COMPUTATIONAL VACCINE DESIGN OF CLASS I FUSION PROTEINS

by

KAREN JULIANA GONZALEZ RESTREPO

(Under the Direction of Eva-Maria Strauch)

ABSTRACT

Class I fusion proteins are critical mediators of viral entry in a range of pathogenic viruses, making them prime candidates for vaccine development. Despite their potential, the design of effective vaccines based on these proteins is hindered by two challenges: structural instability and antigenic variability. To address these limitations, we developed three distinct computational strategies aimed at optimizing the vaccine potential of class I fusion proteins. Our initial two methodologies targeted the enhancement of structural stability. The first approach sought to optimize electrostatic contacts within the protein structure, while the second aimed at the strategic introduction of disulfide bonds. Notably, the electrostatic optimization method was implemented across three different viral families and demonstrated significant effectiveness, requiring only a handful of protein variants for successful stabilization. The structural integrity of the leading candidates was subsequently validated through crystallography and cryo-electron microscopy analyses. In a vaccine application focusing on the respiratory syncytial virus (RSV), our designed protein exhibited protective efficacy in a mouse model comparable to that of a recently approved RSV vaccine. Further stabilization was achieved through the introduction of a single disulfide bond, resulting in a thermal stability increase of approximately 12°C without compromising immunogenicity. Our third computational strategy addressed the issue of antigenic variability by

targeting the hemagglutinin (HA) protein of the H1 influenza virus. Utilizing a hybrid approach that combines consensus design with the presentation of a diverse array of immunogenic epitopes, we engineered a HA variant with broad reactivity. This variant induced a diverse antibody response in a mouse model and demonstrated protective capabilities against heterologous viral challenge. Collectively, our computational approaches offer a comprehensive toolkit for the optimization of class I fusion proteins as vaccine candidates. They present substantial promise for the development of effective preventive measures against a wide spectrum of viral infections.

INDEX WORDS: Respiratory syncytial virus, human metapneumovirus, severe acute respiratory syndrome coronavirus 2, influenza virus, vaccine development, class I fusion protein, protein stabilization, antigenic variability, structure-based computational design, consensus-based sequence design.

COMPUTATIONAL VACCINE DESIGN OF CLASS I FUSION PROTEINS

by

KAREN JULIANA GONZALEZ RESTREPO

BS., University of Quindio, Colombia, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

COMPUTATIONAL VACCINE DESIGN OF CLASS I FUSION PROTEINS

by

KAREN JULIANA GONZALEZ RESTREPO

Major Professor:	Eva-Maria Strauch
Committee:	Jarrod J. Mousa
	Jonathan Arnold
	Natarajan Kannan

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

DEDICATION

To the remarkable woman whose life exemplified the true meaning of hard work and commitment: thank you for being my source of inspiration and motivation. I love you deeply, dear mom.

A la mujer extraordinaria cuya vida ejemplificó el verdadero significado del trabajo arduo y la dedicación: gracias por ser mi fuente de inspiración y motivación. Te quiero inmensamente, mamá.

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to my supervisor, Dr. Eva-Maria Strauch, for the invaluable opportunity to work under her mentorship over the years. Her patience, support, and insightful understanding of both professional and personal matters have been fundamental in my growth. Thanks Dr. Strauch for caring about my future, being attentive to my queries, and providing me with timely advice when I needed it.

I am also grateful to my committee members: Dr. Jarrod J. Mousa, Dr. Jonathan Arnold, and Dr. Natarajan Kannan, for their constructive guidance during my doctorate training. Special acknowledgment goes to Dr. Jarrod Mousa and his research team—Dr. Jiachen Huang, Kaito Nagashima, and Rose Juliana Miller—for their invaluable contributions to my research and their willingness to extend assistance when needed.

A sincere thank you goes to both current and former members of the Strauch Lab for their academic and personal support. I am particularly grateful to Qingfa Hou for being my lifeline on numerous occasions, to Dr. Raulia Syrlybaeva for all the adventures shared, and to Autumn Tobin for her consistent and unhesitating help. Additional thanks go to our technicians Arianna, Hope, and Ellen for their role in protein production.

Further, I would like to express my appreciation to those who have contributed to the success of my research. My sincere thanks go to Dr. Stephen M. Tompkins and his lab members, especially Justin Dean Shepard and Scott K Johnson; Dr. Miria Ferreira Criado; Dr. Ted Ross and his lab, particularly Spencer Reynolds Pierce and Dr. Giuseppe A Sautto; Sigmovir Biosystems; Dr. Dewight Williams; and the Proteomics and Mass Spectrometry Facility at the University of

Georgia, with a special mention to Dr. Chau-wen Chou. I am also grateful to the Institute of Bioinformatics, and especially to Sandra S. Getz and April King Mosley, for their unceasing support throughout my Ph.D. journey. Your presence in our institute is truly a treasure.

Last but not least, my heartfelt thanks go to my family and friends who have made this challenging journey far more manageable. To my mother, Amparo Restrepo Castaño, and my brothers, Mauricio and Alexander Gonzalez Restrepo, whose unconditional support has been the cornerstone of my achievements. To my partner, Samuele Lamon, whose belief in my potential often surpasses my own and has consistently encouraged me to strive for excellence. You are my emotional anchor and the driving force that prevents me from succumbing to failure. Thank you for bringing light and joy into my life. To my friend David Montgomery, for his invaluable help since my arrival in the United States, and whose reliability is a constant source of comfort. Finally, my sincere thanks to my initial supervisor, Dr. Juan B. Gutierrez, for opening the doors to this academic journey. Without you, Dr. Gutierrez, this dream would not have been realized.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Respiratory infections are highly prevalent among humans, posing a significant concern for infants and older adults [1]. In 2020, respiratory infections in the United States accounted for approximately 3.5% of the total causes of death in children aged 1 to 4 years and 13% in individuals aged 75 and older [2]. Unfortunately, these estimates reached even higher levels in low- or middle-income countries, where respiratory infections comprised 18.6% and 37.6% of the total causes of death in children and the elderly, respectively, as reported in Peru [2]. While diverse pathogenic agents can cause respiratory diseases, systematic meta-analyses indicate that acute lower respiratory infections in children are predominantly attributed to viruses [3], [4], and about 25% of cases of community-acquired pneumonia in adults are of viral origin [5].

Even though prophylactic interventions and antiviral drugs offer some degree of protection against various respiratory diseases, durable and effective vaccines continue to be an unmet need for viruses such as influenza virus [6], [7], human metapneumovirus (hMPV) [8], or human parainfluenza viruses (hPIVs) [9]. Interestingly, these viruses, along with other clinically relevant pathogens like respiratory syncytial virus (RSV) and coronaviruses, share infection mechanisms that could unveil generalizable approaches for the development of effective treatments [10]–[13]. Specifically, the dependence on class I fusion proteins for mediating membrane fusion and cell entry has emerged as a crucial target for inhibition. Indeed, class I fusion proteins have proven to be potent immunogens and hold significant promise for developing vaccines against respiratory and non-respiratory diseases [14]–[20].

Despite their potential as vaccine candidates, class I fusion proteins remain challenging immunogens due to their inherent characteristics. Because of their role in the fusion process, class I fusion proteins are naturally designed to irreversibly refold from a highly energetic conformation, called the prefusion state, to a more stable structure known as the postfusion state [21]. This spontaneous transition presents a significant hurdle in vaccine development as the protein's most antigenic epitopes often become concealed in the stable postfusion state [22], [23]. Moreover, due to the selective pressure exerted by the host immune response, some class I fusion proteins continuously accumulate point mutations and generate unique epitopes that evade existing immunity [24]–[26]. Consequently, the structural instabilities of the prefusion conformation and its antigenic variability are critical limitations for designing effective vaccines derived from class I fusion proteins.

To overcome the above setbacks, we sought to develop diverse computational strategies to optimize class I fusion proteins as potential vaccine candidates. First, we proposed a two-step design protocol to increase the conformational stability of the prefusion state. This approach involved identifying beneficial substitutions of the prefusion conformation by comparing the energetic profiles of the pre- and postfusion structures. We hypothesized that mutations decreasing the free energy of the prefusion state, with no benefit in the postfusion state, would create an energy gap that prevented the pre- to postfusion transition. Second, we aimed to further increase the prefusion state's stability by introducing disulfide bonds on essential areas for the conformational switch. Our algorithm automatically detects residue positions capable of accommodating a disulfide bond by evaluating the correctness of their geometry. Furthermore, we suggest that the location of the potential new disulfide bond can provide valuable insights into the level of stability it may confer to the prefusion configuration. Lastly, we addressed the influenza

hemagglutinin (HA) antigenic variability by employing a dual strategy: integrating consensus-based sequence design with the targeted presentation of distinct immunogenic epitopes in areas of low sequence conservation. We hypothesized that the key protective elements within the H1 influenza virus may be found in prior vaccine strains that have demonstrated efficacy over extended periods. Accordingly, summarizing the sequence characteristics of these effective strains via consensus design could reveal conserved, protective epitopes that span multiple viral generations. Moreover, substituting highly variable regions with epitopes sourced from different strains within the long-lasting vaccines, could stimulate an antibody response able to recognize diverse antigenic phenotypes. Therefore, we propose that the synergistic application of both approaches holds the potential to produce immunogens capable of inducing a comprehensive spectrum of protective immune responses.

We have illustrated the value of our computational strategies through their implementation in different viral proteins, including the RSV F, hMPV F, the severe acute respiratory syndrome coronavirus 2 spike (SARS-CoV-2 S), and the influenza HA. Notably, our stabilization algorithms demonstrated high efficiency, requiring testing of only a handful of variants to find successful designs. Moreover, the leading candidates of each design approach induced robust immune responses in animal models, consistent with the intended objectives of their respective strategies. Given the clinical relevance of class I fusion proteins, our computational approaches offer valuable resources for accelerating the discovery of effective treatments to current and future public health challenges.

CHAPTER 2

LITERATURE REVIEW

A part of this chapter was accepted by Viruses. Reprinted here with the permission of the publisher. Narkhede, Y. B., Gonzalez, K. J. & Strauch, E. M. Targeting Viral Surface Proteins through Structure-Based Design. Viruses, Vol. 13, Page 1320 13, 1320 (2021).

Viral Fusion Proteins

Fusion glycoproteins decorate the surface of enveloped viruses and are essential for their cell entry. As critical players in the infection process and primary proteins on the viral surface, fusion proteins are excellent targets for developing antivirals, and they constitute the primary immunogens of different vaccine modalities [15], [19], [27]–[31]. Despite their diverse structural features, fusion proteins adhere to common physical principles during the membrane fusion process. This event relies on a dramatic transformation wherein the protein refolds from a strained and unstable prefusion conformation to a highly stable postfusion conformation [21], [32]. The substantial energetic difference between the initial and final conformations is believed to generate the required free energy for catalyzing the fusion between host and viral membranes [21].

Viral fusion proteins are divided into three classes —class I, II, and III [21], [32]–[34], according to their structure and activation mechanism. In terms of structure, class I fusion proteins are trimeric and predominantly composed of α-helices. In their postfusion conformation, they exhibit a characteristic six-helix bundle comprised of a long α-helical coiled coil surrounded by three C-terminal helices [21], [35]. In contrast, class II fusion proteins exist as homo- or heterodimers in their prefusion state and transform into trimers in their postfusion state. Their structural signature includes β-sheets in both conformations [36]. Finally, class III fusion proteins are trimers in both prefusion and postfusion states and present a combination of α-helical and β-sheet structures [37]. Notably, unlike most viral fusion proteins where membrane fusion implies irreversible conformational changes, class III fusion proteins can achieve a thermodynamic equilibrium between the prefusion and postfusion states, allowing for reversible transitions [39], [40].

Regarding activation mechanism, class I and II fusion proteins require proteolytic processing, either on themselves or on companion protein, to facilitate cell entry. Conversely, class III fusion proteins do not utilize maturation cleavage to initiate membrane fusion [38]. Upon activation, class I fusion proteins can be triggered by various factors, which will be discussed in further detail below, while class II fusion proteins primarily respond to low pH conditions [36]. Class III fusion proteins can be triggered by either low pH or through the interaction of a partner protein with a receptor on the host cell [39].

Class I Fusion Proteins

In this study, we focused on class I fusion proteins as they comprise the fusion machinery of a wide range of clinically significant viral families, such as the *Orthomyxoviridae* (e.g., influenza virus), *Paramyxoviridae* (e.g., human parainfluenza viruses), *Pneumoviridae* (e.g., RSV), *Coronaviridae* (e.g., SARS-CoV-2), *Retroviridae* (e.g., Human immunodeficiency virus (HIV)), and *Filoviridae* (e.g., Ebola) [21], [34]. Class I fusion proteins exhibit a substantial presence of α-helices and are comprised of several key components, including a small hydrophobic fusion peptide, an N-terminal heptad repeat sequence 1 (HR1), a C-terminal heptad repeat sequence 2 (HR2), and a transmembrane domain [41]. As a generalized fusion model, class I fusion proteins are initially synthesized as single chains precursors that become fusion competent by proteolytic maturation [38], [42]–[44]. This state, known as the prefusion state, is characterized by a metastable structure that can be readily triggered to undergo a large and irreversible conformational change, resulting in the formation of the postfusion state (Fig 2.1) [45]–[48]. Triggering factors include the switch to a low pH environment (e.g., influenza HA protein) [21], [34], interactions with co-receptors on the cell surface (e.g., HIV Env protein) [21], and stimulation by a second viral protein (e.g., paramyxovirus F proteins) [49].

Upon their activation, the metastable prefusion state undergoes structural changes in the fusion subunit that result in an intermediate state termed the “prehairpin” state [50]–[53] (Fig 2.1). At this point, the prehairpin structure can revert to its prior state in the absence of any membrane or irreversible transition to the postfusion state [34], [54], [55]. Finally, to prompt the fusion process, the hydrophobic fusion peptide is released to connect with the target membrane. This interaction induces the formation of the six-helical postfusion state that brings the virus and host cell membranes in proximity and drives the membrane fusion (Fig 2.1).

Inhibition of the fusion process, and consequently viral infection, can be achieved by either blocking binding to host membrane receptors or by stabilizing any of the conformational states before the completion of the postfusion stage [56]. These inhibition strategies have been observed in potent neutralizing antibodies, providing valuable insights into the essential components for designing effective vaccines [57]–[59]. As a result, the prefusion conformation of class I fusion proteins has emerged as a highly promising immunogen, offering a multitude of opportunities to effectively halt the fusion process [15]–[18], [23].

In addition to vaccines, potential antivirals have also been developed, drawing inspiration from antibody binding modes. Remarkable examples include the mini proteins HSB.2A and HB36, which effectively hinder the influenza fusion process. HSB.2A achieves this by blocking the HA receptor binding site, while HB36 prevents the refolding of the HA stem [60], [61]. Moreover, inhibitors stabilizing the conformations preceding the postfusion state are exemplified by HR2-mimicking peptides, like CP32M, and small molecules such as tertiary butyl-hydroquinone (TBHQ). CP32M stabilizes the intermediate state of the HIV-1 gp41 protein, impeding the formation of the postfusion six-helix bundle [62], whereas TBHQ acts by stabilizing the HA

prefusion conformation, thus increasing the activation barrier to transition into the postfusion state [63].

The RSV and hMPV Fusion (F) protein

RSV and hMPV viruses belong to the *Pneumoviridae* family and share similar infection mechanisms. Their surfaces feature primary glycoproteins, the G and F proteins, which play crucial roles in viral entry – the G protein functions in attachment, while the F protein mediates membrane fusion [64], [65]. Notably, studies have shown that RSV and hMPV variants lacking the G protein can still replicate both *in vitro* and *in vivo* [66]–[68], suggesting that the G protein is not essential for viral entry. In fact, the F protein alone has been demonstrated to mediate both attachment and infectivity [68], [69], although the RSV F function is enhanced in the presence of the G protein [70].

The F protein is initially synthesized as an inactive monomer known as F_0. The activation of F_0 occurs through proteolytic processing by specific host proteases. In the case of hMPV F, trypsin-like proteases cleave it once [71], whereas RSV F requires two cleavage events by furin-like proteases [72], [73]. As a result of two cleavages, RSV F releases a 27 amino acid peptide known as pep27 [74]. The remaining monomeric polypeptides, which stay connected by disulfide bonds [75], subsequently trimerize and give rise to the metastable prefusion state. While the precise sequence of events regarding trimerization and proteolytic processing is not yet entirely clear, evidence suggests that RSV F trimerization likely takes place after proteolytic processing [76], [77]. Triggering of the prefusion conformation finally allows membrane fusion, following the general mechanism described earlier (see section on Class I fusion proteins).

Triggering factors of the RSV F protein remain largely elusive as fusion seems to occur independently of pH and other helper glycoproteins [69], [78], [79]. Although receptor-induced

triggering is a plausible explanation, it is also believed that RSV lacks precise control over the F protein activation, resulting in a basal rate of conversion to the postfusion state [64]. Similarly, despite some hMPV strains requiring acidic conditions to promote fusion [80], [81], low-pH triggering is not universally applicable to all hMPV strains, indicating the existence of diverse fusion activation pathways [82].

The significant level of conservation exhibited by the F protein across RSV and hMPV strains, coupled with its immunogenicity, renders this protein as a highly promising target for vaccine development. RSV F presents six major antigenic sites with varying neutralization sensitivity, depending on the antibody selectivity for the prefusion conformation [59], [83]. The most powerful neutralizing antibodies focus on the antigenic sites Ø and V, which are exclusively located at the apex of the prefusion structure [58], [84] (Fig 2.2). Similarly, antibodies binding to site III demonstrate potent neutralizing activity, exhibiting a preference for the prefusion state [85] (Fig 2.2). On the other hand, antibodies targeting shared antigenic sites between pre- and postfusion conformations (sites II and IV) usually exhibit moderate neutralizing activity [86], [87], while antibodies predominantly binding to the postfusion state (site I) tend to be poorly neutralizing [83] (Fig 2.2).

RSV and hMPV F proteins may only share ~35% sequence identity, yet they exhibit a considerable level of structural similarity (Fig 2.2). Notably, this sequence and structural resemblance results in several antibodies demonstrating cross-reactivity between both proteins [85], [88], [89]. While antibodies targeting analogous antigenic regions of RSV F have been described for hMPV F [90]–[93], studies on the antibody response induced by hMPV natural infection have unveiled additional antigenic sites that do not overlap with the RSV F epitopes [92], [93]. These epitopes are found within the trimer interface, situated either at the apex, middle, or

membrane-proximal regions [92]–[94](Fig 2.2). In contrast to RSV F, the antibody response to hMPV F shows less dominance against the apex of the prefusion structure, likely due to a substantial glycan shield that masks this region [95]. Consequently, the majority of hMPV F neutralizing antibodies are directed to epitopes exposed in both pre and postfusion conformations, such as the sites II, III, and IV [91], [93] (Fig 2.2).

The influenza A hemagglutinin (HA) protein

The influenza HA protein stands as a pioneering example among class I fusion proteins, being one of the first to have its structure elucidated [96]. This significant achievement laid the foundation for understanding the general fusion mechanisms utilized by this protein family. Within the influenza virus, the HA plays a crucial role as a major surface glycoprotein, facilitating the attachment to the host cell and enabling subsequent membrane fusion [97]. The protein's structure illustrates this functionality, with a head domain responsible for receptor binding, while the stem region oversees the fusion of viral and host cell membranes [97] (Fig 2.3A).

The HA protein is initially translated as an uncleaved HA0 precursor, which promptly assembles into a trimeric structure right after synthesis [98]. Upon trimerization, the protein undergoes glycosylation and proteolytic processing, resulting in the generation of the HA1 and HA2 subunits [99] (Fig 2.3A). This proteolytic cleavage is not only critical for infectivity but has also been shown to be a major determinant of pathogenicity in avian influenza A viruses [100], [101]. Highly pathogenic avian influenza viruses are characterized by polybasic cleavage sites, which are recognized and processed intracellularly. In contrast, human and low-pathogenic avian influenza viruses possess monobasic cleavage sites that undergo processing extracellularly [102], [103].

The HA head domain, or globular domain, contains the majority of the HA1 subunit, including the receptor binding site (RBS), and a vestigial esterase subdomain (VE) [96]. For human viruses, the RBS specifically interacts with α-(2,6)-linked sialic acids present on the host cell surface [104]. While few residues within the RBS are highly conserved to maintain effective receptor binding, the peripheral region of this site is highly variable due to the selective pressure exerted by the host immune response [105] (Fig 2.3A).The key structural elements surrounding the RBS are identified as the 130-loop, 150-loop, 190-helix, and 220-loop [106] (Fig 2.3A), and substitutions within these areas determine the antigenic properties of the protein [107]. The major antigenic regions described in the H1 HA protein, which is studied in Chapter 5, are represented in Fig 2.3B.

The HA stalk, also known as the stem domain, comprises the HA2 subunit, including the fusion peptide, along with the remaining amino acids from HA1 [96]. As the main effector of the fusion, the HA stalk is sensitive to pH changes, triggering its transformation into the postfusion state when exposed to the low-pH environment of cellular endosomes [45]. Notably, when the HA1 subunit is absent, the HA2 has been demonstrated to transition to the postfusion conformation even at neutral pH conditions [47]. This suggests that low pH may not be an absolute prerequisite for the conformational change, but facilitates the overcoming of kinetic constraints to achieve the postfusion conformation [47].

Likely due to its accessibility on the viral surface, the head domain exhibits immunodominance over the stem, and antibodies targeting this region are highly neutralizing [108]–[110]. However, the head's constant antigenic changes make these antibodies typically strain-specific, providing protection only until the next antigenic drift occurs [111]. Nevertheless, some head antibodies do display varying levels of cross-reactivity with different influenza strains,

although they are not commonly found [112]–[115]. In contrast, the HA stalk exhibits higher sequence conservation (Fig 2.3A), enabling antibodies directed to this domain to possess broader reactivity [116]–[119]. Unfortunately, a significant portion of stem-antibodies lack neutralizing capabilities, rendering them ineffective in clearing the virus from the lungs [120]. Despite this, some non-neutralizing stem antibodies have been demonstrated to confer protection *in vivo*, likely through Fc-mediated mechanisms [121].

The SARS-CoV-2 spike (S) protein

The coronavirus spike protein is one of the largest proteins within the class I family, consisting of approximately 1,300 amino acids in each protomer of the trimer structure. Unlike the fusion proteins described so far, coronaviruses have different mechanisms for activating the spike protein and do not universally rely on immediate proteolytic cleavage after protein synthesis. Instead, some spikes undergo priming at later stages of the fusion process [122]. In the case of SARS-CoV-2 S, two proteolytic steps have been identified [123]. The first cleavage occurs during protein biogenesis at the S1/S2 site, resulting in the formation of the S1 and S2 subunits [124], [125] (Fig 2.4). This cleavage is facilitated by furin or furin-like proprotein convertase in the Golgi apparatus [11], [126]. The second cleavage is performed on the S2 subunit upon receptor attachment. This step is carried out by transmembrane proteases, such as the serine protease TMPRSS2, or by lysosomal proteases, like cathepsins, within endosomal compartments [127]–[129]. The latter cleavage, known as the S2’ cleavage, is expected to be a common feature among all coronaviruses, as it enables the release of the fusion peptide and activates the protein for membrane fusion [14].

Like the influenza HA protein, the spike S1 and S2 subunits serve distinct functions during viral entry. The S1 subunit is responsible for viral attachment, mediated by a receptor binding

domain (RBD), while the S2 subunit harbors essential elements for the fusion process, including the fusion peptide and heptad repeat domains [130]. To gain entry into target cells, SARS-CoV-2 and other human coronaviruses depend on the interaction between the RBD and the human angiotensin-converting enzyme 2 (ACE2) [128], [131]–[134]. This interaction is facilitated by the remarkable conformational flexibility of the RBD, which not only benefits binding but also contributes to evade immune recognition [135], [136]. The RBD dynamic behavior allows the receptor binding motif (RBM) to be exposed or concealed in what is referred to as the "up" or "down" states. In the "up" conformation, key residues are exposed, promoting effective binding to the ACE2 receptor, whereas in the "down" conformation, these binding residues are hidden [135], [137] (Fig 2.4).

Structural analysis of ACE2-bound and unbound spikes has provided crucial insights into the viral entry process. Upon binding of the first ACE2, the up RBD undergoes certain conformational changes, shifting it away from the trimer axis. This alteration seems to enable the RBD's progressive opening in the remaining protomers, allowing for interactions with additional ACE2 molecules [138]. As more ACE2 molecules bind, the S1 and S2 subunits experience a substantial reduction in atomic contacts, potentially leading to the dissociation of ACE2-bound S1 monomers [138]. The binding of ACE2 and the subsequent disassembly of S1, which can occur either spontaneously or induced by ACE2, leave the S2′ site exposed for its protease processing. S2' cleavage ultimately triggers a cascade of conformational changes in the S2 subunit that result in the characteristic postfusion state [139].

Due to its essential function in the early stages of the viral entry, the S1 subunit is the main target of neutralizing antibodies [140]–[142]. Within S1, the RBD stands out as immunodominant, and specific epitopes that overlap with the ACE2-binding site exhibit potent neutralizing

sensitivity [142], [143] (Fig 2.4). Additionally, regions outside the RBD, such as the N-terminal domain (NTD), have also been found to trigger protective antibody responses, although their neutralizing potency is generally lower compared to ACE2 binding competitors [144] (Fig 2.4). Unfortunately, despite the RBD's constrained mutability, imposed by its role in receptor binding, escape mutants have rapidly emerged in adjacent areas to the RBM. These mutations have led to the development of viral variants that display reduced sensitivity to existing immunity derived from the original SARS-CoV-2 strain [145]–[148].

Similarly to the influenza HA, antibodies targeting conserved epitopes within the stem S2 subunit do exist, although they are rarely induced upon infection or vaccination [149]. These antibodies confer protection by hampering the fusion process and through Fc-mediated effector functions [149].

Rational structure-based vaccine design of class I fusion proteins

While both the prefusion and postfusion states of the fusion proteins have been studied as vaccine candidates, the prefusion conformation has been shown to elicit more potent antibody responses [17], [22], [23], [59], [150]. Notably, stabilized prefusion proteins have been demonstrated to improve the immunogenicity of diverse vaccine formulations such as protein-based vaccines [19], [29], [151], [152], virus-like particles [153], gene-based vectors [31], and nucleic acid-based vaccines [20], [154]. Moreover, stable prefusion proteins have been essential in the identification of potent neutralizing antibodies that can serve as prophylactic and/or therapeutic agents [58], [59], [155].

The successful application of structure-based design approaches to stabilize the prefusion conformation of the RSV F protein has paved the way for generalized strategies to stabilize class I fusion proteins [19]. This pioneering work, which led to first FDA-approved vaccine against

RSV [156], aimed to design prefusion RSV F variants with a stabilized antigenic site Ø. The strategy involved designing a disulfide bond to prevent the postfusion state, introducing two cavity-filling substitutions to increase favorable interactions and the structural order of the protein, and appending a C-terminal T4-phage fibritin trimerization domain ("foldon") to preserve the protein's trimeric structure (Fig 2.5) [19]. The effectiveness of this approach was further confirmed by the massive stabilization of the prefusion F protein of four different types of human parainfluenza viruses [15].

A closer analysis of the refolding mechanism of RSV F further identified that proline substitutions at hinge loops can halt the transition from the prefusion to the postfusion conformation (Fig 2.5). By disrupting the extension of the central helices within the postfusion helical bundle, proline mutations have demonstrated remarkable potential for stabilizing the prefusion conformation of several fusion proteins, including RSV F [77], hMPV F [95], Middle East respiratory syndrome (MERS) S, SARS-CoV S, and SARS-CoV-2 S [151], [157]–[159]. In fact, the rapid study of the novel SARS-CoV-2 S protein was facilitated by introducing two proline mutations that increased the stability of the protein's prefusion state [160]. Additionally, further improvements in the expression levels and stability of the prefusion SARS-CoV-2 S protein were achieved by incorporating four more proline substitutions. These additional proline mutations rigidified flexible loops and stabilized the N termini of helices in the fusion peptide and regions surrounding it [159]. Notably, although the SARS-CoV-2 S protein was also stabilized by introducing cavity filling substitutions, salt bridges, and disulfide bonds, the most substantial increase in expression and stability was observed in proline-containing variants [158], [159].

The neutralization of charge imbalances, particularly at the interface between protomers, as well as replacing the peptide between maturation cleavage sites with a short linker, represent

additional strategies that have highly contributed to increasing the expression levels of different fusion proteins [77], [161]–[163] (Fig 2.5). Finally, a strategy to promote the trimerization of the soluble protein without appending a foldon domain was successfully designed by introducing a 'cysteine zipper' at the C terminus of the RSV F protein (Fig 2.5). The removal of nonpathogenic motifs such as the foldon is of critical importance in vaccine development to reduce potential off-target reactivity [164].

Vaccine design targeting sequence shape shifters: Influenza

As influenza viruses use antigenic variability to escape from the immune response, one of the major challenges in developing influenza vaccines is conferring effective protection against the diverse antigenic forms of the virus [24]. Since current influenza vaccines are only effective when the circulating strain is antigenically similar to the vaccine strains, the design of broadly reactive or universal influenza vaccines is still an unmet need [165].

During naturally occurring infections or after the influenza vaccination, neutralizing antibodies are mostly produced against the viral HA protein [166], [167]. As explained earlier, the HA protein is structurally composed of two domains known as the head domain, which mediates the attachment of the virus to the host cell receptor, and the stem domain, which promotes the fusion between the viral and the cell membranes [96]. While the head domain denotes the immunodominant region of HA proteins and comprises most of their antigenic variability, the stem subunit is a more conserved subdominant domain [109].

Antigen design strategies focusing on HA proteins have shown promising progress toward the development of broadly reactive influenza vaccines. These design advances are divided into two main avenues: head-based and stem-based approaches.

Head-based vaccine design:

As the head domain contains the receptor-binding site, most neutralizing antibodies target this region to prevent the virus from binding to host cell receptors [168]. Given this immunologic pressure, epitopes at the globular head are continually mutating to avoid antibody recognition. Consequently, targeting the head region poses a significant challenge due to its antigenic variability. Nevertheless, two protein design approaches have emerged as promising solutions to overcome this variability, generating immunogens that can elicit broad immune responses. The first approach involves displaying diverse RBDs on ferritin-based nanoparticles [169] (Fig 2.6). This mosaic array of HA heads is designed to present conserved regions from adjacent heterologous antigens, which can activate cross-reactive B cells. As a result, cross-reactive B cells may gain an avidity advantage over strain-specific B cells [169].

The second head-based approach is called "computationally optimized broadly reactive antigens (COBRA)" [170]–[172]. COBRA technology consists of a multilayer consensus design of HA sequences. The HA sequences are initially grouped by antigenic eras or phylogenetic subclades, followed by different rounds of consensus sequence calculations (Fig 2.6). Notably, the COBRA technology has shown promising results in various influenza isolates, including H5N1, H1N1, and H3N2 strains [170]–[172]. While the precise mechanism by which COBRA immunogens induce broad immunity is still unclear, one possibility is that combining several antigenically dissimilar sequences results in proteins displaying a cocktail of immunogenic epitopes. Alternatively, consensus calculations may unveil highly conserved epitopes shared across different influenza subtypes.

Stem-based vaccine design:

A strategy to overcome antigenic variability is to direct the immune response towards the conserved regions of the target immunogen. In HA proteins, the stem is highly conserved across different influenza subtypes, and antibodies targeting this region have proven to be broadly reactive [116], [173]–[176]. However, due to the immunodominance of the HA globular head, stem-directed immunity is minimally induced by vaccination or exposure to influenza [177]. Different studies have revealed that stem-reactive antibodies can be boosted when the dominant epitopes are inaccessible, or when there is continuous exposure to antigenically divergent HA heads [177]–[181]. Based on these observations, three main strategies have been successfully developed to elicit stem antibodies (Fig 2.6):

(1). *Immunization with antigenically variable globular heads:*

Sequential immunizations with synthetic chimeric HAs (cHAs) containing a conserved stem domain but holding divergent globular heads have been demonstrated to boost stem-specific antibodies [182]–[185]. These chimeric proteins are engineered by combining the stem subunit from one influenza subtype with an irrelevant head from viruses absent in humans. In this approach, B cells recognizing both head and stem epitopes are generated during the primary immunization. However, upon subsequent immunizations, the preexisting memory B cells are recalled for conserved antigens. Since the immunodominant head epitopes are antigenically distinct in each immunization, only stem-specific antibodies are boosted [167]. This chimeric-HA immunization strategy is currently undergoing clinical trials [186]. One variation of this approach has also shown promising results in the development of universal influenza vaccines. In this modification, only the major antigenic sites are replaced by exotic, typically avian, HA sequences

rather than the entire globular head. The resulting mosaic proteins are intended to boost the antibody response for stem antigens and for conserved epitopes at the head [187]–[189].

(2). *Removal of the globular head:*

Analogous to the previous strategy, HA variants lacking the globular head or "headless" proteins have been shown to efficiently elicit anti-stem antibodies [120], [174], [198], [190]–[197]. As the removal of the head destabilizes the stem subunit, the incorporation of a trimerization domain or the fusion of the stem antigen to a self-assembling ferritin nanoparticle has been necessary to ensure the structural integrity of these immunogens. Both stabilized stem immunogens have been proven to elicit broad immunity and are promising candidates as broadly protective vaccines. The ferritin nanoparticles are currently being evaluated in clinical trials [199].

(3). *Glycan-masking of immunodominant epitopes:*

The third approach to redirect the immune response to the HA stem consists of concealing immunodominant epitopes at the globular head by introducing new N-glycosylation sites. The glycosylation of these additional spots obstructs the access of antibodies to the main head epitopes and facilitates the recognition of other immune, subdominant regions [200]–[203]. The capacity to drive the immune response toward specific regions of the HA protein has made this approach an attractive method of identifying novel antigenic epitopes at both the head and the stem domains [202]. In addition to masking immunodominant regions, modifications in the HA glycosylation pattern have also been used to increase the immunogenicity of conserved epitopes. In this regard, the removal of certain glycans around stem epitopes has been found to induce a more potent immune response against homologous, heterologous, and heterosubtypic influenza viruses [204].

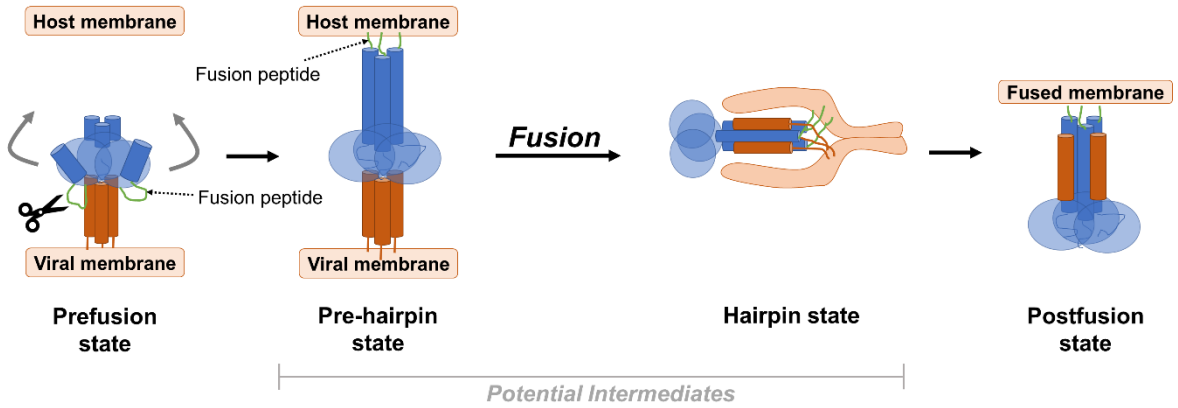


Figure 2. 1 Generalized model for membrane fusion mediated by class I fusion proteins. The protein undergoes activation through proteolysis, leading to the formation of a metastable prefusion state. Upon triggering, this prefusion state initiates a cascade of conformational changes which ultimately facilitate the fusion of viral and cellular membranes.

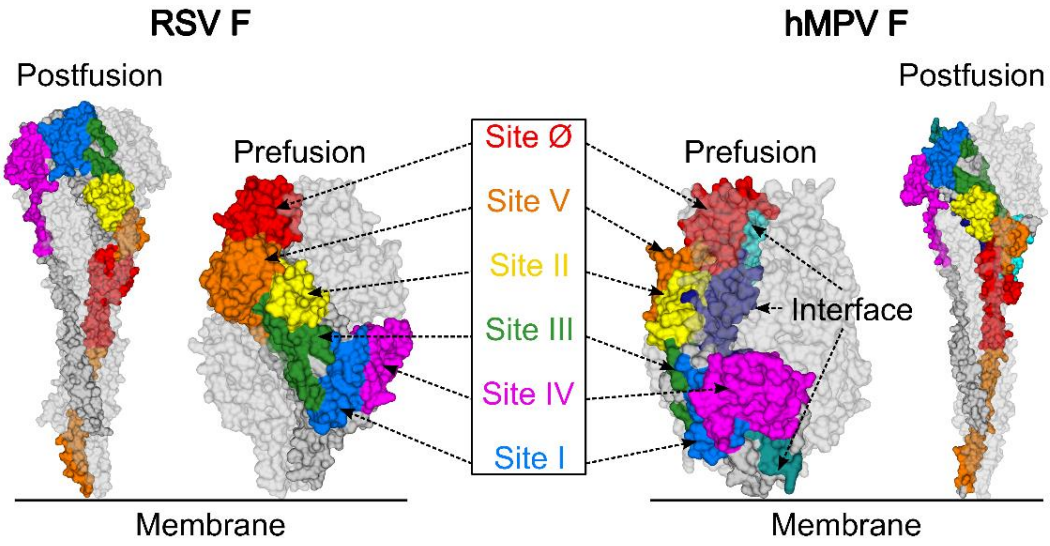


Figure 2. 2 Main antigenic regions of RSV F and hMPV F proteins. Each structure is displayed in their trimeric conformation. Two protomers are shown as grey molecular surfaces, while one protomer highlights the location of antigenic epitopes. The RSV F epitopes were color-coded based on the description in [59], while hMPV F epitopes were assigned according to [93] and [94]. The structures for RSV F correspond to the PDBs 3rrt for the postfusion state [205], and 5w23 for the prefusion state [206]. For hMPV F, the structures are represented by the PDBs 5l1x for the postfusion state [207] and 5wb0 for the prefusion state [95].

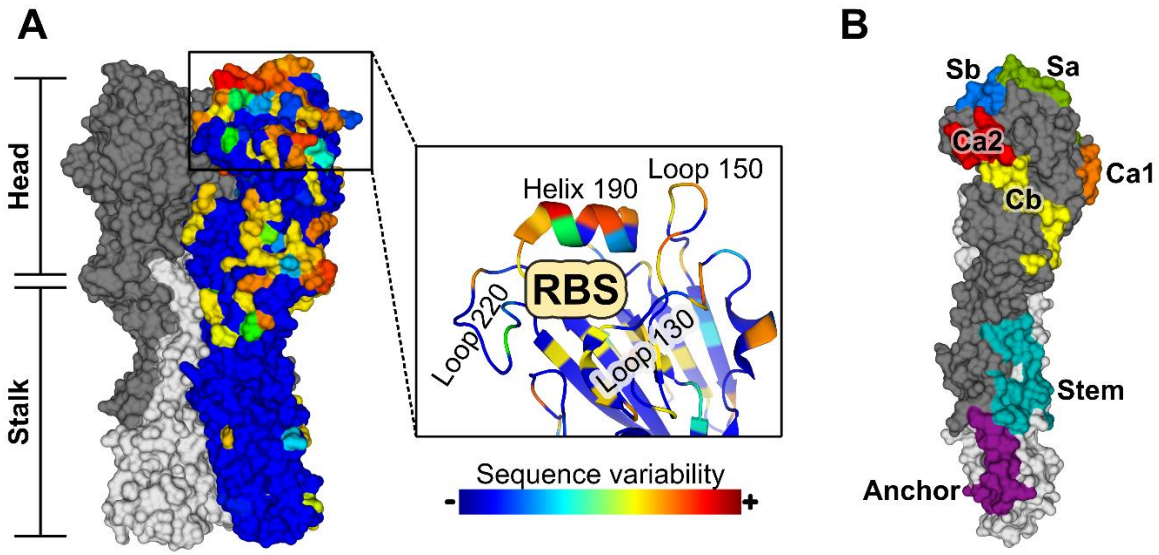


Figure 2. 3 Overview of the influenza H1 HA protein in its prefusion conformation. (A) Protein structure and sequence variability**.** Two protomers are depicted as molecular surfaces. The left protomer showcases the HA1 (dark grey) and HA2 (light grey) subunits, while the right protomer highlights the sequence conservation at each amino acid position. Sequence variability was assessed using Shannon entropy values calculated with the ProDy python package [208]. The analysis included all available full-length human H1 HA sequences from the NCBI Influenza Virus Resource database until 2009 (http://www.fludb.org) [209]. The year 2009 was chosen as the threshold to capture the most significant sequence changes. The highlighted panel emphasizes the receptor binding site, with crucial regions for antigenic drift labeled. **(B)** Main antigenic regions described for the H1 HA protein. The epitopes are visualized on the monomeric structure, with the HA1 subunit depicted in dark grey and the HA2 subunit in light grey. Epitope location was determined based on [210], [116], and [119]. The protein displayed corresponds to the influenza strain A/California/04/2009(H1N1) with PDB 4m4y [112].

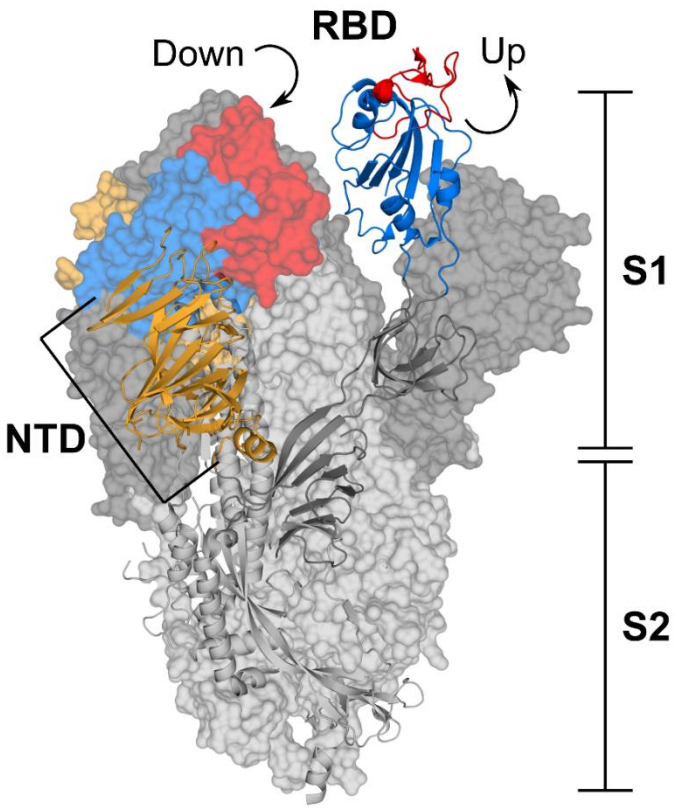


Figure 2. 4 Structural overview of the SARS-CoV-2 S protein in its prefusion state. The protein is displayed in its trimeric configuration, with two protomers presented as molecular surfaces and one protomer depicted using ribbon representation. One protomer exclusively showcases the S1 and S2 subunits, in dark and light grey, respectively. The other protomers highlight the receptor binding domain (blue) in both the up and down states, the ACE-2 binding site (red), and the N-terminal domain (orange). The region assignments were based on [139]. The protein shown corresponds to the PDB 6vyb [211].

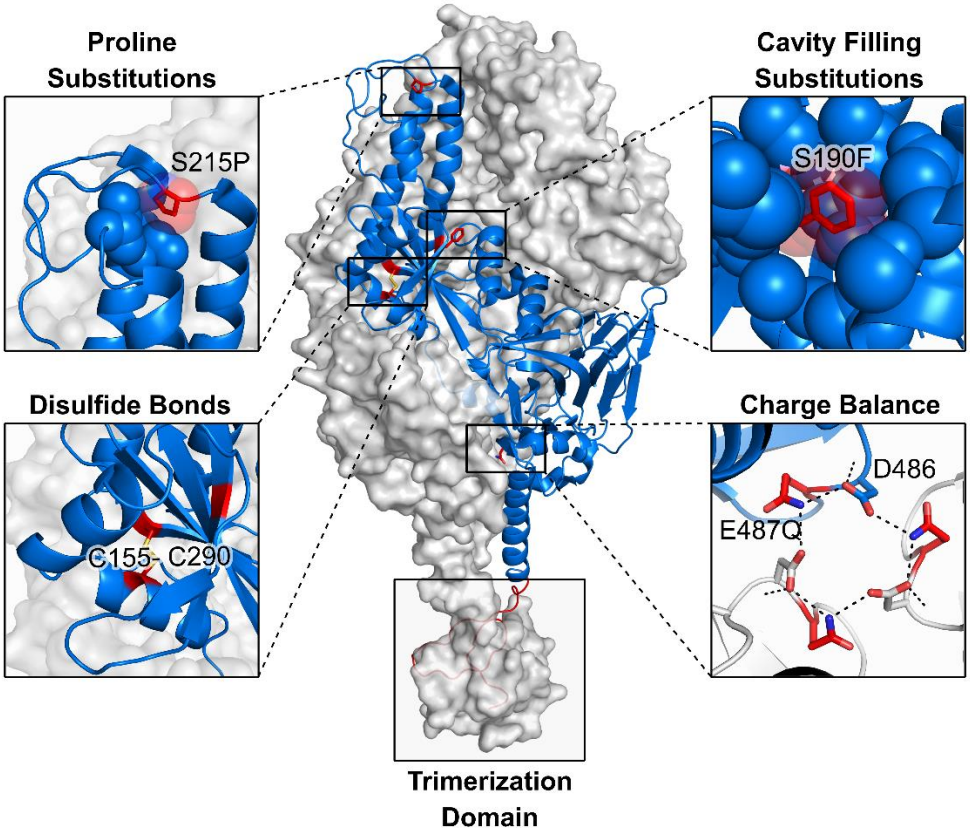


Figure 2. 5 Strategies to stabilize the prefusion conformation of class I fusion proteins. The protein shown corresponds to the trimeric RSV F protein (PDB 4mmv [19] and 5c6b [77]) with two protomers as grey molecular surfaces and one protomer as a blue ribbon. Stabilizing substitutions (S215P, S190F, S155C, S290C, Q487, and a foldon domain) are presented in red, and hydrogen bonds are depicted as black dotted lines. Each panel contains an example of the main stabilization strategies of the prefusion conformation.

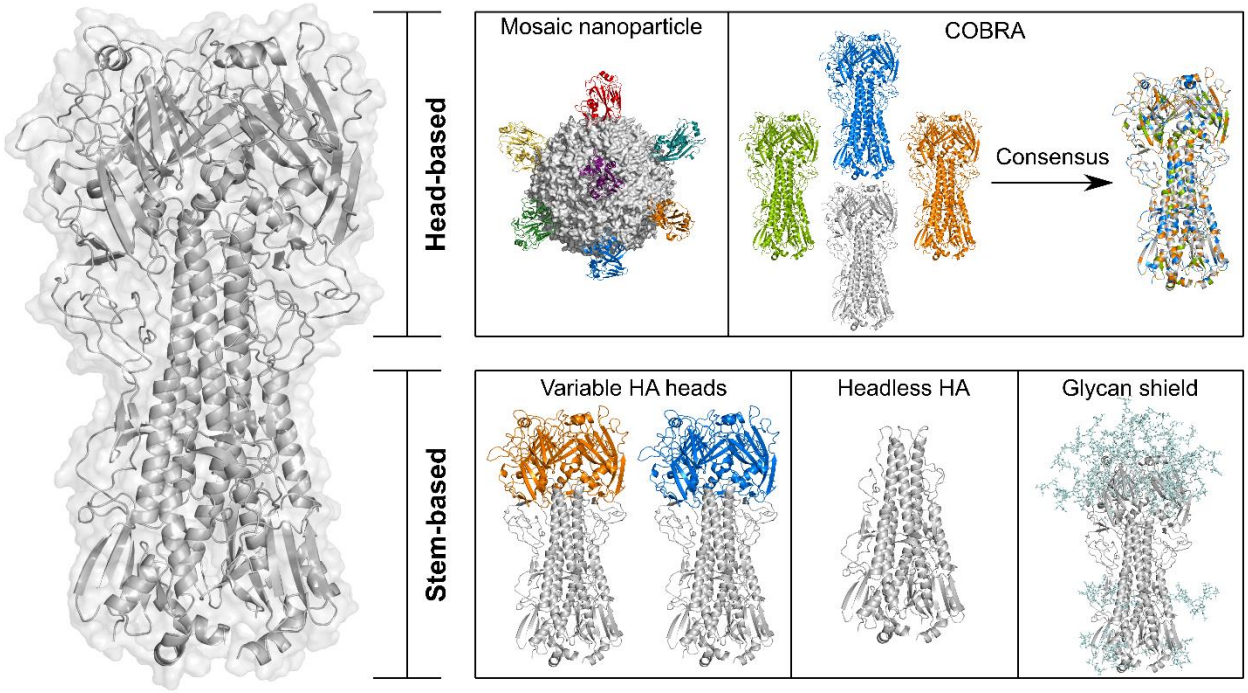


Figure 2. 6 Strategies to design broadly reactive HA-based influenza vaccines. The top panel showcases head-based approaches, whereas the bottom panel presents stem-based approaches. All graphical representations are artificial and for illustrative purposes. The HA proteins displayed correspond to PDB 4m4y [112]. The mosaic nanoparticle was created in PyMol [212] using the ferritin structure from PDB 3bve and the HA head domain from PDB 4m4y. Glycans were drawn using GlyProt [213]. This figure is an adapted version of Figure 6 from our original publication [214].

CHAPTER 3

A GENERAL COMPUTATIONAL DESIGN STRATEGY FOR STABILIZING VIRAL CLASS I FUSION PROTEINS

Gonzalez, K. J., Huang, J., Criado, M. F., Banerjee, A., Tompkins, S., Mousa, J. J. & Strauch, E. M. "A general computational design strategy for stabilizing viral class I fusion proteins."

Submitted to Nature Communications 04/03/2023.

Abstract

Many pathogenic viruses, including influenza virus, Ebola virus, coronaviruses, and Pneumoviruses, rely on class I fusion proteins to fuse viral and cellular membranes. To drive the fusion process, class I fusion proteins undergo an irreversible conformational change from a metastable prefusion state to an energetically more favorable and stable postfusion state. An increasing amount of evidence exists highlighting that antibodies targeting the prefusion conformation are the most potent. However, many mutations must be evaluated before identifying prefusion-stabilizing substitutions. Therefore, we have established a computational design protocol that stabilizes the prefusion state while destabilizing the postfusion conformation. As a proof of concept, we applied this principle to the fusion protein of the RSV, hMPV, and SARS-CoV-2 viruses. For each protein, we tested less than a handful of designs to identify stable versions. Solved structures of designed proteins from the three different viruses evidenced the atomic accuracy of our approach. Furthermore, the immunological response of the RSV F design compared to a current clinical candidate in a mouse model. Given the clinical significance of viruses using class I fusion proteins, our algorithm can substantially contribute to vaccine development by reducing the time and resources needed to optimize these immunogens.

Introduction

Life-threatening viruses such as the human immunodeficiency virus (HIV) [215], Ebola virus [216], Pneumoviruses [12], and the pandemic influenza [10] and coronaviruses[11], use class I fusion proteins to induce the fusion of viral and cellular membranes and infect the host cell. During membrane fusion, class I fusion proteins refold from their metastable conformation (prefusion state) to the highly stable postfusion conformation, likely to provide the energy mediating the fusion reaction [21]. Their essential role in the viral entry as well as their location on the viral surface makes class I fusion proteins one of the major targets of neutralizing antibodies and, thereby, a critical immunogen for vaccination [14]. However, while both pre- and postfusion states are usually immunogenic, the labile prefusion state has been demonstrated to induce a more potent immune response in multiple viral families [15]–[18], [23]. Consequently, the prefusion state has become an attractive vaccine candidate when its conformation can be maintained [15], [19], [20].

Based on structural analyses of the fusion mechanism, the stabilization of the prefusion conformation has been mainly achieved by preventing the release of the fusion peptide or by disrupting the formation of the coiled-coil structure characteristic of the postfusion state [214]. Two strategies have been particularly successful by either designing disulfide bonds at regions undergoing remarkable refolding or introducing proline substitutions to impair the formation of the central postfusion helices [19], [77], [151]. Other stabilization methods have focused on identifying substitutions that increase favorable interactions or rigidify flexible areas in the prefusion structure. These methods either design cavity-filling substitutions [19], [77], [163], neutralize charge imbalances [19], [77], [163], or remove buried charged residues [217]. While the strategies mentioned so far have been effective, the lack of an automated approach has limited the

number of amino acid changes to be assessed and required extensive testing of variants. Notably, more than one hundred different protein variants were evaluated before finding a stable prefusion conformation of the Filovirus GP protein [218], the severe acute respiratory syndrome coronavirus 2 spike protein (SARS-CoV-2 S) [159], and the F protein from Hendra [219], Nipah [18], respiratory syncytial virus (RSV F) [77], [163], human metapneumovirus (hMPV F) [95], and parainfluenza virus types 1-4 [15].

To address these limitations, we developed a general computational approach where the fusion protein's sequence is optimized for the conformation of interest (here, the prefusion state) while destabilizing the other conformation. Our general strategy assumes that conformational rearrangements in class I fusion proteins can be frozen by introducing mutations that reduce the free energy of the prefusion form but do not benefit or better disrupt the postfusion state. While this "negative design" concept has been introduced before in multistate design (MSD) protocols [220], our efforts to implement leading algorithms in class I fusion proteins, such as the MPI_MSD [221], evidenced poor sequence sampling. This is likely due to the extensive sequence-structure search space that needs to be evaluated when modeling both states simultaneously of these large, underpacked proteins. Therefore, we modified the design process by avoiding explicit negative design but using the undesired conformation as a guide to identify suboptimal positions. In a second combinatorial design step, we search for an optimal sequence for the conformation of interest within the subset of substitutions identified to improve the prefusion conformation while disfavoring the postfusion conformation. Using this two-step protocol, we can control the substitution rates by focusing on the most impactful changes. We have successfully stabilized the prefusion state of several large proteins, namely the RSV F, the hMPV F, and the SARS-CoV-2

spike proteins, illustrating the general use of the method. Importantly, only 3-4 variations were necessary to evaluate experimentally, saving a tremendous amount of time and resources.

Results and Discussion

Energy optimization of the prefusion over the postfusion conformation

Fusion proteins must contain various energetically sub-optimal residues for a given conformation to be able to accommodate other conformations required to complete the fusion process [21]. As a first step, we identified these sub-optimal positions for the prefusion conformation based on the protein energetics or its anticipated dynamics (Fig 3.1). For the first approach, mainly used for the stabilization of the RSV F protein (based on the A2 strain, as published under the PDB 5w23 [222]), we uncovered residue positions with contrasting stability between the pre- and postfusion conformations by calculating the energetic contribution of every residue to each state. *In-silico* alanine mutagenesis allowed us to quickly identify contributions towards Gibbs free energy (ΔΔG) of a given residue side chain, approximating the role of the position on the stability of each state [223]. Negative ΔΔG scores (< -1.0 in Rosetta energy units, REU) indicated structural stabilization while positive ΔΔG scores (>1.0 REU) suggested destabilization [223]. Thereby, we created two energetic maps, one for each state, to spotlight residue positions with differential contributions on each conformation (Fig 3.1). In all our examples, about 40 - 50 positions displayed higher stability in the prefusion state than in the postfusion state.

When alanine scanning was not sufficient to locate meaningful designable spots, as defined by a differential of at least 2 REU, we focused on the dynamics of the protein as a second approach to identify sub-optimal positions. For hMPV F and SARS-CoV-2 S proteins, we defined as "designable" all regions undergoing drastic structural rearrangements between states. Highly

movable residues and all positions identified by alanine scanning were exhaustively explored to find substitutions that invert the energetics between states. As the final objective was to find mutations working synergistically rather than individually, all substitutions favoring the prefusion state over the postfusion conformation were assessed in a combinatorial design step (Fig 3.1). Though several sequences were found to lower the prefusion state while increasing the postfusion state energy, the number of mutations introduced was high (~40 substitutions). Therefore, to prevent changes in the immunological properties of the proteins, the number of designable positions was decreased. We aimed to introduce less than 10 mutations by focusing on the lowest energy interactions and the preservation of newly introduced hydrogen bonds and salt bridges in the prefusion state, or substitutions with a strong negative effect on the postfusion structure's energy. The combinatorial design of these substitutions resulted in energetic gaps of at least 119 REU (Fig 3.S1 and Table 3.S1).

Biochemical characterization of RSV F, hMPV F and SARS-CoV-2 S variants

After ranking all designed sequences based on their energy gap between states, the top 3-4 designed variants were expressed and purified. For RSV F, one (R-1b) out of three designs was found to be a monodispersed and trimeric protein, as evaluated by size exclusion chromatography (SEC) (Fig 3.2.A). For the hMPV F and SARS-CoV-2 S, two (M-104 and M-305) out of four, and three (Spk-M, Spk-F, and Spk-R) out of three redesigned proteins behaved similarly, respectively (Fig 3.2.D, G). Remarkably, the hMPV F variant M-104 showed high expression yields, with a 5.5-fold increase with respect to its parent construct the semi-stabilized hMPV F 115-BV [95] (Fig 3.S2). The structural state of these expressed constructs was then evaluated based on their conservation of prefusion-specific epitopes. As predicted, all designs presented a prefusion-like structure as they tightly bound to prefusion-specific binders such as the RSV F antibodies D25

[58], [224] and AM14 [84], [224] (Fig 3.2.B and Fig 3.S3.A), the hMPV F antibodies MPE8 [225] or 465 [94] (Fig 3.2.E and Fig 3.S3.B), or the angiotensin-converting enzyme 2 (ACE2) [226] for the SARS-CoV-2 S protein (Fig 3.2.H).

All expressed proteins showed improved thermal stability compared to their parent prefusion constructs (Table 3.S1). The spike variants Spk-M and Spk-F displayed the highest melting temperature improvement, with ~15°C increment above the current SARS-CoV-2 vaccine, the S-2P construct[130] (Fig 3.2.I). Unlike the vaccine S-2P [130], the design Spk-M preserved the prefusion conformation even after one hour heating at 55°C as evidenced by ACE2 binding at this temperature (Fig 3.S4.A). This stability seems to compare with the highly stable HexaPro construct, which was achieved by introducing several proline substitutions and experimentally evaluating 100 variants [159].

A similar scenario was observed in the hMPV F variant M-104; its stability is comparable to variants containing additional disulfide bonds [227]. In fact, M-104 has a higher melting point (61.5°C, Fig 3.2.F) than the hMPV F variant DS-CavEs (60.7°C)[227] which has two designed disulfide bonds. Additionally, M-104 remains antigenically unaltered after heating at 55°C (Fig 3.S4.B), as has been seen in DS-CavEs2 [227] which contains four new disulfide bonds. Although testing under identical conditions is required to validate these comparisons, these results demonstrate how the correct placement of electrostatic interactions can lead to highly stable proteins.

For RSV F, the improvement of the prefusion state stability cannot be well estimated as we started with the wild-type sequence [222] whose instability substantially impedes its production as an isolated soluble prefusion-state protein [77]; all purified RSV F molecules are found mostly in its postfusion state [228]. Therefore, obtaining the R-1b variant with a melting temperature of 62°C

revealed an effective optimization of the sequence to stay in the prefusion conformation (Fig 3.2.C); especially since no disulfide bonds were introduced and stabilization was achieved through the optimization of non-covalent interactions. Remarkably, design R-1b proved to be antigenically intact even after heating at 55°C (Fig 3.S4.C).

Structure determination of leading RSV F, hMPV F and SARS-CoV-2 S variants

Negative-stain electron microscopy (EM) confirmed homogeneous trimeric prefusion morphology of leading candidates for all three different fusion proteins studied (Fig 3.S5). We therefore proceeded to obtain atomic details by x-ray crystallization and cryo-EM. The crystal structure of the variants R-1b and M-104 confirmed their prefusion conformation at a resolution of 3.1Å and 2.4Å, respectively (Fig 3.3.A-B). The accuracy of our computational predictions was reflected on the high structural similarity between the determined structures and the computational models, with root-mean-square deviations (RMSDs) of only 1.193Å (405 Cα atoms) for R-1b, and 0.53Å (416 Cα atoms) for M-104. The 3D classification performed on the spike cryo-EM images also verified the prefusion structure of the protein with particles displaying one receptor binding domain (RBD) in the up conformation (Fig. S6, Table 3.S6). Solving the structure at 3.7Å resolution, revealed that the S2 subunit, which was the only part engineered, agreed closely with the computational model with a RMSD of only 1.345Å (377 Cα atoms) (Fig 3.3.C).

Although no significant perturbations were observed in all our variants overall, subtle differences at the antigenic site Ø were identified between R-1b and its parent RSV F protein. Specifically, the α4 helix in R-1b is bent towards residue D200 when compared to the parent RSV F protein. As the antigenic site Ø is intrinsically flexible, structural variations in this region are expected and have been reported in several prefusion-stabilized RSV F proteins, including the

clinical candidate DS-Cav1 [19], [77], [163]. For future stabilization efforts, this could be another area of interest for stabilization.

The crystal structures of R-1b and M-104 revealed that despite some deviations between the designed and crystal structure, most introduced substitutions followed their predicted stabilization mechanism by either filing cavities or increasing intra- or interprotomer hydrogen bonds and salt bridges (Fig 3.S7-8 and Table 3.S1). Especially precise rotamer agreement between our computational models and the experimental data was found in cavity-filling mutations, such as E60F in R-1b, and the A159L and V203I in M-104 (Fig 3.3.A-B). Significant rotamer agreement was also observed in substitutions strengthening polar interactions, such as the N380K in R-1b, and the L130D, V430Q, and V449D in M-104 (Fig 3.3.A-B and Fig 3.S8). Other mutations such as the S150E and E487N in R-1b did not interact with the predicted residues but still contributed to the prefusion stability by increasing polar interactions at the protomers' interface (Fig 3.S7). Lastly, we observed the alleviation of buried polar residues by exchanging them for hydrophobic amino acids and improving the overall packing density. This effect was observed, for instance, after swapping the N227 into a leucine residue or removing the unsatisfied hydroxyl of S55 by replacing it with an alanine residue (Fig 3.3.A). Similarly, we observed these mutations for the hMPV F protein. Finally, the design Spk-M was stabilized by four substitutions filling cavities and five substitutions increasing polar interactions at the S2 subunit, three of which were interprotomer contacts (Fig 3.3.C, S9 and Table 3.S1).

Although most of the designed substitutions stabilized the prefusion state, the mutations N175R in R-1b, L130D in M-104, and T941D in Spk-M were intended to disrupt the postfusion conformation (Fig 3.S10). These residues are solvent-accessible in the prefusion state and therefore should not have an impact on the stability of that conformation. However, the postfusion

conformation places these residues at the six-helix bundle where unsatisfied polar residues are highly unfavored and thereby likely to disrupt the core. As we had the postfusion-specific antibody 131-2A [229], we sought to prove this hypothesis and confirm that we had not only stabilized the prefusion state but also indeed destabilized the postfusion state. The diminished binding of the 131-2A antibody to R-1b after heating (60°C) (Fig 3.S4.D), which should have converted the protein into its postfusion state, proved that we fulfilled the design objectives we set out for.

Immunogenicity of design R-1b

We selected the R-1b variant for a vaccination study due to the availability of a highly stable prefusion control, such as the clinical candidate DS-Cav1 [19]. Therefore, to investigate the effect of the introduced mutations on the RSV F immunogenicity, female BALB/c mice were vaccinated twice with either 0.2 or 2 µg of purified R-1b or DS-Cav1 with or without AddaVax adjuvant. Mice were bled at three and nine weeks post-second immunization (Fig 3.4.A). Sera analysis for binding to prefusion RSV F and RSV A2 neutralization revealed that R-1b induced similar levels of RSV F-specific antibody titers (Fig 3.4.B-C and Fig 3.S11) and comparable neutralizing activity related to DS-Cav1 (Fig 3.4.D).

Conclusion

Detailed antibody response studies have illustrated that prefusion-stabilized class I fusion proteins are potent immunogens and promising vaccine candidates. This has been proven to be true for several viruses, including RSV [19], [77], [163], hMPV [227], parainfluenza [15], Nipah [18], MERS-CoV [151], and SARS-CoV-2 [159]. Several of these immunogens have been developed by many steps of iterations of manual structure-based design with experimental evaluation, often testing hundreds of combinations of mutations [15], [18], [77], [95], [159], [163], [217], [219]. To alleviate this laborious exploratory testing, we automated one of the underlying

principles behind their stabilization efforts that considers the biophysics of the fusion protein and its large irreversible switch. We developed a computational approach that seeks to freeze the prefusion conformation by learning about suboptimal contacts from its alternate conformation. Our algorithm allows the automated identification of these regions, and their potential substitutions, based on energy differences and relative motion between the two states. We acknowledge that our computational approach can be limited by the need for both pre- and postfusion structures. However, we believe that the high accuracy of current protein structure prediction algorithms could alleviate this drawback [230]. The efficiency of our method has been demonstrated on three different fusion proteins, such as the RSV F, hMPV F, and the SARS-CoV-2 S, where only 3-4 variants were necessary to successfully find a stable prefusion design. Additionally, we were able to validate the immunogenicity of one design in a mouse model, showing similar in vitro neutralization and specific serum IgG patterns compared to a clinical candidate. Therefore, our algorithm could highly impact the vaccine development field by allowing a rapid optimization of both novel class I fusion proteins as well as leading vaccine immunogens.

Materials and Methods

I. Computational approach to stabilize the prefusion state

All computational analysis were performed with the Rosetta version: 2020.10.post.dev+12.master.c7b9c3e c7b9c3e4aeb1febab211d63da2914b119622e69b

Structure preparation

- *RSV F:*

All computational analyses were performed on the crystal structure of the RSV F protein in the prefusion (PDB:5w23) [222] and postfusion (PDB: 3rrt) [228] conformations. To remove

small clashes in the structures, both conformations were refined with the Rosetta relax application guided by electron density data [231]. Density maps were generated from the corresponding map coefficients files associated to the PDB accession codes. These coefficients were transformed into density maps using the Phenix software version 1.15 [232] and the option "create map from map coefficients" (region padding= 0, and grid resolution factor= 0.3333). To include the electron density in the refinement process, the density energy term was activated in the Rosetta scoring function with a weight of 20. This weight was selected given the low-resolution of the density map and the starting structures. During the relaxation protocol, four rounds of rotamer packing and minimization were performed with gradual increases to the repulsive weight in the scoring function [233]. After 5 cycles of relaxation, the quality of the resulting models was evaluated with the Molprobity web service [234]. The structures with the lowest Rosetta energies and Molprobity scores were used for mutational analysis.

- *hMPV F:*

As described in the RSV F example, the hMPV F prefusion (PDB: 5wb0)[95] and postfusion (PDB: 5l1x) [207] conformations were relaxed using their respective electron density data. Due to the high resolution of the starting prefusion hMPV F structure (2.6Å), the weight of the density energy term was increased to 50 to encourage a good agreement with the density map. All other parameters and postprocessing followed the RSV F example.

- *SARS-CoV-2 S:*

The input SARS-CoV-2 S prefusion and postfusion structures and their corresponding cryo-electron microscopy (cryo-EM) maps were retrieved with the PDB accession codes 6vxx[211] and 6xra[235], respectively. Since both structures were not completely solved, missing regions were modeled using the default comparative modeling protocol available in Rosetta [236].

However, the cryo-EM density maps of the input pre- and postfusion structures were integrated in the modeling process to avoid large deviations from the original configuration. Templates selected for prefusion modeling corresponded to the PDB IDs 6m0j[237] and the initial 6vxx, while the postfusion structure was modeled using 6lxt[238] and the initial 6xra. Model selection was based on overall agreement with the starting structure and templates, as well as a low Rosetta energy score. These homology models were then relaxed with the RosettaScripts [239] framework, incorporating fit-to-density parameters established for cryo-EM density [231]. Due to the high resolution of the starting structures, the refinement process was performed with a density scoring weight of 50 and three cycles of FastRelax [233] in cartesian space. The structures with the lowest Rosetta energies and Molprobity scores were used for mutational analysis.

Selection of target positions to redesign

Residue positions to redesign were selected based on two independent approaches: I) contrasting energetic contributions to the pre- and postfusion conformations, and II) location on regions displaying drastic rearrangements between the pre- and postfusion states.

- *Selection based on energetic contributions*:

Amino acid positions with contrasting energetic contributions between conformations offer an opportunity to manipulate the energetics of the conformational switch by allowing the optimization of one state while the other state is disfavored. Identification of these target spots was done by in-silico alanine mutagenesis where the energetic role of each residue on each conformation was estimated using the change in folding energy upon mutation. Consequently, residue positions displaying simultaneous stabilization of the prefusion state and destabilization of the postfusion state were selected as hotspots to redesign. Details about the selection process are described at the computational alanine scanning section

- *Selection based on protein dynamics:*

When alanine scanning was not enough to identify positions with contrasting stability between conformations, all regions involved in the refolding process were chosen as targets to redesign. To identify these flexible areas, the root mean square deviation (RMSD) of each cα atom was calculated using their corresponding position in the pre- and postfusion structures. Both structures were structurally aligned prior to the analysis, and residue positions displaying motion levels of at least 10Å were selected to redesign. Furthermore, residues flanking the highly movable areas were also considered to be redesigned when their secondary structure differed between the pre- and postfusion states. Flanking residues were included until a set of 8 (SARS-CoV-2 S) or 16 (hMPV F) consecutive residues matched their secondary structure in the pre- and postfusion structures.

Computational alanine scanning

A computational alanine scanning was performed on the pre- and postfusion states of RSV F, hMPV F, and SARS-CoV-2 S proteins to determine the energetic contributions of each amino acid to each conformation. Since the prefusion SARS-CoV-2 S contains domains that are not present in the postfusion conformation, alanine scanning in this protein was limited to shared regions between states. Using the Rosetta ΔΔG protocol "cartesian_ddg" [223], the backbone and sidechains around the position to be mutated were optimized in the cartesian space and the change in folding energy (ΔG) was computed before and after each alanine substitution [223]. The contribution of every residue to stability was calculated in terms of ΔΔG scores (ΔG_{mutant} Δ-$G_{wild\ type}$) where alanine changes holding ΔΔG scores < -1.0 were considered stabilizing substitutions while ΔΔG scores > 1.0 were considered destabilizing changes [223]. All calculations were repeated 3 times and the average value among repetitions was used as the final ΔΔG score. To

enhance the prefusion stability over the postfusion, regions presenting a stabilizing score in the prefusion conformation but not in postfusion conformation were chosen to redesign. Specifically, designable positions were selected based on a) stabilizing ΔΔG score in the prefusion conformation and destabilizing ΔΔG score in the postfusion conformation, b) stabilizing ΔΔG score in the prefusion conformation and neutral ΔΔG score in the postfusion conformation, or c) destabilizing ΔΔG score in the postfusion conformation and neutral ΔΔG score in the prefusion conformation. To restrict the design process to the most relevant spots, positions meeting any of the above criteria were filtered based on an energetic difference of at least 0.7 ($\Delta\Delta G_{postfusion} - \Delta\Delta G_{prefusion}$). Finally, since positions with native alanine are overlooked with this approach, all alanine-bearing spots were also included as targets to redesign.

Computational protein design

To determine the amino acid identities likely to invert the energetics of the pre- and postfusion states, target positions to redesign underwent complete in-silico saturation mutagenesis as described at the computational alanine scanning section. Subsequently, substitutions favoring the prefusion state over the postfusion state were chosen for combinatorial design through Rosetta modeling. To bias the design process towards mutations displaying preference for the prefusion state with a high energetic difference between states, the weight of each substitution was adjusted in the Rosetta energy function according to a fitness score. Our fitness score compiled the stabilization effect of one mutation in both pre- and postfusion states by subtracting the ΔΔG prefusion score from the ΔΔG postfusion score ($\Delta\Delta G_{postfusion} - \Delta\Delta G_{prefusion}$). Mutations favoring the prefusion state over the postfusion conformation were then characterized by positive fitness scores where higher values represented bigger energetic gaps between states. The fitness score was incorporated into the Rosetta score function through a residue-type constraint term derived from

the "FavorSequenceProfile" mover. Since this mover was initially created to re-weight amino acid substitutions depending on their occurrence in a multiple sequence alignment, we have replaced the original position-specific substitution matrix (PSSM) input by a fitness score matrix. To follow the PSSM format, negative fitness scores were replaced by zero and a 0.05 pseudocount was used for the log-odds scores calculation. After tuning the profile weights of each residue, "allowed" mutations at every amino acid position were defined as those with a fitness score greater than or equal to 0.7. For challenging targets such as the SARS-CoV-2 S, the threshold difference was increased to 2 to focus on the most significant substitutions. Likewise, beneficial mutations for both states were allowed only if the stabilization effect in the prefusion state was at least 4 units greater than in the postfusion state. Finally, the combinatorial sequence design was carried out by the FastDesign algorithm [240], [241]. Upon conclusion, further optimization of a specific target spot was optionally done by applying FastDesign (all amino acids allowed) on residues neighboring 6Å around the point of interest and limiting packing and minimization to a 12Å sphere. The design process was initially performed on the prefusion conformation, and the resulting sequences were modeled on the postfusion structure for energetic comparisons.

Selection of top designs

Promising designs were first sorted based on their Rosetta total energy score. Before comparison, the parent pre and postfusion structures were relaxed and energetically minimized using the same protocol as the designed models. Top candidates corresponded to designs showing a lower energy score in the prefusion state compared to the parent pre- and postfusion conformations. Analogously, the designed sequence in the postfusion state had to display a higher energy score than the parent postfusion conformation. The selected designs were then analyzed at the residue level to identify which mutations contributed more to the energetic gap between states.

In this regard, mutations were filtered according to different structural metrics, such as Rosetta per-residue energy, and the number of hydrogen bonds and van der Waals contacts. To improve the prefusion state over the postfusion conformation, mutations showing favorable metrics on the prefusion state and unfavorable values for the postfusion conformation were selected to be tested experimentally.

II. Protein expression and characterization

Protein expression

The top 3-4 RSV F, hMPV F, and SARS-CoV-2 S computational designs as well as the starting constructs hMPV F 115-BV [95], and SARS-CoV-2 S-2P [130], and the control variants RSV F DS-Cav1 [19], postfusion RSV A2 F [205], and postfusion hMPV B2 F [242] were expressed by transient transfection of FreeStyle 293-F cells (Thermo Fisher) with polyethylenimine (PEI) (Polysciences). All computationally designed variants were produced in pCAGGS plasmids encoding the sequence of interest, a C-terminal T4 fibritin trimerization motif (Foldon), and a His6-tag. The designed RSV F constructs contained residues 1-105 and 137-513, and a flexible linker replacing the furin cleavage sites and p27 peptide ("QARGSGSGR")[77]. Likewise, the designed hMPV F sequences included residues 1-95 and 103-472, a modified cleavage site "ENPRRRR", and the A185P mutation [95]. Finally, the designed SARS-CoV-2 S variants followed the semi-stabilized SARS-CoV-2 S-2P protein sequence [130], with two proline substitutions at residues 986 and 987, and a "GSAS" linker replacing the furin cleavage site. All DNA sequences were codon optimized for human expression using the online tool GenSmart Codon Optimization [243]. Cells were cultured at 37°C and 8% CO_2, and the culture supernatant was harvested on the 3rd day after transfection. Proteins were purified by nickel affinity chromatography followed by size-exclusion chromatography (SEC) in phosphate-buffered saline

(PBS) buffer pH 7.4. RSV F and hMPV F variants were SEC purified using a Superdex200 column (Cytiva) while SARS-CoV-2 S was purified with a Superdex6 column (Cytiva).

The angiotensin-converting enzyme 2 (ACE2) was expressed as an Fc-fusion [226] after transient transfection as described above. The protein was purified using a Protein A agarose gravity column (Millipore Sigma) followed by SEC using a S200 column.

Antigenic characterization

Bio-layer interferometry (BLI) was used to evaluate the structural and antigenic conservation of prefusion-specific epitopes. The prefusion-specific binders used for this purpose were the antibodies D25 [58], [224] (Thermo Fisher) and AM14 [84], [224] (Cambridge Biologics) for RSV F, MPE8 [225] and 465 [94] for hMPV, and the angiotensin-converting enzyme 2 (ACE2) [226] for SARS-CoV-2 S. All binders were immobilized on Protein A sensors (GatorBio) at a concentration of 15 nM (RSV F, and hMPV F), or 40 nM (SARS-CoV-2 S). Binding against expressed designs was tested at eight different protein concentrations starting from 200 nM (RSV F, and hMPV F) or 400 nM (SARS-CoV-2 S) and decreasing by 1:2 dilutions. All solutions had a final volume of 200 µL/well using PBS buffer supplemented with 0.02% tween-20 and 0.1mg/mL bovine serum albumin (BLI buffer). Biosensor tips were equilibrated for 20 min in BLI buffer before loading of binders. Loading was then carried out for 180s followed by a baseline correction of 120s. Subsequently, association and dissociation between the binders and designed variants were allowed for 180s each. To validate the BLI results, binding with previous prefusion-stabilized proteins such as the RSV F DS-Cav1[19], hMPV F 115-BV[95], and SARS-CoV-2 S-2P [130] was used as positive controls, and binding with the postfusion constructs RSV A2 F [205], and hMPV B2 F [242] was used as negative controls. All assays were performed using a GatorPrime biolayer interferometry instrument (GatorBio) at a temperature of 30°C and frequency

of 10 Hz. Data analysis was completed with the GatorOne software 1.7.28, using a global association model 1:1.

Thermal stability

The thermal stability of the expressed variants was assessed by differential scanning fluorimetry (DSF). The samples were prepared by creating a solution containing 1.2µL SYPRO orange fluorescent dye (Thermo Fisher) with 3 µL of 100mM $MgCl_2$, 3 µL of 1M KCl, and 3 µL of 1M Tris (pH 7.4). The final solution volume was 60 µL with a protein concentration of 3.5 µM. A negative sample with no protein was also prepared as background control. All measurements were performed by triplicates using 20 µL of sample. The data was collected with a qPCR instrument (CFX Connect, BioRad) and a temperature ramp from 25 to 90°C with 0.5°C increments. The melting temperature was determined based on the lowest point of the negative first derivative of the SYPRO Orange signal.

Antigenic preservation in variants displaying the highest melting temperature was further evaluated after one hour incubation at 55 and 60°C. This process was carried out in a thermocycler with heated lid (T100, BioRad). The conservation of the antigenic sites was determined by binding to prefusion-specific binders as described at the antigenic characterization section. Conversion to the postfusion state was also evaluated for the RSV F variant R-1b using the postfusion-specific antibody 131-2A[229] (Millipore Sigma) and the postfusion RSV A2 F[205] as positive control.

Negative-stain electron microscopy

Purified R-1b, M-104, and Spk-M (buffer-exchanged into 50 mM Tris pH 7.5 and 100 mM NaCl) were applied on carbon-coated copper grids (400 mesh, Electron Microscopy Sciences) using 5 µL of protein solution (10 µg/mL) for 3 min. The grid was washed in water twice and then stained with 0.75% uranyl formate (R-1b) or Nano-W (Nanoprobes) (M-104, and Spk-M) for

1 min. Negative-stain electron micrographs were acquired using a JEOL JEM1011 transmission electron microscope equipped with a high-contrast 2K-by-2K AMT mid-mount digital camera.

X-rays crystallization

The trimeric R-1b and M-104 proteins were concentrated to 14 mg/mL and 13.9 mg/mL, respectively, and crystallization trials were prepared on a TTP LabTech Mosquito Robot in sitting-drop MRC-2 plates (Hampton Research) using several commercially available crystallization screens. R-1b crystals were obtained in the Index HT (Hampton Research) in condition H6 (0.2 M Sodium formate, 20% w/v PEG 3,350), while M-104 crystals were obtained in the Crystal screen (Hampton research) in condition C10 (0.1 M Sodium acetate trihydrate pH 4.6, 2.0 M Sodium formate). Crystals were harvested and cryo-protected with 30% glycerol in the mother liquor before being flash frozen in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source SER-CAT beamLine 21-ID-D. Data were indexed and scaled using XDS[244]. A molecular replacement solution was obtained in Phaser [232] using the prefusion RSV F SC-TM structure (PDB 5c6b)[77] or the prefusion hMPV F 115-BV (PDB 5wb0)[95]. The crystal structures were completed by manually building in COOT [245], followed by subsequent rounds of manual rebuilding and refinement in Phenix [232]. The data collection and refinement statistics are shown in Table 3.S5.

Cryo-electron microscopy

Spk-M cryo-EM density data was obtained by the Eyring Materials Center at Arizona State University (ASU). Purified protein was diluted to a concentration of 0.35 mg/mL in Tris-buffered saline (TBS) and applied to plasma-cleaned CF-300 2/1 grids before being blotted for 3 seconds in a Vitrobot Mark IV (Thermo Fisher) and plunge frozen into liquid ethane. 3,257 micrographs were collected from a single grid using a FEI Titan Krios (Thermo Fisher) equipped with a K2

summit direct electron detector (Gatan, Pleasantville CA.). Data collection was automated with SerialEM with a defocus range of -0.8 to -2.6 um in counting mode on the camera with a 0.2 second frame rate over 8 seconds and total dose of 58.24 electron per angstrom squared. Images were processed using cryoSPARC V3.3.2 [246] (Fig. S11). Micrographs were patch motion corrected. After particle extractions, the blob picker was used, and 1,551,079 and picking was manually adjusted to reduce "blobs" to 1,394,889 particles. After 2D classifications the first 4 classes were ab initio reconstructed and heterogeneously refined. The most populated map was refined with homogenous refinement in cryoSPARC, resulting in a 3.72 Å map. The map was further processed in DeepEM[247]. The resulting final map was aligned with a previously published SARS-CoV-2 S with one RBD domain up (PDB ID 6vyb[211]) using UCSF Chimera-1.15 [248]. Mutations and coordinate fitting were done manually using COOT [245] and structure optimization was achieved by iterative refinement using Phenix real space refinement [232] and COOT. The model and map statistics are presented in Table 3.S6.

III. Animal studies (RSV only)

Mouse immunization

All animal experiments were performed in accordance with the guidelines and approved protocols by the Institutional Animal Care and Use Committee at University of Georgia, Athens, USA. Six-to-eight-week female BALB/c mice were purchased and housed in microisolator cages in the animal facility at University of Georgia. Food and water were provided ad libitum. After acclimation period, 5 mice per group were intramuscularly (i.m.) inoculated with total 100 µL of two different doses (2 µg and 0.2 µg) of either purified DS-Cav1 or R-1b protein with AddaVax adjuvant (50% v/v) or PBS at weeks 0 and 4 (Prime and Boost Vaccination protocol) (Fig 3.4.A).

Bleeds were collected from tail vein pre- and post- immunization (3, 7, and 13 weeks) and sera were analyzed by ELISA and neutralization assay.

Measurement of IgG response by ELISA

Medium binding 96 wells microplates (Greiner Bio-One) were coated with 50 µL per well of DS-Cav1 or R-1b protein at 2 µg/mL at 4°C overnight. Plates were washed in PBS/0.05% Tween 20 (Promega) and then blocked with blocking buffer solution (PBS/0.05% Tween 20 /3% non-fat milk (AmericanBio) /0.5% Bovine Serum Albumin (Sigma)) at room temperature for 2 hours. Pooled serum from each group of mice pre-and post-different stages of immunization or control were inactivated at 56°C per 1 hour for subsequent serial dilution in blocking buffer. 100 µL per well of inactivated diluted sera were incubated in triplicate at room temperature for 2 hours. Subsequently, three washes were performed, and plates were incubated with peroxidase-labeled goat anti-mouse IgG (1:3500) (SeraCare) diluted in blocking buffer. After one-hour incubation at room temperature, plates were washed and TMB substrate working Solution (Vector Laboratories) was added. After 10 min at room temperature, the reaction was stopped by adding 50 µL per well of Stop Solution for TMB ELISA (1N H_2SO_4). Plates were then read on Cytation7 imaging Reader (BioTek) at 450 nm.

RSV neutralization assays

Pooled serum samples from mice in each immunization group (5 animals/group) after vaccination and boost (13 weeks after the beginning of the experiment or prime vaccination/ 9 weeks after Prime and Boost vaccination) were diluted in Opti-MEM media (Thermofisher) in serial 3-fold dilutions. Antibody 101F (provided by Jarrod J. Mousa) was used as a positive control for virus neutralization starting at 20 µg/mL. Further, dilutions were mixed with 120 focus-forming units (FFU) of RSV A virus (strain: rA2 line19F) (kindly provided by Dr. Martin Moore) and

incubated for 1 hour at room temperature. Subsequently, RSV and sera/antibody dilutions were added to Vero E6 (ATCC) monolayer (105 cells/well) in triplicate and incubated for one hour at 37°C, gently rocking the plate every 15 minutes. Following the incubation, cell monolayers were covered with an overlay of 0.75% methylcellulose dissolved in Opti-MEM with 2% Fetal Bovine Serum (FBS) (Thermofisher) and incubated at 37 °C, 5% CO2. After four days, the overlay was removed, and wells fixed with neutral buffered formalin 10% (Sigma) at room temperature for 30 minutes. Further, fixed monolayers were washed with water and dried at room temperature. An FFU assay was performed to identify the percentage of RSV neutralization. Briefly, wells were washed gently with PBS-0.05% Tween-20 (Promega) and incubated per one hour with anti-RSV polyclonal antibody (EMD Millipore) diluted 1:500 in dilution buffer [5% Non-fat dry milk (AmericanBio) in PBS-0.05% Tween-20]. Plates were washed three times with PBS-0.05% Tween-20, followed by 30 minutes incubation of secondary antibody HRP conjugate rabbit anti-goat IgG (Millipore Sigma) diluted 1:500 in dilution buffer. After incubation, wells were washed, and TMB Peroxidase substrate (Vector Laboratories) was added for 1 hour at room temperature. The visualized foci per well were counted under an inverted microscope.

Data and Materials Availability

Atomic coordinates of the reported structures and cryo-EM map were deposited in the Protein Data Bank under accession codes 7TN1, 8E15, and 8FEZ, and in the Electron Microscopy Data Bank under accession code EMD-29035. R-1b, M-104, and Spk-M plasmids are available from EMS under a material transfer agreement with the University of Georgia. Rosetta is available through licensing https://www.rosettacommons.org. Scripts for generating designs will be available on https://github.com/strauchlab/stabilization upon publication.

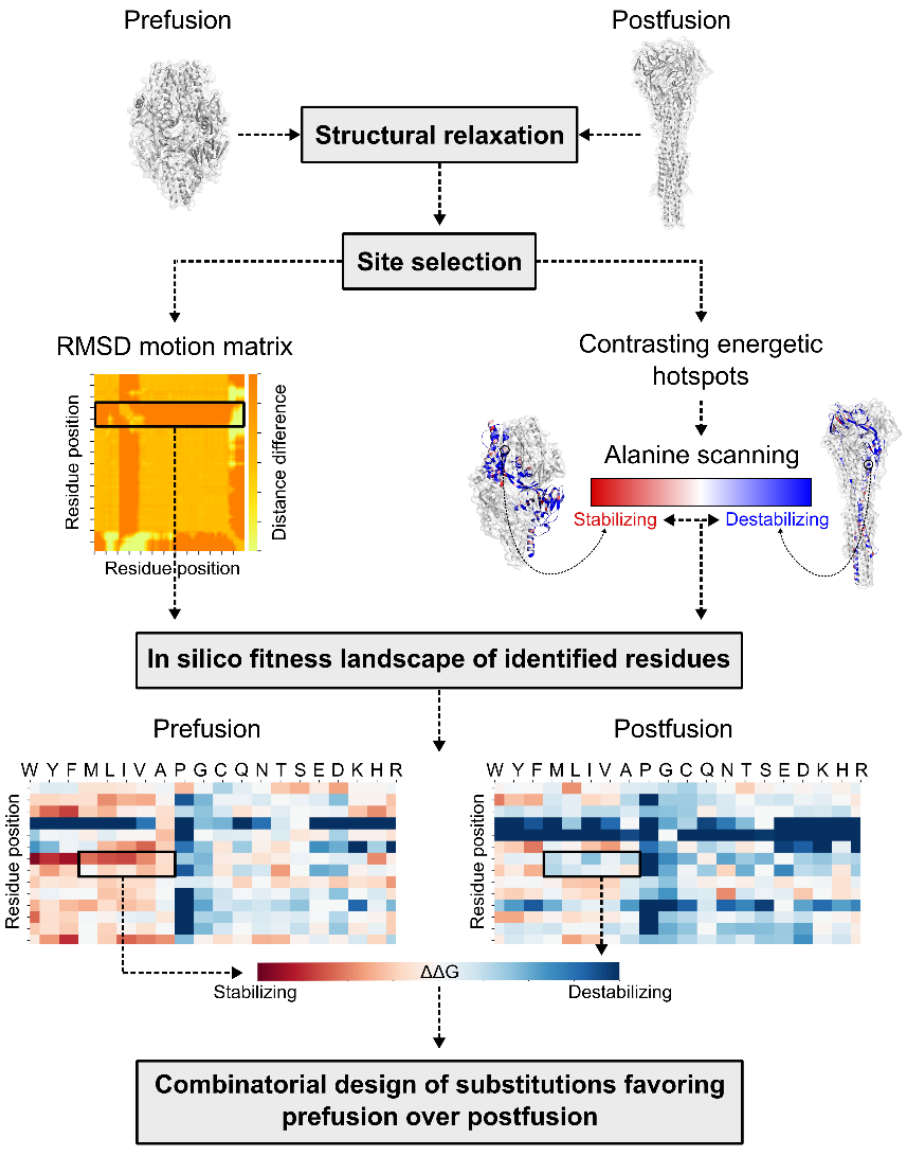


Figure 3. 1 Computational design overview. Step-by-step diagram illustrating the key components of redesigning viral class I fusion proteins. Both pre- and postfusion conformations are input into the pipeline to identify substitutions that favor the prefusion state over the postfusion conformation. RMSD=Root mean square deviation; ΔΔG=change in delta Gibbs free energy.

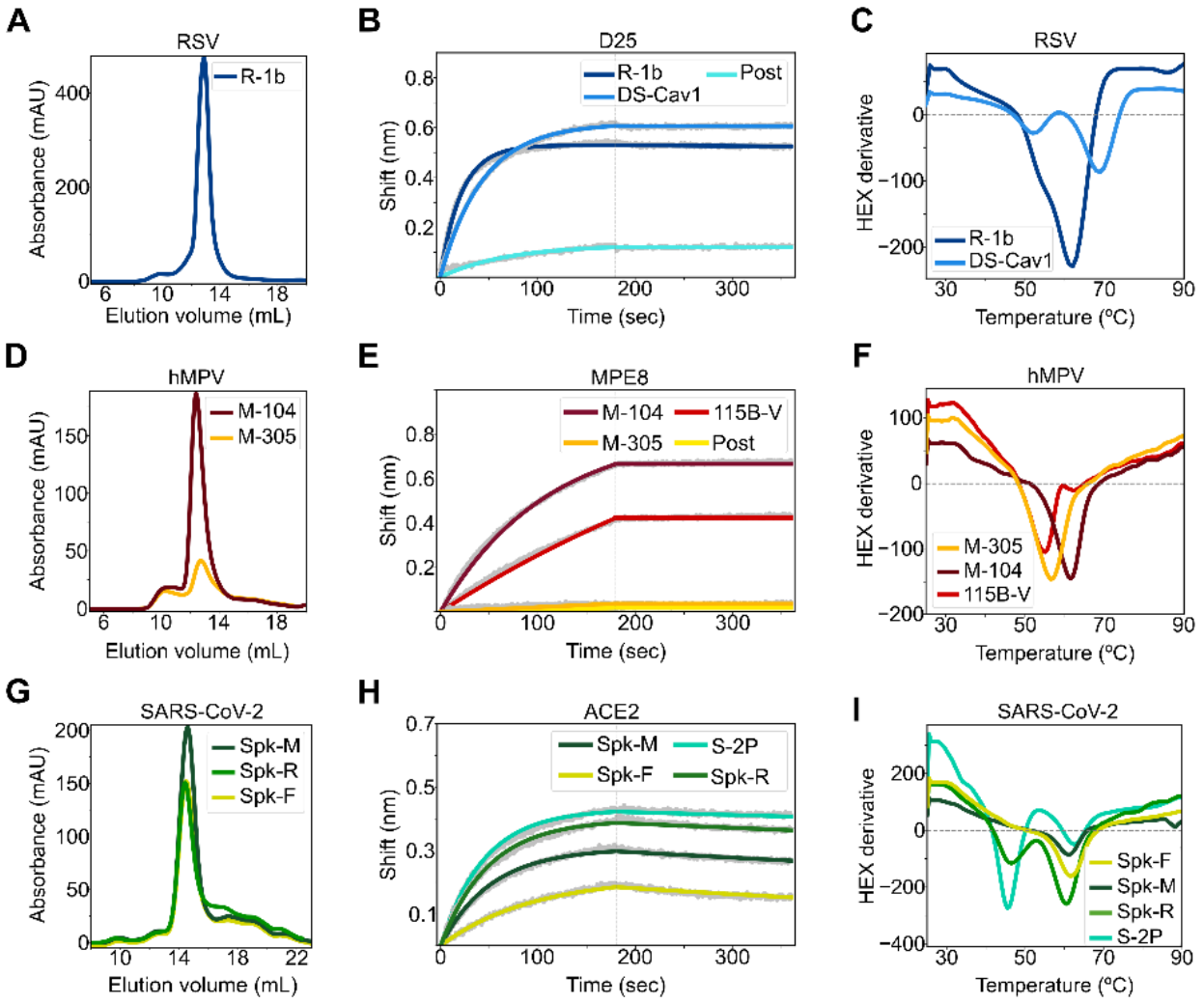


Figure 3. 2 Biochemical characterization of designed variants. (A) Size-exclusion chromatography (SEC) of monodispersed RSV F designs. **(B)** Binding of design R-1b to the prefusion-specific antibody D25 compared to the clinical candidate DS-Cav1 and the postfusion RSV A2 F (post). **(C)** Differential scanning fluorimetry (DSF) of design R-1b and the clinical candidate DS-Cav1. DS-Cav1 was used to compare the stability of R-1b as the parental sequence of the latter is not prefusion-stabilized. **(D)** SEC of monodispersed hMPV F designs. **(E)** Binding of designed hMPV F variants to the prefusion-specific antibody MPE8 compared to their parent prefusion construct 115B-V and the postfusion hMPV B2 F (post). **(F)** DSF of designed hMPV F variants and their parent prefusion construct 115B-V. **(G)** SEC of monodispersed SARS-CoV-2 S

designs. **(H)** Binding of designed SARS-CoV-2 S variants to ACE2 compared to their parent prefusion construct S-2P. **(I)** DSF of designed SARS-CoV-2 S variants and their parent prefusion construct S-2P. Antibody binding assays show in grey the raw data, in colors the fitted curves, and in dotted lines the end of the association time. Binding constants are shown in Tables 3.S2, S3, and S4.

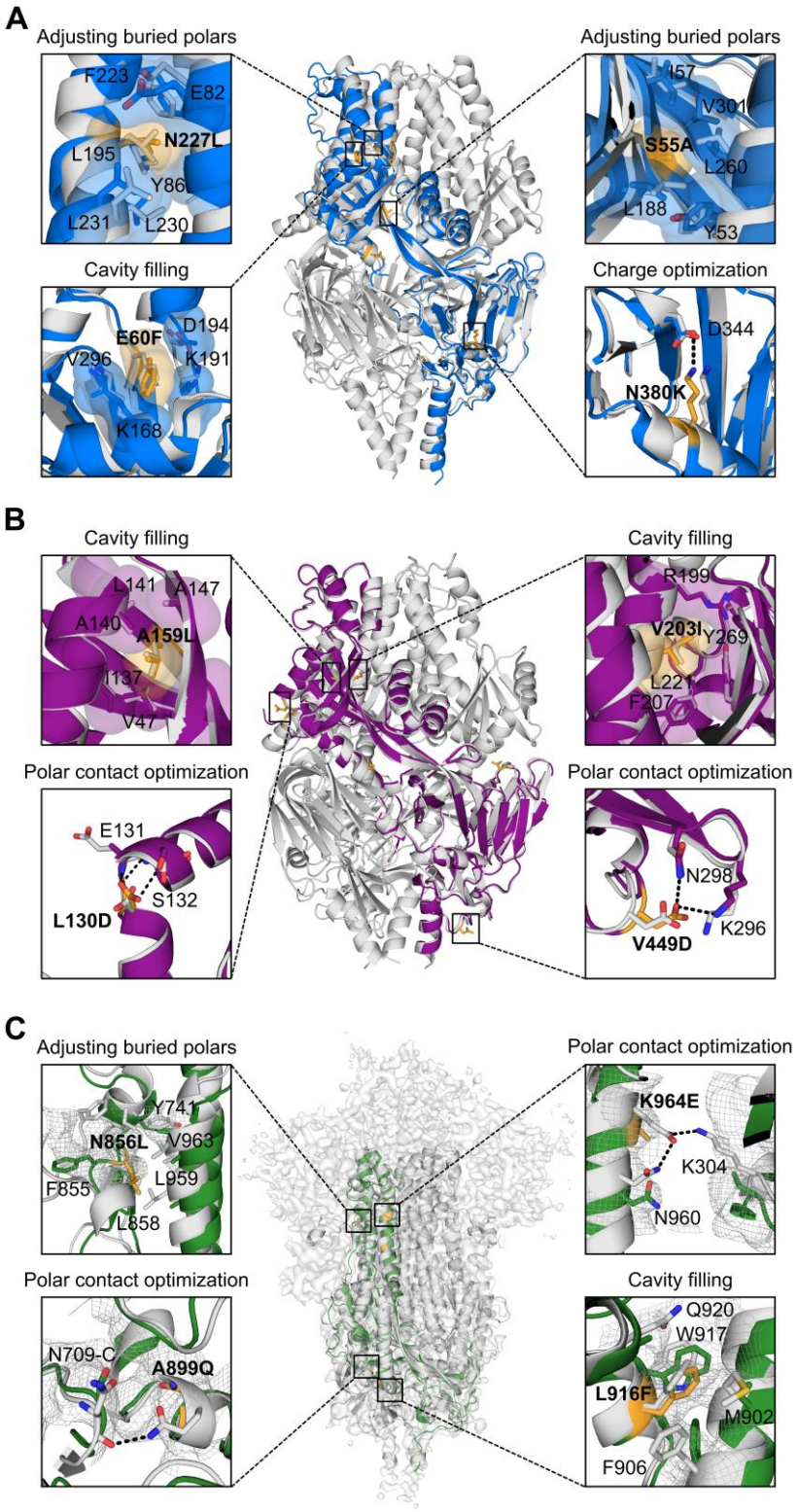


Figure 3. 3 Exemplary stabilizing substitutions of leading designs. (A) R-1b. **(B)** M-104. **(C)** S2 subunit of Spk-M with cryo-EM map. The computational model of each protein is displayed as

a trimeric structure in grey while the crystal structures and cryo-EM reconstruction model are displayed as monomeric structures in blue (RSV), magenta (hMPV), or green (SARS-CoV-2). Each panel shows a magnified view of selected stabilizing substitutions, featured in yellow sticks, aligned with their computational model. Residues involved in packing changes are displayed with translucent molecular surfaces, and black dotted lines represent hydrogen bonds or salt bridges. As density is missing in the overall map to assign the precise location of the side chains, we displayed existing density as mesh representation to compare agreement with the computational model as stabilized regions have more density than the remainder of the map. The stabilization mechanism of all designed substitutions is presented in Table 3.S1.

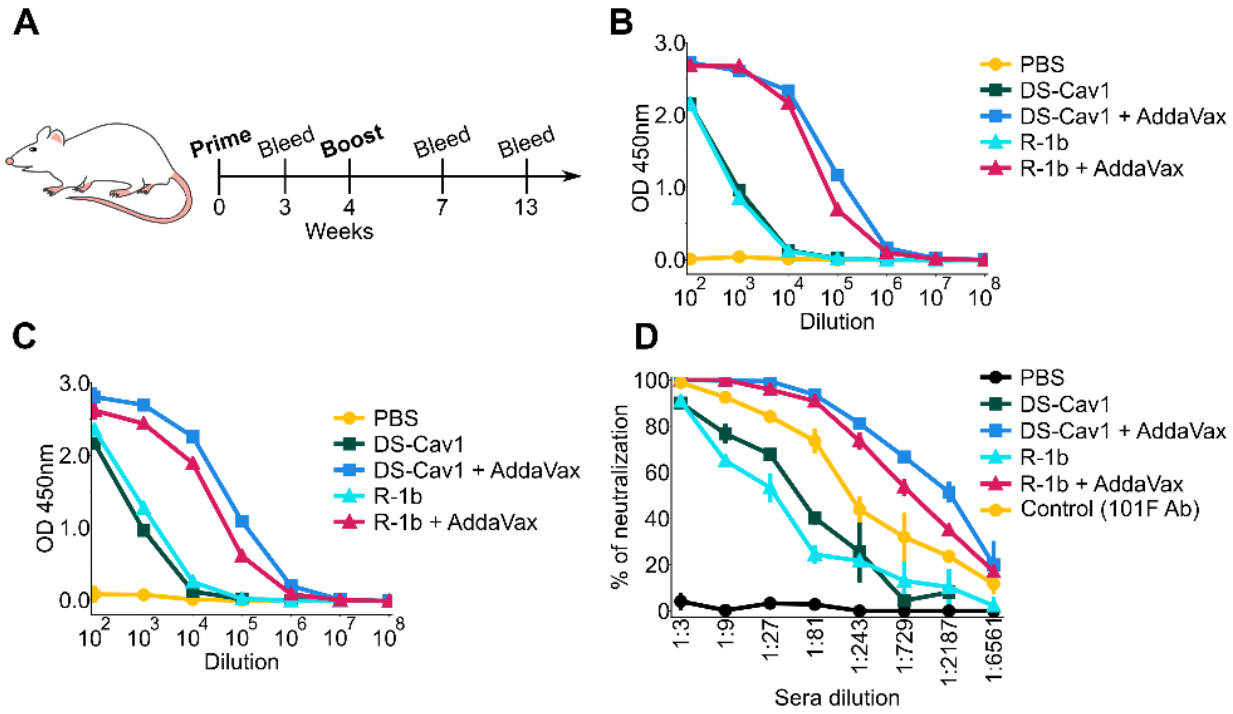


Figure 3. 4 Immunogenicity assessment of R-1b in a mouse model using 0.2ug doses. (A) Schematic diagram of vaccination study schedule. **(B)** Serum RSV-specific IgG measured by ELISA three weeks post-boost. **(C)** Serum RSV-specific IgG measured by ELISA nine weeks post-boost. **(D)** Serum neutralization titers determined using RSV A (rA2 strain L19F) and sera from mice nine weeks post-boost**.** Vertical lines represent the standard deviation of three repetitions using pooled serum samples from mice in each immunization group (5 animals/group).

Supplementary Information

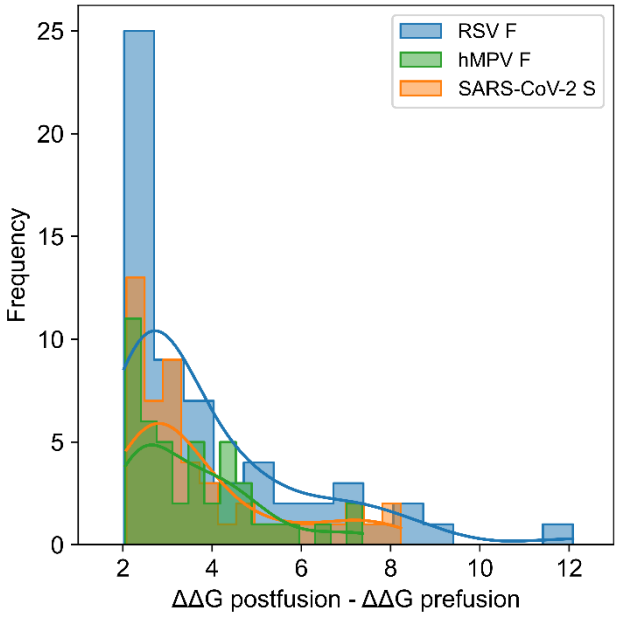


Figure 3.S 1 Gibbs free energy changes (ΔΔG) after alanine scanning in the pre- and postfusion structures of RSV F, hMPV F, and SARS-CoV-2 S proteins. Mutations favoring the prefusion state over the postfusion conformation are characterized by positive scores where higher values represent bigger energetic gaps between states. A solid line represents the kernel density estimate of each histogram. On display are presented only significant ΔΔG changes as defined by a differential of at least 2 units.

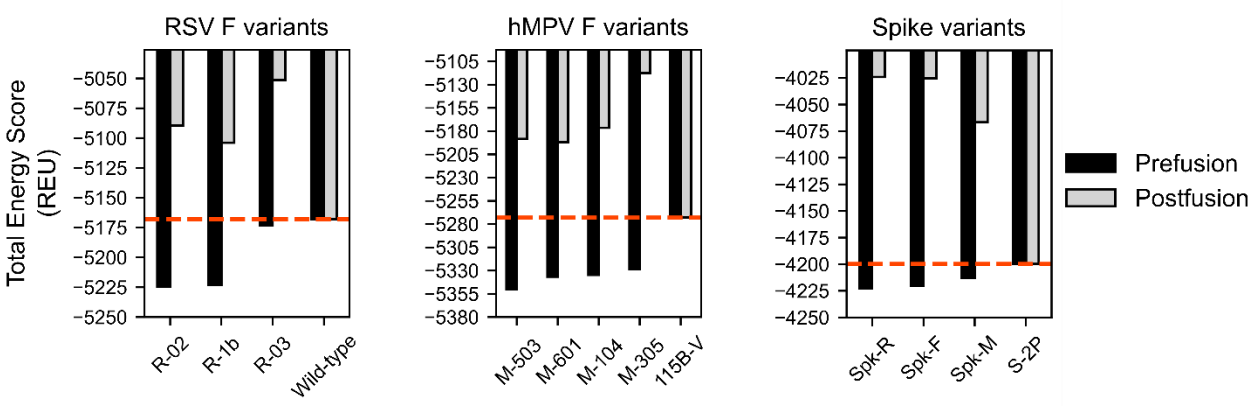


Figure 3.S 2 Energy comparison between prefusion base constructs and designed variants. In black is represented the prefusion conformation while grey depicts the postfusion conformation. An orange dotted line highlights the energy of the starting sequences in Rosetta Energy Units (REU). To compare the energy gain or loss of each variant, all prefusion energies were normalized to the postfusion state using the ratio postfusion-energy_base_construct/prefusion-energy_base_construct. The energy gap between states is shown in Table 3.S1.

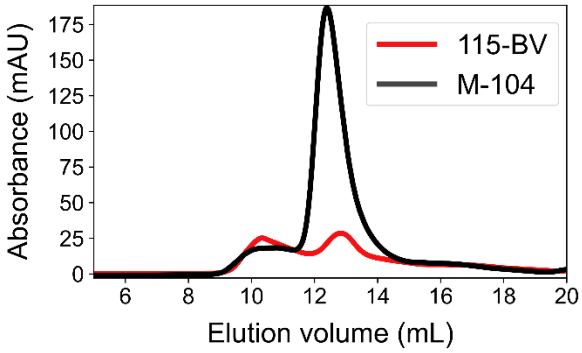


Figure 3.S 3 Size-exclusion chromatography of design M-104 in comparison to its parent construct 115B-V.

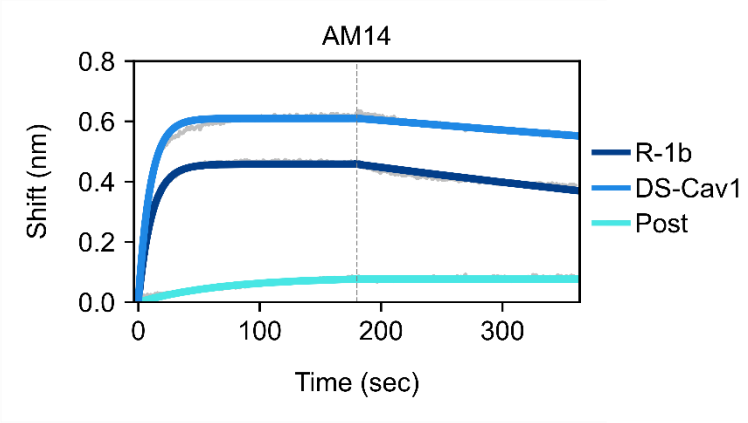


Figure 3.S 4 Binding of RSV F variants to the prefusion-specific antibody AM14. In grey is shown the raw data and in colors the fitted curves. A dotted vertical line represents the end of the association time. Binding constants are shown in Table 3.S2. "Post" stands for postfusion RSV F A2.

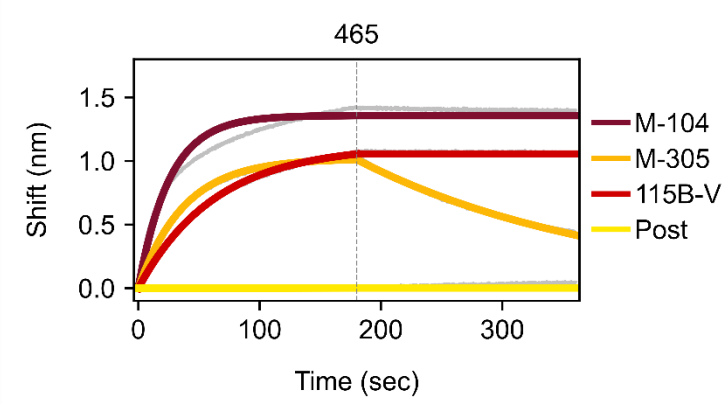


Figure 3.S 5 Binding of hMPV F variants to the prefusion-specific antibody 465. In grey is shown the raw data and in colors the fitted curves. A dotted vertical line represents the end of the association time. Binding constants are shown in Table 3.S3. "Post" stands for postfusion hMPV B2 F.

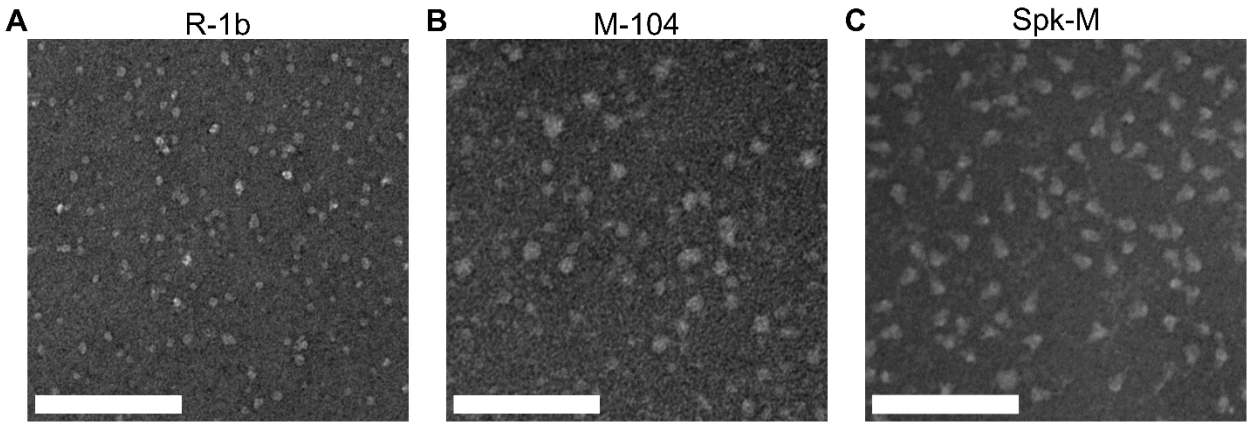


Figure 3.S 6 Negative stain-electron microscopy of designs (A) R-1b, **(B)** M-104, and **(C)** Spk-M. Scale bar: 100 nm.

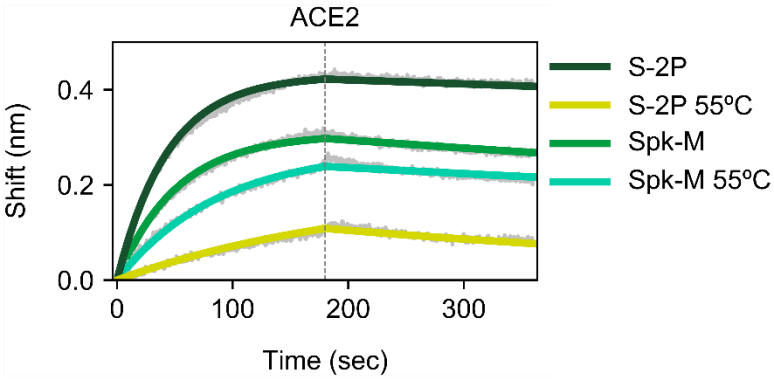


Figure 3.S 7 Binding of design Spk-M and its base construct S-2P to ACE2 after heat treatment. In grey is shown the raw data and in colors the fitted curves. A dotted vertical line represents the end of the association time. Binding constants are shown in Table 3.S4.

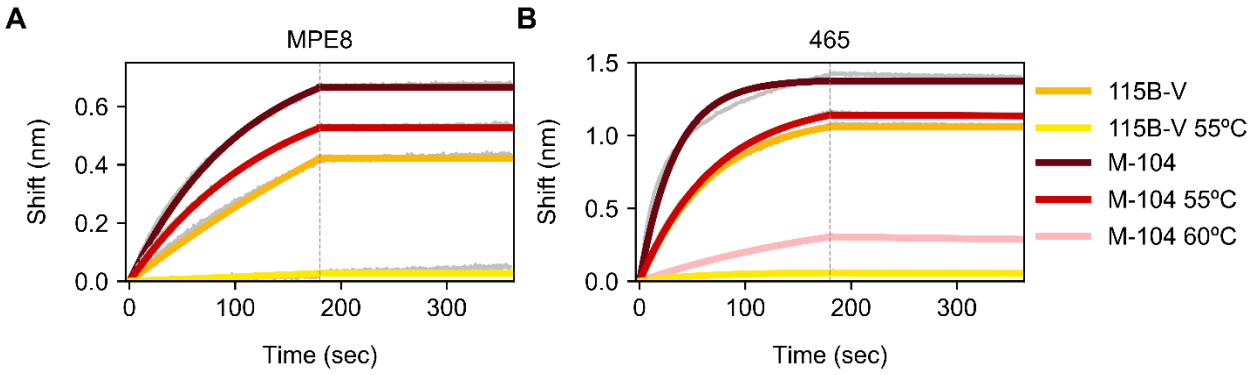


Figure 3.S 8 Binding of design M-104 and its base construct 115B-V to prefusion-specific antibodies after heat treatment. (A) Binding to MPE8. **(B)** Binding to 465. In grey is shown the raw data and in colors the fitted curves. A dotted vertical line represents the end of the association time. Binding constants are shown in Table 3.S3.

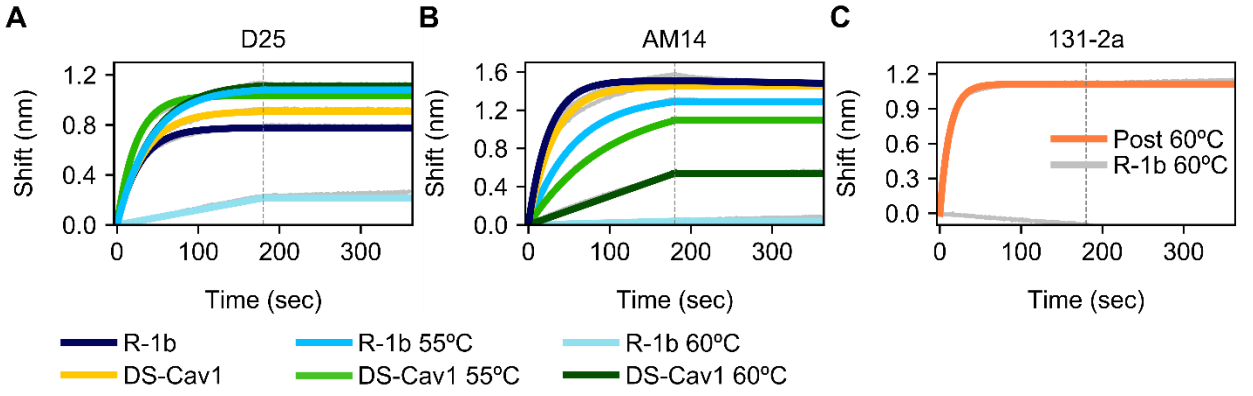


Figure 3.S 9 Binding of RSV F variants to prefusion- and postfusion-specific antibodies after heat treatment. (A) Binding to D25. **(B))** Binding to AM14. **(C)** Binding to 131-2A. "Post" stands for postfusion RSV A2 F. As R-1b does not bind to 131-2A after heating at 60°C, only the raw data is displayed for this protein. In grey is shown the raw data and in colors the fitted curves. A dotted vertical line represents the end of the association time. Binding constants are shown in Table 3.S2.

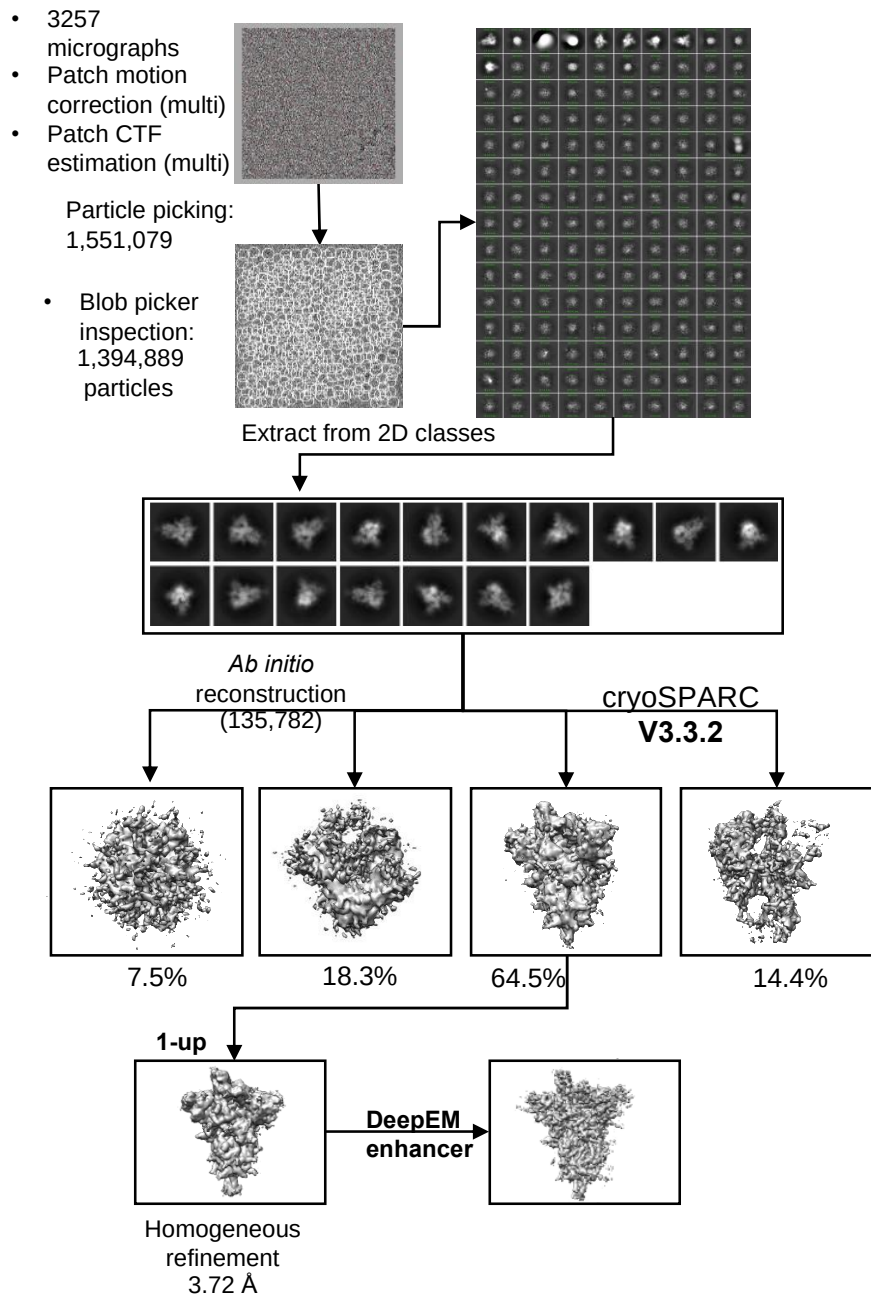


Figure 3.S 10 Summary of cryo-electron microscopy data for design Spk-M. Micrographs were processed in cryoSPARC V3.3.2 and final refinement in DeepEM enhancer.

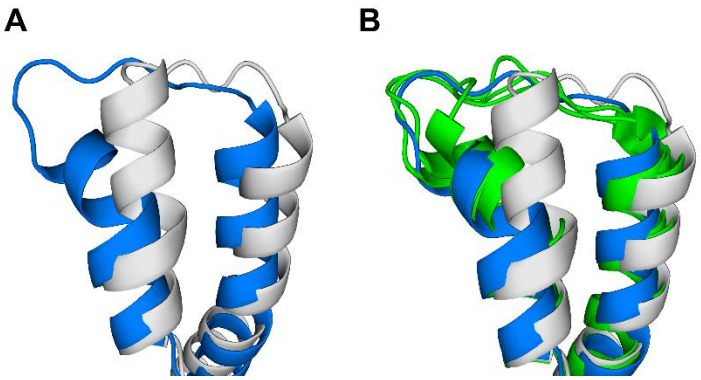


Figure 3.S 11 Structural alignment between head residues of different RSV F variants. **(A)** Comparison between design R-1b (blue) and parent construct PDB 5w23 (grey). **(B)** Comparison shown in A plus three different DS-Cav1 crystal structures in green (PDB 4mmu, 5ea4, 5k6c). On display are residues 195-227.

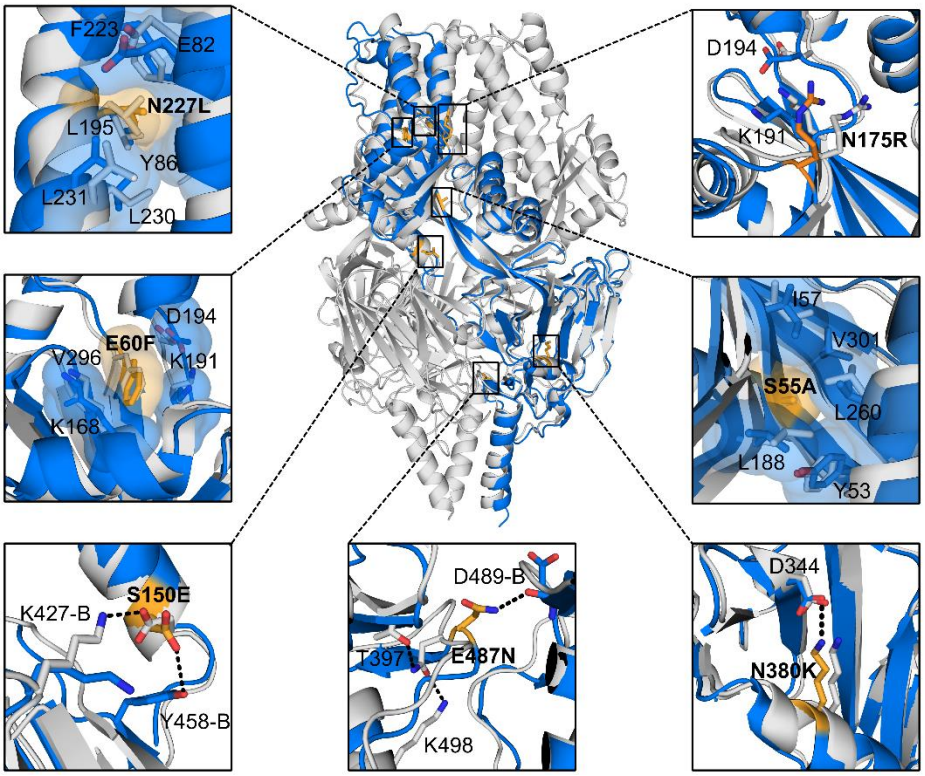


Figure 3.S 12 Atomic interactions of all substitutions introduced in design R-1b compared with a computational model. The computational model of the protein is displayed as a trimeric structure in grey, while the crystal structure is displayed as a monomeric structure in blue, with introduced mutations in yellow. Each panel shows a magnified view of the atomic interactions involving each substitution (in yellow sticks), aligned with their computational model. Residues contributing to packing changes are displayed with translucent molecular surfaces, and black dotted lines represent hydrogen bonds or salt bridges.

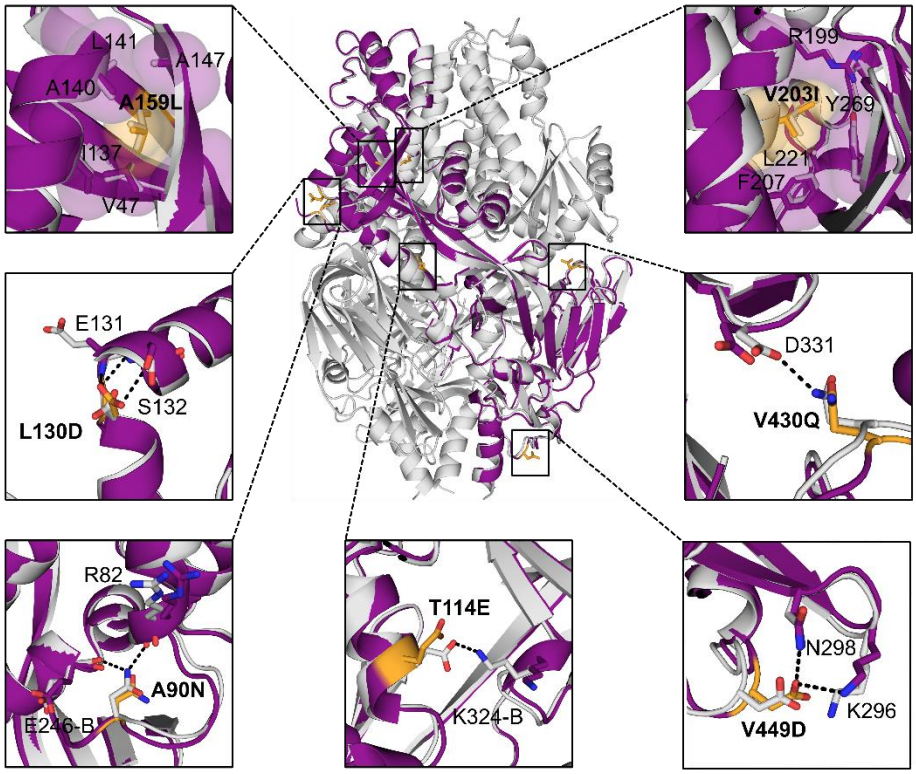


Figure 3.S 13 Atomic interactions of all substitutions introduced in design M-104 compared with a computational model. The computational model of the protein is displayed as a trimeric structure in grey, while the crystal structure is displayed as a monomeric structure in purple with introduced mutations in yellow. Each panel shows a magnified view of the atomic interactions involving each substitution (in yellow sticks), aligned with their computational model. Residues contributing to packing changes are displayed with translucent molecular surfaces, and black dotted lines represent hydrogen bonds or salt bridges.

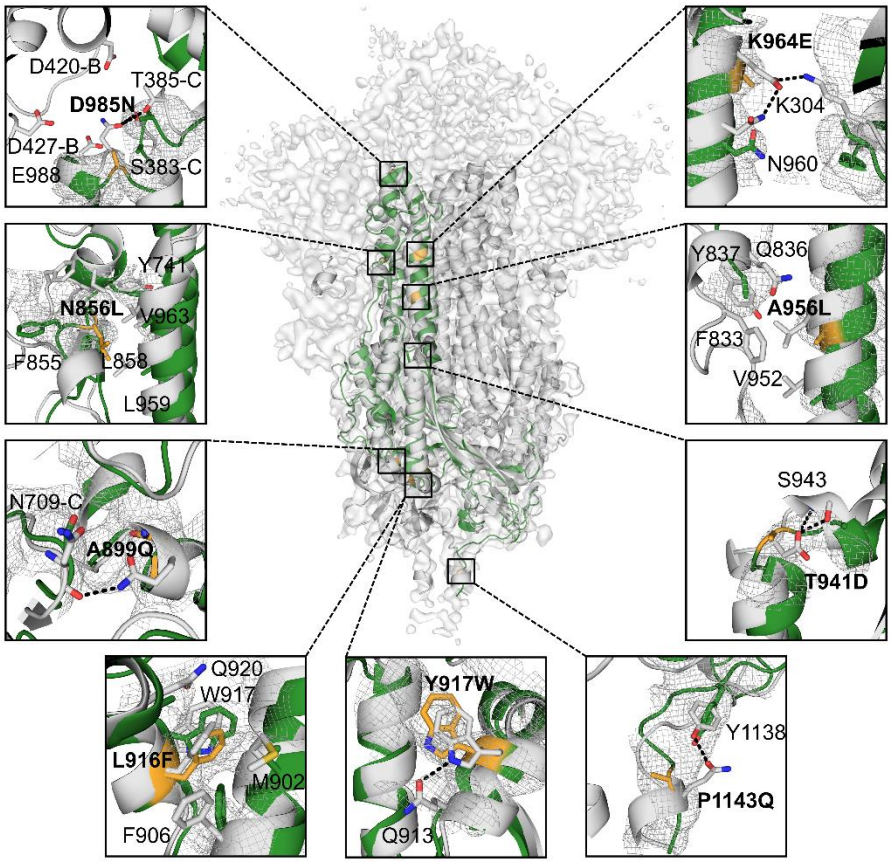


Figure 3.S 14 Predicted atomic interactions of all designed substitutions introduced in the S2 subunit of design Spk-M. The computational model of the protein is displayed as a trimeric structure in grey, while the cryo-EM reconstruction model is displayed as a monomeric structure in green with introduced mutations in yellow. The Spk-M cryo-EM map is shown as a translucent surface in grey. Each panel shows a magnified view of the atomic interactions involving each substitution (in yellow sticks), aligned with their computational model. As density is missing in the overall map to assign the precise location of the side chains, we displayed existing density as a mesh representation to compare agreement with the computational model. Black dotted lines represent hydrogen bonds or salt bridges.

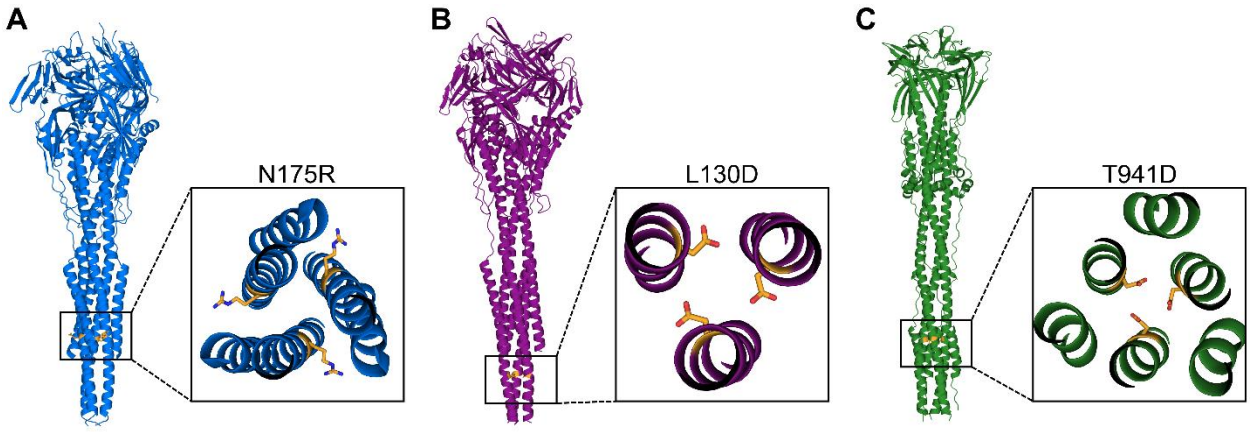


Figure 3.S 15 Computational models of postfusion destabilizing substitutions introduced in (A) R-1b. **(B)** M-104. **(C)** Spk-M. Each panel shows a magnified view of the predicted rotamer configuration of each mutation. All substitutions are represented in yellow sticks.

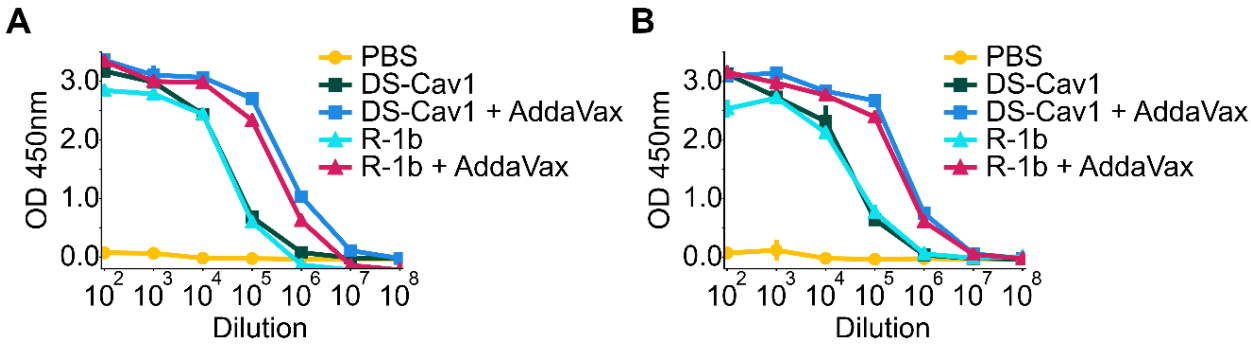


Figure 3.S 16 Immunogenicity assessment of design R-1b in mice using 2ug doses. A) Serum RSV-specific IgG measured by ELISA three weeks post-boost. **(B)** Serum RSV-specific IgG measured by ELISA nine weeks post-boost.

Table 3.S 1 Energetic differences between pre- and postfusion states, stabilizing mutations, and thermal stability of designed fusion proteins.

Virus	*Protein variants*	Energy gap (REU) *Pre -vs- Postfusion*	Stabilizing mutations						Melting temperature (°C)
			Cavity filling	Inter-protomer polar interactions	Intra-protomer polar interactions	Reduction of unsatisfied polars	Decrease charge repulsion	Postfusion destabilizing	
RSV	Base construct (WT)	0*	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	R-1b	119.1	E60F	S150E	N380K	S55A, N227L	E487N	N175R	62
	R-02	135.1	E60F, S150L	N/A	N380K	S55V, N227L	E487V	N175R	N/A
	R-03	121.7	E60F, S150L	N/A	N380K	S55L, N227F	E487V	N175R	N/A
hMPV	Base construct (115B-V)	0*	N/A	N/A	N/A	N/A	N/A	N/A	54.8
	M-104	158.2	A159L, V203I	A90N, T114E	V430Q, V449D	N/A	N/A	L130D	61.5
	M-305	210.7	A159I	A90N	G106R, G277D, A314K, V449D	N/A	E453P	L130D	56.7
	M-503	161.8	G106W, A107F, A159L, V162I, V203I	T114E	V430Q, V449D	N/A	N/A	L130D	N/A
	M-601	144.4	A159L, V191I	T114E, D209E, A216R	V430Q, V449D	S149T	N/A	L130D	N/A

SARS-CoV-2	Base construct (S-2P)	0*	N/A	N/A	N/A	N/A	N/A	N/A	45.7[a],62.5[b]
	Spk-M	146.1	L916F, Y917W, A956L	A899Q	K964E, P1143Q	N856L	D985N	T941D	61
	Spk-F	194.9	A956L, A1016I	E990R	G769K, P1143Q	N/A	N/A	T941D	61.5
	Spk-R	198.4	A1016I	E990R	G744T, G769K, N955D, P1143N	N/A	D985N	T941D	46.5[a], 60.3[b]

*The base construct's pre- and postfusion energies are normalized.

N/A = not applicable

a) First apparent melting temperature

b) Second apparent melting temperature

Table 3.S 2 Binding kinetics of RSV F variants obtained by biolayer interferometry

Protein Variants	Assay temperature (°C)	Antibodies								
		D25			AM14			131-2a		
		koff(1/s)	kon(1/Ms)	KD(M)	koff(1/s)	kon(1/Ms)	KD(M)	koff(1/s)	kon(1/Ms)	KD(M)
R-1b	RT	N/A	1.79E+05 1.51E+05 1.44E+05	N/A	1.24E-04 1.01E-04 1.11E-04	2.19E+05 2.03E+05 2.27E+05	5.68E-10 4.97E-10 4.89E-10	4.00E-04 3.10E-04 3.53E-04	3.37E+04 4.79E+04 4.75E+04	1.19E-08 6.47E-09 7.43E-09
	55	N/A	1.56E+05 1.20E+05 1.21E+05	N/A	N/A 1.15E-05 1.31E-05	8.87E+04 8.63E+04 9.91E+04	N/A 1.34E-10 1.33E-10	N/A	N/A	N/A
	60	N/A	92.7 83.8 54.6	N/A	N.B	N.B	N.B	N.B	N.B	N.B
DS-Cav1	RT	N/A	1.21E+05 1.87E+05 1.49E+05	N/A	N/A	1.61E+05 2.04E+05 1.81E+05	N/A	0.001 3.17E-04 3.07E-04	8.29E+03 9.19E+03 1.76E+04	1.33E-07 3.44E-08 1.74E-08
	55	N/A	2.13E+05 1.60E+05 1.42E+05	N/A	N/A	5.03E+04 5.85E+04 5.01E+04	N/A	N/A	N/A	N/A
	60	N/A	1.15E+05 1.35E+05 1.32E+05	N/A	N/A	3.02E+04 1.31E+04 5.39E+03	N/A	N/A	N/A	N/A
RSV A2 F (postfusion)	RT	N.B	N.B	N.B	N.B	N.B	N.B	N/A	4.00E+05 4.09E+05 4.85E+05	N/A
	60	N/A	N/A	N/A	N/A	N/A	N/A	N/A	5.22E+05 4.82E+05 4.52E+05	N/A

RT= room temperature; N.B = no binding; N/A = not applicable

Table 3.S 3 Binding kinetics of hMPV F variants obtained by biolayer interferometry

Protein Variants	Assay temperature (°C)	Antibodies					
		MPE8			465		
		koff(1/s)	kon(1/Ms)	KD(M)	koff(1/s)	kon(1/Ms)	KD(M)
M-104	RT	N/A	4.01E+04 3.24E+04 4.21E+04	N/A	N/A	1.94E+05 2.00E+05 1.93E+05	N/A
	55	N/A	3.66E+04 2.17E+04 2.51E+04	N/A	6.60E-05 6.85E-05 4.25E-05	7.32E+04 7.21E+04 7.18E+04	9.02E-10 9.51E-10 5.92E-10
	60	N.B	N.B	N.B	2.29E-04 N/A 2.62E-04	2.29E+04 2.29E+04 2.56E+04	1.00E-08 N/A 1.02E-08
M-305	RT	N.B	N.B	N.B	4.91E-03 4.31E-03 4.86E-03	1.08E+05 1.19E+05 1.08E+05	4.52E-08 3.61E-08 4.51E-08
115B-V	RT	N/A	1.45E+04 1.52E+04 1.09E+04	N/A	N/A	7.99E+04 7.91E+04 7.79E+04	N/A

	55	N.B	N.B	N.B	N.B	N.B	N.B
hMPV B2 F (postfusion)	RT	N.B	N.B	N.B	N.B	N.B	N.B

RT= room temperature; N.B = no binding; N/A = not applicable

Table 3.S 4 Binding kinetics of SARS-CoV-2 S variants obtained by biolayer interferometry

Protein Variants	Assay temperature (°C)	Prefusion binder		
		ACE2		
		koff(1/s)	kon(1/Ms)	KD(M)
Spk-M	RT	5.69E-04	4.73E+04	1.20E-08
		5.54E-04	4.89E+04	1.13E-08
		5.82E-04	4.71E+04	1.24E-08
	55	7.12E-04	2.92E+04	2.44E-08
		6.06E-04	2.75E+04	2.21E-08
		5.41E-04	2.72E+04	1.99E-08
Spk-F	RT	1.07E-03	2.43E+04	4.42E-08
		7.14E-04	2.32E+04	3.08E-08
		7.86E-04	2.61E+04	3.01E-08
Spk-R	RT	4.00E-04	4.58E+04	8.73E-09
		4.61E-04	4.61E+04	1.00E-08
		3.43E-04	5.03E+04	6.81E-09
S-2P	RT	N/A	4.26E+04	N/A
		2.04E-04	5.60E+04	3.64E-09
		N/A	3.93E+04	NA
	55	4.70E-04	2.30E+04	2.04E-08
		3.08E-04	1.23E+04	2.51E-08
		1.93E-03	7.03E+03	2.74E-07

RT= room temperature; N/A = not applicable

Table 3.S 5 Data collection and refinement statistics for R-1b and M-104

	R-1b	M-104
PDB ID	7TN1	8E15
Wavelength	1 Å	1 Å
Resolution range	49.3 - 3.1 (3.211 - 3.1)	47.62 - 2.41 (2.496 - 2.41)
Space group	P 41 21 2	I 21 3
Unit cell	170.5 170.5 171.2 90 90 90	178.191 178.191 178.191 90 90 90
Total reflections	286706 (32084)	72725 (7212)
Unique reflections	44788 (4518)	36363 (3606)
Multiplicity	6.4 (7.2)	2.0 (2.0)
Completeness (%)	96.0 (99.47)	99.91 (99.86)
Mean I/sigma(I)	6.3 (1.7)	9.18 (0.78)
Wilson B-factor	78.2	66.92
R-merge	0.268 (1.34)	0.03633 (0.921)
R-meas	0.292 (1.44)	0.05138 (1.302)
R-pim	0.113 (0.523)	0.03633 (0.921)
CC1/2	0.981 (0.518)	0.999 (0.593)
CC*	0.995 (0.826)	1 (0.863)
Reflections used in refinement	44449 (4513)	36339 (3601)
Reflections used for R-free	1986 (208)	1871 (192)
R-work	0.257 (0.297)	0.2036 (0.3056)
R-free	0.315 (0.371)	0.2487 (0.3366)
Number of non-hydrogen atoms	10472	3438
macromolecules	10420	3360
ligands	42	67
solvent	10	11
Protein residues	1361	443
RMS (bonds)	0.010	0.009
RMS (angles)	1.29	1.02
Ramachandran favored (%)	92.3	97.27
Ramachandran allowed (%)	7.5	2.73
Ramachandran outliers (%)	0.22	0.00
Rotamer outliers (%)	0.25	1.08
Clashscore	13.1	6.61
Average B-factor	71.2	70.90
macromolecules	71.1	70.11
ligands	109.5	110.72
solvent	30.0	70.94

Statistics for the highest-resolution shell are shown in parentheses.

Table 3.S 6 Cryo-EM data collection and refinement statistics for Spk-M

	Spk-M
PDB ID	8FEZ
EMDB	EMD-29035
Microscope	Krios
Magnification	22,500
Pixel size (Å)	1.024
Voltage (kV)	300
Electron exposure (e^- /Å^2)	58.24
Frames	8 sec exposure, 0.2 sec frames, 40 total frames
Defocus Range (μm)	-0.8 to -2.6
Number of movies collected	3257
Particles extracted	1,394,889
Symmetry	C1
Map resolution (Å)	3.72
Protein residues	2843
RMS (bonds)	0.018
RMS (angles)	1.749
Ramachandran favored (%)	95.40
Ramachandran allowed (%)	4.60
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.22
Clashscore	7.96
Rama-Z	-1.24

CHAPTER 4

COMPUTATIONAL DISULFIDE DESIGN IN THE PREFUSION RESPIRATORY SYNCYTIAL VIRUS FUSION PROTEIN: EXPLORING THE IMPACT OF PRECISE POSITIONING

Gonzalez, K. J., Yim, K ., Boukhvalova, M ., Kim, S. & Strauch, E. M. “Computational disulfide design in the prefusion respiratory syncytial virus fusion protein: exploring the impact of precise positioning.”

To be submitted to Frontiers in Immunology.

Abstract

The prefusion conformation of class I fusion proteins holds great promise for vaccine development due to its strong immunogenic properties. However, the fragility of this conformational state has posed a significant challenge. Although non-native disulfides have proven effective in preserving the prefusion state's structure, identifying these stabilizing bonds has remained a intricated task. In this study, we present a novel computational approach to identify prefusion-stabilizing disulfides. Our approach not only evaluates the geometric constraints associated with disulfide bonds but also incorporates a ranking system to estimate the bond's potential to stabilize the prefusion conformation. We propose that disulfides restricting the initial stages of the conformational switch can offer higher stability to the prefusion state than those preventing unfolding at a later stage. The implementation of our algorithm on the respiratory syncytial virus (RSV) fusion protein led to the discovery of three new prefusion-stabilizing disulfides, providing evidence that supports our hypothesis. Furthermore, the evaluation of our top design as a vaccine candidate in a cotton rat model demonstrated robust protection against RSV infection.

Introduction

The unpredictable outbreaks of the respiratory syncytial virus (RSV) have demonstrated that enveloped viruses are a constant concern for public health[249]–[253]. Class I fusion proteins are exceptional vaccine candidates in RSV and many other pathogenic viruses[14]–[20], even though their structural instabilities have been challenging to overcome. Because of their function as membrane-fusion mediators, class I fusion proteins are naturally designed to irreversibly refold from an unstable structure, the prefusion state, to a highly stable conformation, the postfusion state[21]. This spontaneous transition has represented a major limitation for vaccine development as potent immunogenic epitopes get usually concealed in the most stable structure of the protein, the postfusion state[22], [23]. Consequently, stabilizing the prefusion conformation has been a continuous pursuit to retain the protein's immunogenicity and exploit its potential as a vaccine.

Understanding the conformational rearrangements of class I fusion proteins was largely possible thanks to the crystallization of the influenza virus hemagglutinin protein[35], [96], [254] and paramyxoviruses fusion proteins[255]–[258]. As a common fusion mechanism, class I fusion proteins are initially maturated by proteolytic cleavage, where a hydrophobic fusion peptide is left unconstrained to connect with the host cell[40], [42]. Following its activation, the fusion-competent yet fragile prefusion state undergoes conformational changes, facilitating the fusion peptide's interaction with the target membrane[35], [45], [47]. As this interaction progresses, the protein's C-terminal relocates, leading to the assembly of a highly stable six-helix bundle, the postfusion state. This structural transition ultimately enables the fusion of the viral and cellular membranes[50], [259], [260].

The structural analysis of the fusion mechanism has guided the prefusion stabilization of various class I fusion proteins by targeting regions prone for refolding[163]. Historically, this goal

has been achieved through an arduous manual exploration of the protein structure and extensive testing of protein mutants [15], [18], [77], [95], [159], [163], [218], [219]. We had previously pioneered a computational approach that significantly streamlined this demanding process, requiring only a few designs to identify stable prefusion versions [261]. However, our algorithm's limitation to optimizing only electrostatic interactions motivated us to develop a complementary approach capable of automatically designing disulfide bonds.

Our computational method traces potential disulfides based on allowed distance and geometry, similar to other software [262]–[264]. However, while correct geometry indicates the likelihood of bond formation, it does not directly reflect its stabilizing effect [262], [265]. Consequently, we sought to tailor the disulfide design process to class I fusion proteins by considering the conformational dynamics of the protein during the selection of prefusion-stabilizing bonds. We hypothesize that disulfides restricting the initial stages of the conformational switch are more impactful at increasing the stability of the prefusion state than those located on regions unfolding last. By combining our previous approach with this disulfide design strategy, we have successfully increased the stability of the RSV fusion (F) protein without compromising its immunogenic properties.

Results

Disulfide bond design strategy

Inspired by the efficacy of non-native disulfides in stabilizing the prefusion conformation of class I fusion proteins [15], [18], [19], [219], we have developed a computational approach to automatically identify these covalent bonds. Building upon a previously proposed concept [227], our computational strategy is centered on identifying novel disulfides within regions undergoing significant relocation during the prefusion-to-postfusion transition (Fig 4.1. A). Using the Rosetta

Disulfidize as the disulfide scanning tool [239], [241], the new potential bonds are first detected by proximity of Cβ atoms. Residues within 5Å are subsequently mutated to cysteine, and energy minimization and protein repacking are carried out in the surrounding radius (Fig 4.1. A). This restricted structural sampling not only reduces computational costs but also guarantees that the introduced disulfide aligns with the protein's initial configuration. Finally, to increase the likelihood of successful disulfide formation, the new bonds undergo a filtering process based on the Rosetta disulfide energy (dslf_fa13) [266], [267], and geometric adequacy (Fig 4.2.A).

To evaluate how disulfide design can complement and enhance our existing computational approach, we implemented the above methodology in our previously stabilized RSV F protein, R-1b [261]. This implementation led to the identification of 12 potential intrachain disulfides within regions involved in the protein's conformational switch, one of which matched the most effective non-native disulfide reported for RSV F (C155-C290)[19] (Fig 4.2.A). Notably, despite the extensive research conducted on RSV F, eight of these newly identified potential disulfides have remained unexplored until now, representing promising avenues for further investigation[19], [161], [163].

After analyzing disulfides that have been unsuccessful at stabilizing the prefusion state of RSV F [19], [161], [163], we observed that protein stabilization is not solely determined by the accuracy of the bond's geometry. Interestingly, a significant number of these disulfides demonstrated appropriate geometric configurations yet failed to enhance protein stability. This suggests that other factors beyond geometric parameters may play a role in determining the stabilizing capabilities of disulfide bonds. Consequently, we sought to augment our geometry analysis by exploring the role of the protein's dynamics in identifying prefusion-stabilizing disulfides. We hypothesized that substitutions restricting the initial conformational changes of the

protein would have a stronger impact on the stability of the prefusion state than those preventing the final stages of the rearrangement.

To integrate this concept into our selection process, we assumed that regions displaying higher root-mean-square-deviation (RMSD), when comparing the prefusion and postfusion structures, are more likely to unfold at an earlier stage. As a result, RMSD served as the quantitative metric to rank disulfide candidates according to their potential for prefusion stabilization (Fig 4.2. B). Notably, given that conformational changes in the refolding region 1 (protein head) likely trigger conformational changes in the refolding region 2 (membrane-proximal domain)[77], our ranking system was applied independently to each refolding region.

To validate our design and ranking approaches, we expressed three distinct disulfide variants, each exhibiting well-differentiated RMSD values (Fig 4.2. B). The selected disulfides included *V157C-N183C* (design SsA1), located at the beginning of the refolding region 1, *A55C-L188C* (design SsA2), found at the middle of the refolding region 1, and *S443C-S466C* (design SsB1), situated at the beginning of the refolding region 2 (Fig 4.1. B).

Biochemical characterization of disulfide candidates

All three designs, SsA1, SsA2, and SsB1, were successfully expressed and purified as trimeric proteins, as confirmed by size exclusion chromatography (SEC) (Fig 4.3. A). Additionally, the expression yield of these variants showed no significant changes when compared to the parent construct R-1b (Fig 4.3. A). As anticipated, the designed proteins exhibited notable improvements in thermal stability compared to R-1b, with melting temperature increments of 2.5°C (SsB1), 5.5°C (SsA2), or 12.5°C (SsA1) (Fig 4.3. B). This enhancement in heat stability aligns well with our proposed hypothesis, emphasizing that disulfides restricting the initial stages of protein unfolding can have a more profound impact on the overall stability of the prefusion

conformation. Remarkably, the thermal stability of design SsA1 surpassed that of the current RSV vaccine DS-Cav1[19] by approximately 6.5°C (Fig 4.3. B). However, it is important to note that these results do not imply that the V157C-N183C disulfide in SsA1 is inherently superior to the C155-C290 disulfide found in DS-Cav1. Instead, the evident improvement observed in SsA1 likely arises from a synergistic effect resulting from the introduction of the new disulfide and the preexisting optimized electrostatic interactions of R-1b [261].

The melting curve of SsA1 and DS-Cav1 presented two possible unfolding stages, characterized by a low-intensity peak at approximately 55°C and a high-intensity peak above 65°C (Fig 4.3. B). To assess the structural significance of these unfolding phases, we conducted binding assays with prefusion-specific antibodies after subjecting the proteins to heat treatment. Our results revealed that SsA1 effectively preserves the prefusion conformation at high temperatures, as evidenced by its strong binding to D25 [58], [224] and MPE8 [225] even after heating at 65°C, and continued binding to D25 even after heating at 70°C (Fig 4.3. C). In contrast, although DS-Cav1 retained binding to D25 at 65°C, the reduced binding to MPE8 indicated the loss of the protein's quaternary structure at this temperature (Fig 4.3. C). These results not only confirmed the increased stability of SsA1 compared to DS-Cav1 but also demonstrated that the apparent unfolding observed at ~55°C has no substantial impact on the conformation of prefusion antigenic regions. Instead, it is the second melting peak that determines the protein's complete unfolding.

Unfortunately, our antibody binding assays also reveal a diminished interaction between SsA1 and AM14 [84], [224] at room temperature, indicating a potential disruption of the antigenic site V (Fig 4.3. C).

Disulfide bond detection in SsA1

To verify the successful formation of the V157C-N183C disulfide bond in design SsA1, we conducted a tandem liquid chromatography mass spectrometry (LC-MS/MS) analysis. To achieve this, the protein underwent two consecutive alkylation reactions, enabling the differentiation of free cysteines from those involved in disulfide bonding. In the first alkylation step, we used iodoacetic acid (IAA) to attach carboxymethyl groups specifically on free cysteines. Subsequently, a second alkylation reaction was carried out with iodoacetamide (IAM) to introduce carbamidomethyl groups on disulfide-bonding cysteines, following the reduction of the disulfide bond. Finally, peptide fragments obtained from trypsin digestion were subjected to LC-MS/MS analysis.

The ion chromatograms of peptides containing cysteines 157 (**C**LHLEGEVNK) and 183 (AVVSLS**C**GVSVLTSK) displayed two distinct retention times, indicating the presence of both carboxymethyl and carbamidomethyl in the peptides (Fig 4.S1). Although this signified that the cysteines 157 and 183 existed in both free and bound states, the predominance of disulfide-bonding cysteines was evident in base peak intensity of the samples' mass spectrum (Fig 4.S1). Specifically, the normalization level (NL) value of each peak indicated that the relative concentration of the peptide **C**LHLEGEVNK in the disulfide-bonding state was ~29 times higher than in the unbound state. Similarly, the concentration of the peptide AVVSLS**C**GVSVLTSK in the disulfide-bonding state was ~15 times higher than in the unbound state. The fragmentation spectrum of both peptides, confirming the disulfide formation (carbamidomethyl labeling), is shown in Fig 4.S2 and 4.S3.

Immunogenicity of the SsA1 variant in cotton rats.

The SsA1 protein was selected for a vaccination study to investigate the potential effects of enhanced prefusion-stability on the immunogenicity of RSV F. Cotton rats were immunized intramuscularly with 10 or 100µg of AddaSO3-adjuvanted R-1b or SsA1 at weeks 0 and 4 (Fig 4.4.A). Three weeks after the booster immunization, sera were collected to examine the IgG response against prefusion and postfusion RSV F proteins, along with their neutralizing activity (Fig 4.4.A). Notably, animals immunized with either R-1b or SsA1 presented a robust RSV F prefusion-specific response, showing equivalent antibody titers against RSV A or RSV B F proteins, as measured by ELISA (Fig 4.4.B and D). Similarly, no significant differences were observed in terms of RSV A2 F postfusion-specific antibodies, although SsA1 vaccination evidenced slightly lower titers (Fig 4.4.C). Moreover, both immunogens induced comparable levels of antibody neutralizing titers, with most doses surpassing the antibody titers generated by natural RSV A2 infection (Fig 4.4.E). Although these data suggest that the enhanced physical stability of SsA1 did not lead to increased protection against RSV, it remains unclear whether the disruption of the antigenic site V hindered such potential improvement.

To gain further insights into the epitope-specific response elicited by SsA1 and R-1b, we conducted antibody binding competition assays against D25, MPE8, and 131-2A [229]. This competition was carried out following the interaction of the R-1b antigen with different concentrations of pooled serum. Consistent with our ELISA findings, the sera from SsA1 or R-1b vaccinated animals exhibited strong competition for binding sites specific to the prefusion state (Fig 4.5). Notably, vaccinations with SsA1 demonstrated a higher prevalence of antibodies targeting the antigenic site Ø. In contrast, antibodies specific to the postfusion state were predominantly observed in animals vaccinated with R-1b, while quaternary antibodies were

equally abundant in both vaccination groups (Fig 4.5). Altogether, our results suggest that even though the overall protection against RSV seems comparable with both immunogens, there might be variations in the specificity of the antibody response based on the physical stability of their respective binding sites.

Finally, RSV A2 challenge was performed to assess the efficacy and safety of our proteins as potential vaccine candidates. Overall, animals vaccinated with either R-1b or SsA1 successfully cleared the virus from the lung and nose and did not exhibit vaccine-enhanced disease, as evidenced by the levels of interleukin (IL) 4, IL-13, and interferon-gamma (IFN-g) (Fig 4.6). Nevertheless, it is worth noting that significant differences between the SsA1 and R-1b immunogens were observed at a dosage of 10µg, with SsA1 vaccinations showing less inflammatory responses (Fig 4.6.A and C).

Discussion

Given the clinical significance of prefusion class I fusion proteins [20], [27], [59], [195], and the proven effectiveness of non-native disulfides in enhancing protein stability [15], [18], [19], [219], we have developed a computational approach to identify prefusion-stabilizing disulfide bonds. Although we previously designed an automated method to increase the stability of these proteins, we recognized the need to complement it due to its primary emphasis on optimizing electrostatic contacts. Indeed, our prefusion-stabilized RSV F protein, R-1b, served as an ideal candidate to investigate the impact of designed covalent bonds, given its initial optimization relying solely on electrostatic interactions [261]. Surprisingly, despite the extensive research on stabilizing the prefusion RVS F [19], [77], [161], [163], our disulfide search uncovered previously unexplored and promising mutations. This observation highlights the comprehensive analysis that

computational design offers, showcasing its potential benefits not only for novel optimizations but also for the enhancement of established vaccine candidates.

While predicting the likelihood of disulfide formation is feasible through the analysis of the bond's geometry [262]–[264], there is currently no metric available to determine whether a designed disulfide will stabilize or destabilize the protein structure. Various studies have suggested using residue B-factors to guide the selection of stabilizing disulfides, as rigidifying highly flexible areas is known to increase protein stability [262], [265]. However, this metric may not fully capture the complexities of the prefusion state in class I fusion proteins, as demonstrated in the case of RSV F. Depending on the crystallization method and quality, the region with the highest B-factors can vary between the head and the membrane-proximal domain, as observed in the PDBs 5c69[77] and 5w23[222], respectively. Nevertheless, the most stabilizing disulfide reported for RSV F (C155-C290)[19] is located in a region with relatively low B-factor. This observation highlights the challenges of predicting disulfide-induced stabilization of the prefusion conformation and emphasizes the need to consider other factors to accurately estimate the success of an engineered disulfide bond.

As class I fusion proteins undergo intricate refolding processes, we investigated the impact of their conformational dynamics on the ability of a non-native disulfide to stabilize the prefusion conformation. Our findings revealed that disulfides placed in proximity to the fusion peptide have a greater effect on the stability of the prefusion state than those located nearer to the membrane-proximal region (Fig 4.3.B). Interestingly, within the refolding region 1, increases in melting temperature were correlated with the positioning of the designed disulfides, with bonds closer to the N-terminal of the F1 subunit providing higher thermal stability (Fig 4.3.B). These results might reflect the dynamics of the protein, suggesting that disulfide bonds restricting the initial stages of

the conformational switch have a more significant impact on the stability of the prefusion state than those hindering the unfolding at later stages. However, the precise sequence of events preceding the translocation of the fusion peptide has not been definitively determined, and it remains unclear whether the release of the fusion peptide promotes the refolding of the heptad repeat A region or if conformational changes at the apex initiate the projection of the fusion peptide [77]. Our findings appear to lend support to the former hypothesis. If conformational shifts were to initiate at the apex and propagate downstream, then our 55-188 disulfide, located closer to the apex, should have exhibited a more significant stabilizing effect by hindering the initial unfolding phase. Contrarily, it was the 157-183 disulfide, situated closer to the fusion peptide, that displayed the most pronounced stabilizing impact. This leads us to hypothesize that the release of the fusion peptide likely drives the subsequent conformational changes and its inhibition results in a more pronounced stabilization of the prefusion state.

We recognize the importance of further studies to validate our ranking method, which relies on the RMSD of the prefusion-to-postfusion transition. Unfortunately, we encountered challenges in fully analyzing the dynamics of the membrane-proximal domain. The disulfides we identified in this region were not suitable for testing due to several constraints, including poor geometry (482-499), proximity to natural disulfides (392-493), and the replacement of introduced mutations on R-1b (397-487). Consequently, it remains uncertain whether disulfides displaying the highest RMSD values within the refolding region 2 confer greater stability than those located in regions with relatively low conformational changes.

Regrettably, we could not establish a clear correlation between the stability of the prefusion state and its immunogenic properties. It is still uncertain whether the disruption of the antigenic site V in SsA1 impeded its potential for enhanced immunogenicity. Prior studies aiming to improve

the DS-Cav1 stability have shown a relationship between physical stability and increased RSV protection, particularly when employing interprotomer disulfides to ensure an appropriate quaternary structure [163]. However, this association is not consistently observed, as other studies incorporating interprotomer disulfides and increased stability have not reported a corresponding increase in immunogenicity [161], [164]. Our observations indicate that the disulfides involved in enhanced immunogenicity are primarily located within antigenic regions. Therefore, we propose that the overall stability of the protein may not necessarily correlate with improved immunogenicity. Instead, it appears that the local stability of immunogenic regions plays a crucial role in determining immunogenicity. In fact, within our study, we did notice that antibodies targeting the antigenic site Ø, which was stabilized by the disulfide V157C-N183C, were slightly booster in vaccinations with SsA1 (Fig 4.5). However, it is important to consider that this effect could be attributed to the disruption of the antigenic site V and a potential reduction in interclonal competition. Nevertheless, despite the complexities surrounding the relationship between stability and immunogenicity, strategic placement of disulfide bonds may offer opportunities to enhance the quality of the antibody response.

Materials and Methods

I. Computational approach to design disulfide bonds

All computational analyses were performed with the Rosetta version: 2020.10.post.dev+12.master.c7b9c3e c7b9c3e4aeb1febab211d63da2914b119622e69b

Structure preparation

The crystal structure of the R-1b protein (PDB: 7TN1) [261] was refined using the Rosetta relax application [233], [268], [269]. This refinement process was guided by electron density data to avoid large deviations from the initial configuration [231]. The protein's density map was

recreated from its map coefficients using the Phenix software version 1.15 [232] and the option "create map from map coefficients" (region padding= 0, and grid resolution factor= 0.3333). The relaxation process was carried out by assigning a weight of 20 to the density energy term of the Rosetta scoring function and performing four rounds of rotamer packing and minimization [233]. The structure with the lowest Rosetta energy and Molprobity [234] scores was selected for downstream analysis.

Design of disulfide bonds

To ensure the RSV F protein is locked in its prefusion state, we have developed a disulfide design strategy focused on restraining the mobility of regions undergoing drastic conformational changes. These flexible areas were automatically identified by calculating the root-mean-square-deviation (RMSD) of each Cα atom in the prefusion conformation (R-1b) compared to the postfusion structure (PDB: 3rrt) [228]. Residue positions displaying motion levels of at least 10Å were considered to be involved in the conformational switch and selected for disulfide scanning. Since the RSV F protein is a large biomolecule, disulfide scanning was performed independently on small regions of the protein to reduce the computational cost of the analysis. Each region included one residue within the conformational switch area and all residues within a 7Å sphere (Fig 4.1.A). Residues to be mutated to cysteine were first identified by Cβ-Cβ distance (<5.0Å), and potential disulfides were formed when the energetic and geometric properties of the new bond fell within the expected Rosetta values for native disulfides[271], [272]. The scanning and modeling process was carried out with the Disulfidize[241] and FastDesign[240], [241] movers available in RosettaScripts [239], [270], limiting packing and minimization to a 12Å area around the region of interest (Fig 4.1.A).

Selection of candidate disulfide bonds

To enhance the likelihood of successful disulfide formation, we employed a two-step filtering process. Initially, all predicted disulfides were filtered based on the Rosetta disulfide energy (dslf_fa13 <0) to ensure favorable conditions for bond formation. Subsequently, the geometric adequacy of the bonds was evaluated. For this, we conducted a thorough analysis of the disulfide bond angles (dihedrals $\chi1$, $\chi2$, $\chi3$, and angle Cβ-S-S) in 300 high-resolution structures (<1.5 Å) obtained from the Protein Data Bank. All structures were relaxed using the FastRelax mover available in RosettaScripts [233]. Our disulfide candidates were then filtered based on agreement with the observed angle ranges in these reference proteins, allowing for a tolerance of up to 10Å.

Finally, we incorporated the conformational dynamics of the RSV F protein to assess the potential prefusion stabilization of the introduced disulfides. For this purpose, we utilized the root-mean-square deviation (RMSD) of the prefusion-to-postfusion transition as a quantitative metric to evaluate conformational dynamics. Our hypothesis suggests that regions with high RMSD are more likely to unfold first during the transition to the postfusion state. Therefore, disulfides hindering these initial rearrangements would be more effective at stabilizing the prefusion state. As a result, we ranked our disulfide candidates based on the highest RMSD observed within the residue pair forming the bond. This ranking was performed independently for the refolding region 1 and 2, as the conformational changes in the refolding region 2 depend on the earlier changes in refolding region 1[77]. The RMSD values were calculated using the same methodology employed for the identification of the conformational switch area (see Design of disulfide bonds).

II. Protein expression and characterization

Protein expression

Three R-1b disulfide-containing variants, namely SsA1 (V157C /N183C), SsA2 (A55C/L188C), and SsB1 (S443C/S466C), and the control R-1b, RSV A/A2 F (postfusion) [205], RSV A/A2 DS-Cav1 [19], and RSV B/ 18537 DS-Cav1 [273] constructs were expressed by transient transfection of FreeStyle 293-F cells (Thermo Fisher) with polyethylenimine (PEI) (Polysciences). All R-1b variants were produced in pCAGGS plasmids encoding the R-1b protein, a C-terminal T4 fibritin trimerization motif (Foldon), and a His6-tag. The sequence of the R-1b protein contained residues 1-105 and 137-513 with the respective substitutions under study, and a short linker replacing the furin cleavage site and the p27 peptide ("QARGSGSGR") [77]. DNA sequences were codon optimized for human expression using the online tool GenSmart Codon Optimization[243]. 293-F cells were incubated at 37°C and 8% CO_2 for three days after transfection, and proteins were purified by nickel affinity chromatography followed by size-exclusion chromatography (SEC). SEC was carried out using a Superdex200 column (Cytiva) and phosphate-buffered saline (PBS) buffer pH 7.4.

Thermal stability

Differential scanning fluorimetry (DSF) was used to monitor protein stability as a function of temperature. The samples to analyze contained 4µM of protein, 5X SYPRO orange fluorescent dye (Thermo Fisher), 5mM $MgCl_2$, 50mM KCl, and 50mM Tris (pH 7.4). All measurements were performed by triplicates using a qPCR instrument (CFX Connect, BioRad) and a temperature gradient from 25 to 90°C with 0.5°C increments. The melting temperature of each protein was estimated based on the lowest point of the negative first derivative of the fluorescence signal.

Antigenic characterization

Antigenic preservation of prefusion epitopes was evaluated by Bio-layer interferometry (BLI) using the prefusion-specific antibodies D25 [58], [224] (Thermo Fisher), AM14 [84], [224] (Cambridge Biologics), and MPE8 [225]. Binding against expressed designs was tested with 15 nM of the antibodies and eight concentrations of the antigens, starting from 200 nM and decreasing by two-fold dilutions. Prior to all binding assays, Protein A biosensors (GatorBio) were equilibrated for 20 min in BLI buffer (PBS buffer supplemented with 0.02% tween-20 (Promega) and 0.1% bovine serum albumin (BSA) (Sigma)). Subsequently, immobilization of the antibodies on the biosensor tips was allowed for 180s, followed by a baseline correction of 120s, and an association and dissociation steps of 180s each. All assays were performed on a GatorPrime BLI instrument (GatorBio) at a temperature of 30°C and frequency of 10 Hz. Binding constants were obtained with the GatorOne software 1.7.28, using a global association model 1:1 for D25 and MPE8, and 2:1 for AM14.

Antigenic preservation after heat treatment was evaluated in the R1-b variant with the highest melting temperature (SsA1) and the control proteins R1-b and RSV A/A2 DS-Cav1. The proteins were incubated for one hour at 65 and 70°C in a thermocycler with heated lid (T100, BioRad) Binding to D25 and MPE8 antibodies was measured afterwards following the protocol described above.

Disulfide bond detection through alkylation and mass spectrometry

Alkylation with iodoacetic acid (IAA) and iodoacetamide (IAM) was used to corroborate the formation of the disulfide bond V157C - N183C in the SsA1 protein. These alkylation reactions were intended to label free cysteines differentially from disulfide-bonding cysteines. Specifically,

carboxymethyl groups were attached to free cysteines (IAA reaction) and carbamidomethyl groups were attached to disulfide-bonding cysteines (IAM reaction), after reduction of the disulfide bond. Following alkylation, the samples were digested, and peptides were analyzed through tandem liquid chromatography mass spectrometry (LC-MS/MS).

The detailed steps of the process are described below:

- *Alkylation reaction:* 20µL of SsA1 protein at 1.5 mg/mL were incubated for 30 minutes in the dark with 2µL of 0.1M iodoacetic acid. The protein sample was denatured and reduced by adding 0.5µL of 0.5M dithiothreitol (DTT). The mixture was heated at 100°C for 5 minutes and after cooling down the reaction was allowed for 30 minutes at room temperature. The second alkylation was carried out by adding 2µL of 0.5M iodoacetamide and incubating for 30 minutes.
- *Deglycosylation:* The alkylated sample was mixed with 100µL of digest buffer (0.2% sodium deoxycholate (SDC) in 50 mM ammonium bicarbonate) and digested with 20U PNGase F (Lectenz Bio, Athens) in Sartorius Vivacon 500 (10K MWCO) for 2.5 hours at 37°C.
- *Trypsin digestion:* The sample in the filter was washed twice with 200µL of 20mM triethylammonium bicarbonate. 0.3µg of Trypsin in 50µL of 20mM triethylammonium bicarbonate were added to the sample in the filter, and the digestion was carried out overnight at 37°C. The next day, the tryptic digests were spun out of the filter, and remaining peptides in the filter were eluted with 100µL of water. The tryptic peptides were dried by Vacufuge.
- *LC-MS/MS:* The mass spectrometry analyses were performed on a Thermo Fisher LTQ Orbitrap Elite Mass Spectrometer coupled with a Proxeon Easy NanoLC system (Waltham, MA) located at Proteomics and Mass Spectrometry Facility, University of Georgia. The enzymatic peptides were loaded into a reversed-phase column (self-packed column/emitter

with Dr. Maisch ReproSil-pur C18AQ 120Å 3uM resin), and directly eluted into the mass spectrometer. Briefly, the two-buffer gradient elution (0.1% formic acid as buffer A and 99.9% acetonitrile with .0.1% formic acid as buffer B) started with 0% buffer B for 2 minutes, and then increased to 40% buffer B for 95 minutes and to 95% buffer B for 10 minutes. The MS data was obtained using the Xcalibur software (version 3.0, Thermo Fisher Scientific) and the data-dependent acquisition (DDA) method. A survey MS scan was acquired first, and then the top 10 ions in the MS scan were selected for following CID (collision-induced dissociation) and HCD (higher energy C trap dissociation) tandem mass spectrometry (MS/MS) analysis. Both MS and MS/MS scans were obtained by Orbitrap at the resolutions of 120,000 and 15,000, respectively. Protein identification and modification characterization were performed using Thermo Proteome Discoverer (version 3.0) with Mascot (Matrix Science) against Uniprot plus the SsA1 sequence, and a modified contaminations database with commonly known contaminating proteins (Mascot).

III. Animal studies

Cotton rat immunization

Inbred 6-8 weeks-old, *Sigmodon hispidus* female and male cotton rats (source: Sigmovir Biosystems, Inc., Rockville MD) were maintained and handled under veterinary supervision in accordance with the National Institutes of Health guidelines and Sigmovir Institutional Animal Care and Use Committee's approved animal study protocol (IACUC Protocol #15). Cotton rats were housed in clear polycarbonate cages and provided with standard rodent chow (Harlan #7004) and tap water *ad lib.* Groups of 5 or 6 animals (3 females/2 males or 3 females/3 males) were immunized intramuscularly with two different doses (10 µg and 100 µg) of either purified R-1b or SsA1 protein with AddaSO3 adjuvant (50% v/v) at weeks 0 and 4 (Prime and Boost) (Fig 4.4.A).

Control experiments were carried out immunizing with only PBS (negative control), 10^5 PFU of RSV A/A2 Live inoculated intranasally on week 0 (positive control for RSV A/A2 neutralization), or formalin-inactivated RSV with a dose of 1:100 in PBS (FI- RSV lot#100) (positive control for vaccine-enhanced disease). Bleeds were collected from the retro-orbital sinus at week 7 (Fig 4.4.A), and sera were analyzed by ELISA, BLI, and neutralization assay.

RSV A/A2 challenge

RSV A/A2 (ATCC, Manassas, VA) was propagated in HEp-2 cells after serial plaque-purification to reduce defective-interfering particles. A pool of virus designated as RSV/A2 Lot# 092215 SSM containing approximately 3.0 x 10^8 pfu/mL in sucrose stabilizing media was used for the *in vivo* experiment. Virus stock was stored at -80ºC and had been characterized *in vivo* using the cotton rat model and validated for upper and lower respiratory tract replication. Vaccinated and control animals were inoculated intranasally at week seven with 0.1 mL of RSV/A2 (Lot# 092215 SSM) at 10^5 PFU per animal. Back titration on the challenge virus was performed to confirm challenge dose. Cotton rats were euthanized on day five post-infection for analysis of viral load and lung mRNA gene expression.

RSV IgG measurement by ELISA

100µL of purified R-1b [261], RSV A/A2 F (postfusion) [205], RSV A/A2 DS-Cav1 [19], and RSV B/ 18537 DS-Cav1 [273] at 2 µg/mL were coated onto 96 well ELISA plates (Immulon 2 HB, Thermo Fisher) at 4°C overnight. Next day, plates were washed with wash buffer (PBS buffer with 0.05% tween-20) and blocked with 200µL/well of blocking buffer (wash buffer supplemented with 3% non-fat milk (LabScientific) and 0.5% BSA). After two hours incubation at room temperature, plates were washed, and 100µL/well of diluted serum from each rat was added and incubated for two hours at room temperature. Serum was serially diluted 1:3 in blocking

buffer starting from a 1:20,000 or 1:60,000 (R1-b binding) dilution until a total of seven dilutions. Once the serum incubation finished, the plates were washed, and 100µL/well of Rabbit anti Cotton Rat IgG (1:1000 in blocking buffer) (Invitrogen) was added and incubated for one hour at room temperature. The plates were washed again and 100µL/well of Goat anti Rabbit IgG-HRP (1:9000 in blocking buffer) (Invitrogen) was added and incubated for one hour at room temperature. Finally, plates were washed and 100µL/well of TMB substrate solution (Fisher Scientific) was added and incubated in the dark for 15 minutes. The reaction was stopped with 100µL/well of 2M H_2SO_4 and the optical density (OD) was measured at 450 nm on a SpectraMax M2 Reader (Molecular Devices). All experiments were performed by duplicates and the averaged OD value was considered for analysis. Serum titers were estimated with GraphPad Prism 9.0 [274] using the area under the curve (AUC) measurement with a baseline of 0.3 absorbance units and $\log_3$-transformed serum dilutions.

Antibody competition assays using BLI

Antibody binding competition against D25, MPE8 and 131-2A [229] (Millipore Sigma) was carried out to estimate epitope-specific responses after vaccination. Prior to data collection, anti-penta-His sensors (GatorBio) were hydrated for 20 minutes in BLI buffer (PBS buffer supplemented with 0.5% BSA and 0.05% tween-20, pH 7.4). The assay was initiated with an equilibration step in BLI buffer for 100s followed by immobilization of the R-1b or RSV A2 F (postfusion) proteins at 20µg/mL, for 50s. The probes were then washed for 200s in blocking buffer (25% ChonBlock buffer (Chondrex Inc.) diluted in BLI buffer) and interaction with pooled serum samples (three weeks post-boost) was allowed for 700s. Sera were serially diluted 1:3 in blocking buffer, starting from a 1:10 dilution until a total of six dilutions. Blocking buffer with no serum and naïve serum from PBS-vaccinated rats (1:10 dilution) were also included as controls.

After serum interaction, a baseline phase was carried out by dipping the probes in blocking buffer for 60s. Finally, association of competing antibodies was performed for 700s. The MPE8 antibody was tested at 18µg/mL while the remaining antibodies were tested at 9µg/mL. All antibodies were diluted in blocking buffer and the antigens were diluted in BLI buffer. Percentage of antibody competition was calculated as 1- (association response vaccinated serum / association response naïve serum) × 100%. Measurements were performed by duplicates and the averaged competition value was considered for analysis. All assays were performed with a GatorPrime BLI instrument at a temperature of 30°C and frequency of 10 Hz.

RSV neutralizing antibody assay (60% reduction)

Heat inactivated serum samples were diluted 1:20 with Eagle's Minimum Essential Medium (EMEM) and serially diluted further 1:4. Diluted serum samples were incubated with RSV A/A2 (25-50 PFU) for one hour at room temperature and inoculated in duplicates onto confluent HEp-2 monolayers in 24 well plates. After one hour incubation at 37°C in a 5% CO2 incubator, the wells were overlayed with 0.75% methylcellulose medium. After four days of incubation, the overlay was removed, and the cells were fixed with 0.1% crystal violet stain for one hour and then rinsed and air dried. The corresponding reciprocal neutralizing antibody titers were determined at the 60% reduction end-point of the virus control. The averaged value for two measurements was considered for analysis.

Lung and nose viral titration

Lung (*en bloc* and tri-sect, left section) and nasal tissue were homogenized in 3 mL of Hanks' Balanced Salt Solution (HBSS) supplemented with 10% Sucrose-Phosphate-Glutamate (SPG). Homogenates were clarified by centrifugation and diluted in EMEM. Confluent HEp-2 monolayers were infected in duplicates with diluted homogenates in 24 well plates. After one hour

incubation at 37°C in a 5% CO2 incubator, the wells were overlayed with 0.75% methylcellulose medium. After 4 days of incubation, the overlay was removed, and the cells were fixed with 0.1% crystal violet stain for one hour and then rinsed and air dried. Plaques were counted and virus titer was expressed as plaque forming units per gram of tissue. The averaged value over two measurements was used for analysis.

Real-time PCR

Total RNA was extracted from homogenized tissue (lingular lobe) or cells using the RNeasy purification kit (QIAGEN). 1µg of total RNA was used to prepare cDNA using Super Script II RT (Invitrogen) and oligo dT primer (1µL, Invitrogen). For the real-time PCR reactions, the Bio-Rad iQ^{TM} SYBR Green Supermix was used in a final volume of 25µL, with final primer concentrations of 0.5µM. Reactions were set up in duplicates in 96-well trays. Amplifications were performed on a Bio-Rad iCycler for 1 cycle of 95ºC for 3 min, followed by 40 cycles of 95ºC for 10s, 60ºC for 10s, and 72ºC for 15s. The baseline cycles and cycle threshold (Ct) were calculated by the iQ5 software in the PCR Base Line Subtracted Curve Fit mode. Relative quantitation of DNA was applied to all samples. The standard curves were developed using serially-diluted cDNA sample most enriched in the transcript of interest (*e.g.*, lungs from 6 hours post RSV infection of FI-RSV-immunized animals). The Ct values were plotted against $\log_{10}$ cDNA dilution factor, and these curves were used to convert the Ct values obtained for different samples to relative expression units. The averaged relative expression units for two measurements were then normalized to the level of β-actin mRNA ("housekeeping gene") expressed in the corresponding sample.

Statistical analysis

All statistical analysis were performed in R studio 4.1.0 using the tidyverse and rstatix packages. ELISA results and vaccine-enhanced disease assessment from R-1b and SsA1 vaccinations were compared applying a two-tailed t-test. Neutralization data was compared with the Mann-Whitney U test. The normality of each group was confirmed by a Shapiro wilk test, and the equality of variances was checked through the Levene's test. The overall level of significance was set at 5%. Boxplot visualization was done with the seaborn package in python 3.7.

Data and Materials Availability

SsA1 plasmid is available from EMS under a material transfer agreement with the University of Georgia. Rosetta is available through licensing https://www.rosettacommons.org. Scripts for generating designs will be available on https://github.com/strauchlab/disulfide_design upon publication.

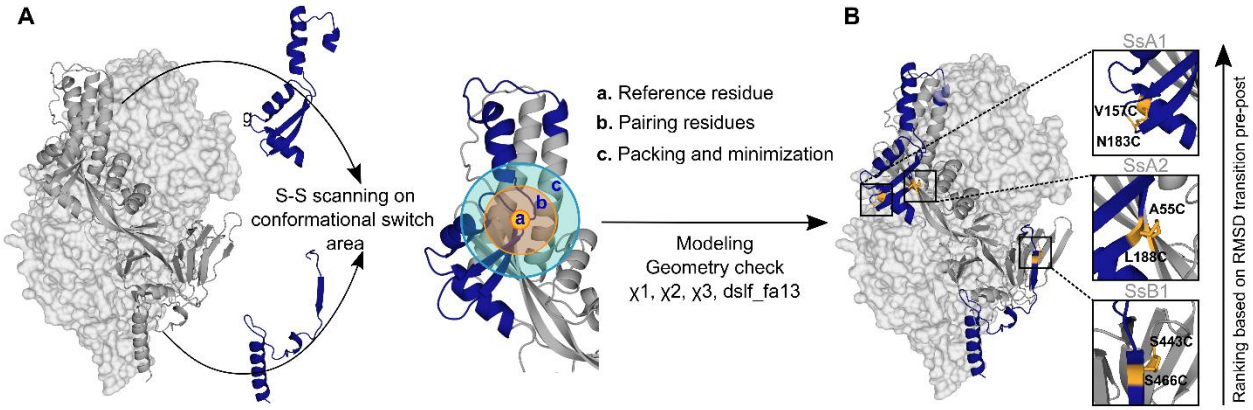


Figure 4. 1 Disulfide design strategy. (A) Disulfide scanning focuses on residues located on the protein's conformational switch area. Potential disulfide bonds are determined independently on small regions of the protein to reduce the computational cost of the analysis. These regions are defined by: (a) Reference residue: An amino acid located on the conformational switch area. (b)

Pairing residues: Neighboring residues within a 7Å sphere that could form a disulfide bond with the reference residue. (c) Packing and minimization area: Region 12Å around the pairing residues allowed to repack and energetically minimize during the modeling process. **(B)** New potential disulfide bonds identified in R-1b. The disulfides were selected based on a correct geometry and ranked according to the protein's conformational dynamics. We propose that disulfide bonds placed in regions of high root-mean-square-deviation (RMSD), when comparing the pre- and postfusion structures, would confer higher stability to the prefusion state. The R-1b protein (PDB: 7TN1) is on display with two protomers as light-grey molecular surfaces and one protomer as a dark-grey ribbon. Regions undergoing drastic conformational changes are highlighted in dark blue, and newly designed disulfide bonds are shown in yellow sticks.

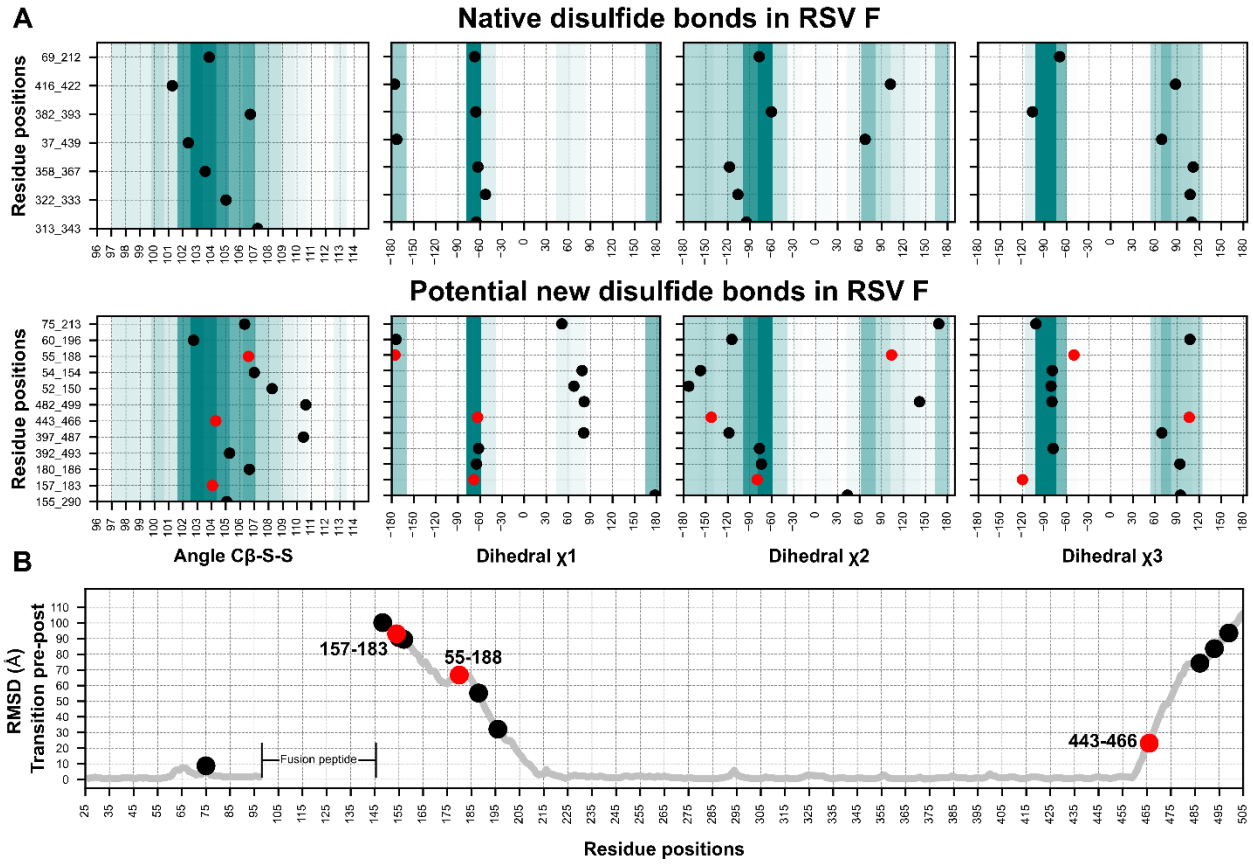


Figure 4. 2 Geometry assessment of the newly identified potential disulfides in RSV F. (A) Geometric parameters used to evaluate the correctness of a disulfide bond. The top panel displays the bond angles sampled by the native disulfides in the RSV F (PDB: 5w23) while the bottom panel presents the bond angles of the new potential disulfides identified in the R-1b protein. The background color scales in each plot represent the frequency of each bond angle, according to a set of 300 high-resolution structures (<1.5Å). Darker colors indicate angles that were more frequent in this reference set, while lighter colors represent angles of low frequency. **(B)** Root-mean-square-deviation (RMSD) mapping of residues forming new potential disulfide bonds. RMSD values were calculated by aligning the prefusion and postfusion structures of the RSV F protein (PDB: 5w23 and 3rrt). The highest RMSD found within the residue pair forming the

disulfide bond determined the ranking of the disulfide and is highlighted in black or red circles. Disulfides selected for experimental validation are highlighted in red in all panels.

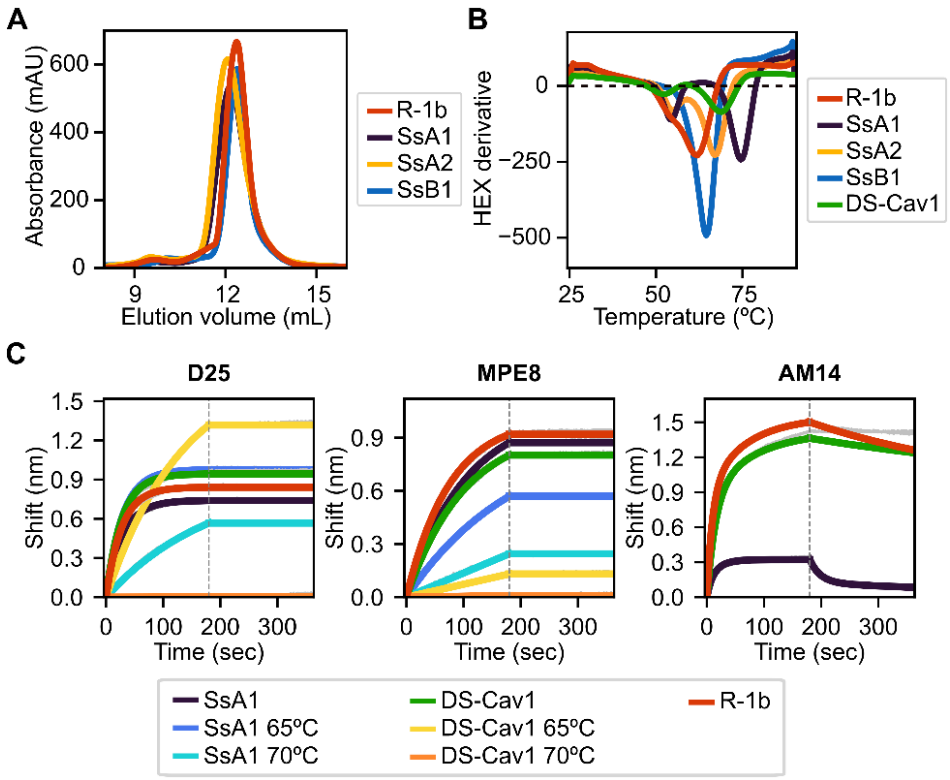


Figure 4. 3 Biochemical characterization of R-1b variants containing designed non-native disulfide bonds. (A) Size-exclusion chromatography. **(B)** Differential scanning fluorimetry of designed variants in comparison with the clinical candidate DS-Cav1[19]. On display are shown averaged values over three replicates. Estimated melting temperatures are R-1b 62°C, SsA1 74.5°C, SsA2 67.5°C, SsB1 64.5°C, and DS-Cav1 68°C. **(C)** Binding of variants SsA1, R-1b, and DS-Cav1 to prefusion antibodies after heat treatment. R-1b was only tested at room temperature as the protein's melting temperature was below the set temperature for the assays. Plotted values correspond to binding at the highest protein concentration (200nM). BLI row data is shown in grey, while fitted curves are shown in colors. The end of the association time is delimited with a dotted line. Binding constants are shown in Table 4.S1.

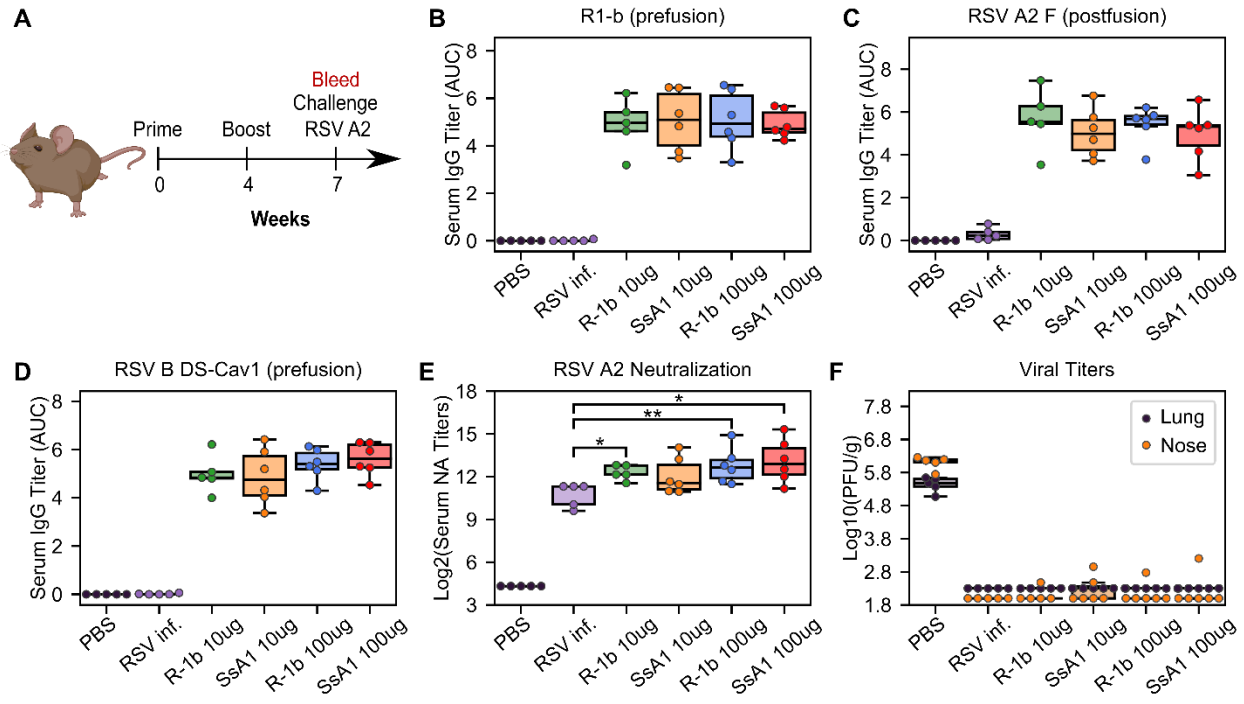


Figure 4. 4 Immunogenicity assessment of R1-b variants in cotton rats. (A) Vaccination study regimen. **(B)** Serum IgG binding against R1-b, **(C)** RSV A/A2 F (postfusion) [205], and **(D)** RSV B/18537 DS-Cav1[273] proteins measured by ELISA three weeks post-boost. Control experiments corresponded to a mock vaccination with PBS and a primary infection with RSV A/A2 on day 0 (RSV inf.). Serum titers were estimated with GraphPad Prism 9.0 [274] using the area under the curve (AUC) measurement with a baseline of 0.3 absorbance units and $\log_3$-transformed serum dilutions. Sera were serially diluted three-fold from 1:60,000 (R1-b binding) or 1:20,000 to a total of 7 dilutions. **(E)** Serum neutralizing (NA) titers against RSV A/A2 using sera three weeks post-boost. **(F)** RSV A/A2 viral titers in lung and nose five days after challenge. Viral titers were determined by plaque forming units (PFU) per gram of tissue. Averaged values from two measurements for each animal are presented in circles while the distribution of the data is shown in boxplots. All boxplots display the median as a central line, lower and upper quartiles as the box limits, and minimum to maximum values as whiskers. Pairwise statistical analyses were performed

with a two-tailed T-test for the ELISA experiments and Mann-Whitney U test for the neutralization data. *, $p \leq 0.05$; **, $p \leq 0.01$. Cotton rat cartoon was created with BioRender.com through the Institute of Bioinformatics licensing at the University of Georgia.

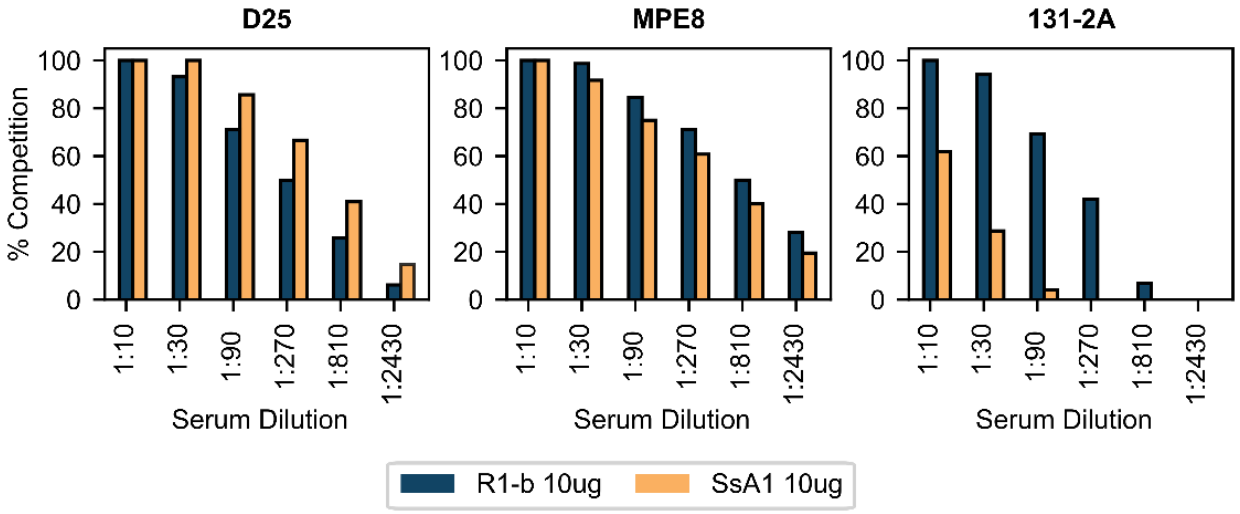


Figure 4. 5 BLI-based antibody binding competition against three weeks post-boost serum. Pooled serum samples from rats vaccinated with 10µg of R-1b or SsA1 were tested for binding competition against D25, MPE8, and 131-2A antibodies. Sera reactivity was first evaluated against immobilized R-1b and the complexes R-1b+serum antibodies were then tested for interaction with the reference antibodies. Percentage of antibody competition was calculated as 1- (association response vaccinated serum / association response naive serum) × 100%. Measurements were performed by duplicates and the averaged value is on display.

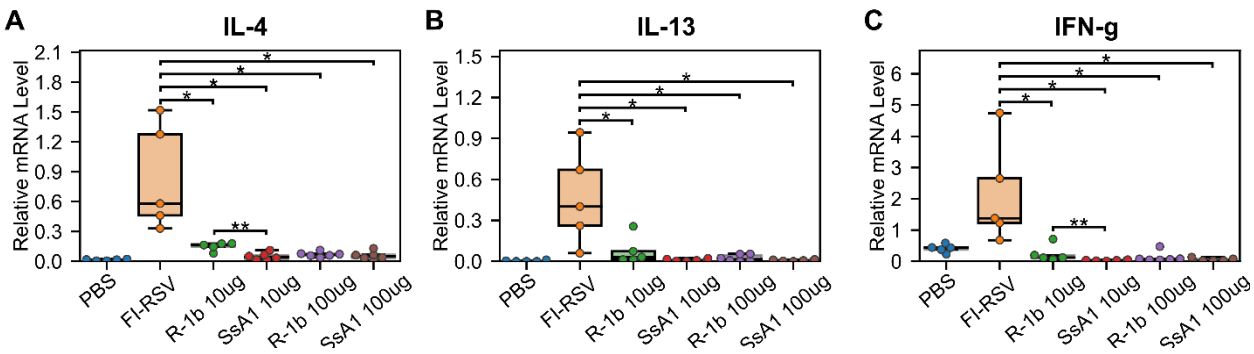


Figure 4. 6 Vaccine-enhanced disease assessment according to mRNA levels of (A) Interleukin-4 (IL-4)**, (B)** Interleukin-13 (IL-13), and **(C)** Interferon gamma (IFN-g). mRNA levels were measured in lung tissue five days post-challenge. A mock vaccination with PBS was used as the negative control, and mice immunized with formalin-inactivated RSV (FI-RSV) as the positive control. The relative expression units were normalized to the level of β-actin mRNA ("housekeeping gene") expressed in the corresponding sample. Averaged values from two measurements for each animal are presented in circles while the distribution of the data is shown in boxplots. All boxplots show the median as a central line, lower and upper quartiles as the box limits, and minimum to maximum values as whiskers. Pairwise statistical analyses were performed with a two-tailed T-test. *, $p \leq 0.05$; **, $p \leq 0.01$.

Supplementary Information

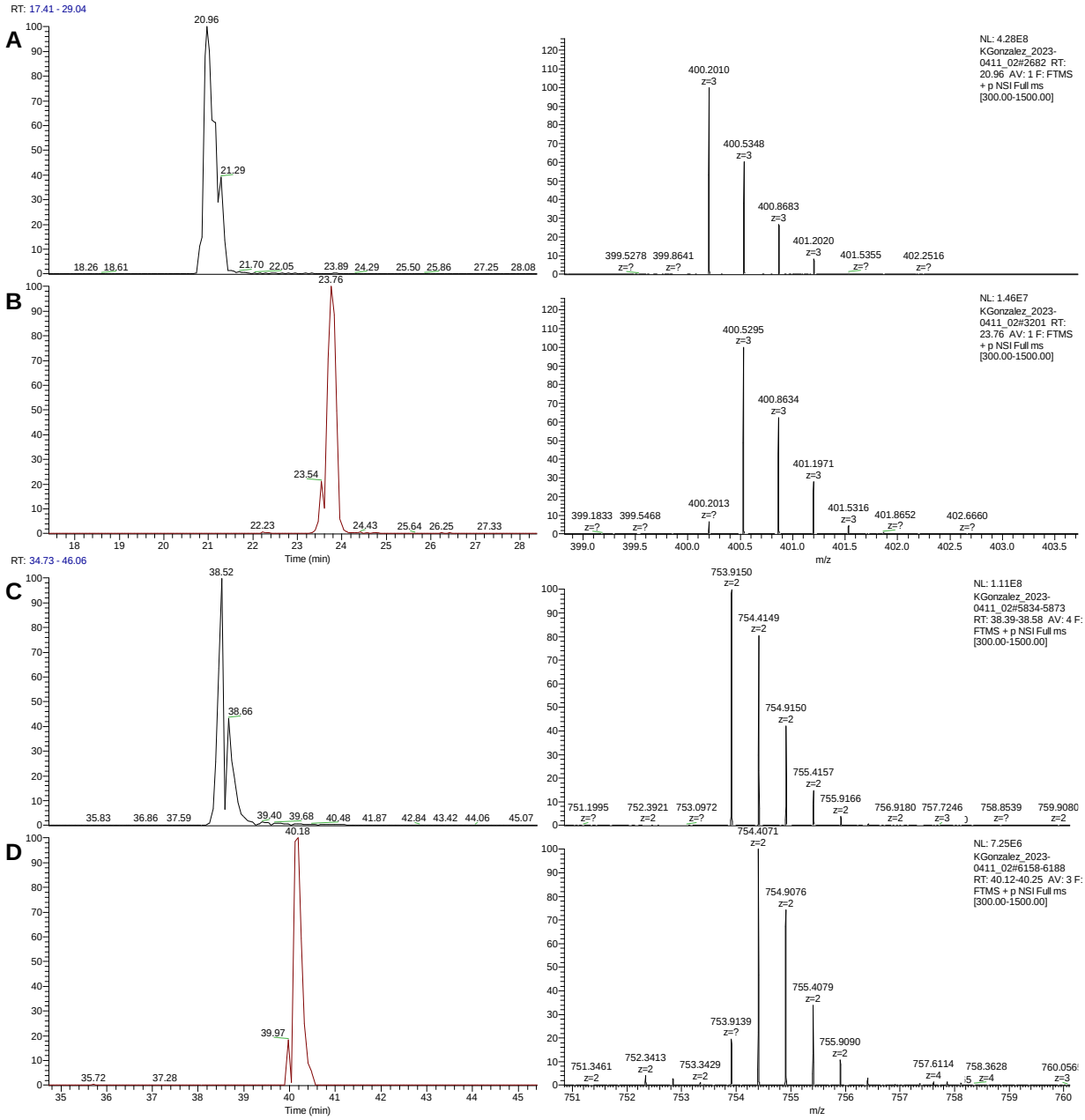


Figure 4.S 1 Tandem liquid chromatography mass spectrometry (LC-MS/MS) analysis of SsA1 peptides containing cysteine 157 (CLHLEGEVNK) and cysteine 183 (AVVSLSCGVSVLTSK). The SsA1 protein was treated with two alkylation procedures to distinguish free cysteine residues from disulfide-bonding cysteines. Alkylation with iodoacetic

acid (IAA) was used to attach carboxymethyl groups to free cysteines while alkylation with iodoacetamide (IAM) was used to label disulfide-bonding cysteines with carbamidomethyl groups, after proper reduction of the disulfide bond. **(A)** IAM-treated (ion m/z 400.2010) and **(B)** IAA-treated (ion m/z 400.5295) spectrums of peptide containing cysteine 157 (**C**LHLEGEVNK). **(C)** IAM-treated (ion m/z 753.9150) and **(D)** IAA-treated (ion m/z 754.4071) spectrums of peptide containing cysteine 183 (AVVSLS**C**GVSVLTSK). Left panels depict ion chromatograms for two types of ions (IAM and IAA treatments) and right panels show the mass spectrum of each sample. Although both peptides were slightly alkylated during the IAA treatment (free-cysteine labeling) as shown in (B) and (D), the base peak intensity of the IAM treatments (disulfide-bonding labeling), dictated by the normalization level (NL) value, evidenced the predominance of disulfide-bonding cysteines. The NL values for IAM and IAA alkylation of peptide **C**LHLEGEVNK were 4.28E8 and 1.48E7, respectively, while peptide AVVSLS**C**GVSVLTSK showed 1.11E8 and 7.25E6, respectively.

#1	b^+	b^{2+}	Seq.	y^+	y^{2+}	#2
1	161.03793	81.02260	C-Carbamidomethyl			10
2	274.12199	137.56463	L	1038.55783	519.78256	9
3	411.18090	206.09409	H	925.47377	463.24052	8
4	524.26496	262.63612	L	788.41486	394.71107	7
5	653.30756	327.15742	E	675.33080	338.16904	6
6	710.32902	355.66815	G	546.28820	273.64774	5
7	839.37161	420.18945	E	489.26674	245.13701	4
8	938.44003	469.72365	V	360.22415	180.61571	3
9	1052.48296	526.74512	N	261.15573	131.08150	2
10			K	147.11280	74.06004	1

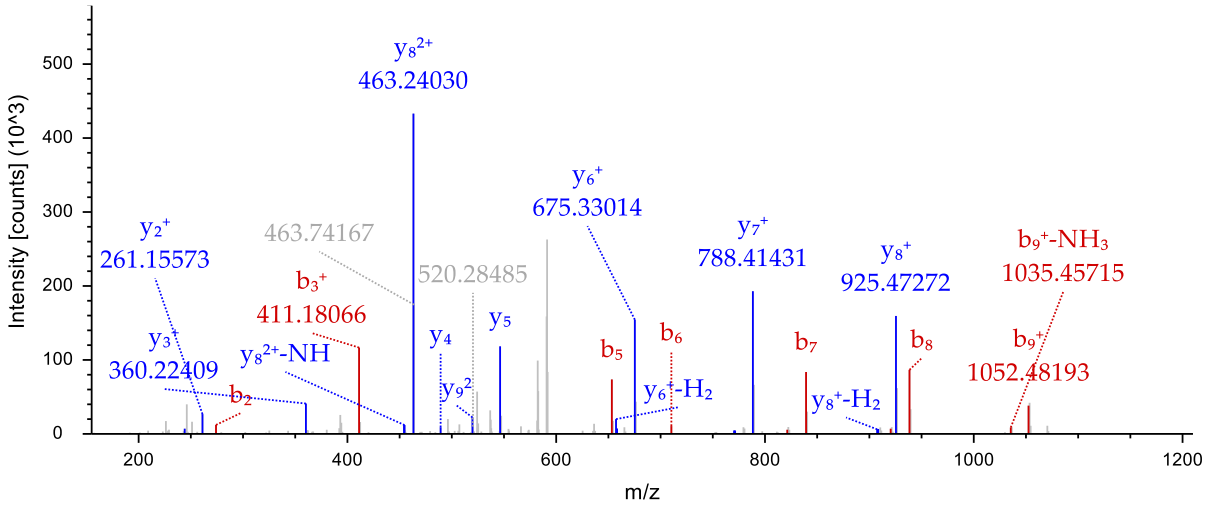


Figure 4.S 2 Fragmentation spectrum of ion containing SsA1 cysteine 157 (CLHLEGEVNK) labeled with carbamidomethyl (IAM treatment). On display is shown the spectrum of +2 charge state product ion, with monoisotopic m/z: 599.79796 Da (+0.08 mmu/+0.14 ppm), MH+: 1198.58865 Da, and retention time: 21.5271 min. The peptide was identified with Sequest HT (v1.17); XCorr:3.37, and a fragment match tolerance of 0.02 Da. Fragments used for search were y-H_2O; y-NH_3; b; b-H_2O; b-NH_3; y.

#1	b^+	b^{2+}	Seq.	y^+	y^{2+}	#2
1	72.04439	36.52583	A			15
2	171.11280	86.06004	V	1435.78249	718.39488	14
3	270.18122	135.59425	V	1336.71408	668.86068	13
4	357.21325	179.11026	S	1237.64566	619.32647	12
5	470.29731	235.65229	L	1150.61364	575.81046	11
6	557.32934	279.16831	S	1037.52957	519.26842	10
7	717.35999	359.18363	C-Carbamidomethyl	950.49754	475.75241	9
8	774.38145	387.69436	G	790.46689	395.73709	8
9	873.44986	437.22857	V	733.44543	367.22635	7
10	960.48189	480.74458	S	634.37702	317.69215	6
11	1059.55031	530.27879	V	547.34499	274.17613	5
12	1172.63437	586.82082	L	448.27657	224.64193	4
13	1273.68205	637.34466	T	335.19251	168.09989	3
14	1360.71408	680.86068	S	234.14483	117.57605	2
15			K	147.11280	74.06004	1

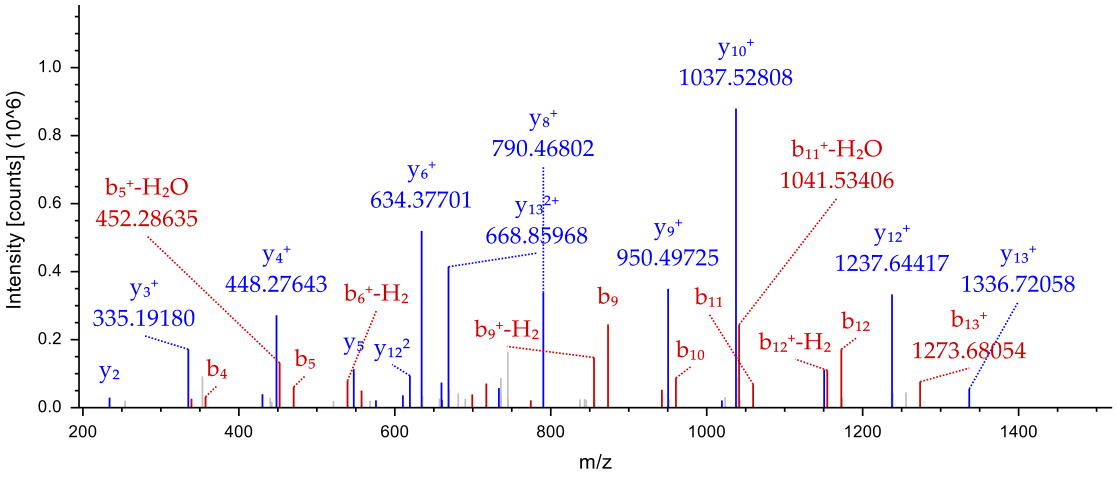


Figure 4.S 3 Fragmentation spectrum of ion containing SsA1 cysteine 183 (AVVSLSCGVSVLTSK) labeled with carbamidomethyl (IAM treatment). On display is shown the spectrum of +2 charge state product ion, with monoisotopic m/z: 753.91394 Da (+0.5 mmu/+0.66 ppm), MH+: 1506.82061 Da, and retention time: 38.9540 min. The peptide was identified with Sequest HT (v1.17); XCorr:3.69, and a fragment match tolerance of 0.02 Da. Fragments used for search: y-H_2O; y-NH_3; b; b-H_2O; y

Table 4.S 1 Binding kinetics of prefusion-stabilized RSV F variants using Bio-layer interferometry.

Protein Variants	Assay temperature (°C)	Antibodies								
		D25			AM14*			MPE8		
		koff (1/s)	kon (1/Ms)	KD (M)	koff (1/s)	kon (1/Ms)	KD (M)	koff (1/s)	kon (1/Ms)	KD (M)
SsA1	RT	N/A	1.79E+05	N/A	0.0541 0.00216	179 6.16E+05	0.000303 3.51E-09	N/A	4.80E+04	N/A
		N/A	1.72E+05	N/A	0.0322 N/A	357 8.91E+06	9.01E-05 N/A	N/A	5.73E+04	N/A
	65	N/A	1.72E+05	N/A	N/A	N/A	N/A	N/A	2.61E+04	N/A
		N/A	1.75E+05	N/A	N/A	N/A	N/A	N/A	3.08E+04	N/A
	70	N/A	3.59E+04	N/A	N/A	N/A	N/A	N/A	198	N/A
		N/A	2.72E+04	N/A	N/A	N/A	N/A	N/A	87.8	N/A
DS-Cav1	RT	N/A	1.57E+05	N/A	0.00162 N/A	5.16E+04 3.26E+05	3.14E-08 N/A	N/A	4.93E+04	N/A
		N/A	1.82E+05	N/A	0.00126 N/A	5.87E+04 3.95E+05	2.14E-08 N/A	N/A	5.10E+04	N/A
	65	N/A	3.48E+04	N/A	N/A	N/A	N/A	N/A	64.7	N/A
		N/A	3.70E+04	N/A	N/A	N/A	N/A	N/A	73.8	N/A
	70	N.B.	N.B.	N.B.	N/A	N/A	N/A	N.B.	N.B.	N.B.
		N.B.	N.B.	N.B.	N/A	N/A	N/A	N.B.	N.B.	N.B.
R-1b	RT	N/A	1.82E+05	N/A	0.00291 N/A	5.85E+04 5.00E+05	4.98E-08 N/A	N/A	6.60E+04	N/A

		N/A	1.82E+05	N/A	0.0034 N/A	6.26E+04 5.72E+05	5.42E-08 N/A	N/A	5.58E+04	N/A

RT= Room Temperature, N.B. = No Binding. N/A= Not Applicable.

* Binding was calculated with a 2:1 binding model. The model assumes two molecules bind with different binding constants. Both constants are presented.

CHAPTER 5

A NEW PROTEIN DESIGN STRATEGY TO ENGINEER HEMAGGLUTININ (HA)-BASED VACCINES WITH POTENTIAL BROAD PROTECTION

Gonzalez, K. J., Shepard, J. D., Shi, H ., Gokanapudi, N., Johnson, S. K., Jones, C., Page, C., Tompkins, S., Ross, T. M & Strauch, E. M. "A new protein design strategy to engineer hemagglutinin (HA)-based vaccines with potential broad protection."

To be submitted to Journal of Virology.

Abstract

The influenza virus effectively evades existing immunity by accumulating amino acid variations in regions targeted by antibodies. To ensure adequate protection, vaccines must stimulate a comprehensive immune response capable of neutralizing a wide range of the virus's antigenic variants. The development of such universally reactive immunogens remains an ongoing challenge in vaccine research, emphasizing the need for innovative design approaches. In this study, we explored the potential of three protein design strategies to engineer broadly protective HA-based immunogens within the H1 influenza subtype. We were successful in expressing and characterizing 14 designed candidates, which were subsequently evaluated for their vaccinal potential in a mouse model. Of these, three proteins demonstrated effective protection against viral challenge with the A/Brisbane/02/2018 (H1N1) strain. Remarkably, one protein elicited a wide variety of antibodies capable of interacting with both pre- and post-pandemic HAs, in addition to H3 subtypes. This immunogen was engineered by replacing all regions outside the head's primary antigenic sites with a consensus sequence derived from long-lasting H1N1 vaccine strains. Meanwhile, the head's antigenic regions maintained a native sequence, each epitope containing a different strain within the long-lasting vaccines. Given the promising broad-spectrum immunological response elicited by this design strategy, our approach may represent a novel avenue for the development of universally reactive influenza vaccines.

Introduction

Although seasonal vaccination is still the most valuable tool to control influenza infections[275], the rapid evolution of the virus has dramatically constrained the efficacy and durability of these vaccines[6], [7]. Unfortunately, the influenza virus proficiently escapes from existing immunity by introducing point mutations that transform its antigenic properties[24], [25]. Since current influenza vaccines are not cross-reactive, vaccination is only effective for specific influenza strains, and vaccines become obsolete once the virus changes[276]. Therefore, ideal influenza vaccines are expected to overcome the antigenic variability of the virus and be broadly protective against seasonal and potential pandemic strains[25], [109].

The hemagglutinin (HA) protein has been the central component of influenza vaccines due to its ability to trigger strong humoral responses[277]. In fact, hemagglutination inhibiting (HAI) antibody titers have been historically used to predict influenza vaccine efficacy and select new vaccine compositions[7]. The HA protein is structurally defined by the head and the stem domains, each playing different roles during the viral infection[96]. While the head region mediates the attachment to host cells through a receptor-binding site (RBS), the stem domain serves as a conformational switch that allows the fusion between the viral and the cell membrane[97]. Likely because of its accessibility on the viral surface, the head domain is immunodominant over the stem, and antibodies binding to this region are more potent at directly neutralizing the virus[110]. Sites of vulnerability are predominantly focused on five antigenic sites, designated as Sa, Sb, Ca1, Ca2, and Cb [278], although loops adjacent to the RBS also play a significant role in the antigenic properties of the protein [107]. As a consequence of the immunodominance of these regions, the HA head undergoes a constant selective pressure that drives it to be antigenically diverse, whereas the stem domain remains fairly conserved [105].

To take advantage of the conservation of the HA stem in enhancing vaccine cross-reactivity, different approaches are under development to focus influenza immune responses towards this domain[279]. Attractive strategies include glycan masking[200], [201], [203] or complete removal of the head[120], [174], [191], [193], and sequential vaccination with chimeric proteins intended to recall preexisting anti-stem antibodies[184], [186], [189]. Other strategies involving conserved sequences have considered both HA head and stem and aimed to summarize in one single protein all common information within influenza subtypes. This idea arose in the computationally optimized broadly reactive antigens (COBRAs) where multiple layers of consensus calculations were able to condense historic amino acid changes undergone by the influenza virus[170]. COBRA proteins have shown favorable reactivity in H5N1, H1N1, and H3N2 influenza subtypes[170]–[172].

Despite significant progress in expanding the protective scope of HA-based vaccines, the development of a universal influenza vaccine remains an unmet need. Consequently, in this study we implemented various protein design methodologies aiming at generating broadly reactive HA immunogens. One particular design strategy yielded significant success, resulting in an HA variant capable of conferring protection against heterologous viral challenge. The diverse antibody response elicited by this protein underscores its potential for broad-spectrum protection, thereby positioning our design approach as a promising new avenue for the development of universal influenza vaccines.

Results

Design of HA immunogens

We evaluated the potential of three protein engineering strategies to generate broadly reactive HA immunogens within the H1 influenza subtype. Initially, we hypothesized that

elements crucial for protection might be found in the vaccine strains recommended by the World Health Organization (WHO) [280]. While seasonal vaccinations are generally required, certain vaccine strains have exhibited multi-seasonal efficacy, suggesting that they contain attributes conserved among diverse viral strains. To explore this, we generated consensus HA sequences from H1N1 vaccine strains, with the objective of unveiling these conserved cross-reactive epitopes. We have termed this approach "consensus vaccines."

Expanding upon the previous concept, our second approach postulates that vaccine strains offering extended periods of protection may contain the most relevant information for eliciting robust immunity. Specifically, the strains Brazil/11/1978 (Brz/78), Singapore/6/1986 (Sing/86), New Caledonia/20/1999 (NC/99), and California/07/2009 (Cal/09) were each shown to confer protection against H1N1 viruses for at least six years [280]. As such, these strains serve as repositories of sequence features that span multiple viral generations. Thus, our second design methodology focused on generating a consensus sequence from the above long-lasting protective strains. However, given that the resulting protein's head would inherently be biased toward pre-pandemic strains—due to their higher prevalence in the input sequences—we opted not to assign the consensus sequence to the immunogenic regions of the protein head. Instead, we constructed a mosaic-like head, featuring a diverse array of antigenic epitopes (Fig 5.S.1). This involved isolating the antigenic head epitopes from each of the long-lasting vaccines, followed by the generation of various combinatorial arrangements across the head domain. With this approach, we aimed to stimulate an antibody response potentially recognizing commonly found phenotypes in both head and stem domains. We refer to this strategy as "mixed epitopes" (Fig 5.S.1).

Finally, motivated by the promise of multi-layered consensus [171], we adopted this approach to generate consensus-based designs with all available H1 HA sequences from human-

isolated influenza viruses up to the year 2019. This strategy was designed with dual objectives: first, to identify highly conserved regions that have persisted throughout the evolution of the virus, and second, to provide a comparative framework for assessing how effectively the consensus sequence of vaccine strains captures this evolutionary history. Contrary to previous technologies [171], which required a careful selection of input sequences for consensus calculations, our multi-layer implementation was intended to be more general and automated. Briefly, all our HA dataset was sorted based on strain isolation year, and a single consensus sequence was computed for each period. After this, sequences were grouped according to phylogenetic trees estimates. To enhance clustering diversity, we employed three phylogenetic algorithms and various sequence inputs (e.g., whole protein sequences, head-only sequences, or pseudo-sequences incorporating the most immunogenic HA head epitopes). Each of these clustering techniques yielded slightly divergent groups, thereby facilitating a broader range of consensus calculations. The multi-layer consensus was then carried out in accordance with the sequence clusters delineated on each phylogenetic tree (Fig 5.S.2 and 5.S.3). Notably, as the initial consensus step based on isolation year substantially reduced the number of input sequences, ambiguous consensus resulted at certain residue positions where multiple residues exhibited the same frequency. In such instances, we employed native sequences possessing pertinent attributes—such as pandemic-like or seasonal-like characteristics—to inform the identity of the ambiguous residues. We hypothesized that the integration of these template sequences would facilitate the display of both pre- and post-pandemic antigenic features, thereby expanding the range and specificity of the elicited antibody response.

We ranked all our designed sequences based on their percentage of sequence similarity with the pandemic Cal/09 strain, and we selected 27 constructs exhibiting varying degrees of sequence identity for further experimental characterization.

Biochemical characterization of designed proteins

Of the selected designs, 14 out of 27 proteins successfully expressed as trimeric complexes in mammalian cells, as confirmed by their optimal size-exclusion chromatography (SEC) profiles (Fig 5.1.A). While a substantial portion of these proteins exhibited high expression levels, the remaining constructs yielded suboptimal results, particularly those originating from mixed-epitope designs (Fig 5.1.B). Given that these mixed-epitope constructs were assembled through a random combination of sequences, their lower expression levels are likely due to the inclusion of elements less favorable in naturally occurring configurations. Nonetheless, assessment of the thermal stability of all expressed proteins showed melting temperatures comparable to those of native HAs, suggesting that their fundamental structural attributes might have remained unperturbed (Fig 5.1.B). Notably, around 50% of our designs, including some based on mixed epitopes, demonstrated an enhanced thermal stability of at least 4°C compared to the Cal/09 HA (Fig 5.1.B).

To characterize the antigenic properties of the expressed designs, we tested binding to a panel of 12 antibodies, each displaying distinct reactivity (Fig 5.2.A). Specifically, we evaluated five broadly-reactive antibodies, including two head-specific, C05[281] and 1F8[115], one stem-specific, FI6v3[118], and two anchor antibodies, P1-03 and P1-05[282]. Additionally, we assessed 4H3, 4G10, and 3D3, which are head-specific antibodies displaying neutralizing activity; 4G10 and 3D3 are effective against pandemic strains only, while 4H3 shows efficacy against both pandemic and post-pandemic strains [115]. To complete the panel, we included four narrowly reactive antibodies that were naturally induced by the 2017–2018 influenza vaccine (Fluzone)[282]. These antibodies target either the RBS surroundings (Ca09-40 and Ca09-16), a lateral patch epitope (Ca09-28), or an undefined region on the head with non-neutralizing activity (Ca09-38)[282].

As anticipated, all designed proteins exhibited high affinity for antibodies targeting the HA2 subunit, such as stem and anchor antibodies (Fig 5.2.A). These results not only confirm the correct structural assembly of the newly engineered proteins but also highlight the high degree of sequence conservation within this subunit. In the context of broadly reactive, head-specific antibodies, 1F8 demonstrated strong binding to nearly all protein designs, with the exceptions of design #15, which showed weak binding, and design #5, which showed no binding. On the other hand, while most interactions with the antibody C05 occurred with proteins that lacked an insertion at position 133a (between HA1 residues 133 and 134), as it is characteristic of this antibody [283], designs #17 and #56 evidenced binding despite the presence of the insertion (Fig 5.2.A). Lastly, interactions with narrowly reactive antibodies were variable, likely reflecting different compositions of seasonal pre-pandemic and pandemic-like epitopes within the designed proteins (Fig 5.2.A). Overall, the diverse antibody binding affinities observed in these characterizations imply that our engineered proteins are antigenically distinct, potentially offering different mechanisms of protection when employed as immunogens.

Given the binding profiles displayed by all our designs, we conducted a Principal Component Analysis (PCA) to better visualize their differences and similarities regarding antigenic features (Fig 5.2.B). Using a preset cluster count of seven, we found that three of our designed proteins clustered with established long-lasting H1N1 vaccine strains, suggesting antigenic resemblance. Specifically, design #6 showed a close alignment with the Cal/09 strain, indicative of the presence of protective, pandemic-like epitopes. Likewise, designs #20 and #56 were closely associated with the A/New Caledonia/20/1999 (NC/99) strain, suggesting they possess epitopes with equivalent pre-pandemic antigenic properties. Additionally, design #57 presented an antigenic profile analogous to the COBRA antigen P1, known for eliciting broad-

spectrum immune responses against both pandemic and pre-pandemic viruses [115], [171], [284]. Given the broad protective capabilities previously demonstrated by the control proteins in each of these antigenic clusters, it appears likely that our vaccine designs with similar antigenic profiles could induce comparable immunogenic responses. Finally, the remaining designed proteins fell into a transitional clustering zone between seasonal pre-pandemic strains and P1. This may be reflective of their limited binding affinity to pandemic and post-pandemic-specific antibodies, despite the incorporation of the K133a insertion—a feature typically associated with pandemic-like strains.

Immunogenicity evaluation in mouse model

To evaluate our designs' ability to elicit broadly protective immune responses, we vaccinated mice using a prime-boost regimen (Fig 5.3.A). Following the last immunization, serum samples were collected to measure the HAI activity of the induced antibody response. Initial testing focused on the pandemic strain Cal/09 and the post-pandemic strain A/Brisbane/02/2018 (Bris/18) to estimate potential protection against current influenza strains (Fig 5.3.B). Consistent with our earlier antibody binding data, the antibodies triggered by design #6 demonstrated robust activity against Cal/09 and moderate reactivity against Bris/18. Conversely, sera from designs #57 and #15 displayed notable HAI activity against both viral strains (Fig 5.3.B). Surprisingly, such reactivity was absent in immunizations with design #5, despite its antigenic similarities to design #15 as per antibody binding-based clustering (Fig 5.2.B). This discrepancy might be a consequence of the divergent affinities between these two designs and the post-pandemic-specific antibody Ca09-40 (Fig 5.2.A). The pronounce binding affinity between design #15 and Ca09-40, which was not observed in design #5 (Fig 5.2.A), suggests the presence of a post-pandemic epitope in the head region of design #15 that may be lacking in design #5. Finally, corroborating our expectations

based on antibody-binding clustering and minor reactivity to antibodies recognizing pandemic-like proteins (Fig 5.2), the remaining designed vaccines were ineffective in eliciting measurable HAI titers against both Cal/09 and Bris/18 strains (Fig 5.3.B).

While HAI^+ titers are correlate of protection [7], many influenza antibodies can confer immunity through alternative mechanisms [115], [121], [285]. Consequently, we further assessed the protective efficacy of our vaccine designs by measuring viral clearance in the lungs following a challenge with the Bris/18 strain. Consistent with the HAI data, immunization with designs #6 and #15 effectively halted viral replication in the lungs, with most viral titers falling below our limit of detection ($\sim 1 \times 10^2$ PFU/mL of lung homogenates) (Fig 5.3.C). In contrast, despite showing promising HAI activity, the immune response elicited by design #57 failed to achieve effective viral clearance (Fig 5.3.B-C). Interestingly, an opposite phenomenon was observed with mice vaccinated with design #56; they exhibited successful protection against viral replication even though the HAI activity was previously determined to be inadequate (Fig 5.3.B-C).

To further explore the breadth of reactivity, we conducted HAI testing against both pre-pandemic and H3 strains. This evaluation focused on vaccinated groups that demonstrated promising viral clearance, as well as design #57. Within this experimental framework, the antibody response elicited by design #56 had notable reactivity, displaying remarkable HAI activity against A/Singapore/6/1986 (Sing/86) (Fig 5.3.D). Furthermore, 50% of mice vaccinated with design #56 generated HAI^+ titers against A/New Caledonia/20/9 (NC/99) (Fig 5.3.D). This result aligns with the potential seasonal attributes previously identified in our antibody binding-based clustering analysis (Fig 5.2.B). Surprisingly, half of the serum samples from animals vaccinated with design #56 also showed HAI activity against the H3 subtype A/Switzerland/9715293/2013 (Swit/13), although they did not surpass the 1:40 threshold (Fig 5.3.D). A similarly pattern of reactivity was

observed for design #6, which induced HAI^+ titers against both Sing/86 and Swit/13 (Fig 5.3.D). Finally, design #57 induced reactive antibodies against pre-pandemic viruses, most notably against Bris/07 (Fig 5.3.D).

In summary, based on HAI titers, vaccination with design #6 generated the most broadly reactive immune response among our panel of vaccine candidates. This design induced antibodies that were reactive against four viruses, including two of pandemic-like nature, one pre-pandemic, and one belonging to the H3 subtype. Following closely, design #56 induced antibodies with HAI^+ activity against two pre-pandemic strains and one H3 subtype. Additionally, design #56 demonstrated protective efficacy against the post-pandemic Bris/18 strain. Conversely, although design #57 successfully generated HAI^+ titers against both pandemic and pre-pandemic strains, its inability to offer protection during viral challenge raises questions about its practical utility.

Sera binding competition

Intrigued by the broadly reactive response generated by design #6 and #56, as well as the protective efficacy of design #15, we chose these candidates for further investigation of their antibody responses. Specifically, we sought to gain a more in-depth understanding of the binding regions targeted by the antibodies elicited through vaccination. To this end, we assessed serum binding competition against a panel of well-characterized, broadly reactive antibodies. These antibodies included the head antibodies C05 and 1F8, the stem antibody CR9114[286], and the anchor antibody P1-05. Our results initially evidenced a pronounced head-specific antibody response coming from vaccinations with design #15. These antibodies demonstrated competition solely with the head-specific antibody 1F8, particularly when reacting with the Bris/18 and Cal/09 HAs (Fig 5.4). Considering that the level of binding competition was moderate, and we only found HAI activity against pandemic-like strains, it is plausible that these antibodies employ a distinct

binding strategy compared to 1F8, rendering them more strain-specific. Collectively, these data suggest that the protection conferred by design #15 was likely mediated through head-specific mechanisms, most probably via HAI activity.

In stark contrast to design #15, vaccination with design #56 elicited the most heterogenous antibody response among our selected candidates (Fig 5.4). A particularly notable aspect was the pronounced competition for stem and anchor epitopes, which was not found in any of the other vaccination groups. This robust stem response was anticipated, given the vaccine's success in achieving viral clearance despite the absence of HAI activity (Fig 5.3. C). Interestingly, we also identified several head-directed antibodies against a diverse panel of HAs, including proteins from H3 strains. The latter was of particular interest, driving us to perform additional HAI assessments. Specifically, we investigated whether the intense antibody competition against C05 when reacting with the Sing/16 HA would translate to HAI activity. Unfortunately, no HAI activity was found against Sing/16 (Fig 5.S.4.A). Consequently, although the protective role of the observed head antibodies lacking HAI activity remains ambiguous, our current data suggest that design #56 has the potential for broad protection. It may confer immunity against some pre-pandemic strains through HAI activity, while neutralizing other strains through Fc-mediated mechanisms [287].

Finally, the antibody binding features resulting from vaccination with design #6 closely aligned with that of the control vaccinations using Bris/18 and Cal/09. These responses were dominated by a substantial population of head-specific antibodies, whereas the induction of stem antibodies was relatively limited (Fig 5.4). When considered in conjunction with the previously described HAI activity, these findings suggest that the protection conferred by design #6 is predominantly mediated by head-specific antibodies.

Interestingly, design #6 and the controls Cal/09 and Bris/18 appeared to differ in the head response competing for the 1F8 binding site, in the context of NC/99 and Sing/86 HAs. Serum from animals vaccinated with design #6 seemed to harbor a greater number of antibodies targeting this region compared to those in the control groups (Fig 5.4). However, the functional implications of this apparent enhancement remain to be fully delineated, pending an assessment of the neutralizing potency of these antibodies.

Second-generation of HA designs

Due to the promising reactivity profiles of designs #6 and #56, we generated second-generation versions to better understand and potentially improve their immunogenic properties. For design #56, we specifically investigated whether HAI activity against pandemic-like strains was constrained by the antigenic site Sa. Given that the original design of #56 incorporated the NC/99 sequence at the Sa site (Fig 5.S.1), we hypothesized that this sequence could be the primary factor contributing to the protein's pre-pandemic attributes. To test this, we substituted the NC/99 sequence at the Sa site with the corresponding sequence from Cal/09. The resulting variant was named "56_Sa."

Regarding design #6, we explored two options. First, sequence alignment revealed that a major distinction between design #6 and Bris/18 lay at residue position 186. While Bris/18 contains a Proline at this position, both design #6 and Cal/09 feature a Serine (Fig 5.S.5). Given the immunodominance of the 190-helix [108] and the apparent establishment of this Proline in newer sequences (Fig 5.S.6), we proposed that incorporating an S186P mutation into design #6 might enhance the vaccine's protective capacity against current and potentially emergent strains. This modified construct was designated as "6_186."

To further diversify the protective potential of design #6, our second variant aimed at enhancing stem-directed antibody responses. We focused our efforts on amino acid position 133a, where evidence indicates that an insertion at this site can substantially shape the specificity of the resulting antibody response [288]. To counteract this, we introduced a putative glycosylation site into design #6 by implementing an N133T substitution. This modification generates an N-X-T motif spanning residues 131-133, which may undergo glycosylation at position 131. The potential glycosylation could sterically mask the 133a insertion, redirecting the immune response towards subdominant regions like the stem domain. This protein variant was named "6_132."

Immunogenicity evaluation of second-generation variants

To evaluate the immunogenic implications of the designed mutations, we performed a second vaccination study following the same regimen described earlier (Fig 5. 3.A). Serum samples were collected two weeks post-second boost and analyzed for HAI activity against Bris/18 and Cal/09. For design 6_186, we postulated that the introduction of a Proline residue at position 186 could potentially amplify the vaccine's effectiveness against newer strains like Bris/18. However, in contrast to our expectations, the antibody responses triggered by 6_186 did not display significant changes in HAI activity as compared to those elicited by the original design #6 (Fig 5.5.A). It is worth noting that while every mouse vaccinated with design #6 showed strong HAI^+ titers against Bris/18, a subset of mice vaccinated with 6_186 failed to produce antibodies with HAI activity. A similar pattern was noted when the antibody responses were tested against the Cal/09 strain (Fig 5.5.A).

For the second variant of design #6, design 6_132, we hypothesized that introducing a glycosylation site at amino acid position 131 would have minimal impact on the HAI^+ response against pandemic-like strains. This assumption was informed by existing research, highlighting

the Sb and Ca2 antigenic sites as the dominant targets of the immune response in the context of post-pandemic vaccinations [289]. Consistent with their findings, serum samples from mice vaccinated with 6_132 did not exhibit significant variations in HAI activity against either the Bris/18 or Cal/09 strains, compared to those vaccinated with the original design #6 (Fig 5.5.A).

The 56_Sa variant was anticipated to enhance head reactivity towards pandemic-like sequences, given that the Sa region in the original design #56 was derived from a pre-pandemic strain (NC/99). Intriguingly, this hypothesis was validated when HAI reactivity was assessed against Bris/18 but not against Cal/09 (Fig 5.5.A). This outcome is particularly unexpected, considering that the head region of the 56_Sa protein shares a higher degree of sequence similarity with Cal/09 (79.6%) than with Bris/18 (75.3%). To further probe the implications of the Sa substitution, we assessed HAI reactivity against the pre-pandemic strain NC/99, but similar unanticipated results were obtained (data not shown). In this experiment, the original design #56 failed to elicit any HAI activity against NC/99, whereas the 56_Sa variant did so in half of the vaccinated animals. Unfortunately, the incongruent nature of our HAI data precludes a definitive conclusion regarding the impact of the Sa region in design #56.

The intricate challenges of influenza vaccine design were further underscored by our viral clearance assays following challenge with the Bris/18 strain. Surprisingly, while all our second-generation vaccine candidates induced antibody responses with significant HAI activity against Bris/18, none succeeded in effectively clearing the virus from the lungs (Fig 5.5.B). Design 56_Sa showed partial viral clearance, yet it failed to provide protection in half of the animals tested.

Sera binding competition of second generation of HA variants

To gain a deeper insight into how the immune response induced by our second-generation vaccines differed from that triggered by the parental antigens, we conducted serum-binding

competition assays against the broadly reactive antibodies C05, 1F8, CR9114, and P1-05 (Fig 5.6). Our results suggest that design 56_Sa induced more head-specific antibody responses in the context of H1 subtypes, as compared to the antibody response generated by the original design #56 (Fig 5.4 and 5.6). Accordingly, sera obtained from mice vaccinated with 56_Sa demonstrated a reduction of at least 19% in competition against the CR9114 stem antibody, across all H1 HAs tested. Conversely, competition for head-binding sites in H3 HAs declined by at least 33% in sera from 56_Sa vaccinations compared to those from design #56 (Fig 5.6). Taken together, these data implies that the antibody response produced by design 56_Sa is more subtype-specific compared to the response triggered by design #56. Furthermore, the design 56_Sa response is possibly more strain-specific within the H1 subtype, given its increased emphasis on head-directed antibodies. These observations are consistent with prior studies highlighting the influence of the Sa region on the specificity and immunodominance of head epitopes [108], [290].

Differences in the antibody response induced by the two variants of design #6 were mostly seen regarding the CR9114 stem epitope and head antibodies potentially binding to pre-pandemic and H3 strains. A significant variation in the sera from mice vaccinated with design 6_186 was the absence of antibodies that compete for the C05 binding site on H3 HAs (Fig 5.6). The head antibodies produced by 6_186 appear to adopt a distinct binding mode compared to those elicited by design #6. This difference is evident in the lack of binding to H3 HAs and in the unique competition observed for both C05 and 1F8 binding sites on the NC/99 HA—a feature not previously seen with sera from design #6 (Fig 5.4 and 5.6). In contrast, antibodies from design #6 predominantly competed for the 1F8 binding site on the NC/99 HA (Fig 5.4). These results suggest that the S186P substitution imposes a structural constraint that alters the binding mode of head antibodies, potentially narrowing their reactivity range. In terms of stem antibodies, we found that

vaccination with design 6_186 generated a greater number of antibodies competing for the CR9114 binding site compared to the original design #6. This was particularly evident when sera were assessed against Bris/18, NC/99, and Sing/86 HAs (Fig 5.6). Interestingly, the overall competition profile of the 6_186 sera resembled more closely the antibody response observed following Bris/18 vaccination (Fig 5.4). Consequently, it is possible that the emergence of P186 in more recent strains aimed at discouraging antibody binding to the immunodominant Sb epitope.

The serum competition profile for 6_132 vaccinations exhibited minimal deviation from that generated by the original design #6 (Fig 5.6). With the 6_132 design, we aimed to enhance stem antibody responses by reducing binding affinity in the vicinity of the 133a insertion. When compared to design #6, sera from 6_132 vaccinations did show a modest increase in competition for the CR9114 stem epitope across most HAs, except for Cal/09, Bris/07, and Swit/13 HAs (Fig 5.6). Interestingly, while sera from 6_132 also presented a relative decrease in C05-competing antibodies in H3 HAs, there was an increased level of competition for this binding site in Bris/07 and NC/99 HAs. This altered competition profile could indicate the masking of the 133a residue, suggesting that antibodies elicited by the 6_132 variant adopt a binding mode that remains unaffected by this insertion within the H1 subtype. Unfortunately, none of the C05-competing antibodies generated by 6_132 exhibited HAI activity against either NC/99 or Bris/07 (Fig 5.S.4. B). Therefore, the precise implications of these apparent increases in stem antibodies, as well as antibodies targeting the head regions of both pre- and post-pandemic strains, will require further exploration, particularly in terms of their neutralization capabilities.

T-cell epitope analysis

Recognizing that protection in mice may not directly correlate with efficacy in humans, we aimed to estimate potential protective effects in humans via T-cell epitope profiling. For this,

designs #6 and #56, as well as native HA sequences representing post-pandemic (Bris/18), pandemic (Cal/09), and pre-pandemic (NC/99) phenotypes were analyzed through diverse computational tools for T-cell epitope prediction [291]–[297]. Our analysis incorporated a broad spectrum of major histocompatibility complex (MHC) class II and class I alleles, collectively covering the genetic diversity of most of the human population [298], [299]. Specifically, we focused on MHC II alleles DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*09:01, DRB1*11:01, DRB1*13:01, and DRB1*15:01, and the MHC-I alleles A*01:01, A*02:01, A*03:01, A*24:02, B*07:02, and B*44:03.

Our predictions indicated that both designs #6 and #56 present potential $CD4^+$ T-cell epitopes across multiple alleles, although with a few differences. Only design #56 exhibited epitopes for allele DRB1*01:01, while design #6 was exclusive in sharing epitopes with pandemic-like sequences for alleles DRB1*03:01 and DRB1*09:01 (Fig 5.S.8). Notably, in most alleles, one or more potential $CD4^+$ epitopes were found at the HA head interface, specifically between residues 190 and 280. These epitopes were not fully conserved, thereby distinguishing pre-pandemic from pandemic-like sequences. In this context, design #6 demonstrated greater sequence similarity with pandemic-like proteins, whereas design #56 more closely resembled pre-pandemic strains (Fig 5.S.8). This outcome was anticipated, given the prevalence of pre-pandemic features in the long-lasting vaccines that served as the basis for design #56. Conversely, epitopes located in the stem domain were highly conserved and well-preserved in both designs #6 and #56 (Fig 5.S.8).

Similar to the MHC-II predictions, both designs #6 and #56 exhibited potential $CD8^+$ T-cell epitopes across the majority of selected alleles (Fig 5.S.8). The only exception was allele A*03:01, for which only design #6 demonstrated epitope sharing with pandemic-like sequences.

Additionally, both vaccine candidates preserved highly conserved $CD8^+$ stem epitopes, several of which coincided with predicted $CD4^+$ epitopes. In terms of epitopes in the head region, the two designs diverged in a manner consistent with the MHC-II predictions: design #6 exhibited greater similarity to pandemic sequences, whereas design #56 aligned more closely with pre-pandemic strains (Fig 5.S.8). Importantly, residues within helix 190 were identified as components of a dual $CD4^+/CD8^+$ epitope. Considering that design #56 features the Cal/09 sequence in this region, this candidate can potentially stimulate a cross-reactive response against pandemic-like strains.

Collectively, our T-cell epitope predictions suggest that neither design #6 nor design #56 disrupts highly conserved HA T-cell epitopes. This implies that both vaccine candidates have the potential to effectively stimulate cross-reactive T-cell responses in humans.

Discussion

Given the significance of conserved sequences in developing vaccines for rapidly mutating viruses such as influenza, we evaluated three consensus-based protein engineering strategies aimed at designing broadly reactive immunogens. To potentially find the minimal sequence inputs capable of revealing shared characteristics among HA proteins, we examined consensus sequences derived from small datasets, including either all recommended H1N1 vaccine strains or exclusively long-lasting vaccine strains, as well as from comprehensive datasets encompassing all available HA H1 human sequences up to 2019. Regrettably, the majority of our engineered proteins were unsuccessful in generating HAI titers against pandemic-like viruses and did not confer protection against a recent influenza strain, such as Bris/18 (Fig 5.3.B-C). These results suggest that the immunodominant regions within the head were either skewed towards pre-pandemic features or underwent sequence recombination that likely retained immunodominance without accurately reflecting either pre- or post-pandemic periods.

We anticipated the first scenario in sequences derived from consensus with vaccine strains, given the overrepresentation of vaccines from pre-pandemic periods. To mitigate this bias in our multi-layer consensus strategy, we sought to maintain a balanced number of sequence representatives across different time periods. However, due to the pronounce differences between pre- and post-pandemic sequences, regions of high immunogenicity often yielded ambiguous consensus results. We initially hypothesized that substituting these ambiguous positions with an amino acid residue not necessarily the most prevalent, but one found in a protective strain from either the pre- or post-pandemic periods, could result in sequences featuring components from both eras. Consequently, the elicited antibody response would confer protection against both pre- and post-pandemic phenotypes. Contrary to our expectations, our data suggest that such recombination did not extend the protective capabilities of the resultant proteins. Plausible explanations to this outcome include either that the antigenic properties of the selected candidates were dominated by the template protein providing the ambiguous positions, or that our extensive sequence recombination generated novel antigenic phenotypes that bear little resemblance to either pre- or post-pandemic strains. Additionally, our candidate selection process may have inadvertently skewed the results towards uncommon features. This bias could arise from our criterion of favoring sequences with at least eight amino acid differences in head immunogenic regions relative to native HAs.

Despite the setback outlined earlier, three of our protein designs—#6, #15, and #56—demonstrated protective efficacy against the Bris/18 virus (Fig 5.3.C). According to HAI and antibody competition assays, the immunological responses generated by designs #6 and #15 were predominantly influenced by antibodies targeting the head region, most notably in the case of design #15 (Fig 5.4). Conversely, design #56 elicited a diverse antibody profile capable of

recognizing HAs from both pre- and post-pandemic periods, as well as proteins from the H3 subtype (Fig 5.4). While pinpointing the effects to specific amino acid substitutions remains elusive due to the extensive divergence between native sequences and our engineered constructs (Fig 5.S.5 and 5.S.7), the methodologies employed in the creation of these designs can provide valuable insights into their antigenic properties.

Designs #15 and #6 were generated using multi-layer consensus methodologies, utilizing either full-length sequences or sequences restricted to antigenic regions, respectively. In contrast, design #56 was formulated through the mixed epitopes approach (Fig 5.S.1–3). Our data suggest that the observed pandemic-like characteristics in design #15 could have been incorporated during the final layer of the consensus-based design (Fig 5.S.3). In this layer, the precursor sequences contributing to design #15 were a composite of consensus sequences from various years: 1918, 1988, and both pre- and post-pandemic periods (Fig 5.S.3). While these sequences apparently balance pandemic-like and seasonal-like features, it is worth noting that pre-pandemic sequences have undergone more antigenic changes compared to their post-pandemic counterparts [300]. When these pre-pandemic sequences are integrated into the consensus calculations—either in multi-layer designs or in per-year consensus summaries—they introduce considerable sequence ambiguity, particularly in regions with antigenic significance. Consequently, the more well-defined post-pandemic sequences could disproportionately influence the antigenic identity in the final consensus, as might be the case of design #15. This bias also extends to our general strategy to resolve ambiguities based on native sequence templates. In instances where the consensus remains highly ambiguous, the identity of the native template could predominate in defining the antigenic features of the resulting design. As such, while consensus-based design excels in capturing essential attributes in low-variability regions, it appears less proficient at

accommodating the complexity of highly variable regions. Indeed, when the original COBRA technology was used exclusively with H1N1 human sequences, era-specific antibody responses were also observed [171]. To address this limitation, alternative strategies could focus on reducing the immunodominance of these highly variable regions, thereby allowing greater responses toward conserved areas.

The broad antibody diversity induced by vaccination with design #56 underscores the potential efficacy of integrated strategies combining consensus design with alternative methods potentially modulating the head's immunodominance by presenting a wide range of antigenic phenotypes. Within our mixed epitopes strategy, this alternative approach involved the simultaneous display of a diverse array of head-derived immunogenic epitopes, all sourced from long-lasting H1N1 vaccine strains. Contrary to recent mosaic design studies that employed exotic avian sequences to eclipse immunodominant regions of the head [187], [188], our approach aimed at assembling a relevant cocktail of human-derived epitopes that could potentially stimulate antibody responses against multiple viral phenotypes. The comprehensive head-directed antibody responses generated by vaccination with design #56 suggest that this goal was largely accomplished (Fig 5.4), although at the expense of some HAI reactivity (Fig 5.3). However, this limited HAI response was not entirely unexpected as existing literature indicates that antibodies targeting specific antigenic regions exhibit distinct functionalities [108]. For instance, antibodies focused on the Sa region are generally more HAI-reactive, whereas those targeting the Sb region own neutralizing capabilities but lack HAI activity [108]. Consistent with these findings, we observed some HAI activity against NC/99 and Cal/09 following vaccination with designs featuring the Sa region from these respective strains (#56 and #48; Fig 5.S.1). Furthermore, it seems plausible that the neutralizing properties of antibodies targeting the Sb region contributed

to the observed viral clearance in a subset of mice vaccinated with designs #56 and #11, both of which included a Cal/09-derived Sb epitope (Fig 5.3.C).

The enhanced neutralizing capabilities of design #56, relative to other mixed-epitope designs, along with its robust stem response (Fig 5.4), additionally suggest that the mixed epitopes strategy can potentially overcome the immunodominance of the head region, albeit dependent on specific sequence combinations. Further investigations quantifying the distribution of antibodies targeting each antigenic domain, including stem regions, are requisite for understanding how epitope immunodominance could have been modulated within design #56. Interestingly, studies addressing the immunodominance hierarchy among the five classic immunogenic sites (Sa, Sb, Ca1, Ca2, Cb) have reported varying classifications in mice, contingent on the viral strain used for analysis. For example, responses to the A/Puerto Rico/8/1934 strain primarily focused on Sb and Sa sites [108], while those to the post-pandemic A/Michigan/45/2015 strain were dominated by Sb and Ca2 sites [210]. These observations raise compelling questions about how immunodominance hierarchies may shift within individual viral strains, and whether such hierarchies can be ignored when multiple dominant epitopes from different strains are simultaneously presented. Answering these questions is crucial for understanding the underlying mechanisms contributing to the broad reactivity observed in design #56 and for assessing the potential of mixed epitopes in modulating immunodominance.

While the precise consequences of substituting the Sa region in design #56—from a pre-pandemic to a pandemic sequence—are yet to be determined, the partial protection conferred by the second-generation antigen (Fig 5.6) implies a certain level of mutational tolerance within this design. This is in stark contrast to design #6, where protective efficacy was entirely abolished by just a single mutation (either S186P or N133T) (Fig 5.5.B). Although the impact of these specific

substitutions within the context of design #56 remains an open question, our current data suggest that the mixed epitopes strategy may possess inherent robustness, not being constrained by a fixed sequence within each antigenic site. This flexibility is particularly significant for the potential extension of this approach to various influenza subtypes. Consequently, the overall design concept may be versatile enough to accommodate not only long-lasting vaccines but also sets of sequences that manifest key antigenic characteristics.

Although further research is needed to evaluate the neutralizing capabilities of design #56 across a broad spectrum of viral strains, our work introduced a novel methodology with promising implications for the development of universal influenza vaccines. Importantly, our results also invite to re-examinate the criteria used to select influenza vaccines, as high HAI titers might not be an accurate representation of protection.

Materials and Methods

I. Sequence design of hemagglutinin (HA) variants

Consensus-based sequence design was used to integrate common features among HA sequences in newly generated proteins. For this, all available H1 HA full-length sequences isolated from human source between 1918 to 2019 were downloaded from the NCBI Influenza Virus Resource database (http://www.fludb.org)[209]. To reduce sequence bias due to unbalanced number of sequences for different time periods, a consensus dataset was calculated based on strain-isolation year. Accordingly, sequences were divided by their isolation year, aligned, and filtered by <99.65% sequence identity. Sequence alignments were performed with MUSCLE 3.8.31[301] and identity filtering was done with the ProDy python package[302]. Single consensus sequences for each year, referred to as per-year consensus sequences, were then estimated using the Bio python package (consensus threshold=50%)[303]. For years where consensus could not be

calculated, e.g., when only two sequences were left after filtering, the original sequences were kept as representatives of that period.

The per-year consensus sequences were input into three different sequence-design strategies, as described below:

- *Consensus based on H1N1 recommended vaccine strains:*

Three consensus sequences were calculated using the vaccine strains recommended by the WHO between 1977 and 2019. The first consensus was estimated using all vaccine strain sequences retrieved from https://www.fludb.org/brc/vaccineRecommend.spg?decorator=influenza. This consensus sequence was estimated with consensus threshold=45% to decrease the ambiguity of the consensus calculation.

Given that the recommended vaccine strains did not cover years before 1977, different historical HA sequences were added to the vaccine strains alignment to have representative sequences of that time. These extra sequences corresponded to South_Carolina/1/1918, Puerto_Rico/8/1934, Weiss/1/1943, Fort_Monmouth/1/1947, Denver/1/1957, New_Jersey/1/1976, and California/10/1978. The dataset with recommended vaccine strains and the historical sequences was used to calculate the second consensus sequence.

Finally, the last consensus was calculated using only the vaccine strains that were protective for more than 6 years. These strains included Brazil/11/1978, Singapore/6/1986, New Caledonia/20/1999, and California/07/2009.

- *Mixture of antigenic epitopes from H1N1 vaccines strains that protected for more than 6 years:*

This strategy was intended to create a "recombination" of the main antigenic sites of the HA proteins in combination with consensus design. The regions to recombine were selected based

on residues frequently interacting with neutralizing antibodies (NAbs) according to a visual inspection of different crystallized HA-NAb complexes. Thus, six antigenic regions were defined as follows:

i. Region 1: Helix 190 from residues 186 to 198.
ii. Region 2: Loop 220 (residues 206-208, 210, 212, 218, 219, 222, 224-228, 242).
iii. Region 3: Loop 150 (residues 155-160, 162, 163).
iv. Region 4: Loop 130/140 (residues 124, 125, 128, 129, 131, 132, 133A, 134, 136, 137, 140-145).
v. Region 5: Residues 149, 169-173, 255, 256.
vi. Region 6: Residues 53, 77-83A, 273, 275, 276, 278.

Residue numbering are based on the PDB 4m4y[112].

The sequences of each antigenic region were individually extracted from vaccine strains that were protective for more than six years (Brazil/11/1978, Singapore/6/1986, New Caledonia/20/1999, California/07/2009), and the recombination was generated by permutating the sequences into the six defined antigenic regions (Fig 5.S.1). The remaining parts of the head and the stem of the protein were obtained from a consensus sequence using only the long-lasting vaccine strains, with ambiguous positions replaced by the most frequent small/uncharged residues.

- *Multiple layers of consensus design:*

To generate a more diverse set of consensus sequences, three sequence-clustering algorithms, neighbor-joining, UPGMA (Unweighted Pair Group Method with Arithmetic Mean), and maximum likelihood, were independently used to cluster the per-year consensus sequences. Neighbor-joining and UPGMA clustering were performed with MUSCLE, while maximum likelihood was executed with RAxML 8.0.056 [304]. Clustering trees were visualized with iTOL

(Interactive Tree Of Life)[305], and single consensus sequences were obtained by multiple layers of consensus calculations on each tree branch. Briefly, consensus sequences were calculated for the most external subgroups of a branch and then input to a second consensus calculation for the next, more internal, subgroup. This process was repeated until all the branches were combined into a single sequence, or the biggest branches, containing sequences from diverse time periods, were summarized into single sequences (Fig 5.S.2 and 5.S.3). To avoid consensus bias by unbalanced number of sequences within each group, we tried to keep equal number of sequence representatives of each subgroup by calculating consensus on groups displaying high number of sequences. The selection of the subgroups to be condensed by consensus calculation was done manually based on the visual representation of the phylogenetic trees.

Phylogenetic-based clustering was implemented on three different sets of sequences. The first group corresponded to full-length per-year consensus sequences, the second set were shortened per-year consensus sequences having only the HA protein head (the immunodominant region of the protein), and the last group involved per-year consensus pseudo-sequences containing only the main antigenic sites at the head of the HA protein. Each set of sequences was aligned with MUSCLE and duplicates were removed prior to clustering. For the RaxML method, the full-length or protein head alignments were refined with Gblocks 0.91b [306] to discard large patches of gaps that could interfere with the distribution of groups.

The protein head sequence was defined according to the crystal structure of A/California/04/2009 H1N1 (PDB 4m4y) as the region comprised between the sequences “LEDK” and “GAIN” (residues 42 – 289 , numbering from PDB 4m4y chain E). Similarly, the HA main antigenic sites at the head were selected based on the classic antigenic sites Sa, Sb, Ca1, Ca2, and Cb as described here [289] plus all residues interacting with NAbs according to different

crystallized HA-NAb complexes. The final set of antigenic regions corresponded to the positions 53,55A,62,77-83A,116C,117,119,122-125,128,129,131,133A,134,136,137,140-145,149,155-160,162-167,169-173,186-198,206-208,210,212,218,219,222,224-228,238-240,242,246,255,256,273,275,276,278 (numbering from PDB 4m4y chain E). The remaining parts of the protein were defined as explained in the next section "Strategies to overcome consensus ambiguity".

Strategies to overcome consensus ambiguity

Consensus sequences from small datasets usually present ambiguity when multiple amino acids have the same frequency at a given position. To define the identity of these ambiguous positions we used both sequence-based and structure-based approaches. For the sequence-based strategy, we replaced all undefined positions with the corresponding residue on a vaccine strain sequence protective for more than 6 years (Brazil/11/1978, Singapore/6/1986, New Caledonia/20/1999, California/07/2009) or by the corresponding residue on the consensus sequence obtained with all recommended H1N1 vaccine strains (see section "*Consensus based on H1N1 vaccine strains"*). Similarly, in the structure-based approach, we threaded all designed sequences onto three different structural templates so that ambiguous regions were replaced with the corresponding amino acids in the template. The structural templates used for this analysis were the California/04/2009 H1N1 (PDB: 4m4y), as a representative of pandemic strains, A/Solomon Islands/3/06 H1N1 (PDB: 6cf7 [307]), as a representative of seasonal strains and strains lacking the 133a insertion, and the A/Almaty/32/1998 H1N1 (PDB: 6mya [283]), as a representative of historical strains.

This approach was also used to complete those sequences containing only the HA head or its antigenic regions (section *"Multiple layers of consensus design").* Thus, for example, some

designed proteins would present a consensus sequence on the head, while the rest of the protein corresponds to a long-lasting vaccine strain.

Selection process

To select sequences that were distinct from natural proteins, all designs were filtered based on at least eight amino acids differences at the head immunogenic regions compared to native HAs from the H1N1 influenza subtype. The residue positions comprising these immunogenic regions are described at the section "*Multiple layers of consensus design."* Subsequently, to ensure all designs were different from each other, a second filtering process was carried out to preserve sequences with at least five amino acids differences at antigenic regions. Finally, all sequences were ranked based on their percentage of sequence identity with the California/04/2009 strain. Sequences selected for experimental validation were chosen based on variable levels of similarity with California/04/2009.

II. Protein expression and characterization

Protein expression

Both first generation and second generation of designed hemagglutinin (HA) proteins were produced using pCAGGS vectors encoding the HA residues 1-520 (H1N1 A/California/04/09 HA numbering), a C-terminal T4 fibritin trimerization motif (Foldon), and a His6-tag. All DNA sequences were optimized for human expression with the GenSmart Codon Optimization tool[243]. The plasmids were transiently transfected in FreeStyle 293-F cells (Thermo Fisher) with polyethylenimine (PEI) (Polysciences), and cells were cultured for three days at 37°C and 8% CO_2. Proteins were purified through nickel affinity chromatography and size-exclusion chromatography (SEC) using a Superdex200 column (Cytiva) and phosphate-buffered saline (PBS) buffer pH 7.4.

Antigenic characterization

Purified proteins displaying an optimal SEC profile were antigenically characterized using Bio-layer interferometry (BLI). Both narrowly reactive and broadly reactive antibodies, such as C05, 1F8, FI6v3, P1-03, P1-05, 4H3, 4G10, 3D3, Ca-28, Ca-38, Ca-16, Ca-40, were used to evaluate if common antigenic sites were not disrupted during the design process. Prior to antibody loading, Protein A biosensors (GatorBio) were equilibrated for 20 min in BLI buffer (PBS buffer supplemented with 0.02% tween-20 and 0.1mg/mL bovine serum albumin (BSA)). Subsequently, the antibodies were immobilized on the biosensor tips for 90s at a concentration of 15nM. A baseline correction was carried out for 60s, and the association and dissociation phases with the designed proteins were allowed for 120s each. Protein solutions were kept at a concentration of 200nM, and all wells had a final volume of 200 µL/well using BLI buffer. Antibody binding against the following recombinant HAs was used as control experiments:

H1 subtype: A/California/07/2009 (Cal/09), A/New Caledonia/20/1999 (NC/99), A/Brisbane/59/2007 (Bris/07), and the COBRA design P1.

H3 subtype: A/Switzerland/9715293/2013 (Swit/13), and A/Singapore/INFIMH-16-0019/2016 (Sing/16).

All assays were performed by triplicates in a GatorPrime biolayer interferometry instrument (GatorBio) at a temperature of 30°C and frequency of 10 Hz. Since only one concentration of protein was measured, data analysis was done qualitatively based on the observed on-rate signal.

Thermal stability

Differential scanning fluorimetry (DSF) was used to determine the thermal stability of all expressed proteins, including the HA control Cal/09. All protein samples contained 3.5 µM of protein, 5X SYPRO orange fluorescent dye (Thermo Fisher), 5mM MgCl2, 50mM KCl, and 50mM Tris (pH 7.4). Background control was included by mixing all reagents with no protein.

Measurements were collected with a qPCR instrument (CFX Connect, BioRad) using a temperature ramp from 25 to 90°C with 0.5°C increments. The melting temperature was determined based on the lowest point of the HEX signal. All experiments were performed by triplicates.

III. Animal studies

Cell Culture

Madin-Darby Kidney (MDCK) cells (International Reagent Resource FR-926) were maintained in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 5% fetal bovine serum (Atlanta Biologics) and incubated at 37°C and 5% CO2.

Viruses

H1N1 strains A/Singapore/06/1986 (Sing/86), A/New Caledonia/20/1999 (NC/99), A/Brisbane/02/2018 (Bris/18), and H3N2 strains A/Singapore/INFIMH-17-0019/2016 (Sing/16) and A/Switzerland/9715293/2013 (Swit/13) were propagated in 9- to 11-day old special pathogen free embryonated chicken eggs (ASVBio) for 3 days at 37°C. Allantoic fluid was harvested and clarified by centrifugation at 3000 rpm for 15 min, then aliquoted and stored at -80°C until further use. A/Brisbane/59/2007 (Bris/07) and A/California/04/2009 (Cal/09) H1N1 were propagated in MDCK cells. Briefly, cells were grown to approximately 80-90% confluency in tissue culture treated flasks (Thermo Scientific) with DMEM and 5% FBS. Cells were washed with 1x PBS and inoculated with A/Brisbane/59/2007 or A/California/04/2009 in DMEM with 1µg/mL of TPCK Trypsin (Worthington Biochem), and incubated at 37°C and 5% CO2 until significant cytopathic effect was observed. Culture supernatants were harvested and pelleted by centrifugation at 3000 rpm for 15 min. Clarified supernatants were aliquoted and stored at -80°C until further use. Plaque assays were performed to determine virus stock titers. 12-well tissue culture treated plates (Thermo

Scientific) were seeded with $5x10^5$ cells per well in DMEM with 5% FBS and incubated at 37°C and 5% CO2. 24h later, cells were washed with 1x PBS (Corning), and inoculated with 10-fold serial dilutions of virus stocks for 1h. After 1h, cells and inoculum were covered with a 2x overlay medium of MEM (Gibco), 1M HEPES (Corning), 20mM L-glutamine (Corning), 7.5% sodium bicarbonate (Corning), 1% penicillin/streptomycin (GenDepot), supplemented with 1ug/ml TPCK Trypsin and incubated at 37°C and 5% CO2 for 72h. Plates were then washed with 1x PBS, fixed with acetone and methanol (80:20), stained with crystal violet, and plaques were counted visually. Stock titers were calculated as plaque forming units per milliliter (PFU/mL).

Mice, Vaccination and Challenge

6-8-week-old female BALB/c mice (Charles River) were housed in special pathogen-free cages and given water and chow ad libitum. Mice were primed on day 0 with a 1:1 ratio of 3µg of HA in 1x PBS and AddaVax (Invivogen) in 50µL doses, or 1x PBS as a negative control, administered intramuscularly (I.M.) in the right hind quarter. Mice were boosted twice, once on day 28, and again on day 42 with the same dose. Cheek bleeds were performed 14 days after the last boost for subsequent hemagglutination inhibition assays. Four weeks after the last boost, mice were anesthetized with isoflurane, and intranasally challenged with 1x104 PFU of A/Brisbane/02/2018 in 30uL doses diluted in 1x PBS. Three days post-challenge, mice were humanely euthanized, and lungs were collected and homogenized. Homogenates were clarified by centrifugation at 3000 rpm for 15 min and stored at -80°C until further use.

Hemagglutination Inhibition

Prior to performing hemagglutination inhibition (HAI) assays, all serum samples were pretreated with receptor destroying enzyme (Denka Seiken Co. Ltd) and heat inactivated at 56°C for 30 minutes. Viruses were first titrated to achieve a reciprocal hemagglutination titer of 8.

Briefly, 2-fold serial dilutions of virus were performed in 1x PBS (Corning), then incubated at room temperature for 30 min with 0.5% turkey red blood cells (tRBC) to determine the HA and appropriate dilution factor for subsequent HAI assays. Serum was then serially diluted 10-fold in 1x PBS. Viruses diluted in 1x PBS to a working stock with an HA titer of 8 were added directly to the serum and incubated for 1h at room temperature. 0.5% tRBCs in 1x PBS was then added to the serum and virus mixture and incubated for 30 min at room temperature. The reciprocal of the highest agglutinating dilution was recorded as the HAI titer for each serum sample.

Viral plaque assay

1×10^6 cells per well of MDCK cells were seeded in a six-well plate one day prior to performing the plaque assay. Next day, lung tissues were homogenized in 1 mL of DMEM (Thermo Fisher). Tissue homogenates were centrifuged at 2000 rpm for 5 min, and the supernatant was serially diluted 10-fold in DMEM supplemented with 1% penicillin-streptomycin (DMEM + P/S) (Thermo Fisher). Once the MDCK cell reached 90% confluency, the plates were washed with 2x DMEM + P/S and infected with 100 µL of each homogenate supernatant. Plates were shaken every 15 min for 1h. The supernatant was then discarded, and plates were washed twice with DMEM + P/S. Subsequently, 2 mL of 2× MEM and 0.8% agarose overlay (Cambrex, East Rutherford, NJ, USA) was added and incubated for 72h at 37 °C with 5% CO_2. The agarose was then removed, and cells were fixed with 10% buffered formalin followed by staining with 1% crystal violet (Fisher Science Education, Waltham, MA, USA) for 15 min. After thorough rinsing with water, plates were allowed to air dry, and the number of plaques were counted. Lung viral titers were estimated as plaque forming units (PFU) per mL of homogenized tissue (PFU/mL).

Focus Forming Unit Assay

$4x10^4$ cells per well of MDCK cells in DMEM and 5% FBS were seeded in 96 well plates (Costar) and incubated at 37°C and 5% CO2. 24h later, cells were washed with 1x PBS and 5-fold dilutions of clarified lung homogenates prepared in inoculation media (DMEM, 1µg/mL TPCK) were added to the cells and incubated for 1h at 37°C and 5% CO2. Following incubation, a 0.8% methylcellulose (Sigma Aldrich) overlay with 1µg/mL of TPCK and 1% penicillin/streptomycin was added to the cells, and incubated at 37°C and 5% CO2 for 18 – 24h. Following incubation, the methylcellulose overlay was aspirated, wells were washed with 1x PBS, and fixed with methanol and acetone (80:20). Cells were then incubated with primary anti-NP antibody (Biovest International Inc., cell line H16-L10-HRS), diluted 1:1000 in a blocking buffer of 5% non-fat dry milk (Rockland Immunochemicals) and 0.1% Tween20 (Promega) in 1x PBS on a platform shaker for 1h at room temperature. Plates were then washed with 1x PBS, and an anti-mouse IgG - HRP conjugated secondary (Thermo Fisher) diluted 1:1000 in blocking buffer was placed onto cells and incubated at room temperature on a platform shaker for 1h at room temperature. Following another wash, TMB substrate (Vector Laboratories) was added to the plate and incubated at room temperature for 20 min. The plate was washed with deionized water, and colorometric foci were counted under a microscope. Titers were calculated as fluorescent focus units per milliliter (FFU/mL).

Antibody competition assays by Bio-layer interferometry

Antibody competition assays against known protective epitopes were performed to get deeper insights into the antibody response elicited. Specifically, sera from mice vaccinated with designs #6, #56, #15, #57, the second-generation antigens 6_186, 6_132, 56_Sa, and the controls Cal/09 and Bris/18 were tested for competition against the antibodies P1-05, CR9114, 1F8, and

C05. All experiments were carried out with anti-penta-His sensors previously hydrated for 20 minutes in BLI buffer (PBS buffer supplemented with 0.5% BSA and 0.05% tween-20, pH 7.4). The BLI assays initiated with an equilibration phase in BLI buffer for 100s followed by antigen loading for 400s. The antigens corresponded to the HA proteins from strains Bris/18, Cal/09, Sing/86, Bris/07, NC/99, Swit/13, Sing/16, and HK/19, and the designs used for vaccination. All proteins were loaded at 40µg/mL diluted in BLI buffer. After loading, a baseline step was performed for 200s in ChonBlock blocking buffer (25% ChonBlock buffer diluted in BLI buffer) (Chondrex Inc.). Subsequently, interaction between the antigens and the sera was allowed for 600s. Sera were tested at 1:10 dilution, using pooled serum from each vaccination group diluted in Chonblock blocking buffer. Wells with no serum but only Chonblock blocking buffer were included to gather the reference antibody binding signal with no competition. After serum interaction, non-specific binding was removed by dipping the probes in Chonblock blocking buffer for 60s. Association of competing antibodies was finally carried out for 600s. Competing antibodies were tested at 20µg/mL (P1-05 and CR9114), 30µg/mL (C05), or 40µg/mL (1F8), diluting in Chonblock blocking buffer. Percentage of antibody competition was calculated as (1 – antibody-association signal in serum samples/ antibody-association signal reference sample (no serum)) × 100%. All assays were performed using a GatorPrime BLI instrument (GatorBio) at a temperature of 30°C and frequency of 10 Hz.

I. Computational analysis

T-cell epitope prediction

Identification of $CD4^+$ T cell epitopes was carried out with NetMHCIIpan- 4.1[291] and MixMHC2pred[292], [293]. The HA sequence from designs #6 and #56, as well as the viral strains Bris/18, Cal/09, and NC/99 were scanned for potential antigen presentation in human major histocompatibility complex (MHC) II alleles DRB1*01:01, DRB1*03:01, DRB1*04:01,

DRB1*07:01, DRB1*08:01, DRB1*09:01, DRB1*11:01, DRB1*13:01, and DRB1*15:01. The scanner was done with a peptide length of 15 residues, and selection threshold of <= 1% ranking score. Peptides predicted by both computational tools were considered as potential $CD4^+$ T cell epitopes.

Peptides likely to interact with MHC class I proteins, specially the human alleles A*01:01, A*02:01, A*03:01, A*24:02, B*07:02, and B*44:03, were scanned using four different computational tools: NetMHCpan-4.1[294], MHCflurry 2.0[295], MixMHCpred 2.2[296], and HLAthena[297]. The analysis was done using a peptide length of 9 residues, and a selection threshold of <= 0.5% ranking score (presentation percentile in MHCflurry predictions). Peptides identified by at least three of the prediction tools were considered as potential $CD8^+$ T cell epitopes.

Data and Materials Availability

Scripts for generating designs will be available on https://github.com/strauchlab/cossid upon publication.

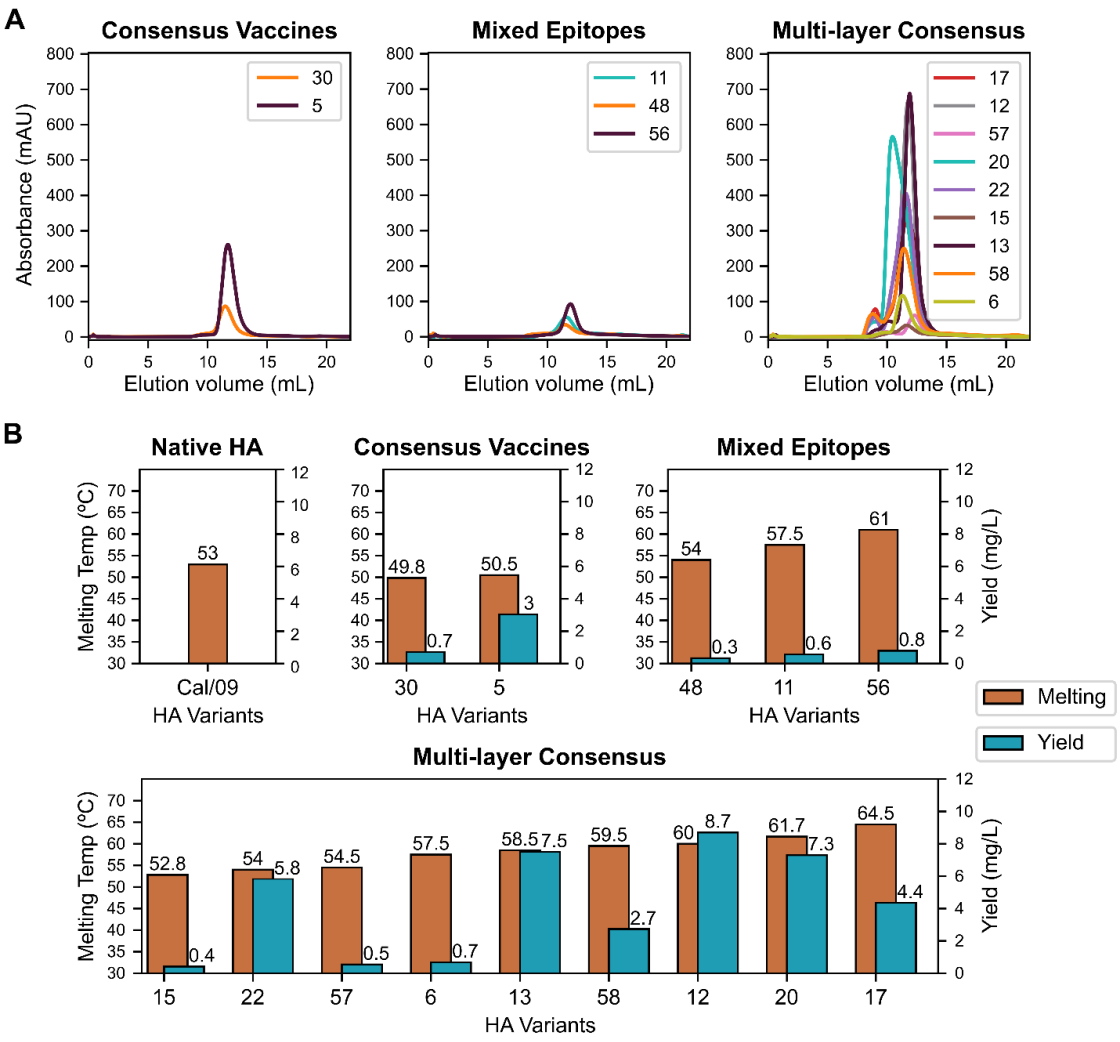


Figure 5. 1 Protein expression and thermal stability of designed HA variants. (A) Size-exclusion chromatography. **(B)** Differential scanning fluorometry and approximated expression yield. Yield values were estimated based on one single experiment while melting temperatures refer to averaged values over three repetitions. On the top of each bar are displayed the corresponding melting temperature or expression yield of each protein. The native protein A/California/04/2009 (Cal09) only displays melting temperature as expression yield was not tested. All panels are divided by the strategy used to generate the HA sequences. Computationally designed HA variants are labeled with numbers ranging from 5 to 63.

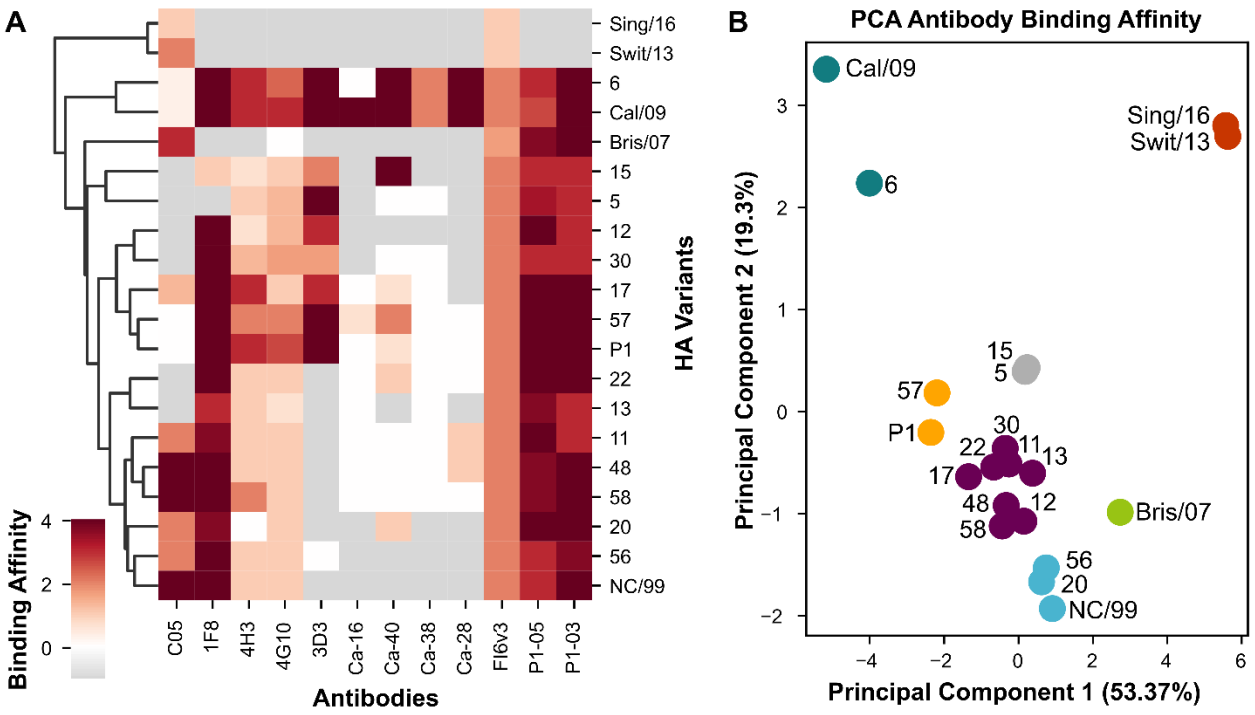


Figure 5. 2 Antibody binding profile of HA variants. (A) Heatmap of antibody binding affinity. Binding affinity is presented as a qualitative value based on the observed on-rate signal measured by Bio-layer interferometry (BLI); higher values (redder colors) indicate higher binding affinity. Zero and negative values (white/grey colors) indicate no binding. Positive and negative controls included the HAs H1 /California/07/2009 (Cal/09), A/New Caledonia/20/1999 (NC/99), A/Brisbane/59/2007 (Bris/07), the HAs H3 A/Switzerland/9715293/2013 (Swit/13), and A/Singapore/INFIMH-16-0019/2016 (Sing/16), and the COBRA antigen P1. **(B)** Principal component analysis (PCA) according to binding affinities presented on (A). Color clustering corresponds to a K-means clustering with predefined number of groups = 7. All plots were generated with Python 3.7.

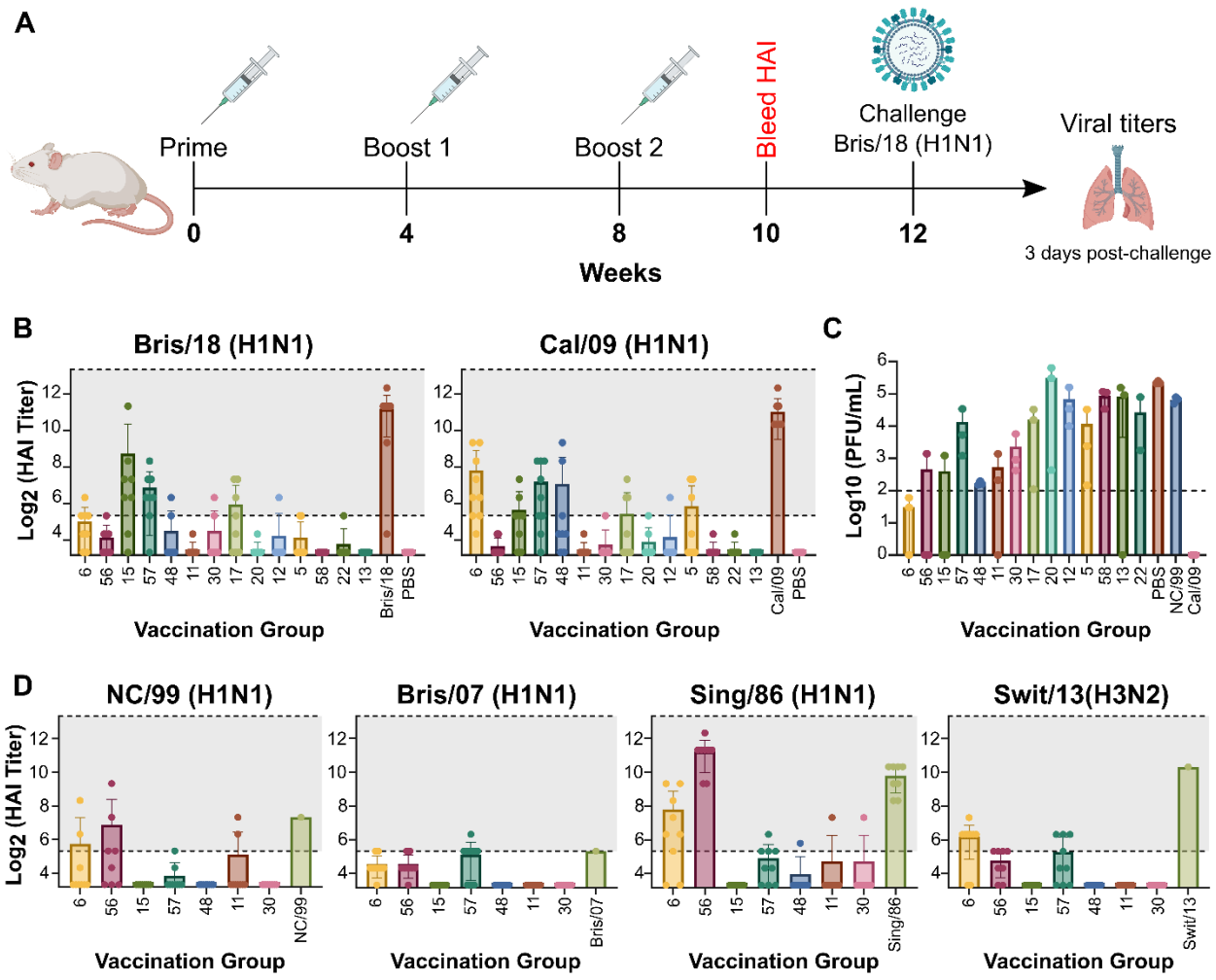


Figure 5. 3 Mice immunization schedule, hemagglutinin inhibition titers (HAI), and viral titers in lung. (A) Vaccination study regimen. BALB/C mice (n=9) were vaccinated three times with 3ug of designed antigens adjuvanted with Addavax. On week 10, mice were bled for analysis of HAI titers. Challenge with A/Brisbane/02/2018 H1N1 (Bris/18) influenza virus was performed on week 12, followed by viral lung titers assessment, three days after challenge. **(B)** and **(D)** HAI titers evaluation two weeks post-second boost. Titers are represented as log_2-transformed reciprocal dilutions. The lower dotted line represents a titer of 1:40 which was selected as the threshold of a positive HAI reaction (HAI^+). The upper dotted line indicates the maximum limit of detection. Each vaccination group displays individual mouse measurements as circles and a bar

plot indicating the mean and standard deviation of the group. The title of each panel specifies the virus used to evaluate HAI activity. Positive controls included sera from vaccinations with the HA antigen of the strain under study. Negative controls corresponded to mock vaccinations with PBS. **(C)** Viral lung titers three days post-challenge (n=3). Viral titers are presented as $\log_{10}$-transformed values of plaque forming units (PFU) per mL of homogenized tissue. The dotted line indicates the limit of detection ($\sim 1 \times 10^2$ PFU/mL) below which viral clearance is considered successful. Sera from animals vaccinated with Cal/09 were used as positive control while vaccinations with NC/99 and PBS were used as negative controls. All bar plots were produced with GraphPad Prism 9.0 [274]. The vaccination schedule was created with BioRender.com through the Institute of Bioinformatics licensing at the University of Georgia. Influenza strains abbreviations correspond to: Cal/09: A/California/04/2009, NC/99: A/New Caledonia/20/9, Bris/07: A/Brisbane/59/2007, Sing/86: A/Singapore/6/1986, Swit/13: A/Switzerland/9715293/2013.

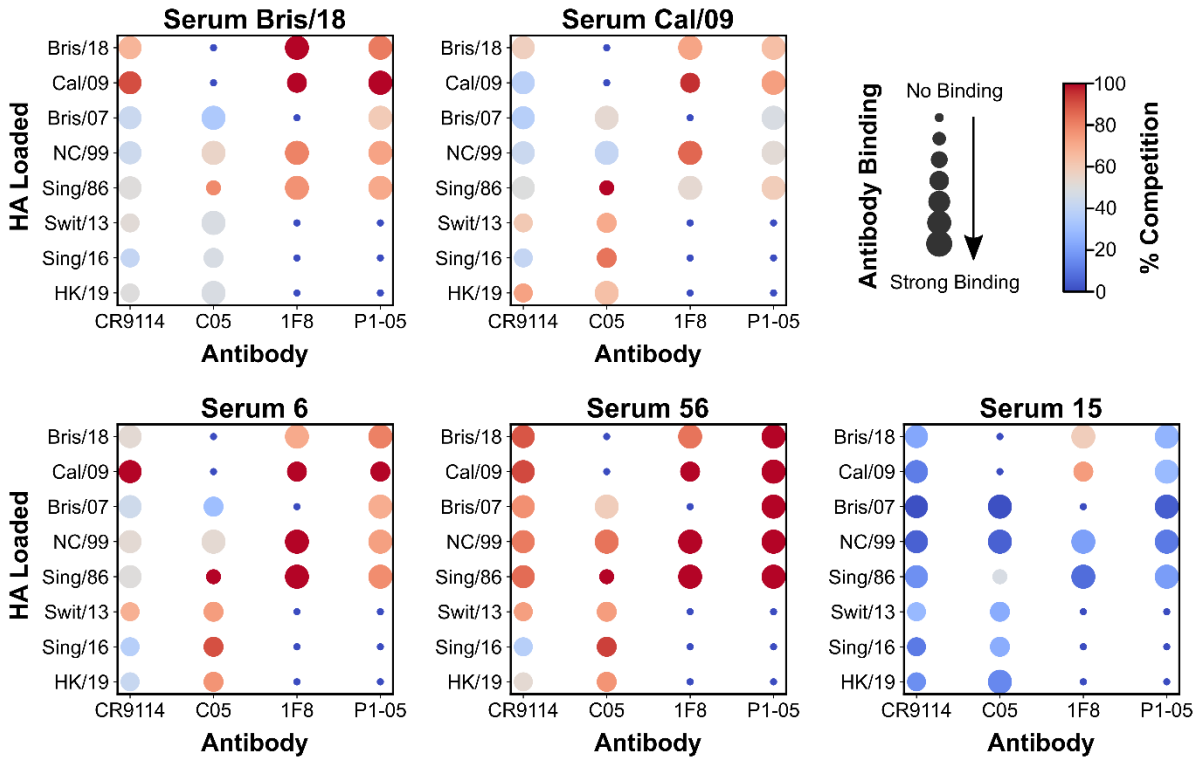


Figure 5. 4. BLI-based antibody binding competition. Pooled sera from vaccination groups displaying broad HAI activity or promising viral clearance were tested for binding competition against head (C05 and 1F8), stem (CR9114), and anchor (P1-05) antibodies. The Y-axis contains the HA proteins loaded into the BLI probes while the X-axis shows the competing antibody. The color code represents the level of competition while the circle's size shows the affinity between the loaded HA and the competing antibody in control samples with no serum. % of competition was calculated as (1 – antibody-association signal in serum samples/ antibody-association signal reference sample (no serum)) × 100%. All plots were generated with Python 3.7. Influenza strains abbreviations correspond to: Bris/18: A/Brisbane/02/2018 H1N1, Cal/09: A/California/04/2009, NC/99: A/New Caledonia/20/9, Bris/07: A/Brisbane/59/2007, Sing/86: A/Singapore/6/1986, Swit/13: A/Switzerland/9715293/2013, Sing/16: A/Singapore/INFIMH-16-0019/2016, HK/19:A/Hong Kong/45/2019.

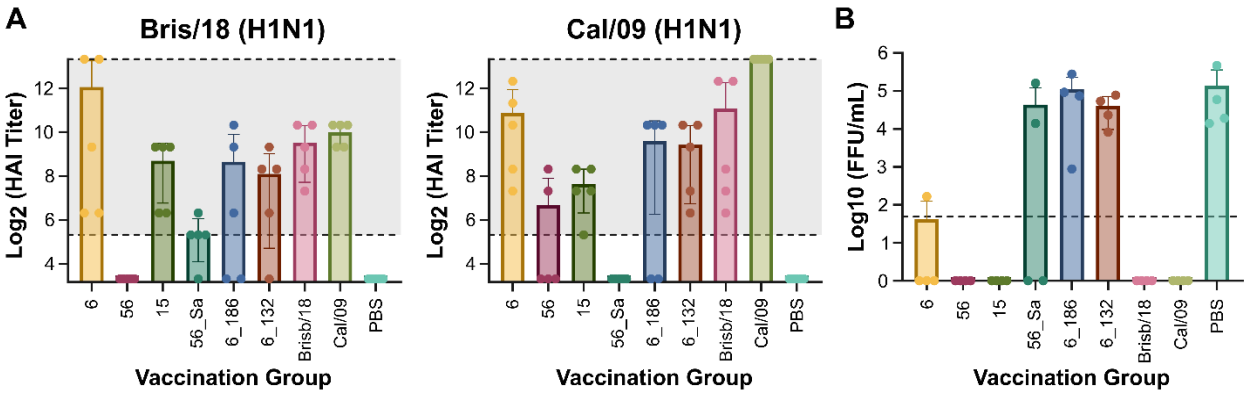


Figure 5. 5 HAI titers and viral lung titers in vaccinations with second generation of #6 and #56 antigens. (A) HAI titers evaluation two weeks post-second boost. Titers are represented as $\log_2$-transformed reciprocal dilutions. The lower dotted line represents a titer of 1:40 which was selected as the threshold of a positive HAI reaction (HAI^+). The upper dotted line indicates the maximum limit of detection. Each vaccination group displays individual mouse measurements as circles and a bar plot indicating the mean and standard deviation of the group. The title of each panel specifies the virus used to evaluate HAI activity. **(B)** Viral lung titers three days post-challenge (n=4). Viral titers are presented as $\log_{10}$-transformed values of the number of fluorescent focus units (FFU) per mL of homogenized tissue. The dotted line indicates the limit of detection ($\sim 1 \times 10^2$ FFU/mL) below which viral clearance is considered successful. Positive controls in both experiments included sera from vaccinations with Bris/18 or Cal/09, and negative controls corresponded to mock vaccinations with PBS. Pairwise statistical analyses with a two-tailed T-test indicated no significant differences between HAI titers of 6 vs 6_186 and 6 vs 6_132, in both Bris/18 and Cal/09 assays. All graphs and statistical analyses were generated with GraphPad Prism 9.0 [274].

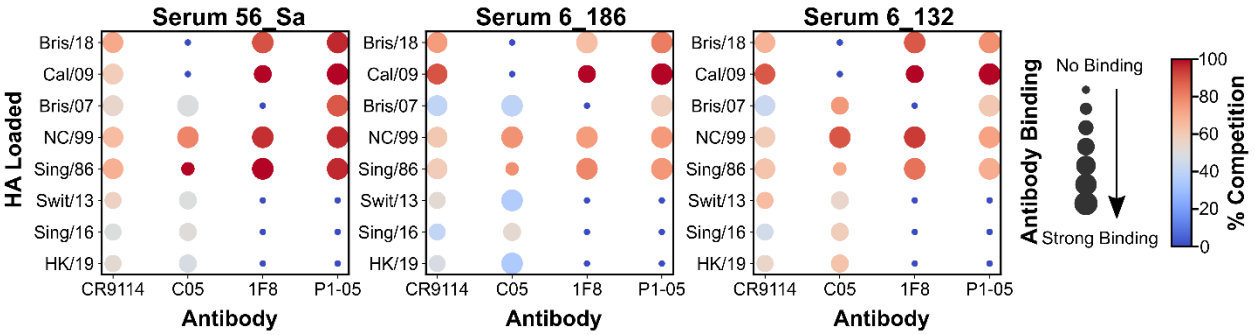
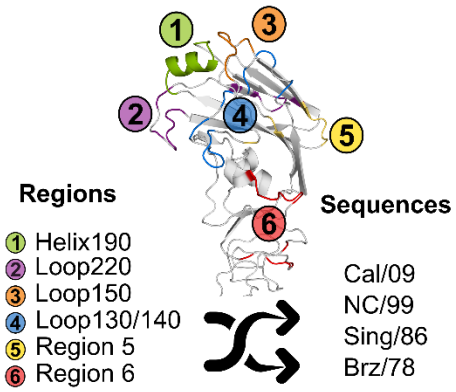


Figure 5. 6 BLI-based antibody competition against sera from vaccinations with second generation of designs #56 and #6. Pooled sera were tested for binding competition against head (C05 and 1F8), stem (CR9114), and anchor (P1-05) antibodies. The Y-axis contains the HA proteins loaded into the BLI probes while the X-axis shows the competing antibody. The color code represents the level of competition while the circle's size shows the affinity between the loaded HA and the competing antibody in control samples with no serum. % of competition was calculated as (1 – antibody-association signal in serum samples/ antibody-association signal reference sample (no serum)) × 100%. All plots were generated with Python 3.7. Influenza strains abbreviations correspond to: Bris/18: A/Brisbane/02/2018 H1N1, Cal/09: A/California/04/2009, NC/99: A/New Caledonia/20/9, Bris/07: A/Brisbane/59/2007, Sing/86: A/Singapore/6/1986, Swit/13: A/Switzerland/9715293/2013, Sing/16: A/Singapore/INFIMH-16-0019/2016, HK/19:A/Hong Kong/45/2019.

Supplementary Information



Design	Regions					
	1	2	3	4	5	6
56	Cal/09	Brz/78	NC/99	Sing/86	Cal/09	Cal/09
48	Brz/78	Sing/86	Cal/09	NC/99	Cal/09	Brz/78
11	Cal/09	Sing/86	Brz/78	NC/99	Cal/09	Cal/09

Figure 5.S 1 Mixed epitopes design strategy. Six immunogenic regions were defined at the HA head according to the crystal structures of HA proteins in complex with neutralizing antibodies. The sequence of these regions was extracted from H1N1 vaccine strains that were protective for more than six years (Brazil/11/1978 (Brz/78), Singapore/6/1986 (Sing/86), New Caledonia/20/1999 (NC/99), California/07/2009 (Cal/09)) and combined in a permutation fashion. The remaining portions of the protein, represented in grey, were replaced by the consensus sequence among the long-lasting vaccines. Expressed designs from the mixed epitopes strategy are shown on the right table, showcasing the vaccine strain from which each antigenic region sequence was obtained. The protein on display corresponds to the PDB 4m4y.

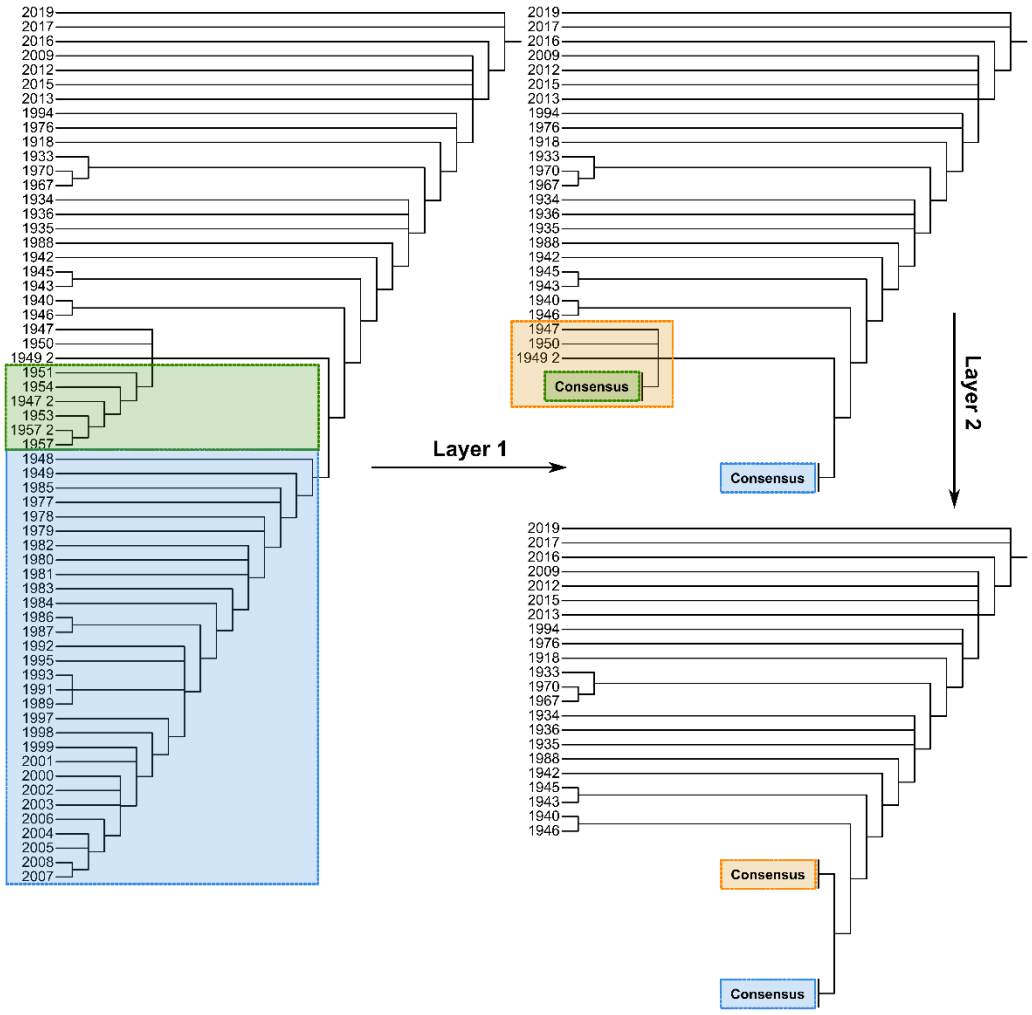


Figure 5.S 2 Multi-layer consensus strategy to design HA variant #6. Sequence clustering was carried out with RAxML [304] using per-year consensus pseudo-sequences containing only the HA head's antigenic sites. To avoid consensus bias by unbalanced number of sequences, consensus calculations are performed on tree branches displaying a high number of sequences. The resulting consensuses can be part of subsequent consensus calculation if the external branch containing them also displays a high number of sequences. Each set of consensus calculations is referred to as layers. The process is repeated until all tree branches have about the same number of sequences. At this point the final consensus is obtained. For design #6, the regions outside the antigenic sites were extracted from the Cal/09 sequence. The tree representation was drawn with iTOL [305].

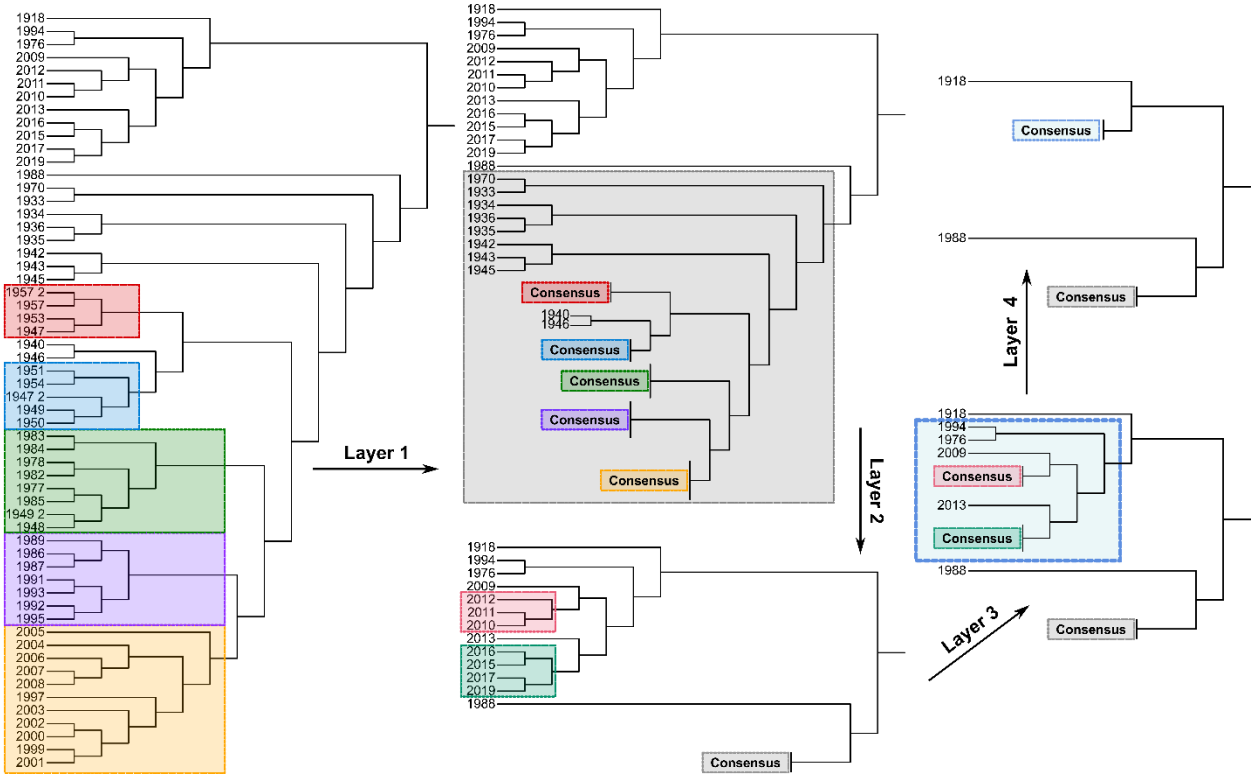


Figure 5.S 3 Multi-layer consensus strategy to design HA variant #15. Sequence clustering was carried out with MUSCLE -UPGMA [301] using full-length per-year consensus sequences. To avoid consensus bias by unbalanced number of sequences, consensus calculations are performed on tree branches displaying a high number of sequences. The resulting consensuses can be part of subsequent consensus calculation if the external branch containing them also displays a high number of sequences. Each set of consensus calculations is referred to as layers. The process is repeated until all tree branches have about the same number of sequences. At this point the final consensus is obtained. Since the final consensus sequence of design #15 contained some positions with ambiguous consensus, these positions were replaced by a consensus sequence obtained with all recommended H1N1 vaccine strains. The tree representation was drawn with iTOL [305].

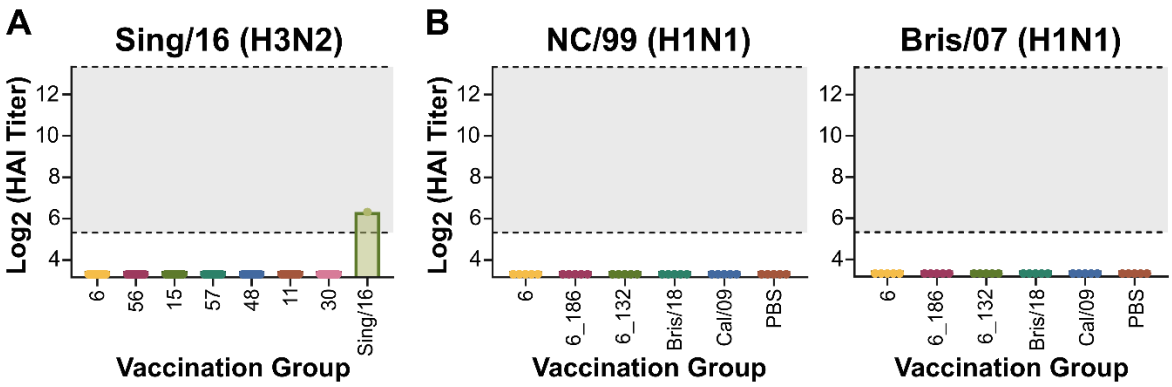


Figure 5.S 4 HAI titers two weeks post-second boost. (A) First generation HA variants. **(B)** Second generation of design #6. Titers are represented as $\log_2$-transformed reciprocal dilutions. The lower dotted line represents a titer of 1:40 which was selected as the threshold of a positive HAI reaction (HAI^+). The upper dotted line indicates the maximum limit of detection. Each vaccination group displays individual mouse measurements as circles and a bar plot indicating the mean and standard deviation of the group. The title of each panel specifies the virus used to evaluate HAI activity. All plots were produced with GraphPad Prism 9.0 [274]. Influenza strains abbreviations correspond to: NC/99: A/New Caledonia/20/9, Bris/07: A/Brisbane/59/2007, Sing/16: A/Singapore/INFIMH-16-0019/2016.

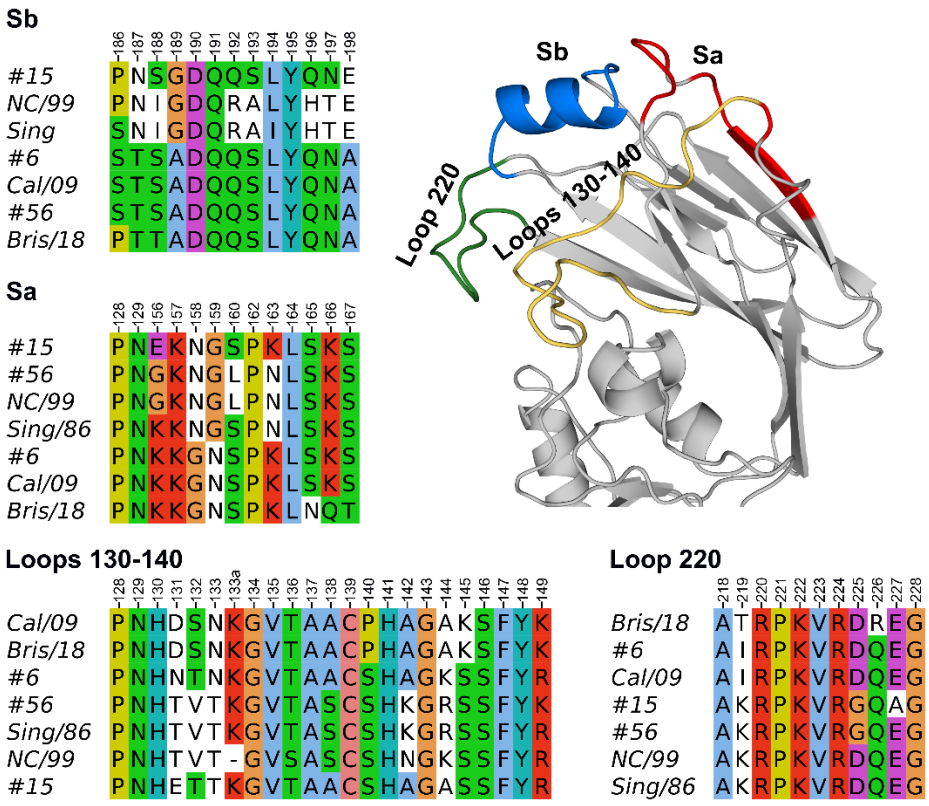


Figure 5.S 5 Sequence comparison of protective designs and native HA sequences at the HA head's main immunogenic sites. Sequence comparison was performed with Jalview [308] using the clustal color scheme. The protein on display corresponds to the PDB 4m4y and residue numbering is based on this structure.

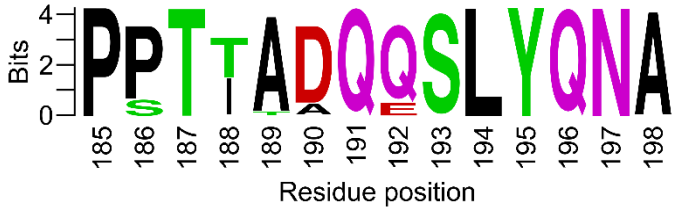


Figure 5.S 6 Sequence changes of H1 HA's Sb region from 2018 to 2022. Sequences were obtained from the GISAID database[309] and filtered based on % identity < 99.7%. Sequence logo was generated with WebLogo[310].

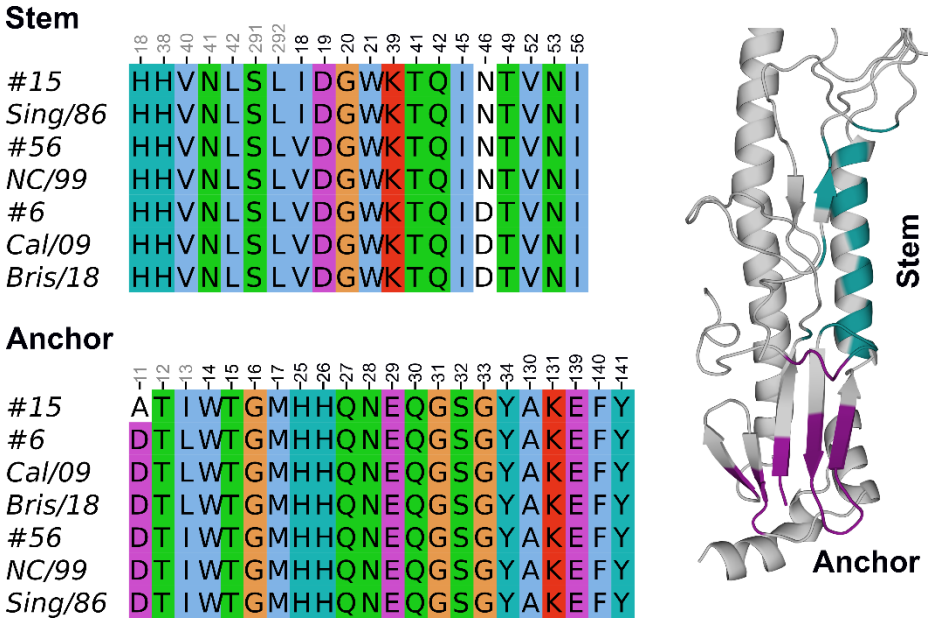
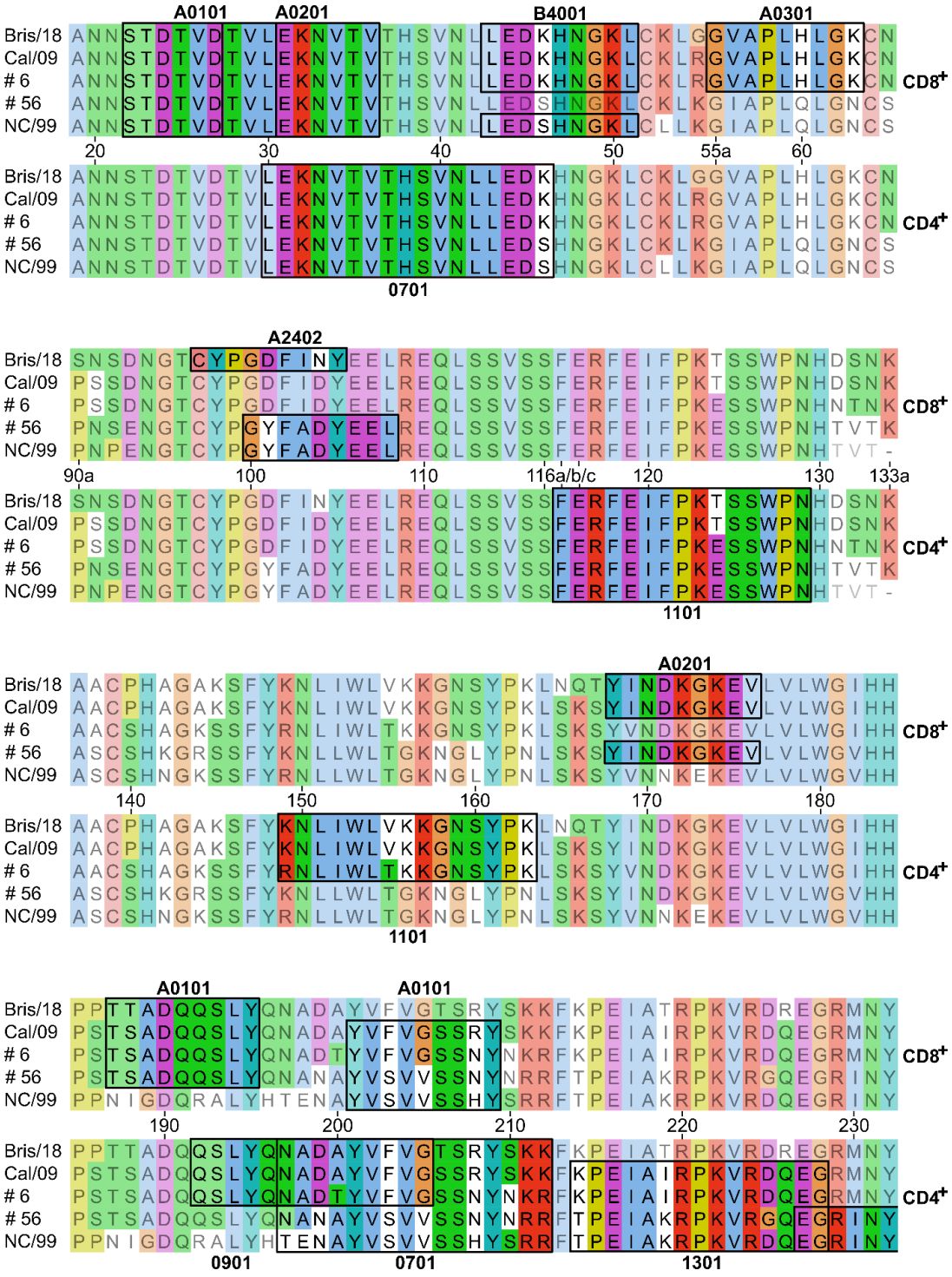


Figure 5.S 7 Sequence comparison of protective designs and native HA sequences at the HA stem immunogenic sites. Sequence comparison was performed with Jalview [308] using the clustal color scheme. The protein on display corresponds to the PDB 4m4y and residue numbering is based on this structure.



Continuation figure 5.S.8

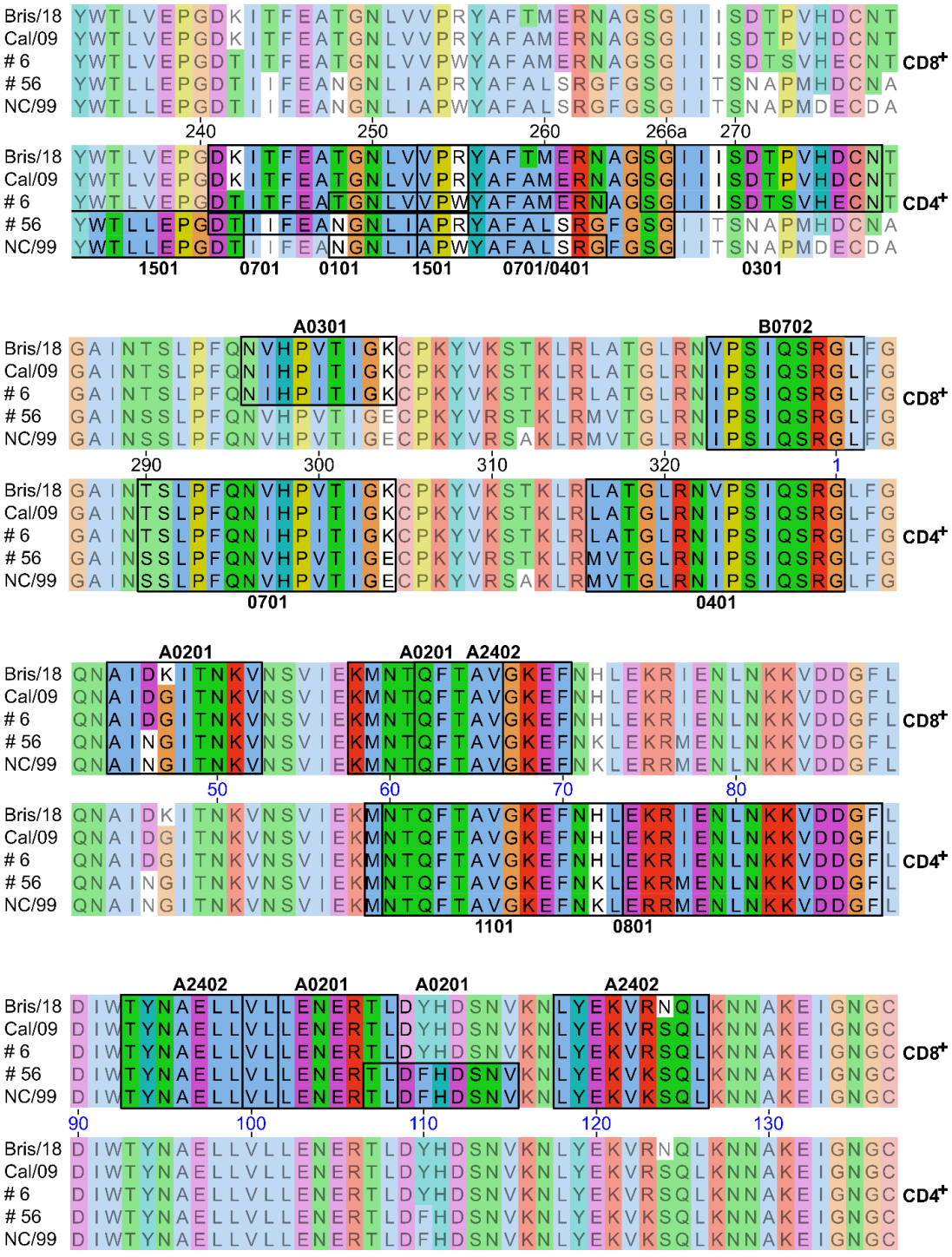


Figure 5.S 8 T-cell epitope content prediction of HA variants. Only sequence segments containing a predicted T-cell epitope are displayed. The segments are shown in blocks where the top sequence presents epitopes potentially presented by the MCH-I complex ($CD8^+$) while the

bottom sequence shows epitopes potentially presented by the MCH-II complex ($CD4^+$). The peptides are highlighted in black squares with the corresponding human allele for which the prediction was made at the top ($CD8^+$) or the bottom ($CD4^+$). All MHC-II alleles correspond to the subtype DRB1. Residue positions are shown in between sequence blocks, with black numbers representing the HA1 subunit and blue numbers representing the HA2 subunit. Residue numbers are based on the PDB 4m4y. Sequence representation was performed with Jalview using the clustal color scheme.

CHAPTER 6

SUMMARY AND CONCLUDING REMARKS

Class I fusion proteins play a pivotal role in a broad spectrum of viral infections, including those responsible for devasting pandemics. An increasing amount of evidence exist underscoring their robust immunogenic properties and consequent promise for vaccine development. Despite of this, the practical application of these proteins in vaccine formulations often encounters significant obstacles, such as structural instability and accelerated antigenic evolution. To address these challenges, we have developed three computational strategies with the aim of optimizing the attributes of class I fusion proteins for vaccine candidacy.

In our initial approach, we centered on enhancing the structural stability of the protein's most immunogenic state—the prefusion conformation. Utilizing electrostatic interaction optimization as a guiding principle, we assessed the potential energetic contributions of individual amino acid residues across various structural stages of the protein. Specifically, we introduced substitutions that favorably contributed to the energy landscape of the prefusion state, while providing no benefit to the postfusion conformation. This approach aimed to create an energetic gap between the two states, thereby inhibiting conformational transitions. The strategy was effectively implemented across a range of viral families, such as RSV, hMPV, and SARS-CoV-2. The elucidation of high-resolution structures of select candidates provided compelling validation for the atomic accuracy of our computational models. Furthermore, we demonstrated the vaccine potential of our RSV design through mouse immunization studies, which revealed protection levels comparable with those offered by a recently approved RSV vaccine.

In our second approach, we complemented our initial electrostatic optimization by strategically incorporating disulfide bonds into the design process. Using the RSV F protein as a model system, we focused on constraining the mobility of regions susceptible to conformational alterations during the transition from the pre- to postfusion state. Integration of this disulfide-bond framework into our previously electrostatically-optimized construct resulted in a thermal stability increase of approximately 12°C. Further evaluation in a cotton rat model corroborated that the protein's immunogenicity was preserved and suggested potential enhancements in the quality of the elicited antibody response.

Our third and final strategy addressed the issue of antigenic variability in the HA protein of the H1 influenza virus. Given that current influenza vaccines offer protection only when there is antigenic congruence between the circulating virus and the vaccine, we aimed to design immunogens capable of eliciting a broadly reactive antibody response. Accordingly, our most successful construct offered robust protection against a heterologous viral challenge in a mouse model and induced antibodies reactive with HA proteins from both pre- and post-pandemic periods as well as H3 subtypes. To obtain this construct, we employed consensus design coupled with the presentation of a heterogenous set of immunogenic epitopes. Specifically, we substituted regions outside of the immunogenic epitopes of HA's head with a consensus sequence derived from H1N1 vaccines known for extended protective efficacy. The remaining segments of the protein, characterized by their substantial variability, were each replaced with a unique epitope sourced from the panel of long-lasting vaccines. The resultant diversified antibody response suggests that this approach holds significant promise for the development of more universally effective influenza vaccines.

In summary, our findings indicate that our computational design strategies successfully met their intended objectives, yielding a set of class I fusion protein variants with enhanced characteristics. The application of these algorithms, either singularly or synergistically, holds potential for mitigating the inherent limitations of these proteins as vaccine candidates—primarily their structural instability and restricted antigenic reactivity. We recognize, however, that additional validation studies are needed to confirm the robustness of our computational methodologies. For instance, our structural stabilization approaches are contingent upon the availability of resolved protein structures. It remains an open question whether our algorithms would produce similarly accurate results when applied to computational models in the absence of crystallographically determined structures. Likewise, our strategies targeting antigenic variability are dependent on the availability of representative sequences that exhibit significant protective attributes. The generalizability of this approach to other influenza subtypes could be conditional on the existence of a relevant panel of antigenic sequences displaying key immunogenic features. Nonetheless, this study serves as a foundational platform for the development of more advanced methods, contributing valuable insights for the ongoing quest to improve vaccine design against a multitude of viral infections.

References

[1] H. R. van Doorn and H. Yu, “Viral Respiratory Infections,” in *Hunter’s Tropical Medicine and Emerging Infectious Diseases.*, 2020, pp. 284–288.

[2] WHO, “WHO Mortality Database- Respiratory Infections.” https://platform.who.int/mortality/themes/theme-details/topics/topic-details/MDB/respiratory-infections.

[3] Y. Li *et al.*, “Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis,” *Lancet*, vol. 399, no. 10340, pp. 2047–2064, 2022, doi: 10.1016/S0140-6736(22)00478-0.

[4] N. I. Mazur *et al.*, “Global Respiratory Syncytial Virus-Related Infant Community Deaths,” *Clin. Infect. Dis.*, vol. 73, pp. S229–S237, 2021, doi: 10.1093/cid/ciab528.

[5] M. Burk, K. El-Kersh, M. Saad, T. Wiemken, J. Ramirez, and R. Cavallazzi, “Viral infection in community-acquired pneumonia: A systematic review and meta-analysis,” *Eur. Respir. Rev.*, vol. 25, no. 140, pp. 178–188, 2016, doi: 10.1183/16000617.0076-2015.

[6] Centers for Disease Control and Prevention, “Vaccine Effectiveness: How Well Do Flu Vaccines Work?,” *2023*. https://www.cdc.gov/flu/vaccines-work/vaccineeffect.htm.

[7] K. Houser and K. Subbarao, “Influenza Vaccines: Challenges and Solutions,” *Cell Host Microbe*, vol. 17, no. 3, pp. 295–300, 2015, doi: doi:10.1016/j.chom.2015.02.012.

[8] CDC, “Human Metapneumovirus (HMPV),” 2023. https://www.cdc.gov/ncird/human-metapneumovirus.html#:~:text=Currently%2C there is no specific,no vaccine to prevent HMPV.

[9] CDC, “Human Parainfluenza Viruses (HPIVs),” 2022. https://www.cdc.gov/parainfluenza/prevention-treatment.html#:~:text=Treatment-,Prevention,human parainfluenza viruses (HPIV).

[10] A. Pabis, R. J. Rawle, and P. M. Kasson, “Influenza hemagglutinin drives viral entry via two sequential intramembrane mechanisms,” *PNAS*, 2020, doi: 10.1101/2020.01.30.926816.

[11] J. Shang *et al.*, “Cell entry mechanisms of SARS-CoV-2,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 21, pp. 11727–11734, May 2020, doi: 10.1073/pnas.2003138117.

[12] A. Chang and R. E. Dutch, “Paramyxovirus fusion and entry: multiple paths to a common end.,” *Viruses*, vol. 4, no. 4, pp. 613–36, 2012, doi: 10.3390/v4040613.

[13] M. Aggarwal and R. K. Plemper, “Structural insight into paramyxovirus and pneumovirus entry inhibition,” *Viruses*, vol. 12, no. 3. 2020, doi: 10.3390/v12030342.

[14] F. A. Rey and S. M. Lok, “Common Features of Enveloped Viruses and Implications for Immunogen Design for Next-Generation Vaccines,” *Cell*, vol. 172, no. 6. Cell Press, pp. 1319–1334, Mar. 08, 2018, doi: 10.1016/j.cell.2018.02.054.

[15] G. B. E. 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.*, vol. 115, no. 48, pp. 12265–12270, Nov. 2018, doi: 10.1073/pnas.1811980115.

[16] H. V. Dang *et al.*, “An antibody against the F glycoprotein inhibits Nipah and Hendra virus infections,” *Nat. Struct. Mol. Biol.*, vol. 26, no. 10, pp. 980–987, Oct. 2019, doi: 10.1038/s41594-019-0308-9.

[17] J. O. Ngwuta *et al.*, “Prefusion F-specific antibodies determine the magnitude of RSV

neutralizing activity in human sera," *Sci. Transl. Med.*, vol. 7, no. 309, p. 309ra162, Oct. 2015, doi: 10.1126/scitranslmed.aac4241.

[18] R. J. Loomis *et al.*, "Structure-Based Design of Nipah Virus Vaccines: A Generalizable Approach to Paramyxovirus Immunogen Development," *Front. Immunol.*, vol. 11, no. June, pp. 1–16, 2020, doi: 10.3389/fimmu.2020.00842.

[19] J. S. McLellan *et al.*, "Structure-Based Design of a Fusion Glycoprotein Vaccine for Respiratory Syncytial Virus," *Science*, vol. 342, no. 6158, p. 592, Nov. 2013, doi: 10.1126/SCIENCE.1243283.

[20] L. R. Baden *et al.*, "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine," *N. Engl. J. Med.*, vol. 384, no. 5, pp. 403–416, Feb. 2021, doi: 10.1056/NEJMoa2035389.

[21] S. C. Harrison, "Viral membrane fusion," *Nat. Struct. Mol. Biol.*, vol. 15, no. 7, pp. 690–698, 2008, doi: doi:10.1038/nsmb.1456.

[22] B. S. Graham, "Vaccine development for respiratory syncytial virus," *Current Opinion in Virology*, vol. 23. NIH Public Access, pp. 107–112, 2017, doi: 10.1016/j.coviro.2017.03.012.

[23] J. Falloon *et al.*, "An Adjuvanted, Postfusion F Protein–Based Vaccine Did Not Prevent Respiratory Syncytial Virus Illness in Older Adults," *J. Infect. Dis.*, vol. 216, no. 11, pp. 1362–1370, Dec. 2017, doi: 10.1093/infdis/jix503.

[24] B. Schweiger, I. Zadow, and R. Heckler, "Antigenic drift and variability of influenza viruses," *Med. Microbiol. Immunol.*, vol. 191, no. 3–4, pp. 133–138, 2002, doi: 10.1007/s00430-002-0132-3.

[25] R. Servín-Blanco, R. Zamora-Alvarado, G. Gevorkian, and K. Manoutcharian, "Antigenic

variability: Obstacles on the road to vaccines against traditionally difficult targets," *Human Vaccines and Immunotherapeutics*, vol. 12, no. 10. Taylor and Francis Inc., pp. 2640–2648, Oct. 02, 2016, doi: 10.1080/21645515.2016.1191718.

[26] G. I. López-Cortés *et al.*, "The Spike Protein of SARS-CoV-2 Is Adapting Because of Selective Pressures," *Vaccines*, vol. 10, no. 6, 2022, doi: 10.3390/vaccines10060864.

[27] L. Du, Y. He, Y. Zhou, S. Liu, B. J. Zheng, and S. Jiang, "The spike protein of SARS-CoV - A target for vaccine and therapeutic development," *Nature Reviews Microbiology*, vol. 7, no. 3. Nature Publishing Group, pp. 226–236, Feb. 09, 2009, doi: 10.1038/nrmicro2090.

[28] Y. He, Y. Zhou, P. Siddiqui, and S. Jiang, "Inactivated SARS-CoV vaccine elicits high titers of spike protein-specific antibodies that block receptor binding and virus entry," *Biochem. Biophys. Res. Commun.*, vol. 325, no. 2, pp. 445–452, Dec. 2004, doi: 10.1016/j.bbrc.2004.10.052.

[29] J. Marcandalli *et al.*, "Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus," *Cell*, vol. 176, no. 6, pp. 1420-1431.e17, Mar. 2019, doi: 10.1016/j.cell.2019.01.046.

[30] A. B. Vogel *et al.*, "Self-Amplifying RNA Vaccines Give Equivalent Protection against Influenza to mRNA Vaccines but at Much Lower Doses," *Mol. Ther.*, vol. 26, no. 2, pp. 446–455, Feb. 2018, doi: 10.1016/j.ymthe.2017.11.017.

[31] M. N. Widjojoatmodjo *et al.*, "Recombinant low-seroprevalent adenoviral vectors Ad26 and Ad35 expressing the respiratory syncytial virus (RSV) fusion protein induce protective immunity against RSV infection in cotton rats," *Vaccine*, vol. 33, no. 41, pp. 5406–5414, Oct. 2015, doi: 10.1016/j.vaccine.2015.08.056.

[32] M. Kielian and F. A. Rey, “Virus membrane-fusion proteins: More than one way to make a hairpin,” *Nat. Rev. Microbiol.*, vol. 4, no. 1, pp. 67–76, 2006, doi: 10.1038/nrmicro1326.

[33] A. Albertini, S. Bressanelli, J. Lepault, and Y. Gaudin, *Structure and Working of Viral Fusion Machinery*, vol. 68. Academic Press Inc., 2011.

[34] D. J. Schibli and W. Weissenhorn, “Class I and class II viral fusion protein structures reveal similar principles in membrane fusion,” *Mol. Membr. Biol.*, vol. 21, no. 6, pp. 361–371, 2004, doi: 10.1080/09687860400017784.

[35] P. A. Bullough, F. M. Hughson, J. J. Skehel, and D. C. Wiley, “Structure of influenza haemagglutinin at the pH of membrane fusion,” *Nature*, vol. 371, no. 6492, pp. 37–43, 1994, doi: 10.1038/371037a0.

[36] M. Kielian, “Class II virus membrane fusion proteins,” *Virology*, vol. 344, no. 1. Academic Press, pp. 38–47, Jan. 2006, doi: 10.1016/j.virol.2005.09.036.

[37] B. Podbilewicz, “Virus and Cell Fusion Mechanisms,” *Annu. Rev. Cell Dev. Biol*, vol. 30, pp. 111–139, 2014, doi: 10.1146/annurev-cellbio-101512-122422.

[38] Y. Gaudin, “Reversibility in Fusion Protein Conformational Changes The Intriguing Case of Rhabdovirus-Induced Membrane Fusion,” in *Fusion of Biological Membranes and Related Problems*, Boston: Kluwer Academic Publishers, 2000, pp. 379–408.

[39] M. Backovic and T. S. Jardetzky, “Class III viral membrane fusion proteins,” in *Current Opinion in Structural Biology*, vol. 19, no. 2, Dordrecht: NIH Public Access, 2009, pp. 189–196.

[40] J. M. White, S. E. Delos, M. Brecher, and K. Schornberg, “Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme,” *Critical Reviews*

in Biochemistry and Molecular Biology, vol. 43, no. 3. Taylor and Francis Inc., pp. 189–219, 2008, doi: 10.1080/10409230802058320.

[41] J. M. White, S. E. Delos, M. Brecher, and K. Schornberg, "Structures and Mechanisms of Viral Membrane Fusion Proteins," *Crit. Rev. Biochem. Mol. Biol.*, vol. 43, no. 3, pp. 189–219, 2008, doi: 10.1080/10409230802058320.Structures.

[42] G. Izaguirre, "The Proteolytic Regulation of Virus Cell Entry by Furin and Other Proprotein Convertases," *Viruses*, vol. 11, no. 9. NLM (Medline), Sep. 2019, doi: 10.3390/v11090837.

[43] S. Belouzard, V. C. Chu, and G. R. Whittaker, "Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 14, pp. 5871–5876, Apr. 2009, doi: 10.1073/pnas.0809524106.

[44] W. Weissenhorn, A. Dessen, S. C. Harrison, J. J. Skehel, and D. C. Wiley, "Atomic structure of the ectodomain from HIV-1 gp41," *Nature*, vol. 387, no. 6631, pp. 426–430, May 1997, doi: 10.1038/387426a0.

[45] J. J. Skehel *et al.*, "Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus mediated membrane fusion," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 79, no. 4 I, pp. 968–972, 1982, doi: 10.1073/pnas.79.4.968.

[46] R. W. H. Ruigrok, S. R. Martin, S. A. Wharton, J. J. Skehel, P. M. Bayley, and D. C. Wiley, "Conformational changes in the hemagglutinin of influenza virus which accompany heat-induced fusion of virus with liposomes," *Virology*, vol. 155, no. 2, pp. 484–497, 1986, doi: 10.1016/0042-6822(86)90210-2.

[47] J. Chen *et al.*, "A soluble domain of the membrane-anchoring chain of influenza virus hemagglutinin (HA2) folds in Escherichia coli into the low-pH-induced conformation,"

Proc. Natl. Acad. Sci. U. S. A., vol. 92, no. 26, pp. 12205–12209, Dec. 1995, doi: 10.1073/pnas.92.26.12205.

[48] J. Chen, K. H. Lee, D. A. Steinhauer, D. J. Stevens, J. J. Skehel, and D. C. Wiley, "Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation," *Cell*, vol. 95, no. 3, pp. 409–417, Oct. 1998, doi: 10.1016/S0092-8674(00)81771-7.

[49] T. C. Marcink *et al.*, "Subnanometer structure of an enveloped virus fusion complex on viral surface reveals new entry mechanisms," *Sci. Adv.*, vol. 9, no. 6, 2023, doi: 10.1126/sciadv.ade2727.

[50] D. M. Eckert and P. S. Kim, "Mechanisms of viral membrane fusion and its inhibition," *Annual Review of Biochemistry*, vol. 70. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, pp. 777–810, Nov. 2001, doi: 10.1146/annurev.biochem.70.1.777.

[51] T. Matthews, M. Salgo, M. Greenberg, J. Chung, R. DeMasi, and D. Bolognesi, "Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes," *Nature Reviews Drug Discovery*, vol. 3, no. 3. Nature Publishing Group, pp. 215–225, 2004, doi: 10.1038/nrd1331.

[52] I. Muñoz-Barroso, K. Salzwedel, E. Hunter, and R. Blumenthal, "Role of the Membrane-Proximal Domain in the Initial Stages of Human Immunodeficiency Virus Type 1 Envelope Glycoprotein-Mediated Membrane Fusion," *J. Virol.*, vol. 73, no. 7, pp. 6089–6092, Jul. 1999, doi: 10.1128/jvi.73.7.6089-6092.1999.

[53] C. Wild, T. Greenwell, and T. Matthews, "A Synthetic Peptide from HIV-1 gp41 Is a Potent

Inhibitor of Virus-Mediated Cell—Cell Fusion," *AIDS Research and Human Retroviruses*, vol. 9, no. 11. pp. 1051–1053, Mar. 1993, doi: 10.1089/aid.1993.9.1051.

[54] E. Leikina, C. Ramos, I. Markovic, J. Zimmerberg, and L. V. Chernomordik, "Reversible stages of the low-pH-triggered conformational change in influenza virus hemagglutinin," *EMBO J.*, vol. 21, no. 21, pp. 5701–5710, Nov. 2002, doi: 10.1093/emboj/cdf559.

[55] W. Weissenhorn, A. Hinz, and Y. Gaudin, "Virus membrane fusion," *FEBS Lett.*, vol. 581, no. 11, pp. 2150–2155, 2007, doi: 10.1016/j.febslet.2007.01.093.

[56] H. Y. Liu and P. L. Yang, "Small-Molecule Inhibition of Viral Fusion Glycoproteins," *Annual Review of Virology*, vol. 8. pp. 459–489, 2021, doi: 10.1146/annurev-virology-022221-063725.

[57] D. C. Ekiert *et al.*, "Cross-neutralization of influenza A viruses mediated by a single antibody loop," *Nature*, vol. 489, no. 7417, pp. 526–532, Sep. 2012, doi: 10.1038/nature11414.

[58] J. S. McLellan *et al.*, "Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody," *Science (80-.).*, vol. 340, no. 6136, pp. 1113–1117, May 2013, doi: 10.1126/SCIENCE.1234914/SUPPL_FILE/MCLELLAN.SM.PDF.

[59] I. Rossey, J. S. McLellan, X. Saelens, and B. Schepens, "Clinical Potential of Prefusion RSV F-specific Antibodies," *Trends in Microbiology*, vol. 26, no. 3. Elsevier Current Trends, pp. 209–219, Mar. 01, 2018, doi: 10.1016/j.tim.2017.09.009.

[60] E. M. Strauch *et al.*, "Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site," *Nat. Biotechnol.*, vol. 35, no. 7, pp. 667–671, 2017, doi: 10.1038/nbt.3907.

[61] S. J. Fleishman *et al.*, “Computational design of proteins targeting the conserved stem region of influenza hemagglutinin,” *Science (80-.).*, vol. 332, no. 6031, pp. 816–821, May 2011, doi: 10.1126/science.1202617.

[62] X. Yao *et al.*, “Structural basis of potent and broad HIV-1 fusion inhibitor CP32M,” *J. Biol. Chem.*, vol. 287, no. 32, pp. 26618–26629, 2012, doi: 10.1074/jbc.M112.381079.

[63] R. J. Russell *et al.*, “Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 46, pp. 17736–17741, Nov. 2008, doi: 10.1073/PNAS.0807142105/SUPPL_FILE/0807142105SI.PDF.

[64] M. B. Battles and J. S. McLellan, “Respiratory syncytial virus entry and how to block it,” *Nature Reviews Microbiology*, vol. 17, no. 4. Nature Publishing Group, pp. 233–245, Apr. 01, 2019, doi: 10.1038/s41579-019-0149-x.

[65] R. G. Cox and J. V. Williams, “Breaking in: Human metapneumovirus fusion and entry,” *Viruses*, vol. 5, no. 1. Multidisciplinary Digital Publishing Institute (MDPI), pp. 192–210, Jan. 16, 2013, doi: 10.3390/v5010192.

[66] R. A. Karron *et al.*, “Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vitro: Clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 25, pp. 13961–13966, Dec. 1997, doi: 10.1073/pnas.94.25.13961.

[67] J. E. Crowe *et al.*, “Live subgroup B respiratory syncytial virus vaccines that are attenuated, genetically stable, and immunogenic in rodents and nonhuman primates,” *J. Infect. Dis.*, vol. 173, no. 4, pp. 829–839, 1996, doi: 10.1093/infdis/173.4.829.

[68] S. Biacchesi *et al.*, “Recombinant Human Metapneumovirus Lacking the Small

Hydrophobic SH and/or Attachment G Glycoprotein: Deletion of G Yields a Promising Vaccine Candidate," *J. Virol.*, vol. 78, no. 23, pp. 12877–12887, 2004, doi: 10.1128/jvi.78.23.12877-12887.2004.

[69] S. A. Feldman, S. Audet, and J. A. Beeler, "The Fusion Glycoprotein of Human Respiratory Syncytial Virus Facilitates Virus Attachment and Infectivity via an Interaction with Cellular Heparan Sulfate," *J. Virol.*, vol. 74, no. 14, pp. 6442–6447, Jul. 2000, doi: 10.1128/JVI.74.14.6442-6447.2000/ASSET/64279ACC-0E41-47EB-B41D-6D22A9947394/ASSETS/GRAPHIC/JV1400217006.JPEG.

[70] B. R. Heminway *et al.*, "Analysis of respiratory syncytial virus F, G, and SH proteins in cell fusion," *Virology*, vol. 200, no. 2, pp. 801–805, 1994, doi: doi: 10.1006/viro.1994.1245.

[71] Y. Shirogane *et al.*, "Efficient Multiplication of Human Metapneumovirus in Vero Cells Expressing the Transmembrane Serine Protease TMPRSS2," *J. Virol.*, vol. 82, no. 17, pp. 8942–8946, Sep. 2008, doi: 10.1128/jvi.00676-08.

[72] G. Bolt, L. Ø. Pedersen, and H. H. Birkeslund, "Cleavage of the respiratory syncytial virus fusion protein is required for its surface expression: Role of furin," *Virus Res.*, vol. 68, no. 1, pp. 25–33, Jun. 2000, doi: 10.1016/S0168-1702(00)00149-0.

[73] L. González-Reyes *et al.*, "Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 17, pp. 9859–9864, Aug. 2001, doi: 10.1073/pnas.151098198.

[74] M. Begoña Ruiz-Argüello *et al.*, "Effect of proteolytic processing at two distinct sites on shape and aggregation of an anchorless fusion protein of human respiratory syncytial virus and fate of the intervening segment," *Virology*, vol. 298, no. 2, pp. 317–326, Jul. 2002, doi:

10.1006/viro.2002.1497.

[75] N. D. Day *et al.*, “Contribution of cysteine residues in the extracellular domain of the F protein of human respiratory syncytial virus to its function,” *Virol. J.*, vol. 3, p. 34, May 2006, doi: 10.1186/1743-422X-3-34.

[76] M. S. A. Gilman *et al.*, “Characterization of a Prefusion-Specific Antibody That Recognizes a Quaternary, Cleavage-Dependent Epitope on the RSV Fusion Glycoprotein,” *PLoS Pathog.*, vol. 11, no. 7, 2015, doi: 10.1371/journal.ppat.1005035.

[77] A. Krarup *et al.*, “A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism,” *Nat. Commun.*, vol. 6, p. 8143, Sep. 2015, doi: 10.1038/ncomms9143.

[78] N. Srinivasakumar, P. L. Ogra, and T. D. Flanagan, “Characteristics of fusion of respiratory syncytial virus with HEp-2 cells as measured by R18 fluorescence dequenching assay,” *J. Virol.*, vol. 65, no. 8, pp. 4063–4069, 1991, doi: 10.1128/jvi.65.8.4063-4069.1991.

[79] J. S. Kahn, M. J. Schnell, L. Buonocore, and J. K. Rose, “Recombinant vesicular stomatitis virus expressing respiratory syncytial virus (RSV) glycoproteins: RSV fusion protein can mediate infection and cell fusion,” *Virology*, vol. 254, no. 1, pp. 81–91, Feb. 1999, doi: 10.1006/viro.1998.9535.

[80] R. M. Schowalter, S. E. Smith, and R. E. Dutch, “Characterization of Human Metapneumovirus F Protein-Promoted Membrane Fusion: Critical Roles for Proteolytic Processing and Low pH,” *J. Virol.*, vol. 80, no. 22, pp. 10931–10941, 2006, doi: 10.1128/JVI.01287-06.

[81] M. Schowalter, S. E. Smith, and R. E. Dutch, “Low-pH-Induced Membrane Fusion

Mediated by Human Metapneumovirus F Protein Is a Rare, Strain-Dependent Phenomenon," *J. Virol.*, vol. 82, pp. 10931–10941, 2008, doi: 10.1128/JVI.00472-08.

[82] J. T. Kinder, E. M. Klimyte, A. Chang, J. V Williams, and R. E. Dutch, "Human metapneumovirus fusion protein triggering: Increasing complexities by analysis of new HMPV fusion proteins," 2019, doi: 10.1016/j.virol.2019.03.003.

[83] M. S. A. Gilman *et al.*, "Rapid profiling of RSV antibody repertoires from the memory B cells of naturally infected adult donors," *Sci. Immunol.*, vol. 1, no. 6, Dec. 2016, doi: 10.1126/SCIIMMUNOL.AAJ1879.

[84] M. S. A. A. Gilman *et al.*, "Characterization of a Prefusion-Specific Antibody That Recognizes a Quaternary, Cleavage-Dependent Epitope on the RSV Fusion Glycoprotein," *PLoS Pathog.*, vol. 11, no. 7, pp. 1–17, Jul. 2015, doi: 10.1371/journal.ppat.1005035.

[85] D. Corti *et al.*, "Cross-neutralization of four paramyxoviruses by a human monoclonal antibody," *Nature*, vol. 501, no. 7467, pp. 439–443, 2013, doi: 10.1038/nature12442.

[86] J. S. McLellan, M. Chen, A. Kim, Y. Yang, B. S. Graham, and P. D. Kwong, "Structural basis of respiratory syncytial virus neutralization by motavizumab," *Nat. Struct. Mol. Biol.*, vol. 17, no. 2, pp. 248–250, Feb. 2010, doi: 10.1038/nsmb.1723.

[87] 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.*, vol. 84, no. 23, pp. 12236–12244, Dec. 2010, doi: 10.1128/jvi.01579-10.

[88] X. Xiao *et al.*, "Characterization of potent RSV neutralizing antibodies isolated from human memory B cells and identification of diverse RSV/hMPV cross-neutralizing epitopes," *MAbs*, vol. 11, no. 8, pp. 1415–1427, 2019, doi: 10.1080/19420862.2019.1654304.

[89] J. E. Schuster *et al.*, “A broadly neutralizing human monoclonal antibody exhibits in vivo efficacy against both human metapneumovirus and respiratory syncytial virus,” in *Journal of Infectious Diseases*, 2015, vol. 211, no. 2, pp. 216–225, doi: 10.1093/infdis/jiu307.

[90] M. Sun, L. Guo, L. Li, L. Liu, and T. Zhang, “Neutralising antibodies against human metapneumovirus,” 2023, doi: 10.1016/S2666-5247(23)00134-9.

[91] C. G. Rappazzo *et al.*, “Potently neutralizing and protective anti-human metapneumovirus antibodies target diverse sites on the fusion glycoprotein,” *Immunity*, vol. 55, no. 9, pp. 1710-1724.e8, Sep. 2022, doi: 10.1016/J.IMMUNI.2022.07.003.

[92] X. Xiao *et al.*, “Profiling of hMPV F-specific antibodies isolated from human memory B cells,” *Nat. Commun.*, vol. 13, no. 1, 2022, doi: 10.1038/s41467-022-30205-x.

[93] S. A. Rush *et al.*, “Characterization of prefusion-F-specific antibodies elicited by natural infection with human metapneumovirus,” *Cell Rep.*, vol. 40, no. 12, p. 111399, Sep. 2022, doi: 10.1016/j.celrep.2022.111399.

[94] J. Huang, D. Diaz, and J. J. Mousa, “Antibody recognition of the Pneumovirus fusion protein trimer interface,” *bioRxiv*, vol. 30602, no. 1, p. 2020.05.20.107508, 2020, doi: 10.1101/2020.05.20.107508.

[95] M. B. Battles *et al.*, “Structure and immunogenicity of pre-fusion-stabilized human metapneumovirus F glycoprotein,” *Nat. Commun.*, vol. 8, no. 1, pp. 1–11, Dec. 2017, doi: 10.1038/s41467-017-01708-9.

[96] I. A. Wilson, J. J. Skehel, and D. C. Wiley, “Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution,” *Nat. 1981 2895796*, vol. 289, no. 5796, pp. 366–373, 1981, doi: 10.1038/289366a0.

[97] T. Stegmann, “Membrane fusion mechanisms: The influenza hemagglutinin paradigm and its implications for intracellular fusion,” *Traffic*, vol. 1, no. 8, pp. 598–604, 2000, doi: 10.1034/J.1600-0854.2000.010803.X.

[98] C. S. Copeland, R. W. Doms, E. M. Bolzau, R. G. Webster, and A. Helenius, “Assembly of influenza hemagglutinin trimers and its role in intracellular transport,” *J. Cell Biol.*, vol. 103, no. 4, pp. 1179–1191, 1986, doi: 10.1083/jcb.103.4.1179.

[99] I. A. Wilson, J. J. Skehel, and D. C. Wiley, “Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution,” *Nature*, vol. 289, no. 5796, pp. 366–373, 1981, doi: 10.1038/289366a0.

[100] S. G. Lazarowitz and P. W. Choppin, “Enhancement of the infectivity of influenza A and B viruses by proteolytic cleavage of the hemagglutinin polypeptide,” *Virology*, vol. 68, no. 2, pp. 440–454, 1975, doi: 10.1016/0042-6822(75)90285-8.

[101] D. A. Steinhauer, “Role of hemagglutinin cleavage for the pathogenicity of influenza virus,” *Virology*, vol. 258, no. 1, pp. 1–20, 1999, doi: 10.1006/viro.1999.9716.

[102] H. Kido *et al.*, “Isolation and characterization of a novel trypsin-like protease found in rat bronchiolar epithelial Clara cells. A possible activator of the viral fusion glycoprotein,” *J. Biol. Chem.*, vol. 267, no. 19, pp. 13573–13579, 1992, doi: 10.1016/s0021-9258(18)42250-8.

[103] H. Limburg *et al.*, “TMPRSS2 Is the Major Activating Protease of Influenza A Virus in Primary Human Airway Cells and Influenza B Virus in Human Type II Pneumocytes,” *J. Virol.*, vol. 93, no. 21, pp. 649–668, 2019, doi: 10.1128/jvi.00649-19.

[104] W. Weis, J. H. Brown, S. Cusack, J. C. Paulson, J. J. Skehel, and D. C. Wiley, “Structure

of the influenza virus haemagglutinin complexed with its receptor, sialic acid," *Nature*, vol. 333, pp. 426–431, 1988.

[105] E. Kirkpatrick, X. Qiu, P. C. Wilson, J. Bahl, and F. Krammer, "The influenza virus hemagglutinin head evolves faster than the stalk domain," *Sci. Reports 2018 81*, vol. 8, no. 1, pp. 1–14, Jul. 2018, doi: 10.1038/s41598-018-28706-1.

[106] N. C. Wu and I. A. Wilson, "Structural biology of influenza hemagglutinin: An amaranthine adventure," *Viruses*, vol. 12, no. 9. 2020, doi: 10.3390/v12091053.

[107] B. F. Koel *et al.*, "Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution," *Science (80-.).*, vol. 342, no. 6161, pp. 976–979, 2013, doi: 10.1126/science.1244730.

[108] D. Angeletti *et al.*, "Defining B cell immunodominance to viruses," *Nat. Immunol.*, vol. 18, no. 4, pp. 456–463, Mar. 2017, doi: 10.1038/ni.3680.

[109] G. A. Sautto, G. A. Kirchenbaum, and T. M. Ross, "Towards a universal influenza vaccine: Different approaches for one goal," *Virology Journal*, vol. 15, no. 1. BioMed Central Ltd., Jan. 19, 2018, doi: 10.1186/s12985-017-0918-y.

[110] S. R. Christensen *et al.*, "Assessing the Protective Potential of H1N1 Influenza Virus Hemagglutinin Head and Stalk Antibodies in Humans," *J. Virol.*, vol. 93, no. 8, 2019.

[111] J. Wrammert *et al.*, "Rapid cloning of high-affinity human monoclonal antibodies against influenza virus," *Nature*, vol. 453, no. 7195, pp. 667–671, 2008, doi: 10.1038/nature06890.

[112] M. Hong *et al.*, "Antibody Recognition of the Pandemic H1N1 Influenza Virus Hemagglutinin Receptor Binding Site," *J. Virol.*, vol. 87, no. 22, pp. 12471–12480, Nov. 2013, doi: 10.1128/jvi.01388-13.

[113] D. C. Ekiert *et al.*, “Cross-neutralization of influenza A viruses mediated by a single antibody loop,” *Nature*, vol. 489, no. 7417, pp. 526–532, Sep. 2012, doi: 10.1038/nature11414.

[114] J. R. R. Whittle *et al.*, “Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 34, pp. 14216–14221, 2011, doi: 10.1073/pnas.1111497108.

[115] G. A. Sautto *et al.*, “A Computationally Optimized Broadly Reactive Antigen Subtype–Specific Influenza Vaccine Strategy Elicits Unique Potent Broadly Neutralizing Antibodies against Hemagglutinin,” *J. Immunol.*, vol. 204, no. 2, pp. 375–385, 2020, doi: 10.4049/jimmunol.1900379.

[116] D. C. Ekiert *et al.*, “Antibody recognition of a highly conserved influenza virus epitope,” *Science (80-.).*, vol. 324, no. 5924, pp. 246–251, Apr. 2009, doi: 10.1126/science.1171491.

[117] J. Sui *et al.*, “Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses,” *Nat. Struct. Mol. Biol.*, vol. 16, no. 3, pp. 265–273, 2009, doi: 10.1038/nsmb.1566.

[118] D. Corti *et al.*, “A Neutralizing Antibody Selected from Plasma Cells That Binds to Group 1 and Group 2 Influenza A Hemagglutinins,” *Science (80-.).*, vol. 333, no. August, pp. 850–856, 2011.

[119] J. J. Guthmiller *et al.*, “Broadly neutralizing antibodies target a haemagglutinin anchor epitope,” *Nature*, vol. 602, no. 7896, pp. 314–320, Dec. 2022, doi: 10.1038/s41586-021-04356-8.

[120] S. A. Valkenburg *et al.*, “Stalking influenza by vaccination with pre-fusion headless HA

mini-stem," *Sci. Rep.*, vol. 6, Mar. 2016, doi: 10.1038/srep22666.

[121] D. J. Dilillo, G. S. Tan, P. Palese, and J. V. Ravetch, "Broadly neutralizing hemagglutinin stalk-specific antibodies require FcR interactions for protection against influenza virus in vivo," *Nat. Med.*, vol. 20, no. 2, pp. 143–151, 2014, doi: 10.1038/nm.3443.

[122] B. J. Bosch and P. J. M. Rottier, "Nidovirus Entry into Cells," in *Nidoviruses*, 2008, pp. 157–178.

[123] D. Bestle *et al.*, "TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells," *Life Sci. Alliance*, vol. 3, no. 9, 2020, doi: 10.26508/LSA.202000786.

[124] B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N. G. Seidah, and E. Decroly, "The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade," *Antiviral Res.*, vol. 176, p. 104742, Apr. 2020, doi: 10.1016/J.ANTIVIRAL.2020.104742.

[125] T. P. Peacock *et al.*, "The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets," *Nat. Microbiol.*, vol. 6, no. 7, pp. 899–909, 2021, doi: 10.1038/s41564-021-00908-w.

[126] M. Hoffmann, H. Kleine-Weber, and S. Pöhlmann, "A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells," *Mol. Cell*, vol. 78, no. 4, p. 779, May 2020, doi: 10.1016/J.MOLCEL.2020.04.022.

[127] J. Shang *et al.*, "Cell entry mechanisms of SARS-CoV-2," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 21, 2020, doi: 10.1073/pnas.2003138117.

[128] M. Hoffmann *et al.*, "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is

Blocked by a Clinically Proven Protease Inhibitor," *Cell*, vol. 181, no. 2, pp. 271-280.e8, 2020, doi: 10.1016/j.cell.2020.02.052.

[129] A. Bayati, R. Kumar, V. Francis, and P. S. McPherson, "SARS-CoV-2 infects cells after viral entry via clathrin-mediated endocytosis," *J. Biol. Chem.*, vol. 296, p. 100306, Jan. 2021, doi: 10.1016/j.jbc.2021.100306.

[130] D. Wrapp *et al.*, "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation," *Science (80-.).*, vol. 367, no. 6483, pp. 1260–1263, 2020, doi: 10.1126/science.abb2507.

[131] F. Li, W. Li, M. Farzan, and S. C. Harrison, "Structure of SARS coronavirus spike receptor-binding domain complexed with receptor," *Science (80-.).*, vol. 309, no. 5742, pp. 1864–1868, 2005, doi: 10.1126/science.1116480.

[132] M. Letko, A. Marzi, and V. Munster, "Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses," *Nat. Microbiol.*, vol. 5, no. 4, pp. 562–569, 2020, doi: 10.1038/s41564-020-0688-y.

[133] W. Li *et al.*, "Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus," *Nature*, vol. 426, no. 6965, pp. 450–454, Nov. 2003, doi: 10.1038/nature02145.

[134] H. Hofmann, K. Pyrc, L. Van Der Hoek, M. Geier, B. Berkhout, and S. Pöhlmann, "Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 22, p. 7988, May 2005, doi: 10.1073/PNAS.0409465102.

[135] R. Henderson *et al.*, "Controlling the SARS-CoV-2 spike glycoprotein conformation," *Nat.*

Struct. Mol. Biol. 2020 2710, vol. 27, no. 10, pp. 925–933, Jul. 2020, doi: 10.1038/s41594-020-0479-4.

[136] C. Xu *et al.*, "Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM," *Sci. Adv.*, vol. 7, no. 1, 2021, doi: 10.1126/sciadv.abe5575.

[137] Y. Cai *et al.*, "Distinct conformational states of SARS-CoV-2 spike protein," *Science (80-.).*, vol. 369, no. 6511, Sep. 2020, doi: 10.1126/science.abd4251.

[138] D. J. Benton *et al.*, "Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion," *Nature*, vol. 588, no. 7837, pp. 327–330, 2020, doi: 10.1038/s41586-020-2772-0.

[139] C. B. Jackson, M. Farzan, B. Chen, and H. Choe, "Mechanisms of SARS-CoV-2 entry into cells," *Nat. Rev. Mol. Cell Biol.*, vol. 23, no. 1, pp. 3–20, 2022, doi: 10.1038/s41580-021-00418-x.

[140] L. Premkumar *et al.*, "The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients," *Sci. Immunol.*, vol. 5, no. 48, p. 8413, Jun. 2020, doi: 10.1126/SCIIMMUNOL.ABC8413.

[141] C. O. Barnes *et al.*, "SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies," *Nature*, vol. 588, no. 7839, pp. 682–687, Oct. 2020, doi: 10.1038/s41586-020-2852-1.

[142] L. Piccoli *et al.*, "Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology," *Cell*, vol. 183, no. 4, pp. 1024-1042.e21, Nov. 2020, doi: 10.1016/j.cell.2020.09.037.

[143] P. J. M. Brouwer *et al.*, “Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability,” *Science (80-.).*, vol. 369, no. 6504, pp. 643–650, Aug. 2020, doi: 10.1126/science.abc5902.

[144] L. Liu *et al.*, “Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike,” *Nature*, vol. 584, no. 7821, pp. 450–456, Jul. 2020, doi: 10.1038/s41586-020-2571-7.

[145] E. C. Thomson *et al.*, “Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity,” *Cell*, vol. 184, no. 5, pp. 1171-1187.e20, Mar. 2021, doi: 10.1016/j.cell.2021.01.037.

[146] Y. Weisblum *et al.*, “Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants,” *Elife*, vol. 9, p. 1, Oct. 2020, doi: 10.7554/eLife.61312.

[147] A. J. Greaney *et al.*, “Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition,” *Cell Host Microbe*, vol. 29, no. 1, pp. 44-57.e9, Jan. 2021, doi: 10.1016/j.chom.2020.11.007.

[148] K. M. Hastie *et al.*, “Defining variant-resistant epitopes targeted by SARS-CoV-2 antibodies: A global consortium study,” *Science (80-.).*, vol. 374, no. 6566, pp. 478–472, Oct. 2021, doi: 10.1126/science.abh2315.

[149] D. Pinto *et al.*, “Broad betacoronavirus neutralization by a stem helix–specific human antibody,” *Science (80-.).*, vol. 373, no. 6559, pp. 1109–1116, Sep. 2021, doi: 10.1126/science.abj3321.

[150] B. Liang *et al.*, “Enhanced Neutralizing Antibody Response Induced by Respiratory Syncytial Virus Prefusion F Protein Expressed by a Vaccine Candidate,” *J. Virol.*, vol. 89,

no. 18, pp. 9499–9510, Sep. 2015, doi: 10.1128/jvi.01373-15.

[151] J. Pallesen *et al.*, "Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 35, pp. E7348–E7357, Aug. 2017, doi: 10.1073/pnas.1707304114.

[152] C. Keech *et al.*, "Phase 1–2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine," *N. Engl. J. Med.*, vol. 383, no. 24, pp. 2320–2332, Dec. 2020, doi: 10.1056/NEJMoa2026920.

[153] L. M. G. Cullen, J. C. G. G. Blanco, and T. G. Morrison, "Cotton rat immune responses to virus-like particles containing the pre-fusion form of respiratory syncytial virus fusion protein," *J. Transl. Med.*, vol. 13, no. 1, p. 350, Nov. 2015, doi: 10.1186/s12967-015-0705-8.

[154] F. P. Polack *et al.*, "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine," *N. Engl. J. Med.*, vol. 383, no. 27, pp. 2603–2615, Dec. 2020, doi: 10.1056/NEJMoa2034577.

[155] M. P. Griffin *et al.*, "Safety, tolerability, and pharmacokinetics of MEDI8897, the respiratory syncytial virus prefusion F-targeting monoclonal antibody with an extended half-life, in healthy adults," *Antimicrob. Agents Chemother.*, vol. 61, no. 3, Mar. 2017, doi: 10.1128/AAC.01714-16.

[156] FDA, "FDA Approves First Respiratory Syncytial Virus (RSV) Vaccine," 2023. https://www.fda.gov/news-events/press-announcements/fda-approves-first-respiratory-syncytial-virus-rsv-vaccine.

[157] R. N. Kirchdoerfer *et al.*, "Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11,

Dec. 2018, doi: 10.1038/s41598-018-34171-7.

[158] J. Juraszek *et al.*, “Stabilizing the closed SARS-CoV-2 spike trimer,” *Nat. Commun.*, vol. 12, no. 1, pp. 1–8, Dec. 2021, doi: 10.1038/s41467-020-20321-x.

[159] C. L. Hsieh *et al.*, “Structure-based design of prefusion-stabilized SARS-CoV-2 Spikes,” *Science (80-.).*, vol. 369, no. 6509, pp. 1501–1505, Sep. 2020, doi: 10.1101/2020.05.30.125484.

[160] D. Wrapp *et al.*, “Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation,” 2019.

[161] L. Zhang *et al.*, “Design and characterization of a fusion glycoprotein vaccine for Respiratory Syncytial Virus with improved stability,” *Vaccine*, vol. 36, no. 52, pp. 8119–8130, 2018, doi: 10.1016/j.vaccine.2018.10.032.

[162] R. N. Kirchdoerfer *et al.*, “Pre-fusion structure of a human coronavirus spike protein,” *Nature*, vol. 531, no. 7592, pp. 118–121, Mar. 2016, doi: 10.1038/nature17200.

[163] M. G. Joyce *et al.*, “Iterative structure-based improvement of a fusion-glycoprotein vaccine against RSV,” *Nat. Struct. Mol. Biol.*, vol. 23, no. 9, pp. 811–820, Sep. 2016, doi: 10.1038/nsmb.3267.

[164] G. B. E. E. Stewart-Jones *et al.*, “A cysteine zipper stabilizes a pre-fusion F glycoprotein vaccine for respiratory syncytial virus,” *PLoS One*, vol. 10, no. 6, p. e0128779, Jun. 2015, doi: 10.1371/journal.pone.0128779.

[165] C. J. Wei, M. C. Crank, J. Shiver, B. S. Graham, J. R. Mascola, and G. J. Nabel, “Next-generation influenza vaccines: opportunities and challenges,” *Nature Reviews Drug Discovery*, vol. 19, no. 4. Nature Research, pp. 239–252, Apr. 01, 2020, doi:

10.1038/s41573-019-0056-x.

[166] A. E. Fiore, C. B. Bridges, and N. J. Cox, “Seasonal influenza vaccines,” *Current Topics in Microbiology and Immunology*, vol. 333, no. 1. Springer Verlag, pp. 43–82, 2009, doi: 10.1007/978-3-540-92165-3_3.

[167] C. L. D. D. McMillan, P. R. Young, D. Watterson, and K. J. Chappell, “The next generation of influenza vaccines: Towards a universal solution,” *Vaccines*, vol. 9, no. 1. MDPI AG, pp. 1–20, Jan. 01, 2021, doi: 10.3390/vaccines9010026.

[168] S. Hiroi *et al.*, “Human monoclonal antibodies neutralizing influenza virus A/H1N1pdm09 and seasonal A/H1N1 strains – Distinct Ig gene repertoires with a similar action mechanism,” *Immunobiology*, vol. 223, no. 3, pp. 319–326, Mar. 2018, doi: 10.1016/j.imbio.2017.10.040.

[169] M. Kanekiyo *et al.*, “Mosaic nanoparticle display of diverse influenza virus hemagglutinins elicits broad B cell responses,” *Nat. Immunol.*, vol. 20, no. 3, pp. 362–372, 2019, doi: 10.1038/s41590-018-0305-x.

[170] B. M. Giles and T. M. Ross, “A computationally optimized broadly reactive antigen (COBRA) based H5N1 VLP vaccine elicits broadly reactive antibodies in mice and ferrets,” *Vaccine*, vol. 29, no. 16, pp. 3043–3054, 2011, doi: 10.1016/j.vaccine.2011.01.100.

[171] D. M. Carter *et al.*, “Design and Characterization of a Computationally Optimized Broadly Reactive Hemagglutinin Vaccine for H1N1 Influenza Viruses,” *J. Virol.*, vol. 90, no. 9, pp. 4720–4734, 2016, doi: 10.1128/jvi.03152-15.

[172] T. M. Wong *et al.*, “Computationally Optimized Broadly Reactive Hemagglutinin Elicits Hemagglutination Inhibition Antibodies against a Panel of H3N2 Influenza Virus

Cocirculating Variants," *J. Virol.*, vol. 91, no. 24, pp. 1581–1598, Dec. 2017, doi: 10.1128/jvi.01581-17.

[173] F. Krammer and P. Palese, "Influenza virus hemagglutinin stalk-based antibodies and vaccines," *Current Opinion in Virology*, vol. 3, no. 5. Elsevier B.V., pp. 521–530, 2013, doi: 10.1016/j.coviro.2013.07.007.

[174] A. Impagliazzo *et al.*, "A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen," *Science (80-.).*, vol. 349, no. 6254, pp. 1301–1306, Sep. 2015, doi: 10.1126/science.aac7263.

[175] M. Throsby *et al.*, "Heterosubtypic Neutralizing Monoclonal Antibodies Cross-Protective against H5N1 and H1N1 Recovered from Human IgM+ Memory B Cells," *PLoS One*, vol. 3, no. 12, p. e3942, Dec. 2008, doi: 10.1371/journal.pone.0003942.

[176] J. Sui *et al.*, "Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses," *Nat. Struct. Mol. Biol.*, vol. 16, no. 3, pp. 265–273, 2009, doi: 10.1038/nsmb.1566.

[177] A. H. Ellebedy *et al.*, "Induction of broadly cross-reactive antibody responses to the influenza HA stem region following H5N1 vaccination in humans," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 36, pp. 13133–13138, Sep. 2014, doi: 10.1073/pnas.1414070111.

[178] F. Krammer, N. Pica, R. Hai, G. S. Tan, and P. Palese, "Hemagglutinin Stalk-Reactive Antibodies Are Boosted following Sequential Infection with Seasonal and Pandemic H1N1 Influenza Virus in Mice," *J. Virol.*, vol. 86, no. 19, pp. 10302–10307, Oct. 2012, doi: 10.1128/jvi.01336-12.

[179] M. S. Miller *et al.*, "1976 and 2009 H1N1 Influenza Virus Vaccines Boost Anti-

Hemagglutinin Stalk Antibodies in Humans," *J. Infect. Dis.*, vol. 207, no. 1, pp. 98–105, Jan. 2013, doi: 10.1093/infdis/jis652.

[180] J. Sui *et al.*, "Wide prevalence of heterosubtypic broadly neutralizing human anti-influenza a antibodies," *Clin. Infect. Dis.*, vol. 52, no. 8, pp. 1003–1009, Apr. 2011, doi: 10.1093/cid/cir121.

[181] K. E. Neu, C. J. Henry Dunand, and P. C. Wilson, "Heads, stalks and everything else: how can antibodies eradicate influenza as a human disease?," *Current Opinion in Immunology*, vol. 42. Elsevier Ltd, pp. 48–55, 2016, doi: 10.1016/j.coi.2016.05.012.

[182] R. Nachbagauer *et al.*, "A chimeric haemagglutinin-based influenza split virion vaccine adjuvanted with AS03 induces protective stalk-reactive antibodies in mice," *npj Vaccines*, vol. 1, no. 1, p. 1, Sep. 2016, doi: 10.1038/npjvaccines.2016.15.

[183] M. E. Ermler *et al.*, "Chimeric Hemagglutinin Constructs Induce Broad Protection against Influenza B Virus Challenge in the Mouse Model," *J. Virol.*, vol. 91, no. 12, Jun. 2017, doi: 10.1128/jvi.00286-17.

[184] F. Krammer, N. Pica, R. Hai, I. Margine, and P. Palese, "Chimeric Hemagglutinin Influenza Virus Vaccine Constructs Elicit Broadly Protective Stalk-Specific Antibodies," *J. Virol.*, vol. 87, no. 12, pp. 6542–6550, Jun. 2013, doi: 10.1128/jvi.00641-13.

[185] R. Hai *et al.*, "Influenza Viruses Expressing Chimeric Hemagglutinins: Globular Head and Stalk Domains Derived from Different Subtypes," *J. Virol.*, vol. 86, no. 10, pp. 5774–5781, May 2012, doi: 10.1128/jvi.00137-12.

[186] R. Nachbagauer *et al.*, "A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled

phase I trial," *Nat. Med.*, vol. 27, no. 1, pp. 106–114, Jan. 2021, doi: 10.1038/s41591-020-1118-7.

[187] F. Broecker *et al.*, "A mosaic hemagglutinin-based influenza virus vaccine candidate protects mice from challenge with divergent H3N2 strains," *npj Vaccines*, vol. 4, no. 1, pp. 1–9, Dec. 2019, doi: 10.1038/s41541-019-0126-4.

[188] W. Sun *et al.*, "Development of Influenza B Universal Vaccine Candidates Using the 'Mosaic' Hemagglutinin Approach," *J. Virol.*, vol. 93, no. 12, Apr. 2019, doi: 10.1128/jvi.00333-19.

[189] F. Krammer and P. Palese, "Universal Influenza Virus Vaccines That Target the Conserved Hemagglutinin Stalk and Conserved Sites in the Head Domain," *J. Infect. Dis.*, vol. 219, no. Supplement_1, pp. S62--S67, Apr. 2019, doi: 10.1093/infdis/jiy711.

[190] K. S. Corbett *et al.*, "Design of nanoparticulate group 2 influenza virus hemagglutinin stem antigens that activate unmutated ancestor B cell receptors of broadly neutralizing antibody lineages," *MBio*, vol. 10, no. 1, pp. 2810–2828, Jan. 2019, doi: 10.1128/mBio.02810-18.

[191] J. Steel *et al.*, "Influenza virus vaccine based on the conserved hemagglutinin stalk domain," *MBio*, vol. 1, no. 1, May 2010, doi: 10.1128/mBio.00018-10.

[192] G. Bommakanti *et al.*, "Design of Escherichia coli-Expressed Stalk Domain Immunogens of H1N1 Hemagglutinin That Protect Mice from Lethal Challenge," *J. Virol.*, vol. 86, no. 24, pp. 13434–13444, Dec. 2012, doi: 10.1128/jvi.01429-12.

[193] V. V. A. A. Mallajosyula *et al.*, "Influenza hemagglutinin stem-fragment immunogen elicits broadly neutralizing antibodies and confers heterologous protection," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 25, pp. E2514--E2523, Jun. 2014, doi: 10.1073/pnas.1402766111.

[194] V. V. A. Mallajosyula *et al.*, "Hemagglutinin sequence conservation guided stem immunogen design from influenza A H3 subtype," *Front. Immunol.*, vol. 6, no. JUN, p. 329, Jun. 2015, doi: 10.3389/fimmu.2015.00329.

[195] H. M. Yassine *et al.*, "Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection," *Nat. Med.*, vol. 21, no. 9, pp. 1065–1070, Sep. 2015, doi: 10.1038/nm.3927.

[196] T. J. Wohlbold, R. Nachbagauer, I. Margine, G. S. Tan, A. Hirsh, and F. Krammer, "Vaccination with soluble headless hemagglutinin protects mice from challenge with divergent influenza viruses," *Vaccine*, vol. 33, no. 29, pp. 3314–3321, Jun. 2015, doi: 10.1016/j.vaccine.2015.05.038.

[197] S. C. Wang, H. Y. Liao, J. Y. Zhang, T. J. R. Cheng, and C. H. Wong, "Development of a universal influenza vaccine using hemagglutinin stem protein produced from Pichia pastoris," *Virology*, vol. 526, pp. 125–137, Jan. 2019, doi: 10.1016/j.virol.2018.10.005.

[198] Y. Lu, J. P. Welsh, and J. R. Swartz, "Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 1, pp. 125–130, Jan. 2014, doi: 10.1073/pnas.1308701110.

[199] National Institute of Allergy and Infectious Diseases (NIAID), "Influenza HA Ferritin Vaccine, Alone or in Prime-Boost Regimens With an Influenza DNA Vaccine in Healthy Adults," 2017. https://clinicaltrials.gov/ct2/show/NCT03186781.

[200] S.-C. Lin, Y.-F. Lin, P. Chong, and S.-C. Wu, "Broader Neutralizing Antibodies against H5N1 Viruses Using Prime-Boost Immunization of Hyperglycosylated Hemagglutinin DNA and Virus-Like Particles," *PLoS One*, vol. 7, no. 6, p. e39075, Jun. 2012, doi:

10.1371/journal.pone.0039075.

[201] S.-C. Lin, W.-C. Liu, J.-T. Jan, and S.-C. Wu, “Glycan Masking of Hemagglutinin for Adenovirus Vector and Recombinant Protein Immunizations Elicits Broadly Neutralizing Antibodies against H5N1 Avian Influenza Viruses,” *PLoS One*, vol. 9, no. 3, p. e92822, Mar. 2014, doi: 10.1371/journal.pone.0092822.

[202] G. Bajic *et al.*, “Influenza Antigen Engineering Focuses Immune Responses to a Subdominant but Broadly Protective Viral Epitope,” *Cell Host Microbe*, vol. 25, no. 6, pp. 827--835.e6, Jun. 2019, doi: 10.1016/j.chom.2019.04.003.

[203] D. Eggink, P. H. Goff, and P. Palese, “Guiding the Immune Response against Influenza Virus Hemagglutinin toward the Conserved Stalk Domain by Hyperglycosylation of the Globular Head Domain,” *J. Virol.*, vol. 88, no. 1, pp. 699–704, Jan. 2014, doi: 10.1128/jvi.02608-13.

[204] W.-C. Liu, J.-T. Jan, Y.-J. Huang, T.-H. Chen, and S.-C. Wu, “Unmasking Stem-Specific Neutralizing Epitopes by Abolishing N-Linked Glycosylation Sites of Influenza Virus Hemagglutinin Proteins for Vaccine Design,” *J. Virol.*, vol. 90, no. 19, pp. 8496–8508, Oct. 2016, doi: 10.1128/jvi.00880-16.

[205] J. S. Mclellan, Y. Yang, B. S. Graham, and P. D. Kwong, “Structure of Respiratory Syncytial Virus Fusion Glycoprotein in the Postfusion Conformation Reveals Preservation of Neutralizing Epitopes †,” *J. Virol.*, vol. 85, no. 15, pp. 7788–7796, 2011, doi: 10.1128/JVI.00555-11.

[206] D. Tian *et al.*, “Structural basis of respiratory syncytial virus subtype-dependent neutralization by an antibody targeting the fusion glycoprotein,” *Nat. Commun.*, vol. 8, no.

1, pp. 1–7, Dec. 2017, doi: 10.1038/s41467-017-01858-w.

[207] V. Más *et al.*, "Engineering, Structure and Immunogenicity of the Human Metapneumovirus F Protein in the Postfusion Conformation," *PLOS Pathog.*, vol. 12, no. 9, p. e1005859, Sep. 2016, doi: 10.1371/journal.ppat.1005859.

[208] A. Bakan *et al.*, "Evol and ProDy for bridging protein sequence evolution and structural dynamics," *Bioinformatics*, vol. 30, no. 18, pp. 2681–2683, 2014, doi: 10.1093/bioinformatics/btu336.

[209] Y. Zhang *et al.*, "Influenza Research Database: An integrated bioinformatics resource for influenza virus research," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D466–D474, Jan. 2017, doi: 10.1093/NAR/GKW857.

[210] S. T. H. Liu *et al.*, "Antigenic sites in influenza H1 hemagglutinin display species-specific immunodominance," *J. Clin. Invest.*, vol. 128, no. 11, pp. 4992–4996, Nov. 2018, doi: 10.1172/JCI122895.

[211] A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein," *Cell*, vol. 181, no. 2, pp. 281-292.e6, Apr. 2020, doi: 10.1016/J.CELL.2020.02.058.

[212] L. Schrödinger, "The PyMOL Molecular Graphics System, Version~1.8." Schrodinger LLC, New York, NY, USA, Nov. 2015.

[213] A. Bohne-Lang and C. W. Von der Lieth, "GlyProt: In silico glycosylation of proteins," *Nucleic Acids Res.*, vol. 33, no. SUPPL. 2, p. W214, Jul. 2005, doi: 10.1093/nar/gki385.

[214] Y. B. Narkhede, K. J. Gonzalez, and E. M. Strauch, "Targeting Viral Surface Proteins through Structure-Based Design," *Viruses 2021, Vol. 13, Page 1320*, vol. 13, no. 7, p. 1320,

Jul. 2021, doi: 10.3390/V13071320.

[215] R. W. Doms and J. P. Moore, “HIV-1 membrane fusion: Targets of opportunity,” Rockefeller University Press, Oct. 2000. doi: 10.1083/jcb.151.2.f9.

[216] S. Moller-Tank and W. Maury, “Ebola Virus Entry: A Curious and Complex Series of Events,” *PLoS Pathog.*, vol. 11, no. 4, p. e1004731, Apr. 2015, doi: 10.1371/journal.ppat.1004731.

[217] J. P. M. L. Lucy Rutten, Morgan S.A. Gilman, Sven Blokland, Jarek Juraszek, Jason S. McLellan, “Structure-Based Design of Prefusion-Stabilized Filovirus Glycoprotein Trimers,” no. January, pp. 19–21, 2020.

[218] L. Rutten, M. S. A. Gilman, S. Blokland, J. Juraszek, J. S. McLellan, and J. P. M. Langedijk, “Structure-Based Design of Prefusion-Stabilized Filovirus Glycoprotein Trimers,” *Cell Rep.*, vol. 30, no. 13, p. 4540, Mar. 2020, doi: 10.1016/J.CELREP.2020.03.025.

[219] J. J. W. Wong, R. G. Paterson, R. A. Lamb, and T. S. Jardetzky, “Structure and stabilization of the Hendra virus F glycoprotein in its prefusion form,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 4, pp. 1056–1061, Jan. 2016, doi: 10.1073/pnas.1523303113.

[220] J. A. Davey and R. A. Chica, “Multistate approaches in computational protein design,” *Protein Sci.*, vol. 21, no. 9, pp. 1241–1252, Sep. 2012, doi: 10.1002/PRO.2128.

[221] A. Leaver-Fay, R. Jacak, P. B. Stranges, and B. Kuhlman, “A generic program for multistate protein design,” *PLoS One*, vol. 6, no. 7, 2011, doi: 10.1371/journal.pone.0020937.

[222] D. Tian *et al.*, “Structural basis of respiratory syncytial virus subtype-dependent neutralization by an antibody targeting the fusion glycoprotein,” *Nat. Commun.*, vol. 8, no. 1, pp. 1–7, Dec. 2017, doi: 10.1038/s41467-017-01858-w.

[223] H. Park *et al.*, “Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules,” *J. Chem. Theory Comput.*, vol. 12, no. 12, pp. 6201–6212, Dec. 2016, doi: 10.1021/acs.jctc.6b00819.

[224] M. J. Kwakkenbos *et al.*, “Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming,” *Nat. Med.*, vol. 16, no. 1, pp. 123–128, Jan. 2010, doi: 10.1038/nm.2071.

[225] D. Corti *et al.*, “Cross-neutralization of four paramyxoviruses by a human monoclonal antibody,” *Nat. 2013 5017467*, vol. 501, no. 7467, pp. 439–443, Aug. 2013, doi: 10.1038/nature12442.

[226] K. K. Chan *et al.*, “Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2,” *Science (80-.).*, vol. 369, no. 6508, pp. 1261–1265, Sep. 2020, doi: 10.1126/SCIENCE.ABC0870.

[227] C.-L. Hsieh *et al.*, “Structure-based design of prefusion-stabilized human metapneumovirus fusion proteins,” *Nat. Commun. 2022 131*, vol. 13, no. 1, pp. 1–11, Mar. 2022, doi: 10.1038/s41467-022-28931-3.

[228] J. S. McLellan, Y. Yang, B. S. Graham, and P. D. Kwong, “Structure of Respiratory Syncytial Virus Fusion Glycoprotein in the Postfusion Conformation Reveals Preservation of Neutralizing Epitopes,” *J. Virol.*, vol. 85, no. 15, pp. 7788–7796, 2011, doi: 10.1128/jvi.00555-11.

[229] L. J. Anderson, P. Bingham, and J. C. Hierholzer, “Neutralization of respiratory syncytial virus by individual and mixtures of F and G protein monoclonal antibodies.,” *J. Virol.*, vol. 62, no. 11, p. 4232, Nov. 1988, doi: 10.1128/JVI.62.11.4232-4238.1988.

[230] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nat. 2021 5967873*, vol. 596, no. 7873, pp. 583–589, Jul. 2021, doi: 10.1038/s41586-021-03819-2.

[231] R. Y. R. Wang, Y. Song, B. A. Barad, Y. Cheng, J. S. Fraser, and F. DiMaio, “Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta,” *Elife*, vol. 5, no. September2016, Sep. 2016, doi: 10.7554/ELIFE.17219.

[232] P. D. Adams *et al.*, “PHENIX: a comprehensive Python-based system for macromolecular structure solution,” *urn:issn:0907-4449*, vol. 66, no. 2, pp. 213–221, Jan. 2010, doi: 10.1107/S0907444909052925.

[233] M. D. Tyka *et al.*, “Alternate states of proteins revealed by detailed energy landscape mapping,” *J. Mol. Biol.*, vol. 405, no. 2, pp. 607–618, Jan. 2011, doi: 10.1016/j.jmb.2010.11.008.

[234] V. B. Chen *et al.*, “MolProbity : all-atom structure validation for macromolecular crystallography,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 1, pp. 12–21, Jan. 2010, doi: 10.1107/s0907444909042073.

[235] Y. Cai *et al.*, “Distinct conformational states of SARS-CoV-2 spike protein,” *Science (80-.).*, vol. 369, no. 6511, Sep. 2020, doi: 10.1126/science.abd4251.

[236] Y. Song *et al.*, “High-Resolution Comparative Modeling with RosettaCM,” *Structure*, vol. 21, no. 10, pp. 1735–1742, Oct. 2013, doi: 10.1016/J.STR.2013.08.005.

[237] J. Lan *et al.*, “Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor,” *Nat. 2020 5817807*, vol. 581, no. 7807, pp. 215–220, Mar. 2020, doi: 10.1038/s41586-020-2180-5.

[238] S. Xia *et al.*, “Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly

potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion," *Cell Res. 2020 304*, vol. 30, no. 4, pp. 343–355, Mar. 2020, doi: 10.1038/s41422-020-0305-x.

[239] S. J. Fleishman *et al.*, "Rosettascripts: A scripting language interface to the Rosetta Macromolecular modeling suite," *PLoS One*, vol. 6, no. 6, p. e20161, Jun. 2011, doi: 10.1371/journal.pone.0020161.

[240] J. B. Maguire *et al.*, "Perturbing the energy landscape for improved packing during computational protein design," *Proteins Struct. Funct. Bioinforma.*, vol. 89, no. October, pp. 1–14, Apr. 2020, doi: 10.1002/prot.26030.

[241] G. Bhardwaj *et al.*, "Accurate de novo design of hyperstable constrained peptides," *Nature*, vol. 538, no. 7625, pp. 329–335, Sep. 2016, doi: 10.1038/nature19791.

[242] J. Huang *et al.*, "Structure, Immunogenicity, and Conformation-Dependent Receptor Binding of the Postfusion Human Metapneumovirus F Protein STRUCTURE AND ASSEMBLY," *J. Virol. jvi.asm.org*, vol. 95, pp. 593–614, 2021, doi: 10.1128/JVI.00593.

[243] L. Fang, "Codon Optimization," WO2020024917A1, 2020.

[244] W. Kabsch, "XDS," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. Pt 2, p. 125, Feb. 2010, doi: 10.1107/S0907444909047337.

[245] P. Emsley and K. Cowtan, "Coot: model-building tools for molecular graphics," *urn:issn:0907-4449*, vol. 60, no. 12, pp. 2126–2132, Nov. 2004, doi: 10.1107/S0907444904019158.

[246] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination," *Nat. Methods 2017 143*, vol. 14, no.

3, pp. 290–296, Feb. 2017, doi: 10.1038/nmeth.4169.

[247] R. Sanchez-Garcia, J. Gomez-Blanco, A. Cuervo, J. M. Carazo, and J. Vargas, “DeepEMhacer: a deep learning solution for cryo-EM volume post-processing,” doi: 10.1101/2020.06.12.148296.

[248] T. D. Goddard, C. C. Huang, and T. E. Ferrin, “Visualizing density maps with UCSF Chimera,” *J. Struct. Biol.*, vol. 157, no. 1, pp. 281–287, Jan. 2007, doi: 10.1016/J.JSB.2006.06.010.

[249] R. Jia *et al.*, “Resurgence of Respiratory Syncytial Virus Infection During COVID-19 Pandemic Among Children in Shanghai, China,” *Front. Microbiol.*, vol. 13, no. July, 2022, doi: 10.3389/fmicb.2022.938372.

[250] M. Ujiie, S. Tsuzuki, T. Nakamoto, N. Iwamoto, and M. Ujiie, “Resurgence of respiratory syncytial virus infections during covid-19 pandemic, tokyo, japan,” *Emerg. Infect. Dis.*, vol. 27, no. 11, pp. 2969–2970, 2021, doi: 10.3201/eid2711.211565.

[251] E. J. Chow, T. M. Uyeki, and H. Y. Chu, “The effects of the COVID-19 pandemic on community respiratory virus activity,” *Nat. Rev. Microbiol. 2022 213*, vol. 21, no. 3, pp. 195–210, Oct. 2022, doi: 10.1038/s41579-022-00807-9.

[252] D. A. Foley *et al.*, “The Interseasonal Resurgence of Respiratory Syncytial Virus in Australian Children Following the Reduction of Coronavirus Disease 2019–Related Public Health Measures,” *Clin. Infect. Dis.*, vol. 73, no. 9, pp. e2829–e2830, Nov. 2021, doi: 10.1093/CID/CIAA1906.

[253] S. Hamid *et al.*, “Seasonality of Respiratory Syncytial Virus-United States, 2017-2023,” *MMWR Morb Mortal Wkly Rep*, vol. 72, pp. 355–361, 2023, Accessed: Apr. 13, 2023.

[Online]. Available: https://www.cdc.gov/mmwr/mmwr_continuingEducation.html.

[254] J. Chen, K. H. Lee, D. A. Steinhauer, D. J. Stevens, J. J. Skehel, and D. C. Wiley, "Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation," *Cell*, vol. 95, no. 3, pp. 409–417, Oct. 1998, doi: 10.1016/S0092-8674(00)81771-7.

[255] L. Chen *et al.*, "The structure of the fusion glycoprotein of Newcastle disease virus suggests a novel paradigm for the molecular mechanism of membrane fusion," *Structure*, vol. 9, no. 3, pp. 255–266, Mar. 2001, doi: 10.1016/S0969-2126(01)00581-0.

[256] K. Swanson, X. Wen, G. P. Leser, R. G. Paterson, R. A. Lamb, and T. S. Jardetzky, "Structure of the Newcastle disease virus F protein in the post-fusion conformation," *Virology*, vol. 402, no. 2, pp. 372–379, Jul. 2010, doi: 10.1016/J.VIROL.2010.03.050.

[257] H. S. Yin, R. G. Paterson, X. Wen, R. A. Lamb, and T. S. Jardetzky, "Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 26, pp. 9288–9293, Jun. 2005, doi: 10.1073/PNAS.0503989102/SUPPL_FILE/03989FIG6.PDF.

[258] H.-S. Yin, X. Wen, R. G. Paterson, R. A. Lamb, and T. S. Jardetzky, "Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation," 2006, doi: 10.1038/nature04322.

[259] L. J. Earp, S. E. Delos, H. E. Park, and J. M. White, "The Many Mechanisms of Viral Membrane Fusion Proteins," *Curr. Top. Microbiol. Immunol.*, vol. 285, pp. 25–66, 2004, doi: 10.1007/3-540-26764-6_2.

[260] T. S. Jardetzky and R. A. Lamb, "Activation of Paramyxovirus Membrane Fusion and Virus

Entry," *Curr. Opin. Virol.*, vol. 0, no. 1, p. 24, 2014, doi: 10.1016/J.COVIRO.2014.01.005.

[261] K. J. Gonzalez *et al.*, "A general computational design strategy for stabilizing viral class I fusion proteins," doi: 10.1101/2023.03.16.532924.

[262] D. B. Craig and A. A. Dombkowski, "Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins," *BMC Bioinformatics*, vol. 14, no. 1, pp. 0–6, 2013, doi: 10.1186/1471-2105-14-346.

[263] N. K. Salam, M. Adzhigirey, W. Sherman, D. A. Pearlman, and D. Thirumalai, "Structure-based approach to the prediction of disulfide bonds in proteins," *Protein Eng. Des. Sel.*, vol. 27, no. 10, pp. 365–374, Oct. 2014, doi: 10.1093/PROTEIN/GZU017.

[264] X. Gao, X. Dong, X. Li, Z. Liu, and H. Liu, "Prediction of disulfide bond engineering sites using a machine learning method," *Sci. Reports 2020 101*, vol. 10, no. 1, pp. 1–9, Jun. 2020, doi: 10.1038/s41598-020-67230-z.

[265] Q. A. T. Le, J. C. Joo, Y. J. Yoo, and Y. H. Kim, "Development of thermostable Candida antarctica lipase B through novel in silico design of disulfide bridge," *Biotechnol. Bioeng.*, vol. 109, no. 4, pp. 867–876, 2012, doi: 10.1002/bit.24371.

[266] R. F. Alford *et al.*, "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design," *J. Chem. Theory Comput.*, vol. 13, no. 6, pp. 3031–3048, 2017, doi: 10.1021/acs.jctc.7b00125.

[267] M. J. O'Meara *et al.*, "Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta," *J. Chem. Theory Comput.*, vol. 11, no. 2, pp. 609–622, 2015, doi: 10.1021/ct500864r.

[268] F. Khatib *et al.*, "Algorithm discovery by protein folding game players," *Proc. Natl. Acad.*

Sci. U. S. A., vol. 108, no. 47, pp. 18949–18953, 2011, doi: 10.1073/pnas.1115898108.

[269] P. Conway, M. D. Tyka, F. DiMaio, D. E. Konerding, and D. Baker, "Relaxation of backbone bond geometry improves protein energy landscape modeling," *Protein Sci.*, vol. 23, no. 1, pp. 47–55, 2014, doi: 10.1002/pro.2389.

[270] R. F. Alford *et al.*, "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design," *J. Chem. Theory Comput.*, vol. 13, no. 6, pp. 3031–3048, 2017, doi: 10.1021/acs.jctc.7b00125.

[271] M. J. O'Meara *et al.*, "Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta," *J. Chem. Theory Comput.*, vol. 11, no. 2, pp. 609–622, 2015, doi: 10.1021/ct500864r.

[272] K. A. Barlow *et al.*, "Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation," *J. Phys. Chem. B*, vol. 122, no. 21, pp. 5389–5399, May 2018, doi: 10.1021/acs.jpcb.7b11367.

[273] M. G. Joyce *et al.*, "Crystal structure and immunogenicity of the ds-cav1-stabilized fusion glycoprotein from respiratory syncytial virus subtype b authors," *Pathog. Immun.*, vol. 4, no. 2, pp. 294–323, Dec. 2019, doi: 10.20411/pai.v4i2.338.

[274] B. Chen *et al.*, " A Chimeric Protein of Simian Immunodeficiency Virus Envelope Glycoprotein gp140 and Escherichia coli Aspartate Transcarbamoylase ," *J. Virol.*, vol. 78, no. 9, pp. 4508–4516, 2004, doi: 10.1128/jvi.78.9.4508-4516.2004.

[275] V. Restivo *et al.*, "Influenza vaccine effectiveness among high-risk groups: A systematic literature review and meta-analysis of case-control and cohort studies," *Hum. Vaccin. Immunother.*, vol. 14, 2018, doi: 10.1080/21645515.2017.1321722.

[276] M. Kubo and K. Miyauchi, “Breadth of Antibody Responses during Influenza Virus Infection and Vaccination,” *Trends Immunol.*, vol. 41, no. 5, pp. 394–405, 2020, doi: 10.1016/j.it.2020.03.005.

[277] S. S. Wong and R. J. Webby, “Traditional and new influenza vaccines,” *Clin. Microbiol. Rev.*, vol. 26, no. 3, pp. 476–492, Jul. 2013, doi: 10.1128/CMR.00097-12.

[278] W. Gerhard, J. Yewdell, M. E. Frankel, and R. Webster, “Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies,” *Nature*, vol. 290, no. 5808, pp. 713–717, 1981, doi: 10.1038/290713a0.

[279] N. R. Mathew and D. Angeletti, “Recombinant Influenza Vaccines: Saviors to Overcome Immunodominance,” *Frontiers in Immunology*, vol. 10. p. 2997, 2020, doi: 10.3389/fimmu.2019.02997.

[280] WHO, “Recommendations for influenza vaccine composition,” 2023. https://www.who.int/teams/global-influenza-programme/vaccines/who-recommendations (accessed Nov. 08, 2023).

[281] D. C. Ekiert *et al.*, “Cross-neutralization of influenza A viruses mediated by a single antibody loop,” *Nature*, vol. 489, no. 7417, pp. 526–532, 2012, doi: 10.1038/nature11414.

[282] K. Nagashima *et al.*, “The Pre-Existing Human Antibody Repertoire to Computationally Optimized Influenza H1 Hemagglutinin Vaccines,” *J. Immunol.*, vol. 209, no. 1, pp. 5–15, Jul. 2022, doi: 10.4049/jimmunol.2101171.

[283] S. M. Ghafoori *et al.*, “Structural characterisation of hemagglutinin from seven Influenza A H1N1 strains reveal diversity in the C05 antibody recognition site,” *Sci. Rep.*, vol. 13, p. 6940, 2023, doi: 10.1038/s41598-023-33529-w.

[284] G. A. Sautto, G. A. Kirchenbaum, J. W. Ecker, A.-G. Bebin-Blackwell, S. R. Pierce, and T. M. Ross, “Elicitation of Broadly Protective Antibodies following Infection with Influenza Viruses Expressing H1N1 Computationally Optimized Broadly Reactive Hemagglutinin Antigens,” *ImmunoHorizons*, vol. 2, no. 7, pp. 226–237, Aug. 2018, doi: 10.4049/IMMUNOHORIZONS.1800044.

[285] S. Bangaru *et al.*, “A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface Article A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface,” *Cell*, vol. 177, 2019, doi: 10.1016/j.cell.2019.04.011.

[286] C. Dreyfus *et al.*, “Highly conserved protective epitopes on influenza B viruses,” *Science (80-.).*, vol. 337, no. 6100, pp. 1343–1348, 2012, doi: 10.1126/science.1222908.

[287] H. A. Vanderven and S. J. Kent, “The protective potential of Fc-mediated antibody functions against influenza virus and other viral pathogens,” *Immunol. Cell Biol.*, vol. 98, no. 4, pp. 253–263, Apr. 2020, doi: 10.1111/IMCB.12312.

[288] Y. Li *et al.*, “Immune history shapes specificity of pandemic H1N1 influenza antibody responses,” *J. Exp. Med.*, vol. 210, no. 8, pp. 1493–1500, 2013, doi: 10.1084/jem.20130212.

[289] S. T. H. Liu *et al.*, “Antigenic sites in influenza H1 hemagglutinin display species-specific immunodominance,” *J. Clin. Invest.*, vol. 128, no. 11, pp. 4992–4996, 2018, doi: 10.1172/JCI122895.

[290] A. Yasuhara *et al.*, “Diversity of antigenic mutants of influenza A(H1N1)pdm09 virus escaped from human monoclonal antibodies,” *Sci. Rep.*, vol. 7, no. 1, p. 17735, 2017, doi: 10.1038/s41598-017-17986-8.

[291] B. Reynisson *et al.*, “Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data,” *J. Proteome Res.*, vol. 19, no. 6, pp. 2304–2315, Jun. 2020, doi: 10.1021/ACS.JPROTEOME.9B00874/ASSET/IMAGES/LARGE/PR9B00874_0006.JPEG.

[292] J. Racle *et al.*, “Machine learning predictions of MHC-II specificities reveal alternative binding mode of class II epitopes,” *Immunity*, vol. 56, no. 6, pp. 1359-1375.e13, Jun. 2023, doi: 10.1016/J.IMMUNI.2023.03.009.

[293] J. Racle *et al.*, “Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes,” *Nat. Biotechnol. 2019 3711*, vol. 37, no. 11, pp. 1283–1286, Oct. 2019, doi: 10.1038/s41587-019-0289-6.

[294] B. Reynisson, B. Alvarez, S. Paul, B. Peters, and M. Nielsen, “NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data,” *Nucleic Acids Res.*, vol. 48, pp. 449–454, 2020, doi: 10.1093/nar/gkaa379.

[295] T. J. O’Donnell, A. Rubinsteyn, and U. Laserson, “MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing,” *Cell Syst.*, vol. 11, no. 1, pp. 42-48.e7, Jul. 2020, doi: 10.1016/J.CELS.2020.06.010.

[296] D. Gfeller *et al.*, “Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8 + T-cell epitopes ll Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8 + T-cell epitopes,” doi:

10.1016/j.cels.2022.12.002.

[297] S. Sarkizova *et al.*, “A large peptidome dataset improves HLA class I epitope prediction across most of the human population,” *Nat. Biotechnol.*, vol. 38, no. 2, pp. 199–209, Dec. 2020, doi: 10.1038/s41587-019-0322-9.

[298] A. Sette and J. Sidney, “Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism,” *Immunogenetics*, vol. 50, no. 3–4, pp. 201–212, 1999, doi: 10.1007/s002510050594.

[299] S. Southwood *et al.*, “Several Common HLA-DR Types Share Largely Overlapping Peptide Binding Repertoires,” *J. Immunol.*, vol. 160, no. 7, pp. 3363–3373, Apr. 1998, doi: 10.4049/jimmunol.160.7.3363.

[300] J. K. Taubenberger, J. C. Kash, and D. M. Morens, “The 1918 influenza pandemic: 100 years of questions answered and unanswered,” *Sci. Transl. Med.*, vol. 11, no. 502, 2019, doi: 10.1126/scitranslmed.aau5485.

[301] R. C. Edgar, “MUSCLE: Multiple sequence alignment with high accuracy and high throughput,” *Nucleic Acids Res.*, vol. 32, no. 5, pp. 1792–1797, 2004, doi: 10.1093/nar/gkh340.

[302] A. Bakan *et al.*, “Evol and ProDy for bridging protein sequence evolution and structural dynamics,” *Bioinformatics*, vol. 30, no. 18, pp. 2681–2683, Sep. 2014, doi: 10.1093/bioinformatics/btu336.

[303] P. J. A. Cock *et al.*, “Biopython: Freely available Python tools for computational molecular biology and bioinformatics,” *Bioinformatics*, vol. 25, no. 11, pp. 1422–1423, Jun. 2009, doi: 10.1093/bioinformatics/btp163.

[304] A. Stamatakis, “RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies,” *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, May 2014, doi: 10.1093/bioinformatics/btu033.

[305] I. Letunic and P. Bork, “Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation,” *Nucleic Acids Res.*, vol. 49, no. W1, pp. W293–W296, Jul. 2021, doi: 10.1093/NAR/GKAB301.

[306] G. Talavera and J. Castresana, “Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments,” *Syst. Biol.*, vol. 56, no. 4, pp. 564–577, Aug. 2007, doi: 10.1080/10635150701472164.

[307] M. J. P. Van Dongen *et al.*, “A small-molecule fusion inhibitor of influenza virus is orally active in mice,” *Science (80-.).*, vol. 363, no. 6431, Mar. 2019, doi: 10.1126/science.aar6221.

[308] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton, “Jalview Version 2-A multiple sequence alignment editor and analysis workbench,” *Bioinformatics*, vol. 25, no. 9, pp. 1189–1191, 2009, doi: 10.1093/bioinformatics/btp033.

[309] Y. Shu and J. McCauley, “GISAID: Global initiative on sharing all influenza data – from vision to reality,” *Eurosurveillance*, vol. 22, no. 13. European Centre for Disease Prevention and Control (ECDC), p. 30494, Mar. 30, 2017, doi: 10.2807/1560-7917.ES.2017.22.13.30494.

[310] G. E. Crooks, G. Hon, J. M. Chandonia, and S. E. Brenner, “WebLogo: A sequence logo generator,” *Genome Res.*, vol. 14, no. 6, pp. 1188–1190, 2004, doi: 10.1101/gr.849004.