* **185**, 186-192 (2014).

47. S. Liu *et al.*, Nonnatural protein-protein interaction-pair design by key residues grafting. *Proc Natl Acad Sci U S A* **104**, 5330-5335 (2007).

48. S. Berger *et al.*, Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. *Elife* **5**, (2016).

49. E. Procko *et al.*, A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* **157**, 1644-1656 (2014).

50. E. Procko *et al.*, Computational design of a protein-based enzyme inhibitor. *J Mol Biol* **425**, 3563-3575 (2013).

51. A. Chevalier *et al.*, Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74-79 (2017).

52. S. J. Fleishman *et al.*, Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816-821 (2011).

53. M. L. Azoitei *et al.*, Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *J Mol Biol* **415**, 175-192 (2012).

54. M. L. Azoitei *et al.*, Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science* **334**, 373-376 (2011).

55. B. E. Correia *et al.*, Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. *Structure* **18**, 1116-1126 (2010).

56. G. Ofek *et al.*, Elicitation of structure-specific antibodies by epitope scaffolds. *Proc Natl Acad Sci U S A* **107**, 17880-17887 (2010).

57. S. Jones, J. M. Thornton, Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93**, 13-20 (1996).

58. N. D. Rubinstein *et al.*, Computational characterization of B-cell epitopes. *Mol Immunol* **45**, 3477-3489 (2008).

59. B. E. Correia *et al.*, Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201-206 (2014).

60. N. F. Polizzi *et al.*, De novo design of a hyperstable non-natural protein-ligand complex with sub-A accuracy. *Nat Chem* **9**, 1157-1164 (2017).

61. A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, D. N. Woolfson, Installing hydrolytic activity into a completely de novo protein framework. *Nat Chem* **8**, 837-844 (2016).

62. D. A. Silva *et al.*, De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186-191 (2019).

63. B. Greenwood, The contribution of vaccination to global health: past, present and future. *Philos Trans R Soc Lond B Biol Sci* **369**, 20130433 (2014).

64. G. J. Nabel, Designing tomorrow's vaccines. *N Engl J Med* **368**, 551-560 (2013).

65. B. B. Finlay, G. McFadden, Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* **124**, 767-782 (2006).

66. W. Dejnirattisai *et al.*, Cross-reacting antibodies enhance dengue virus infection in humans. *Science* **328**, 745-748 (2010).

67. A. Takada, Y. Kawaoka, Antibody-dependent enhancement of viral infection: molecular mechanisms and in vivo implications. *Rev Med Virol* **13**, 387-398 (2003).

68. R. Rappuoli, M. J. Bottomley, U. D'Oro, O. Finco, E. De Gregorio, Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *J Exp Med* **213**, 469-481 (2016).

69. D. Angeletti *et al.*, Defining B cell immunodominance to viruses. *Nat Immunol* **18**, 456-463 (2017).

70. D. C. Ekiert *et al.*, Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246-251 (2009).

71. J. Sui *et al.*, Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol* **16**, 265-273 (2009).

72. A. Impagliazzo *et al.*, A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. *Science* **349**, 1301-1306 (2015).

73. H. M. Yassine *et al.*, Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. *Nat Med* **21**, 1065-1070 (2015).

74. J. O. Ngwuta *et al.*, Prefusion F-specific antibodies determine the magnitude of RSV neutralizing activity in human sera. *Sci Transl Med* **7**, 309ra162 (2015).

75. J. C. Boyington *et al.*, Structure-Based Design of Head-Only Fusion Glycoprotein Immunogens for Respiratory Syncytial Virus. *PLoS One* **11**, e0159709 (2016).

76. D. Eggink, P. H. Goff, P. Palese, Guiding the immune response against influenza virus hemagglutinin toward the conserved stalk domain by hyperglycosylation of the globular head domain. *J Virol* **88**, 699-704 (2014).

77. R. P. Ringe *et al.*, Reducing V3 Antigenicity and Immunogenicity on Soluble, Native-Like HIV-1 Env SOSIP Trimers. *J Virol* **91**, (2017).

78. D. W. Kulp *et al.*, Structure-based design of native-like HIV-1 envelope trimers to silence non-neutralizing epitopes and eliminate CD4 binding. *Nat Commun* **8**, 1655 (2017).

79. S. C. Harrison, Viral membrane fusion. *Virology* **479-480**, 498-507 (2015).

80. M. Magro *et al.*, Neutralizing antibodies against the preactive form of respiratory syncytial virus fusion protein offer unique possibilities for clinical intervention. *Proc Natl Acad Sci U S A* **109**, 3089-3094 (2012).

81. J. S. McLellan, Y. Yang, B. S. Graham, P. D. Kwong, Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. *J Virol* **85**, 7788-7796 (2011).

82. J. S. McLellan *et al.*, Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* **342**, 592-598 (2013).

83. A. Krarup *et al.*, A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism. *Nat Commun* **6**, 8143 (2015).

84. M. G. Joyce *et al.*, Iterative structure-based improvement of a fusion-glycoprotein vaccine against RSV. *Nat Struct Mol Biol* **23**, 811-820 (2016).

85. B. Zhang *et al.*, Protection of calves by a prefusion-stabilized bovine RSV F vaccine. *NPJ Vaccines* **2**, 7 (2017).

86. Y. D. Kwon *et al.*, Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat Struct Mol Biol* **22**, 522-531 (2015).

87. R. W. Sanders *et al.*, Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *J Virol* **76**, 8875-8889 (2002).

88. N. Blais *et al.*, Characterization of Pre-F-GCN4t, a Modified Human Respiratory Syncytial Virus Fusion Protein Stabilized in a Noncleaved Prefusion Conformation. *J Virol* **91**, (2017).

89. P. J. Klasse *et al.*, Influences on trimerization and aggregation of soluble, cleaved HIV-1 SOSIP envelope glycoprotein. *J Virol* **87**, 9873-9885 (2013).

90. I. S. Georgiev *et al.*, Single-Chain Soluble BG505.SOSIP gp140 Trimers as Structural and Antigenic Mimics of Mature Closed HIV-1 Env. *J Virol* **89**, 5318-5329 (2015).

91. M. B. Battles *et al.*, Structure and immunogenicity of pre-fusion-stabilized human metapneumovirus F glycoprotein. *Nat Commun* **8**, 1528 (2017).

92. R. N. Kirchdoerfer *et al.*, Pre-fusion structure of a human coronavirus spike protein. *Nature* **531**, 118-121 (2016).

93. J. Pallesen *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).

94. R. W. Sanders, J. P. Moore, Native-like Env trimers as a platform for HIV-1 vaccine design. *Immunol Rev* **275**, 161-182 (2017).

95. A. B. Ward, I. A. Wilson, The HIV-1 envelope glycoprotein structure: nailing down a moving target. *Immunol Rev* **275**, 21-32 (2017).

96. S. W. de Taeye *et al.*, Stabilization of the gp120 V3 loop through hydrophobic interactions reduces the immunodominant V3-directed non-neutralizing response to HIV-1 envelope trimers. *J Biol Chem* **293**, 1688-1701 (2018).

97. S. W. de Taeye *et al.*, Immunogenicity of Stabilized HIV-1 Envelope Trimers with Reduced Exposure of Non-neutralizing Epitopes. *Cell* **163**, 1702-1715 (2015).

98. J. Jardine *et al.*, Rational HIV immunogen design to target specific germline B cell receptors. *Science* **340**, 711-716 (2013).

99. F. Klein *et al.*, Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell* **153**, 126-138 (2013).

100. X. Xiao *et al.*, Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens. *Biochem Biophys Res Commun* **390**, 404-409 (2009).

101. J. G. Jardine *et al.*, HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science* **351**, 1458-1463 (2016).

102. X. Wu *et al.*, Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856-861 (2010).

103. J. G. Jardine *et al.*, HIV-1 VACCINES. Priming a broadly neutralizing antibody response to HIV-1 using a germline-targeting immunogen. *Science* **349**, 156-161 (2015).

104. D. Sok *et al.*, Priming HIV-1 broadly neutralizing antibody precursors in human Ig loci transgenic mice. *Science* **353**, 1557-1560 (2016).

105. M. Tian *et al.*, Induction of HIV Neutralizing Antibody Lineages in Mice with Diverse Precursor Repertoires. *Cell* **166**, 1471-1484 e1418 (2016).

106. B. Briney *et al.*, Tailored Immunogens Direct Affinity Maturation toward HIV Neutralizing Antibodies. *Cell* **166**, 1459-1470 e1411 (2016).

107. M. Medina-Ramirez *et al.*, Design and crystal structure of a native-like HIV-1 envelope trimer that engages multiple broadly neutralizing antibody precursors in vivo. *J Exp Med* **214**, 2573-2590 (2017).

108. J. M. Steichen *et al.*, HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. *Immunity* **45**, 483-496 (2016).

109. J. Guenaga *et al.*, Heterologous epitope-scaffold prime:boosting immuno-focuses B cell responses to the HIV-1 gp41 2F5 neutralization determinant. *PLoS One* **6**, e16074 (2011).

110. J. S. McLellan *et al.*, Design and characterization of epitope-scaffold immunogens that present the motavizumab epitope from respiratory syncytial virus. *J Mol Biol* **409**, 853-866 (2011).

111. M. Medina-Ramirez, R. W. Sanders, Q. J. Sattentau, Stabilized HIV-1 envelope glycoprotein trimers for vaccine use. *Curr Opin HIV AIDS* **12**, 241-249 (2017).

112. G. B. E. Stewart-Jones *et al.*, Structure-based design of a quadrivalent fusion glycoprotein vaccine for human parainfluenza virus types 1-4. *Proc Natl Acad Sci U S A* **115**, 12265-12270 (2018).

113. C. Zhu *et al.*, Rationally designed carbohydrate-occluded epitopes elicit HIV-1 Env-specific antibodies. *Nat Commun* **10**, 948 (2019).

114. B. Rima *et al.*, ICTV Virus Taxonomy Profile: Pneumoviridae. *J Gen Virol* **98**, 2912-2913 (2017).

115. R. T. Stein *et al.*, Respiratory syncytial virus hospitalization and mortality: Systematic review and meta-analysis. *Pediatr Pulmonol* **52**, 556-569 (2017).

116. N. Haber, Respiratory syncytial virus infection in elderly adults. *Med Mal Infect* **48**, 377-382 (2018).

117. J. Chin, R. L. Magoffin, L. A. Shearer, J. H. Schieble, E. H. Lennette, Field evaluation of a respiratory syncytial virus vaccine and a trivalent parainfluenza virus vaccine in a pediatric population. *Am J Epidemiol* **89**, 449-463 (1969).

118. J. Loebbermann, L. Durant, H. Thornton, C. Johansson, P. J. Openshaw, Defective immunoregulation in RSV vaccine-augmented viral lung disease restored by selective chemoattraction of regulatory T cells. *Proc Natl Acad Sci U S A* **110**, 2987-2992 (2013).

119. B. R. Murphy, E. E. Walsh, Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient in fusion-inhibiting activity. *J Clin Microbiol* **26**, 1595-1597 (1988).

120. I. Widjaja *et al.*, Characterization of Epitope-Specific Anti-Respiratory Syncytial Virus (Anti-RSV) Antibody Responses after Natural Infection and after Vaccination with Formalin-Inactivated RSV. *J Virol* **90**, 5965-5977 (2016).

121. S. S. Whitehead, J. E. Blaney, A. P. Durbin, B. R. Murphy, Prospects for a dengue virus vaccine. *Nat Rev Microbiol* **5**, 518-528 (2007).

122. S. C. Harrison, Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690-698 (2008).

123. B. S. Graham, M. S. A. Gilman, J. S. McLellan, Structure-Based Vaccine Antigen Design. *Annu Rev Med* **70**, 91-104 (2019).

124. A. Mejias, O. Ramilo, Review of palivizumab in the prophylaxis of respiratory syncytial virus (RSV) in high-risk infants. *Biologics* **2**, 433-439 (2008).

125. E. Goodwin *et al.*, Infants Infected with Respiratory Syncytial Virus Generate Potent Neutralizing Antibodies that Lack Somatic Hypermutation. *Immunity* **48**, 339-349 e335 (2018).

126. R. Rappuoli, C. W. Mandl, S. Black, E. De Gregorio, Vaccines for the twenty-first century society. *Nat Rev Immunol* **11**, 865-872 (2011).

127. S. A. Plotkin, Correlates of protection induced by vaccination. *Clin Vaccine Immunol* **17**, 1055-1065 (2010).

128. D. R. Burton, L. Hangartner, Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. *Annu Rev Immunol* **34**, 635-659 (2016).

129. P. S. Lee, I. A. Wilson, Structural characterization of viral epitopes recognized by broadly cross-reactive antibodies. *Curr Top Microbiol Immunol* **386**, 323-341 (2015).

130. J. S. McLellan, Neutralizing epitopes on the respiratory syncytial virus fusion glycoprotein. *Curr Opin Virol* **11**, 70-75 (2015).

131. G. Barba-Spaeth *et al.*, Structural basis of potent Zika-dengue virus antibody cross-neutralization. *Nature* **536**, 48-53 (2016).

132. D. F. Robbiani *et al.*, Recurrent Potent Human Neutralizing Antibodies to Zika Virus in Brazil and Mexico. *Cell* **169**, 597-609 e511 (2017).

133. M. Xu *et al.*, A potent neutralizing antibody with therapeutic potential against all four serotypes of dengue virus. *NPJ Vaccines* **2**, 2 (2017).

134. Z. A. Bornholdt *et al.*, Isolation of potent neutralizing antibodies from a survivor of the 2014 Ebola virus outbreak. *Science* **351**, 1078-1083 (2016).

135. S. Chandramouli *et al.*, Structural basis for potent antibody-mediated neutralization of human cytomegalovirus. *Sci Immunol* **2**, (2017).

136. K. M. Hastie *et al.*, Structural basis for antibody-mediated neutralization of Lassa virus. *Science* **356**, 923-928 (2017).

137. D. Corti *et al.*, A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* **333**, 850-856 (2011).

138. D. R. Burton, Antibodies, viruses and vaccines. *Nat Rev Immunol* **2**, 706-713 (2002).

139. D. R. Burton, What Are the Most Powerful Immunogen Design Vaccine Strategies? Reverse Vaccinology 2.0 Shows Great Promise. *Cold Spring Harb Perspect Biol* **9**, (2017).

140. S. F. Andrews *et al.*, Immune history profoundly affects broadly protective B cell responses to influenza. *Sci Transl Med* **7**, 316ra192 (2015).

141. V. I. Zarnitsyna *et al.*, Masking of antigenic epitopes by antibodies shapes the humoral immune response to influenza. *Philos Trans R Soc Lond B Biol Sci* **370**, (2015).

142. V. I. Zarnitsyna, J. Lavine, A. Ellebedy, R. Ahmed, R. Antia, Multi-epitope Models Explain How Pre-existing Antibodies Affect the Generation of Broadly Protective Responses to Influenza. *PLoS Pathog* **12**, e1005692 (2016).

143. G. D. Victora, M. C. Nussenzweig, Germinal centers. *Annu Rev Immunol* **30**, 429-457 (2012).

144. M. Silva *et al.*, Targeted Elimination of Immunodominant B Cells Drives the Germinal Center Reaction toward Subdominant Epitopes. *Cell Rep* **21**, 3672-3680 (2017).

145. G. A. Kirchenbaum, D. M. Carter, T. M. Ross, Sequential Infection in Ferrets with Antigenically Distinct Seasonal H1N1 Influenza Viruses Boosts Hemagglutinin Stalk-Specific Antibodies. *J Virol* **90**, 1116-1128 (2016).

146. K. Van Reeth *et al.*, Heterologous prime-boost vaccination with H3N2 influenza viruses of swine favors cross-clade antibody responses and protection. *NPJ Vaccines* **2**, (2017).

147. C. J. Wei *et al.*, Induction of broadly neutralizing H1N1 influenza antibodies by vaccination. *Science* **329**, 1060-1064 (2010).

148. P. L. Herve *et al.*, RSV N-nanorings fused to palivizumab-targeted neutralizing epitope as a nanoparticle RSV vaccine. *Nanomedicine* **13**, 411-420 (2017).

149. X. Roux *et al.*, Sub-nucleocapsid nanoparticles: a nasal vaccine against respiratory syncytial virus. *PLoS One* **3**, e1766 (2008).

150. R. G. Tawar *et al.*, Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* **326**, 1279-1283 (2009).

151. B. E. Correia *et al.*, Computational protein design using flexible backbone remodeling and resurfacing: case studies in structure-based antigen design. *J Mol Biol* **405**, 284-297 (2011).

152. B. Kuhlman, D. Baker, Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **97**, 10383-10388 (2000).

153. H. Wu *et al.*, Development of motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract. *J Mol Biol* **368**, 652-665 (2007).

154. J. S. McLellan *et al.*, Structural basis of respiratory syncytial virus neutralization by motavizumab. *Nat Struct Mol Biol* **17**, 248-250 (2010).

155. J. L. Xu, M. M. Davis, Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* **13**, 37-45 (2000).

156. J. F. Scheid *et al.*, Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633-1637 (2011).

157. S. Friedensohn, T. A. Khan, S. T. Reddy, Advanced Methodologies in High-Throughput Sequencing of Immune Repertoires. *Trends Biotechnol* **35**, 203-214 (2017).

158. N. L. Kallewaard *et al.*, Structure and Function Analysis of an Antibody Recognizing All Influenza A Subtypes. *Cell* **166**, 596-608 (2016).

159. F. Sesterhenn, J. Bonet, B. E. Correia, Structure-based immunogen design-leading the way to the new age of precision vaccines. *Curr Opin Struct Biol* **51**, 163-169 (2018).

160. N. Jaberolansar *et al.*, Induction of high titred, non-neutralising antibodies by self-adjuvanting peptide epitopes derived from the respiratory syncytial virus fusion protein. *Sci Rep* **7**, 11130 (2017).

161. R. K. Abbott *et al.*, Precursor Frequency and Affinity Determine B Cell Competitive Fitness in Germinal Centers, Tested with Germline-Targeting HIV Vaccine Immunogens. *Immunity* **48**, 133-146 e136 (2018).

162. G. D. Victora, P. C. Wilson, Germinal Center Selection and the Antibody Response to Influenza. *Cell* **163**, 545-548 (2015).

163. P. Tsui *et al.*, Isolation of a neutralizing human RSV antibody from a dominant, non-neutralizing immune repertoire by epitope-blocked panning. *Journal of Immunology* **157**, 772-780 (1996).

164. S. F. Andrews *et al.*, High preexisting serological antibody levels correlate with diversification of the influenza vaccine response. *J Virol* **89**, 3308-3317 (2015).

165. K. Xu *et al.*, Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1. *Nat Med* **24**, 857-867 (2018).

166. Y. Li *et al.*, Immune history shapes specificity of pandemic H1N1 influenza antibody responses. *J Exp Med* **210**, 1493-1500 (2013).

167. D. C. Ekiert *et al.*, A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science* **333**, 843-850 (2011).

168. D. D. Raymond *et al.*, Conserved epitope on influenza-virus hemagglutinin head defined by a vaccine-induced antibody. *Proc Natl Acad Sci U S A* **115**, 168-173 (2018).

169. W. Dejnirattisai *et al.*, Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. *Nat Immunol* **17**, 1102-1108 (2016).

170. L. C. Katzelnick *et al.*, Antibody-dependent enhancement of severe dengue disease in humans. *Science* **358**, 929-932 (2017).

171. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

172. N. Castagne *et al.*, Biochemical characterization of the respiratory syncytial virus P-P and P-N protein complexes and localization of the P protein oligomerization domain. *J Gen Virol* **85**, 1643-1653 (2004).

173. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).

174. J. M. de la Rosa-Trevin *et al.*, Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* **195**, 93-99 (2016).

175. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).

176. T. A. Khan *et al.*, Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Sci Adv* **2**, e1501371 (2016).

177. S. Friedensohn *et al.*, Synthetic Standards Combined With Error and Bias Correction Improve the Accuracy and Quantitative Resolution of Antibody Repertoire Sequencing in Human Naive and Memory B Cells. *Front Immunol* **9**, 1401 (2018).

178. Z. Gu, L. Gu, R. Eils, M. Schlesner, B. Brors, circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).

179. J. Ou, S. A. Wolfe, M. H. Brodsky, L. J. Zhu, motifStack for the analysis of transcription factor binding site evolution. *Nat Methods* **15**, 8-9 (2018).

180. P. S. Huang *et al.*, De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat Chem Biol* **12**, 29-34 (2016).

181. M. Mravic *et al.*, Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science* **363**, 1418-1423 (2019).

182. G. L. Holliday, J. D. Fischer, J. B. Mitchell, J. M. Thornton, Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. *FEBS J* **278**, 3835-3845 (2011).

183. N. S. Laursen, I. A. Wilson, Broadly neutralizing antibodies against influenza viruses. *Antiviral Res* **98**, 476-483 (2013).

184. D. Sok, D. R. Burton, Recent progress in broadly neutralizing antibodies to HIV. *Nat Immunol* **19**, 1179-1188 (2018).

185. F. Sesterhenn *et al.*, Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen. *PLoS Biol* **17**, e3000164 (2019).

186. D. Tian *et al.*, Structural basis of respiratory syncytial virus subtype-dependent neutralization by an antibody targeting the fusion glycoprotein. *Nat Commun* **8**, 1877 (2017).

187. T. A. Whitehead *et al.*, Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* **30**, 543-548 (2012).

188. P. Kristensen, G. Winter, Proteolytic selection for protein folding using filamentous bacteriophages. *Fold Des* **3**, 321-328 (1998).

189. M. D. Finucane, M. Tuna, J. H. Lees, D. N. Woolfson, Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry* **38**, 11604-11612 (1999).

190. A. M. Watkins, P. S. Arora, Anatomy of beta-strands at protein-protein interfaces. *ACS Chem Biol* **9**, 1747-1754 (2014).

191. E. M. Strauch *et al.*, Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site. *Nat Biotechnol* **35**, 667-671 (2017).

192. J. Lee *et al.*, Persistent Antibody Clonotypes Dominate the Serum Response to Influenza over Multiple Years and Repeated Vaccinations. *Cell Host Microbe* **25**, 367-376 e365 (2019).

193. H. F. Moffett *et al.*, B cells engineered to express pathogen-specific antibodies protect against infection. *Sci Immunol* **4**, (2019).

194. J. Zhou, G. Grigoryan, Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci* **24**, 508-524 (2015).

195. X. Hu, H. Wang, H. Ke, B. Kuhlman, High-resolution design of a protein loop. *Proc Natl Acad Sci U S A* **104**, 17668-17673 (2007).

196. P. Conway, M. D. Tyka, F. DiMaio, D. E. Konerding, D. Baker, Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci* **23**, 47-55 (2014).

197. V. Mas *et al.*, Engineering, Structure and Immunogenicity of the Human Metapneumovirus F Protein in the Postfusion Conformation. *PLoS Pathog* **12**, e1005859 (2016).

198. M. D. Tyka *et al.*, Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* **405**, 607-618 (2011).

199. G. Chao *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* **1**, 755-768 (2006).

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

201. M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* **34**, 93-158 (1999).

202. T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *Journal of Biomolecular Nmr* **24**, 171-189 (2002).

203. T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology* **319**, 209-227 (2002).

204. D. Gottstein, D. K. Kirchner, P. Guntert, Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space. *J Biomol NMR* **52**, 351-364 (2012).

205. Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular Nmr* **56**, 227-241 (2013).

206. W. Kabsch, Xds. *Acta Crystallogr D* **66**, 125-132 (2010).

207. A. J. Mccoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

208. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D* **66**, 486-501 (2010).

209. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213-221 (2010).

210. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).

211. D. Baker, What has de novo protein design taught us about protein folding and biophysics? *Protein Sci* **28**, 678-683 (2019).

212. G. J. Rocklin *et al.*, Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168-175 (2017).

213. E. Marcos *et al.*, De novo design of a non-local beta-sheet protein with high stability and accuracy. *Nat Struct Mol Biol* **25**, 1028-1034 (2018).

214. W. M. Dawson, G. G. Rhys, D. N. Woolfson, Towards functional de novo designed proteins. *Curr Opin Chem Biol* **52**, 102-111 (2019).

215. F. Sesterhenn *et al.*, Trivalent cocktail of de novo designed immunogens enables the robust induction and focusing of functional antibodiesin vivo. *bioRxiv*, 685867 (2019).

216. S. E. Boyken *et al.*, De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658-664 (2019).

217. N. H. Joh *et al.*, De novo design of a transmembrane Zn(2)(+)-transporting four-helix bundle. *Science* **346**, 1520-1524 (2014).

218. J. Dou *et al.*, De novo design of a fluorescence-activating beta-barrel. *Nature* **561**, 485-491 (2018).

219. R. A. Langan *et al.*, De novo design of bioactive protein switches. *Nature* **572**, 205-210 (2019).

220. D. A. Silva, B. E. Correia, E. Procko, Motif-Driven Design of Protein-Protein Interfaces. *Methods Mol Biol* **1414**, 285-304 (2016).

221. H. Lechner, N. Ferruz, B. Hocker, Strategies for designing non-natural enzymes and binders. *Curr Opin Chem Biol* **47**, 67-76 (2018).

222. S. O. Fedechkin, N. L. George, J. T. Wolff, L. M. Kauvar, R. M. DuBois, Structures of respiratory syncytial virus G antigen bound to broadly neutralizing antibodies. *Sci Immunol* **3**, (2018).

223. C. Castro, M. Gourley, Diagnostic testing and interpretation of tests for autoimmunity. *J Allergy Clin Immunol* **125**, S238-247 (2010).

224. C. Ohst *et al.*, Reliable Serological Testing for the Diagnosis of Emerging Infectious Diseases. *Adv Exp Med Biol* **1062**, 19-43 (2018).

225. R. Arts *et al.*, Detection of Antibodies in Blood Plasma Using Bioluminescent Sensor Proteins and a Smartphone. *Anal Chem* **88**, 4525-4532 (2016).

226. C. W. Wood *et al.*, CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics* **30**, 3029-3035 (2014).

227. J. Beran *et al.*, Safety and Immunogenicity of 3 Formulations of an Investigational Respiratory Syncytial Virus Vaccine in Nonpregnant Women: Results From 2 Phase 2 Trials. *J Infect Dis* **217**, 1616-1625 (2018).

228. P. S. Kulkarni, J. L. Hurwitz, E. A. F. Simoes, P. A. Piedra, Establishing Correlates of Protection for Vaccine Development: Considerations for the Respiratory Syncytial Virus Vaccine Field. *Viral Immunol* **31**, 195-203 (2018).

229. E. Phung *et al.*, Epitope-Specific Serological Assays for RSV: Conformation Matters. *Vaccines (Basel)* **7**, (2019).

230. R. Griss *et al.*, Bioluminescent sensor proteins for point-of-care therapeutic drug monitoring. *Nat Chem Biol* **10**, 598-603 (2014).

231. D. Angeletti, J. W. Yewdell, Is It Possible to Develop a "Universal" Influenza Virus Vaccine? Outflanking Antibody Immunodominance on the Road to Universal Influenza Vaccination. *Cold Spring Harb Perspect Biol* **10**, (2018).

232. F. Klein *et al.*, HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature* **492**, 118-122 (2012).

233. W. R. Taylor, A 'periodic table' for protein structures. *Nature* **416**, 657-660 (2002).

234. J. B. Spangler, I. Moraga, J. L. Mendoza, K. C. Garcia, Insights into cytokine-receptor interactions from cytokine engineering. *Annu Rev Immunol* **33**, 139-167 (2015).

CURRICULUM VITAE

FABIAN SESTERHENN

PERSONAL INFORMATION

Address	Avenue de la Dôle 2, 1005 Lausanne
Nationality	German
Date of birth	June 9, 1990 in Koblenz, Germany
Phone	+41 76 393 9690
E-Mail	fabian.sesterhenn@gmail.com

EDUCATION

04/2013 – 07/2015
Albert-Ludwigs-University, Freiburg, Germany
Master of Science in Molecular Medicine
Final Grade: 1.0 (ranging from 1.0 (best) to 4.0 (worst))

10/2009 – 09/2012
Albert-Ludwigs-University, Freiburg, Germany
Bachelor of Science in Molecular Medicine
Final Grade: 1.2 (ranging from 1.0 (best) to 4.0 (worst))

ACADEMIC RESEARCH EXPERIENCE

since 10/2015
PhD at École Polytechnique fédérale de Lausanne
Doctoral Program of Biotechnology and Bioengineering
Advisor: Prof. Bruno Correia
Title: Expanding beyond the natural protein repertoire to engineer targeted vaccines and diagnostics

06/2014 – 04/2015
The Scripps Research Institute, La Jolla, CA, USA
Advisor: Prof. William Schief
Master’s thesis: Using directed evolution to characterize HIV broadly neutralizing antibody requirements, improve functionality and to design immunogens

SCHOLARSHIPS AND AWARDS

BRIDGE Proof-of-concept fellowship
Postdoctoral fellowship awarded by the Swiss National Science Foundation and Innosuisse (Nov 2019 - Oct 2020)
Awarded funding: CHF 129,000

Finn Wold Travel Award, July 2018
32nd Annual Symposium of the Protein Society, Boston, USA

SystemsX.ch Travel Grant 2018

Otto Bayer Fellowship, 2014-2015
Bayer Science and Education Foundation

German Business Foundation Fellowship Program, 2010-2015

PATENTS

"Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen". Bruno Correia, **Fabian Sesterhenn**, 2018. European patent application - EP18197259.

"Trivalent cocktail of *de novo* designed epitope-focused immunogens to induce neutralizing antibodies". Bruno Correia, **Fabian Sesterhenn**, Che Yang, Jaume Bonet, 2019. European patent application - EP19183026.

ORAL PRESENTATIONS

"Trivalent cocktail of *de novo* designed immunogens enables the robust induction and focusing of functional antibodies *in vivo"* – RosettaCon2019, Leavenworth, USA - *selected from submitted abstracts.*

"Engineering targeted vaccines through computationally designed proteins" - Synthims2019 (Synthetic and Systems Immunology), Ascona, Switzerland – *selected from submitted abstracts.*

SCIENTIFIC PUBLICATIONS

- **F. Sesterhenn#**, C. Yang# et al., "Trivalent cocktail of *de novo* designed immunogens enables the robust induction and focusing of functional antibodies *in vivo*". bioRxiv 2019, doi: 10.1101/685867.
- **F. Sesterhenn** *et al.*, "Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen". PLoS Biol. 2019; 17(2):e3000164.
- **F. Sesterhenn**, J. Bonet, B. E. Correia, "Structure-based immunogen design-leading the way to the new age of precision vaccines". Review. Curr Opin Struct Biol 51, 163-169 (2018).
- J. Bonet, Z. Harteveld, **F. Sesterhenn**, A. Scheck, B. E. Correia, rstoolbox - a Python library for large-scale analysis of computational protein design data and structural bioinformatics. BMC Bioinformatics 20, 240 (2019).
- Bonet J, Wehrle S, Schriever K, Yang C, Billet A, **Sesterhenn F**, et al. "Rosetta FunFolDes - A general framework for the computational design of functional proteins". PLoS Comput Biol 14(11):e1006623.
- Briney, B., D. Sok, J. G. Jardine, D. W. Kulp, P. Skog, S. Menis, R. Jacak, O. Kalyuzhniy, N. de Val, **F. Sesterhenn**, et al. (2016). "Tailored Immunogens Direct Affinity Maturation toward HIV Neutralizing Antibodies." Cell 166(6): 1459-1470 e1411.
- Jardine, J. G., D. W. Kulp, C. Havenar-Daughton, A. Sarkar, B. Briney, D. Sok, **F. Sesterhenn**, et al. (2016). "HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen." Science 351(6280): 1458-1463.
- Jardine, J. G., [...], **F. Sesterhenn**, et al. (2016). "Minimally Mutated HIV-1 Broadly Neutralizing Antibodies to Guide Reductionist Vaccine Design." PLoS Pathog 12(8): e1005815.

# equal contribution

Ce document a été imprimé au Centre d'impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.